Wheat grain protein content assessment via plant traits retrieved from airborne hyperspectral and satellite remote sensing imagery

Andrew Robert Longmire B. Sc., M. Agric. Sc.

ORCID ID 0000-0002-4212-7371

Submitted in total fulfilment of the degree of

Doctor of Philosophy

School of Agriculture, Food and Ecosystem Sciences

Faculty of Science

The University of Melbourne

October 2023

To Sacha and Oscar, and their generation, who face the consequences of land management choices past, present, and future.

Declaration of published work

I hereby declare that no material included in this thesis has been accepted for the award of any degree at any university or like institution. I further declare that, to my knowledge, this thesis contains no material written or previously published by any person other than myself except where due reference and attribution are made in the text. This thesis is fewer than 100,000 words long. All chapters contain included material from the following multi-author papers which have been published or accepted for publication:

Longmire, A.R., Poblete, T., Hunt, J.R., Chen, D., Zarco-Tejada, P.J., 2022. Assessment of crop traits retrieved from airborne hyperspectral and thermal remote sensing imagery to predict wheat grain protein content. *ISPRS Journal of Photogrammetry and Remote Sensing 193, 284–298*. https://doi.org/10.1016/j.isprsjprs.2022.09.015

Longmire, A.R., Poblete, T., Hornero, A., Chen, D., Zarco-Tejada, P.J., (accepted 2023-10-27). Estimation of grain protein content in commercial bread and durum wheat fields via traits inverted by radiative transfer modelling from Sentinel-2 timeseries. *ISPRS Journal of Photogrammetry and Remote Sensing*.

The central theme of this work is the estimation of grain protein content in wheat crops based on plant traits retrieved from remotely sensed data. For the component papers, the project, and the thesis itself, the concept, experimental design, development and implementation of research, and data analyes were my own responsibility, with guidance and support from my supervisors. I am the primary author of the component papers.

I worked under the supervision of Prof Deli Chen, School of Agriculture, Food and Ecosystem Sciences (SAFES), Faculty of Science) and Prof Pablo Zarco-Tejada, Faculty of Engineering and Information Technology / SAFES.

Andrew Robert Longmire

31 October 2023

Abstract

Wheat (*Triticum* spp.) is crucial to food security. The source of a major proportion of humans' total dietary carbohydrates and protein, it is among the world's most widely grown crops and receives concomitantly large quantities of nitrogen (N) fertiliser. Wheat grain protein content (GPC; %) is a key to food quality, determining the baking quality of bread, the cooking quality of pasta, and the nutritional value of food products. For these reasons, wheat is classified and growers are typically paid predominantly on the basis of GPC, setting its farm income value. Global population growth encourages a justified focus on increasing yields. However, because grain proteins are diluted by carbohydrate (CHO) additions in the latter part of growing seasons, GPC is in an inverse relationship with yield: Improved yields are attended by the risk of reducing GPC. Moreover, GPC is influenced by interacting genetic and agronomic factors, soil properties and weather conditions that affect crops' physiological status and stress levels and can therefore exhibit great spatial variability.

Of the vast quantities of nitrogen (N) applied to wheat crops, a variable but substantial proportion is lost, inducing environmental damage and financial costs, which should be averted. Accurate GPC prediction could reduce N losses, assist in crop management decisions, and improve farm incomes. Nitrogen is central to proteins and can be strategically supplied to crops in order to achieve GPC benchmarks a precision agriculture (PA) approach. In such scenarios, estimates of GPC potential in advance of harvest could guide fertiliser dosing, improving fertiliser efficiency and potentially reducing costs. In contrast, where strategic fertiliser applications are not favoured, crop management could benefit from prior knowledge through strategic harvesting aimed at maximising payments per unit of grain at receival.

However, GPC is a complex variable, influenced by multiple plant traits, themselves affected by soil and moisture conditions and whose effects change through the growing season. While remote sensing (RS) is likely the only practicable method of estimating GPC during seasons, and shows potential, prediction is complex and success has been limited. To make progress, it is necessary to more robustly identify imaging spectroscopy-based physiological traits closely associated with GPC.

Traits with known physiological links to GPC, and which can be retrieved from imaging spectroscopy, include leaf area index (LAI) and chlorophyll (C_{a+b}). Further inspection of these and other RS traits may advance research relevant to PA. The inverse relationship of GPC to CHO assimilation permits the hypothesis that indicators of plant stress can improve GPC estimation. Such stress indicators, including the pigments anthocyanins and carotenoids, can be accurately retrieved along with other biophysical and biochemical quantities from hyperspectral (HS) remote sensing but their relationship to GPC had yet to be tested. Solar-induced fluorescence (SIF), emitted from the photosystems in proportion to instantaneous photosynthetic rate, was also untested as a GPC predictor. Moreover, in addition to the traits themselves, retrieval of plant traits by inversion of radiative transfer models (RTM) also remained untried for GPC estimation. Finally, the crop water stress indicator (CWSI), a proxy for evapotranspiration and hence carbon assimilation, should also show an association with GPC.

Because a large majority of GPC studies have been conducted exclusively in the context of experimental plots, it is appropriate to extend research into the commercial cropping domain, populated to date by only two previous studies. This expansion is facilitated by the recent advent of spatially explicit GPC monitoring during crop harvests. While lacking the ultra-high spectral and spatial resolution of airborne HS sensing, satellite RS, in particular the Sentinel-2 (S2) platforms, possess advantages with respect to broadacre PA. These include a

focus on reflectance bands adapted to vegetation sensing, appropriate spatial resolution, and frequent return times.

This thesis presents results from piloted HS flights and ground campaigns at two dryland field experiments with divergent water supply and wide-ranging N fertiliser treatments, and from HS flights over 17 commercial fields planted to either bread (*T. aestivum* L.) or durum (*T. turgidum* subsp. *durum* (Desf.) Husn.) wheat, across two years in the southern Australian wheat belt. Imagery was acquired with airborne hyperspectral and thermal sensors, with spatial resolutions of approx. 0.3 m and 0.5 m for experimental plots and 1 m / 1.7 m in commercial fields. Leaf clip measurements, leaf and grain samples were collected from plots and through a transect in one field. In commercial fields, ~40,000 records obtained from harvester-mounted protein monitors. CWSI, SIF, vegetation indices and PRO4SAIL RTM inverted parameters were retrieved for each plot and GPC record location. Sentinel-2 (S2) timeseries (TS) were subsequently acquired for > 6,000 ha of commercial dryland wheat fields, inclusive of those included in HS campaigns, also in south-east Australia and through two consecutive years of dissimilar rainfall. In this case, growers provided ~92,000 GPC data points from harvester-mounted protein monitors. For each, C_{a+b}, leaf dry matter, leaf water content (C_w) and LAI were retrieved from the S2 images by radiative transfer model inversion. A gradient boosted machine learning algorithm was applied to analyse these traits' importance to GPC and to predict GPC in 30% of samples unseen by the algorithm in training.

From HS analyses, the photochemical reflectance index (PRI) related to xanthophyll pigments was consistently associated with GPC at both leaf and canopy scale in the plots and transect. In the commercial crops, a gradient boosted machine learning algorithm (GBM) ranked CWSI as the strongest indicator of GPC under severe water stress, while SIF, PRI and inverted biochemical constituents anthocyanins and carotenoids were consistently important

under more moderate growing conditions. Structural parameters inverted from HS were not prominent except under severe drought when CWSI was omitted from models. **S**tatistically significant results were obtained for GPC estimation in unseen samples, with best relationships between predicted and observed GPC of $R^2 = 0.80$ and RMSE = 0.62 % GPC in a model built with thermal and physiological traits obtained from HS and thermal imagery.

Trait importance in S2 analyses was consistent with that seen from HS, in that the rankings of physiological, structural and water stress indicators were aligned: severe drought increased the importance of water stress measures relative to other traits, but in milder conditions physiological traits were emphasised. Airborne SIF added substantially to model skill from single-image S2, particularly in moderate conditions. While coefficients of determination varied substantially according to water stress, error metrics invariably sat within a tight range, under 1 % GPC. Overall, these predictive modelling results, obtained at within-field scale and in challenging conditions, place the current study among others in the same research domain, most of which consider either plot or regional scales. The strongest relationships between predicted and observed GPC ($R^2 = 0.86$, RMSE = 0.56 %), in a model built from five S2 images across a season, were better than those from single-date hyperspectral (HS). In severe water stress, LAI was the main predictor of GPC early in the season, but this switched to C_w later. In milder conditions, importance was more evenly distributed both through the years and between traits, and predictive skill was lower. S2 TS had a clear accuracy advantage over single-date S2, and approached that of HS, especially in benign conditions, emphasising its previously unexplored potential for large-scale GPC monitoring. The methods developed are a novel contribution and can be proposed as a useful basis for future research.

Thesis structure

This thesis comprises the following chapters:

Chapter 1: Introduction, background, motivations, objectives and significance of the PhD research.

Chapter 2: Literature review of methods of estimating GPC and physiologically linked plant/canopy traits from remotely sensed data

Chapter 3: Methods

Chapter 4: Results

Chapter 5: Discussion

Chapter 6: Conclusions, findings of the project, its limitations and scientific contribution, and proposals for future research.

Appendix A: Tables of all vegetation indices considered in analyses

Appendix B: Associated publication 1

Appendix C: Associated publication 2

List of associated publications and proceedings

Longmire, A.R., Poblete, T., Hunt, J.R., Chen, D., Zarco-Tejada, P.J., 2022. Assessment of crop traits retrieved from airborne hyperspectral and thermal remote sensing imagery to predict wheat grain protein content. *ISPRS Journal of Photogrammetry and Remote Sensing 193, 284–298*. https://doi.org/10.1016/j.isprsjprs.2022.09.015

Longmire, A.R., Poblete, T., Hornero, A., Chen, D., Zarco-Tejada, P.J., (accepted 2023-10-27). Estimation of grain protein content in commercial bread and durum wheat fields via traits inverted by radiative transfer modelling from Sentinel-2 timeseries. *ISPRS Journal of Photogrammetry and Remote Sensing*.

Longmire, A.R., Poblete, T., Hunt, J.R., Chen, D., Zarco-Tejada, P.J., 2022. Variability in wheat grain protein estimated from airborne hyperspectral and thermal remote sensing imagery, in: *GRDC Update Papers*. Presented at the GRDC Southern Updates, Grains Research and Development Corporation, Bendigo.

Longmire, A.R., Poblete, T., Hunt, J.R., Chen, D., Zarco-Tejada, P.J., 2022. Wheat grain protein content estimated by machine learning from hyperspectral and thermal remote sensing images, in: *Proceedings of the XXI International Nitrogen Workshop: Halving Nitrogen Waste by 2030*. Presented at the Proceedings of the XXI International Nitrogen Workshop: Halving nitrogen waste by 2030, School of Agricultural, Food and Biosystems Engineering - Universidad Politécnica de Madrid, Madrid, p. 330.

Acknowledgements

I appreciate enormously the patience, generosity and professionalism of my supervisors, Professors Deli Chen and Pablo Zarco-Tejada. While Deli and Pablo brought very different approaches and expertise to my project, each exemplifies radical success in pursuit of excellent science, driven by innate curiosity. I feel honoured to have had the tuition, in its many forms, of two people who are such giants in their fields, and hope, in turn, to have honoured their efforts. With their backing and guidance, I was in good hands for my own pursuit of a childhood dream.

It is thanks to an enthusiastic phone call from Deli that I took the PhD plunge. From that time on, he has offered mentoring, support and encouragement to indulge in academic thought, and to formulate and pursue my own goals.

Pablo's statement, in our first meeting, that one should enjoy every day of one's PhD, set a bar that was not easy to achieve, but suggested a widely applicable approach. Meetings with Pablo always stretched my technical skill and understanding of the concepts at hand, while also building my confidence. His exacting standards in research and publications encourage me to aim high, be critical of my own work, and to strive for accuracy, clarity and elegance.

I am grateful to all of my co-authors on the project's publications, but especially to Dr Tomás Poblete. Tomás' technical expertise and solid contributions to methodological discussions and implementation were key to my success and boosted my understanding of the tools of remote sensing, data science and machine learning.

Thank you also Dr Tony Weatherley, my advisory committee chair, for sage advice and a steady hand, particularly when I felt the project lacked momentum.

Thanks to Dr Anirudh Belwalkar and Dr Adrián Gracia Romero, who in now distant carefree, pre-pandemic days, their first in Australia, saddled up for jaunts to Birchip and Yarrawonga. Adrián, nearing his own PhD completion, brought experience in sampling procedures, while Anirudh, also starting his PhD, already seemed to know plenty about the instruments and data; both were methodical and reliable. Thanks, too, to Wey Yao Wong, for patiently getting me started as a complete beginner in `R` coding, then, much later when I was still pretty novice, Cameron Patrick of The University of Melbourne Statistical Consulting Centre for many more explanations of curly code. I also admire my studious and hardworking colleagues in both the Soils Group and the HyperSens labs at unimelb, who provide many examples of good science done creatively.

Without the Grains Research and Development Corporation's generous graduate research scholarship, I would not have commenced this endeavour. I acknowledge the Birchip Cropping Group (BCG), Riverine Plains Incorporated and the Foundation for Arable Research, who facilitated access to plot trials and, in the case of BCG, to laboratory and other facilities, logistical support and advice. Thanks also to the farm families who supplied the grain protein data on which the project relies, especially Jonathan Dyer, who was generous with time and advice. Such people do the hard work and long hours that keep us all fed, yet also find time to embrace new methods and offer their wealth of practical knowledge to research efforts.

I thank the Dharmananda community, where I live, for their interest and support, including many dinners and cold swims, and particularly Dr Leigh Davison for clarity and perspective.

Thanks to my family for genuinely expecting that I would get to this point – completion – and for gently expressing this belief when I had doubts, and, finally, to Oscar and Sacha for their boundless wisdom and enthusiasm.

Acknowledgement of country

During my candidature, I have worked and lived on the traditional lands of:

- The Wurundjeri Woi-wurrung clans of the Kulin nation, on whose country the principal campus of The University of Melbourne stands.
- The Barengi Gadji peoples, whose territories incorporate much of the Wimmera and Mallee, and the Yorta Yorta people, traditional custodians of the country around Yarrawonga; these are the places where field work and crop studies were done.
- Bundjalung country, home of the Widjabul-Wia-bul, where I now live and where most of the processing, data analysis, and drafting of this thesis and its component papers was done.

I acknowledge the traditional owners, past and present, who managed these respective homelands since time immemorial, did not cede the lands upon European colonisation of the Australian continent and have not been offered just terms for their many losses.

Table of contents

List of figures

List of tables

List of abbreviations

The following abbreviations are used in this thesis and its associated publications. With minor exceptions, neither vegetation indices (VIs) nor machine/deep learning algorithms are listed immediately below. These are provided inline, and a complete listing of VIs cited or used, with their formulae and citations, is provided in Appendix A.

Abbreviation	In full
AAR	average annual rainfall
AGL	above ground level
Anth	anthocyanins
AOD	aerosol optical depth
APSIM	Agricultural Production Systems sIMulator
ASC	Australian soil classification
AUD	Australian dollars
BoM	Bureau of Meteorology
C_{a+b}	chlorophyll a + b
CCC	canopy chlorophyll content
CHO	carbohydrate(s)
C_m	leaf structural thickness
cv	cultivar
CV	cross-validation
C_w	leaf water thickness
CWSI	crop water stress index
C_{x+c}	carotenes / xanthophylls
CZ	cropping zone(s)
DAS	days after sowing
DDAS	degree days after sowing
DL	deep learning
DSS	decision support systems
E	irradiance
ES	experimental site
FLD	Fraunhofer line depth
FWHM	full width at half maximum
GBM	gradient boosted/boosting machine
GDDAS	growing degree days after sowing
GHG	greenhouse gas(es)
GN	grain number
GOME-2	Global Ozone Monitoring Experiment 2
GPC	grain protein content
GPS	global positioning system
GSD	ground sampling distance
GSR	growing season rainfall

Ha	hectare
HS	hyperspectral
IR	infrared
k	extinction coefficient
L	radiance
LAI	leaf area index
LIDFa	leaf inclination distribution function
LNC	leaf nitrogen content
LOV	leaf orientation value
LOO	leave-one-out
LUT	look-up table
M	million
Mha	Megahectare; million hectares
ML	machine learning
MLRA	machine learning regression algorithm
MODIS	Moderate Resolution Imaging Spectroradiometer
MSI	multispectral instruments
Mt	megatonne (10^{12} g)
N	nitrogen
NA	not available
NBI	nitrogen balance index (Dualex proprietary measure)
NBHI	narrow-band hyperspectral index/indices
NDVI	normalised difference vegetation index
NIR	near infrared
N_{leaf}	leaf nitrogen
N_r	reactive nitrogen
NUE	nitrogen use efficiency
OCO-2	Orbiting Carbon Observatory-2
OOS	out-of-sample
PA	precision agriculture
PARGE	Parametric Geocoding and Orthorectification for Airborne Optical Scanner Data
PAW	plant-available water
PAWC	plant-available water capacity
PRI	photochemical reflectance index
PROSAIL/ PRO4SAIL	coupled PROSPECT + SAIL radiative transfer models
R	reflectance
R^2	coefficient of determination
RE	red edge
RGB	red-green-blue
RMSE	root mean square error
$RMSE_{cv}$	root mean square error of cross-validation
ROI	region of interest
rRMSE	relative root mean square error
RS	remote sensing
RTM	radiative transfer model
S2	Sentinel-2

SAIL	Scattering by Arbitrarily Inclined Leaves
SCIAMACHY	SCanning Imaging Absorption SpectroMeter for Atmospheric CHartographY
SGD	stochastic gradient descent
SIF	solar-induced fluorescence
SMARTS	Simple Model of Atmospheric Radiative Transfer of Sunshine
SNR	signal-to-noise ratio
SPAD	single photon avalanche diode
SSI	spectral sampling interval
SSM	site-specific management
SVM/R	support vector machine/regression
SWIR	short-wave infrared
t	threshold
T_a	air temperature
T_{amb}	ambient temperature
T_c	canopy temperature
Tg	teragram (10^{12} g)
TOA	top of atmosphere
TROPOMI	TROPOspheric Monitoring Instrument
TS	timeseries
UAV	unpiloted aerial vehicle
USD	United States dollars
VI	vegetation index
VIF	variance inflation factor
VNIR	visible and near infrared
VPD	vapour pressure deficit
VRT	variable rate technology/ies
WRB	World Reference Base for soil resources
WS	water stress
Z	Zadoks growth stage

1. Introduction

This chapter contains included material from:

Longmire, A.R., Poblete, T., Hunt, J.R., Chen, D., Zarco-Tejada, P.J., 2022. Assessment of crop traits retrieved from airborne hyperspectral and thermal remote sensing imagery to predict wheat grain protein content. *ISPRS Journal of Photogrammetry and Remote Sensing 193, 284–298*. https://doi.org/10.1016/j.isprsjprs.2022.09.015

Longmire, A.R., Poblete, T., Hornero, A., Chen, D., Zarco-Tejada, P.J., (accepted 2023-10-27). Estimation of grain protein content in commercial bread and durum wheat fields via traits inverted by radiative transfer modelling from Sentinel-2 timeseries. *ISPRS Journal of Photogrammetry and Remote Sensing*.

—

1.1. Research background

Wheat (*Triticum spp.)* has played a prominent role in human history. Since it was domesticated in the Fertile Crescent around 10,000 years BP, wheat has grown to occupy more of the world's arable lands than any other single crop (Preece *et al*., 2017; Shewry, 2009). Globally, more than 240 million hectares (Mha) are currently under wheat cultivation, from 67° north to 45° south, and over 900 Tg (10^{12} g; = megatonne (Mt)) of grain was produced in 2021 (FAO, 2022). Wheat is the most traded cereal in the world, with ~ 200 Tg, worth USD55 billion, shipped internationally in 2021. This constitutes approximately 22 % of total annual production. The economic importance of wheat is also linked to its crucial role in our diets and cultures. More than 20 % of humans' intake of both carbohydrates (CHO) and protein come from a combination of bread (*T. aestivum* L.) and durum (*Triticum turgidum* subsp. *durum* (Desf.) Husn.) cultivars (FAO, 2022; Shiferaw *et al*., 2013). In confirmation of its centrality to human life, wheat also imprints cultural and faith traditions, representing sacred entities or practices in major religions (Shewry, 2009).

Aside from its role as a nutritional mainstay, wheat protein determines the baking quality of flour and hence, the rheological properties of leavened bread, and the cooking quality of pasta products (Cubadda *et al*., 2007; Panozzo *et al*., 2014). Therefore, grain protein content (GPC; %) is critically important to both the nutritional value of one of *Homo sapiens'* staple crops and to our experience of food. Commensurately, GPC determines the final purpose of wheat and higher protein content yields higher unit prices at grain receival (Apan *et al*., 2006). Protein content varies according to cultivar and growing conditions and can vary from 7—22 % but in commercial cultivation is usually between 10—15 % (Shewry, 2009); among the many threats to agriculture from climate change, the net influence of atmospheric CO_2 enrichment on GPC is likely to be negative (Högy and Fangmeier, 2008).

Global population growth has increased wheat demand, causing concerns over food security and equity of economic access to food. To address food demand, and growing economic pressure on farms and primary producers, much effort has been invested in increasing wheat yields. Such efforts have led to a 38 % increase in global wheat production volumes since the 1990s, with no increase in the area sown to wheat (Erenstein *et al*., 2022). While the focus on yield is justified, the inverse correlation of yield to GPC means that any action taken to increase one, can have the opposite effect on the other (Barraclough *et al*., 2010; Bogard *et al*., 2010; Zörb *et al*., 2018). Thus, as grain yields increase, we must also ensure that quality is maintained, given the importance of wheat protein.

Nitrogen (N) is a central component of amino acids, and hence also of proteins and chlorophyll. After carbon (C), hydrogen, and oxygen, N is the most abundant element in plants but is frequently also in limited supply. With insufficient N, crop growth is constrained because less chlorophyll can be synthesised, limiting crops' capacity to harvest sunlight for further growth. N limitation also restricts root development and hence plants' capacity to forage for water and nutrients, and promotes early senescence (Kant *et al*., 2011; Paul and

Driscoll, 1997). Within plants and crops, GPC is influenced by interacting genetic, edaphic, agronomic, and meteorological factors, which control N uptake and partitioning, protein translocation and dilution. The amount of protein ultimately accumulated in wheat grains is influenced primarily by the quantity of N in the plant at anthesis (Zadoks (Z) growth stage Z65; Giuliani *et al*., 2011; Lopez-Bellido *et al*., 2004; Masoni *et al*., 2007; Zadoks *et al*., 1974), and the amount of N plants can extract from the soil during grain filling (Gooding *et al*., 2007; Jamieson and Semenov, 2000; Ottman *et al*., 2000; Taulemesse *et al*., 2016). As such, soil properties exert considerable influence over GPC and can result in great GPC variability over short distances, most significantly due to differential moisture availability. Also closely tied to soil moisture, the rate and duration of photosynthesis during grain filling (Z70—79) determines how much protein is diluted by new assimilates. Grain yield increases with greater assimilation, but this results in lower GPC as a proportion of grain mass. This long-recognised inverse yield~GPC relationship reflects the association of plant stress with GPC (McNeal *et al*., 1978; Terman, 1979). Hence factors which promote profitability via increased production volumes have the opposite effect through quality. Adding complexity in water-limited cropping, early-season N supply bears strongly upon moisture conservation: sufficient N to achieve quick canopy closure can limit soil drying, but overapplication risks driving excessive early vigour, depleting moisture, and exacerbating later water stress. This effect, known in Australia as 'haying off' (Angus and Fischer, 1991; van Herwaarden *et al*., 1998), has strong negative yield effects, including the potential for total crop loss. Aversion to such effects deters growers from N applications sufficient to prevent mining (Angus and Grace, 2017).

In our quest to produce sufficient volumes of nutritious food, humans encounter two urgent but sharply contrasting environmental issues in which N plays a central role. Soils in many low-input agricultural systems are affected by N mining (Chen *et al*., 2016; Novelli *et al*.,

2023), wherein removals in harvested products exceed inputs over a sustained period, resulting in soil N depletion. This is problematic because soil N is crucial not only to crop plants, but also to soil microbiota and the overall health, C content, water holding and productivity of soils; each of these is reduced with soil N depletion. Conversely, in high-input agricultural systems, the overapplication of N fertilisers and the failure of crops to remove them from the soil in commensurate quantities, allows large amounts of reactive nitrogen (Nr) to be released to the environment, causing a cascade of impacts as outlined below (e.g. Andrews and Lea, 2013; Fowler *et al*., 2015; Galloway *et al*., 2003; Vitousek *et al*., 1997). Given these conflicting yet intertwined problems, the indispensability of fertilisers to food production, and the unlimited amount it is possible to fix as plant-available, reactive forms, the N dilemma can be counted among the *'wicked problems'* facing humankind in the early- to mid- 21st century.

An estimated 8 % of all N removed from soil by global wheat crops is mined from soils (Ladha *et al*., 2016); because of the astronomical quantities of wheat grown and the spatial extent of cultivation, this is problematic. Due to historic and ongoing land use and management choices, which promoted both bulk soil loss and degradation (Koch *et al*., 2015), soil N depletion is widespread in Australia, and many soils will need higher N inputs to maintain their productivity in the medium term (Angus and Grace, 2017; Chen *et al*., 2016). Many Australian soils were mined of N through continuous cropping or crop-fallow rotations from the time of their conversion until the 1950s, when the use of legumes to fix N in bioavailable forms in ley pastures began (Henzell, 2007, p. 19). Overall, 20—25% of the N present in Australian soils at the time of colonisation has been mined or otherwise lost (Angus and Grace, 2017), and this is ongoing across large areas (Harries *et al*., 2021). By a separate measure, which also attests to declining soil quality, Sanderman *et al*. (2010) estimated that 40—60 % of pre-colonisation soil C has been lost where and since agriculture

was imposed upon Australian soils. Indirectly, this loss of C also implies large losses of N (Murty *et al*., 2002).

Before synthetic fertiliser became widely available, farmers often simply moved on to new areas when native soil N was exhausted. For many decades prior to the introduction of any system to manage N, increases in production relied on 'new' land being brought under cultivation in this way (Henzell, 2007, p. 14). Thus were large areas of Australia colonised, and some subsequently abandoned, when cereal yields declined. This pattern was repeated globally (Scholes and Scholes, 2013): long-term soil N mining due to wheat production is reported in North America (Rasmussen and Parton 1994), Africa (Ntinyari *et al*., 2022; Sanchez, 2002) and South America (Novelli *et al*., 2023). In Australia and elsewhere, the challenge of soil mining remains, and requires a strong focus on N management to prevent further declines in the ability of soils to sustain crops and life (Liang *et al*., 2022).

So, because cropping depletes soil N, the human response is to add it back; both as synthetic fertiliser and through cultivated biological fixation (BNF), in which legumes are included in crop rotations. Synthetic fertiliser now contributes half of the Nr in human diets (Liang *et al*., 2021), and more than 80 % of fertiliser N (Galloway *et al*., 2017; Ladha *et al*., 2016). Globally, 17 % of all N fertiliser is applied to wheat (Heffer and Prud'homme, 2020), much of it as urea. Our collective Nr production doubled between 1980 and 2010 (Fowler *et al*., 2015) and we have infinite capacity to produce more, simply by invoking the Haber-Bosch process, which fixes atmospheric N as Nr and has operated at industrial scale since 1913. However, of all N supplied to crops, around 50 % is lost as environmental pollution (Andrews and Lea, 2013) and as little as 12 % is ultimately consumed by humans (Smil, 2002). In Australia, only around 40 % of the fertiliser N applied to wheat is assimilated into crops in the year of application and the rest is susceptible to loss, largely through ammonia

volatilization and denitrification, which cause combined losses equivalent to harvest removals (Angus and Grace, 2017; Chen *et al*., 2008).

The term 'reactive nitrogen' (Nr) captures any nitrogenous compound except dinitrogen (N_2): the plant-available nitrates (NO_3^-) and ammonium (NH_4^+); nitrites (NO_2^-) and ammonia gas (NH_3), which are not generally available to plants but are large contributors to pollution; oxides of nitrogen (NOx), and the potent greenhouse gas (GHG) nitrous oxide (N_2O; (Stein and Klotz, 2016). Nr species accumulate in the environment because the rate of anthropogenic additions far exceeds that of their removal in crops and through denitrification, and are unique as environmental pollutants because their high mobility and reactivity enable movement between ecosphere domains (Fowler *et al*., 2015; Galloway *et al*., 2017). In this 'nitrogen cascade' (Galloway *et al*., 2003), Nr pollutants in one domain are transformed and transported, and their damage multiplied in another. Stemming from our overuse of Nr and assisted by this cascade, humanity has overstepped the 'safe operating space' for Nr releases, by a factor of two (Steffen *et al*., 2015).

The effects of excessive N use are numerous and costly: Nr species contribute to stratospheric ozone depletion and increase tropospheric smog, with direct damage to human health through PM2.5 pollution, closely linked to NH_3 emissions. Nitrous oxide (N_2O) is a greenhouse gas with global warming potential ~ 300 times greater than that of CO_2 and of which 50% of emissions come from agricultural N use (Shcherbak *et al*., 2014). In the hydrosphere, Nr from agriculture causes eutrophication, algal blooms, hypoxia and biodiversity at extreme scales in marine and aquatic environments, for example in the Gulf of Mexico (McLellan *et al*., 2018), on Australia's Great Barrier Reef (De'ath *et al*., 2012), in almost all major Chinese lakes (Gao and Zhang, 2010), and as nitrate pollution in ground- and drinking water (Ju and Zhang, 2017). In terrestrial ecosystems, Nr is implicated in soil acidification, especially in agricultural soils, native grassland and forest dieback, and biodiversity loss (Liang *et al*.,

2021). Environmental damage arising from fertiliser supply chains further emphasises the need to reduce waste (Galloway *et al*., 2017; Van Grinsven *et al*., 2013); GHG emissions from fertiliser production doubles those from its application (Chen *et al*., 2014) and N fertiliser can contribute half of all energy inputs in conventional wheat cropping (Colaço *et al*., 2012).

'Leaky' systems of crop nutrition contribute to annual global wastage of 18 Tg Nr, worth USD200 billion (Sutton *et al*., 2020). The net cost to society of lost Nr in Europe is more than twice the fertilisers' contributions via farm incomes (Sutton *et al*., 2011), because of Nr-related costs to human health, the wider environment and the feedbacks to agriculture. The 2008 cost of fertiliser overuse to the European Union was estimated at EUR230—485 billion (USD338—712 billion), of which approximately half came from agriculture (Van Grinsven *et al*., 2013). Nitrogen is also the costliest input for many crops worldwide (Xu *et al*., 2012) and in Australia, is the highest variable cost for grain growers. Given its low use efficiency here, annual losses may total AUD500 million (USD 425 M), a significant sunk cost at farm and broader levels (Monjardino *et al*., 2015, 2013) and a large, risky investment for farmers already operating under difficult and variable climatic and economic conditions. Furthermore, in the two years to April 2022, the global urea price increased four-fold (Baffes and Koh, 2022) and given increasing fertiliser demand and likely tightening supplies (Cross and Gruère, 2022), significant long-term price relief is unlikely. These dynamics directly increase crop production costs.

Nitrogen use efficiency (NUE) is a key concept with many definitions, emphases and purposes, recently surveyed by Congreves *et al*. (2021). Here, NUE is taken to mean harvested N divided by total N input; the latter includes native or residual soil N plus fertiliser inputs (Gu *et al*. 2023). NUE and GPC are inextricably linked, as most of the N ultimately removed in a crop – the numerator of the NUE equation – resides in proteins. But

while high GPC demonstrates – among other factors – crop N recovery, NUE is highly dependent on its denominator, of which in turn only fertiliser applications can be estimated with any accuracy, and even then, not always.

Crop NUE is low for reasons including the dynamic nature and multiple loss pathways of Nr species and a failure to match the spatial and temporal variability of soil N, including fertiliser, against crop needs. Low NUE is implicated in both soil mining and elevated Nr emissions, but much can be done to improve NUE through nutrient management (e.g. Chen *et al*., 2014, 2011). In low-input cropping systems, common in Australia, underuse of fertiliser N is the general case: here, selectively increasing in-season applications may offer NUE, soil health and/or profit advantages (Angus and Grace, 2017; Monjardino *et al*., 2015, 2013). While there is great variability between countries, regions and crop types, recent studies estimate current global NUE at around 0.35, and propose that this figure would need to rise to around 0.77 to return humanity within a safe range (Omara *et al*., 2019; Schulte-Uebbing *et al*., 2022).

Nitrogen fertiliser rate decisions are "*a multivariate, and often complex, optimization problem, with weather and soil properties as critical regulators*" (Colaço *et al*., 2021), made in the context of economic risk assessments (Monjardino *et al*., 2013). Weather varies in time while soil properties vary across space, but the two also interact; both under- and overapplication of N are responses to incomplete knowledge of these dynamics, the resultant variability in crop performance and the fate of applied N. Currently, due to the methods on which they commonly rely, decision support systems (DSS) to assist farmers with rate decisions fail to improve NUE and suffer from poor rates of uptake (Schwenke *et al*., 2019).

Preseason soil N measurements are slow, labour-intensive, relatively expensive and provide limited actionable information, yet remain a standard method of assessing fertiliser needs

(Schwenke *et al*., 2019). Further, the information they supply is commonly extrapolated across large spatial extents from a few points, causing further error, and testing frequency is low, even among the ~ 25 % of adopters (Lobry de Bruyn and Andrews, 2016). Methods common in high-input systems are also criticised because they are often associated with a single heavy fertiliser application in the preseason, and later, lighter applications. This either ignores temporal variability in crop N uptake or presumes N will remain in the soil profile, and worsens asynchrony between uptake and availability, missing an opportunity to improve NUE (Raun *et al*., 2002; Schwenke *et al*., 2019; Shanahan *et al*., 2008). Moreover, these methods also generally fail to account for in-season mineralisation, disregarding a potentially substantial source of plant-available N (Shanahan *et al*., 2008). Despite progress, such imprecise methods remain common because the issues at hand are complex and alternatives are seen as equally so, expensive and often unproven.

The recognition that uniform N treatments on spatially variable soils contribute to low NUE encouraged the development of precision agriculture (PA) techniques including variable rate technologies (VRT), and of many sensor-based approaches for N status assessment (Adamchuk, 2013). These efforts generally focus on reducing the amount of N applied, acting therefore on the denominator of the NUE equation. On-combine yield and/or protein monitor data, farmer knowledge, soil data from spatially sparse or dense sources, information from optical sensors and an increasing range of other origins can be used individually or in various combinations to delineate soil management zones used to guide fertiliser VRT. However yield and protein maps are necessarily produced *post hoc* and currently available N prescription maps are unresponsive to rainfall variability (Shanahan *et al*., 2008).

Sensor-guided VRT have been associated with NUE improvements without negative yield effects, but a profit advantage remains to be sufficiently demonstrated (Cassman *et al*., 2002; Colaço and Bramley, 2018). This detracts from their appeal, and the adoption of VRT,

whether guided by sensors or not, is patchy at best: Barnes *et al*. (2019) report that the proportion of wheat growers who use it for fertiliser ranges from 2—47 % among selected European countries, while the United States' equivalent is 11 % (Lowenberg-DeBoer and Erickson, 2019). Across all crop types, approximately 20 % of Australian farmers had adopted VRT for N by 2018, an increase from approximately 15 % in 2012 (Colaço and Bramley, 2018; Llewellyn and Ouzman, 2014), although Schwenke *et al*. (2019) reported that well less than 10 % of eastern-Australian grain farmers used any kind of in-season sensing. While the economic advantages of VRT will grow as fertiliser costs rise and could compound by reducing energy inputs per unit of yield by up to 20 % (Colaço *et al*., 2012), a solid demonstration of in-season GPC estimation and its applicability on farms would be a valuable contribution.

An opportunity exists to advance the development of VRT to better respond to both spatial and temporal variations in crop performance. In-season N applications at the right time can elicit positive GPC responses, without negative effects on yield (Giordano *et al*., 2023; Zebarth *et al*., 2007; Zhang *et al*., 2022). Early GPC estimation, based on crop traits observed in-season, could permit VRT fertiliser applications, potentially improving GPC and reducing the risk of waste. This approach may best apply where late-season N applications already occur, as these rely on field trafficability and the likelihood of fertiliser wash-in, so are region- and weather dependent. However, GPC predictions could also guide strategic harvesting or product blending for better access to premium grain prices (Apan *et al*., 2006; Sowers *et al*., 1994; Zhao *et al*., 2005), perhaps most relevantly in low-input systems. It is necessary to test the capacity of recent advances in remote sensing to address these aims.

Due to the many factors that influence protein sink and source sizes, translocation and dilution, GPC estimation from RS data is more complex than yield estimation, and success in GPC estimation has been limited (Freeman *et al*., 2003; Rodrigues *et al*., 2018; Zhao *et al*.,

2019). Moreover, relationships between plant traits and GPC may be indirect, vary with phenology and interact. Substantial research details the retrieval of other traits linked to GPC, including leaf and canopy chlorophyll (C_{a+b}) and N content, leaf area index (LAI), biomass and water stress. Efforts to estimate GPC via these physiological precursors or directly, employing traditional empirical methods or complex statistical / machine learning techniques designed to elucidate complex non-linear relationships are surveyed in the literature review below.

Most work on GPC estimation to date has relied on vegetation indices (VI) which, though they may accurately indicate individual plant traits, lack robustness across agronomic settings, seasons and locations (Clevers and Kooistra, 2012; Jacquemoud *et al*., 1995). Such inaccuracy should be excluded if robust GPC estimation methods are to be developed for application in wide-ranging conditions. To this end, plant trait retrieval by radiative transfer models (RTM) are an alternative, elucidating plant traits from reflectance spectra in their real-world units with good accuracy (e.g. Bacour *et al.*, 2002; Féret *et al.*, 2008; Jacquemoud *et al.*, 2009, 1995; Li *et al.*, 2015). Such methods may improve understanding of mechanistic linkages between precursor traits and GPC which otherwise remain obscure. Further, notwithstanding the strong physiological link, through reduced CHO assimilation during grain filling, between stress and GPC, consideration of this link is to date sparse in published works.

Contemporaneously, the multitude of plant traits that can be estimated from HS data permits new approaches by which to investigate spatially variable plant stress and photosynthetic rate, hence carbon assimilation, across plant canopies. While mild or temporary plant stresses, which nevertheless have cumulative effects on GPC, are often poorly detectable through estimates of state variables including C_{a+b}, LAI, and their proxies, other RS spectroscopic quantities do provide insight into such short-term and/or mild stress. Solar

induced fluorescence (SIF) fluctuates in parallel with instantaneous photosynthetic rate (Flexas et al., 2002; Hikosaka and Noda, 2019; Zarco-Tejada et al., 2013a) and offers a more direct method of estimating actual photosynthesis, as compared to pigment quantities (Grace *et al*., 2007). The inclusion of SIF with other RS plant traits has strongly improved C_{a+b} and N estimation in wheat leaves and canopies (Camino *et al*., 2018; Jia *et al*., 2021), and has allowed accurate estimation of assimilation rates across N- and water stress levels, also in wheat (Camino *et al*., 2019). Despite these successes in other aspects of wheat performance, to date the SIF contribution to GPC represents a knowledge gap.

To date, all other studies of GPC have relied on experimental plots, limited layouts or sampling within commercial fields or regional crop statistics at far coarser scales than the fields in which wheat is actually grown. While GPC estimation at plot scale has sometimes seen moderate success, its application to commercial production scales is problematic. Likewise, in studies based on regional statistics, the sparsity of ground-truth observations impedes downscaling. Though this did not appear to be the intention of those studies, it is a crucial step toward finding practical solutions. Also, despite the many factors involved with GPC physiology, research focused on the comparative magnitude of plant trait contributions, and the temporal dynamics of such relationships is lacking; to advance, knowledge of what matters, when, to wheat GPC should be investigated. Finally, studies demonstrating wheat GPC estimation from image timeseries remain exceedingly rare, and the publications to date show low estimation accuracy (Rodrigues *et al*., 2018; Wang *et al*., 2014). Each of these studies uses reflectance indices, which impedes the transferability of their methods across agronomic settings. There is potential that methods based on RTM inversion, image timeseries, or both, will allow progress in estimation of the spatial patterns of GPC and permit improvement of N fertiliser DSS; such approaches should be tested.

Many authors assert that responsive calibration of the location, amount and timing of fertiliser applications, aided by sensor-based technologies, has great potential to improve N management (e.g. Gebbers and Adamchuk, 2010). By implication, this group includes the many researchers engaged in the refinement of sensor-based methods related to crop N status, GPC estimation and yield prediction, whose publications reflect a vast research effort in the nexus between food security and N stewardship. Nevertheless, the practical and financial benefits of such approaches remain to be sufficiently proven, even where PA adoption is relatively widespread (Colaço and Bramley, 2018; Robertson *et al*., 2012), and the application of RS/PA methods to GPC specifically are as yet immature.

Just as the problems caused by inappropriate N management vary, so must responses be defined according to heterogeneity of impacts, agricultural systems and economic development: while large NUE improvements are needed in some regions, others still substantially lack access to N (Schulte-Uebbing *et al*., 2022). Some authors warn that agricultural technology may provide a "*false sense of security*" against unsustainable soil quality decline, particularly nutrient mining (Scholes and Scholes, 2013). Such misgivings are justified, particularly where N is underapplied and this situation contributes to food shortages, for example in sub-Saharan Africa (Xia and Yan, 2023), and in Australia where historic and ongoing soil mining are problematic (Liang *et al*., 2022). To avoid further N mining while increasing food production in less developed countries, access to N fertiliser and N management methods must improve in parallel. Others assert that tech-heavy, sensor-based PA, is ill-adapted to use in developing countries (Chen *et al*., 2014). While this also appears reasonable, on a presumed basis of restricted access to improved technology and practices, the geographical origin and focus of much current research is in precisely those places where N applications are instead excessive. On smallholder farms in China and India, for example, better stewardship can achieve large N efficiency gains (Gu *et al*., 2023; Xia *et*

al., 2017). Countries further along the development continuum have witnessed the damage caused by poor N husbandry and have relatively easy access to advanced technologies and methods, although implementation challenges remain. It is crucial that we continue to pursue and share such approaches so that the damage caused by poor N husbandry can be repaired at home and avoided in developed countries. In countries where agriculture is already highly mechanised and technological, we need to improve our engagement with N to maintaining food quality and keep cropping financially viable.

Given the depth and urgency of the soil N mining, Nr, and food security problems, it is prudent to assess a range of approaches. Much of the very substantial research effort detailed in the attached literature review demonstrates both the intent and the achievements of the international RS community in developing remote sensing and PA tools applicable in this domain. It is hoped that this thesis can contribute to rapid and cost-effective GPC detection, and that this may in turn contribute to improving the conversion of N into grain protein.

1.2. Research motivations

The central contention of this thesis is that robust identification of plant traits which are influential on wheat GPC, and retrievable from imaging spectroscopy, is incomplete. This situation impedes the improvement of RS-based GPC estimation; considering the potential of RS-based PA methods to contribute to N fertiliser optimisation, this is a substantial knowledge gap. Reliable GPC estimation prior to harvest could improve farmers' ability to increase fertiliser use efficiency, and to attain grain quality benchmarks; each of these would assist in maintaining the profitability of wheat cropping. To advance GPC estimation from RS, it is important to know which plant traits, retrievable from airborne hyperspectral and satellite multispectral RS, are most informative for GPC estimation. Given the association of wheat GPC with stress, this should include robust inspection of imaging spectroscopy-based physiological traits. The stability of traits' importance to GPC estimation, and their influence

on prediction accuracy, should be tested across experimental and commercial contexts and in diverse conditions.

To advance precision agriculture, then, there is a need to improve RS-based GPC prediction within fields but across large extents. In a novel approach, this research links GPC data collected at high spatial resolution by on-combine GPC monitors during commercial harvests with techniques from airborne hyperspectral, thermal and satellite multispectral RS. The component studies extract the reflectance spectra associated with tens of thousands of commercial GPC points, and with experimental plots, then use both empirical techniques and physiologically based mechanistic methods to retrieve the underlying plant traits. Using equivalent methods, a carefully chosen machine learning algorithm is applied to traits retrieved from both airborne and S2 satellite images. This allows transparent analysis of the links between crop traits and GPC and permits evaluation of the degree to which S2 trait importances align with those from hyperspectral. Because crop performance varies across time and space, the relative influence of plant traits on GPC is also assessed between seasons and across timeseries within seasons. In each analysis, the algorithm's ability to predict GPC is tested; these metrics are also compared between data sources and agronomic contexts.

1.3. Research questions

Given the research motivations, and after an extensive and critical literature review, presented in Chapter 2, the following research questions were defined:

1) Which imaging spectroscopy-based plant physiological traits, retrievable from hyperspectral and thermal airborne remote sensing, are most closely associated with GPC?
2) Are these traits stable across experimental and commercial contexts, different locations and contrasting seasonal weather conditions?
3) Do plant traits retrieved from Sentinel-2 satellite images show similarities to those from hyperspectral images as predictors of GPC?

The application of equivalent methods to Sentinel-2 satellite data allows further inquiry:

4) What is the effect on model predictive skill of bandset reduction from hyperspectral to Sentinel-2 when traits from similar time points in each season are retrieved from the respective datasets?
5) Crop water stress index is a proxy for transpiration and therefore atmospheric gas exchange, while SIF is proportional to photosynthetic rate. Each should therefore add predictive power for GPC estimation via the carbohydrate dilution principle, but neither is available from Sentinel-2. What are the contributions of CWSI and/or SIF to model predictive skill when added to S2 traits?

And using plant traits retrieved from Sentinel-2 timeseries images:

6) What plant trait dynamics are evident within and across seasons?
7) What is the effect on model predictive skill of including TS elements in predictive models?

2. Literature review

This literature review considers, first, the environmental and agronomic factors that influence wheat GPC, and the physiological mechanisms involved. This necessarily involves some discussion of yield because of its intimate links with GPC, and because of the far stronger focus in the literature on yield than GPC. A brief discussion of precision agriculture (PA) is offered, to locate the current work at the nexus of RS and PA in the commercial cropping context. On-combine monitors are concisely introduced, as they are the basis of ground-level data collection for most of the project. RS data sources relevant to the project are surveyed: passive sensing, hyperspectral, Sentinel-2 multispectral, and thermal; proximal sensing is largely avoided. Methods of retrieving plant traits from these sources are then discussed, first in broad terms then with specific reference to GPC-linked structural and physiological traits, particularly related to plant stress, and water stress itself. Examples of the application of these sources and techniques for GPC estimation are presented. In its latter subsections, this chapter covers other material central to the thesis project: machine learning, narrowing to a focus on gradient boosting; and the treatment of multicollinearity. Finally, current challenges and research gaps are identified through a detailed survey of previous work. A strong focus on wheat is prioritised throughout; minor exceptions to this are clearly stated.

2.1. Physiological and agronomic influences on GPC

Wheat growers often target GPC to reach a class associated with higher unit price, while also optimising yield. Yield is usually negatively correlated with GPC, in a relationship characterised by dilution: the amount of protein in grains is established at flowering but reduces as a proportion of grain mass as new carbohydrates are assimilated or translocated from other tissues. Hence when photosynthesis and / or solute transport are constrained during grain filling, limited dilution results in lower yield but higher protein. The dilution principle, its physiological basis and relationship to abiotic factors are borne out in many

studies of wheat and wheat products (Brooks *et al*., 1982; Flagella *et al*., 2010; Gooding *et al*., 2003; Ozturk and Aydin, 2004; Zhao *et al*., 2005, 2022). With a principal focus on water, N, and ambient temperature (T_{amb}) stresses, these are surveyed in the following paragraphs.

Yield and GPC are influenced by interacting genetic, environmental and management factors, often characterised as G * E * M (Hatfield and Walthall, 2015; Russell *et al*., 2017). Plant genetics – the cultivar sown – are uniform within fields, chosen by managers in consideration of operational needs; the reliance of the Yield~GPC relationship on genetics is contested (Kibite and Evans, 1984; Lollato *et al*., 2021). Environmental factors are disaggregated, for the current analysis, into weather and soil. Seasonal precipitation can be presumed uniform within fields, and although some aspects of weather, such as frost severity, may vary within fields, neither can be influenced in dryland cropping. Soil properties can also vary greatly within fields and, through their effect on water and N availability, exert strong control over yield and GPC. Site-specific management (SSM), a component of PA, attempts to compensate for the differential effects of soil conditions.

The soils of the Wimmera and Mallee regions of the southern Australian wheatbelt, where this study was undertaken, often show marked variability over short distances. The dominant Wimmera soils are vertosols, but both sodosols and chromosols are also common; these may occur interspersed with each other, and each is commonly affected by subsoil alkalinity, sodicity, salinity and boron (B; Nuttall and Armstrong, 2010; Weiss *et al*., 2022). In the Mallee, cropping fields were laid out on former dune fields. In these systems, inter-dune (swale) soils are typically shallow and clay rich. Their often shallow, rocky hardpan and B toxicity impede wheat root growth and penetration, and while they can hold substantial amounts of water and native N, this may be unavailable to plants because of soil texture and osmotic effects (Hoffmann *et al*., 2016; Kirkegaard and Lilley, 2007; Nuttall *et al*., 2003). Dune crests, by contrast, have deep, loose sand and clay aggregates, low organic matter and

hence native N, few impediments to root penetration, and allow rainwater to percolate to depth, where it is protected from evaporation and accessible to roots (Kirkegaard and Lilley, 2007; Llewellyn *et al*., 2008). These dynamics are known to affect wheat yields, which can be highly variable within fields and across seasons and will similarly disrupt GPC. The Wimmera and Mallee regions are referred to in the experimental chapters below as cropping zones (CZ) 1 and 2, respectively.

2.1.1. Water / N stress

The effects of water and N stress can be difficult to separate, and in arid and semiarid regions such as those considered here, wheat production may be co-limited by N and water stress (WS) (Pancorbo *et al*., 2021; Sadras, 2004). Soil moisture management is by definition unavailable in dryland agriculture, and therefore has limited relevance to the current work. However, apart from their direct effects on plant performance, soil moisture and WS affect N uptake: solid N fertiliser must be dissolved and washed into the root zone; moisture sufficient for stomatal opening is needed for nutrient transport into roots and within the plant. The amount of N plants can extract from soil during grain filling is a strong influence on GPC, (Gooding *et al.*, 2007; Jamieson and Semenov, 2000; Ottman *et al.*, 2000). Moreover, any WS-induced reduction in growth rate also lowers crop N demand (Gonzalez-Dugo *et al*., 2015). The effects of water and N stress on GPC can be substantial and take various pathways; their timing matters to harvest outcomes, but in general, an increase in either stress increases GPC (Daniel and Triboï, 2002; Zhao *et al*., 2005). Overall, GPC is positively correlated with N supply, and this association is modified by WS (Campbell *et al*., 1981).

Like the factors that determine their status within plants, the relative contributions to GPC of water- and N stress interact and vary through seasons. Late season WS shortens grain filling and promotes premature senescence; these effects reduce CHO accumulation and hence GPC

dilution (Brooks *et al*., 1982; Gooding *et al*., 2003). While full-season WS had a very strong positive effect on GPC, the effect of deficits was far stronger later than early in the season (Ozturk and Aydin, 2004); the positive effect of WS on protein is also noted in wheat products (Flagella *et al*., 2010; Guttieri *et al*., 2000; Saint Pierre *et al*., 2008). Each of water- and N stress has increased GPC, with interaction such that GPC augmentation from WS was greater under supplementary N (Klem *et al*., 2018). Similarly, Raya-Sereno *et al*. (2021) found higher GPC under WS, that the magnitude of this effect depended on the amount of supplementary N and that the effect of N supplements was consistent only when water was sufficient. However, nor is the link between crop N status and GPC direct: correlations between anthesis leaf N content (LNC) and GPC have been found to be weak, though significant, and show the reliance of the LNC~GPC link on both efficient protein translocation and CHO assimilation late in seasons (Zhao *et al*., 2005, 2019). Indeed, WS can disrupt the translocation of both proteins and CHO (Gooding *et al*., 2007), and is a strong determinant of the fate of N (Klem *et al*., 2018).

Water and N stresses induce changes that can be detected by remote sensing, including reduced leaf or canopy chlorophyll content (Angus and Fischer, 1991), and reduced biomass, often synonymous with LAI. These effects of these compound, further reducing light interception and growth potential as well as total CHO and protein stores. Plants under photooxidative, water and heat stress also increase their production of the stress response pigments Anthocyanins (Anth) and carotenoids (C_{x+c}; Chalker-Scott, 1999; Groth *et al*., 2020; Janeczko *et al*., 2018; Naing and Kim, 2021). Sensitive to C_{x+c}, the photochemical reflectance index (PRI; (Gamon *et al.*, 1992)) reacts to water stress changes and serves as a proxy for photosynthetic rate (Feng *et al*., 2017; Magney *et al*., 2016b). PRI is proposed as an alternative to thermal RS for water stress detection (Kohzuma *et al.*, 2021; Suárez *et al.*, 2008) and through this has improved wheat yield estimates (Feng *et al.*, 2017; Magney *et al.*,

2014). Photosynthesis reductions caused by N stress are also detectable via reduced chlorophyll fluorescence (Camino *et al*., 2018; Jia *et al*., 2021). Water stress, though not directly N status, can be detected through RS as low leaf water content, decreased transpiration and increased leaf and canopy temperature; these are covered below.

2.1.2. N management

The responses of yield and GPC to increasing soil N availability can typically be broken into two phases along a N continuum. In the first phase, consistent with low N, yield responds strongly to extra N while GPC may decline, especially if N supply remains below demand. In the second phase, at higher N, the yield response flattens as it approaches its water-limited maximum, while the GPC curve approaches exponentiality. This phenomenon is long recognised and response curves resembling Fig. 2.1 appear widely in the literature (Fischer *et al*., 1993; Fowler, 2003; Pan *et al*., 2020; Pancorbo *et al*., 2023; Russell, 1963; Wright *et al*., 2004; Zebarth *et al*., 2007).

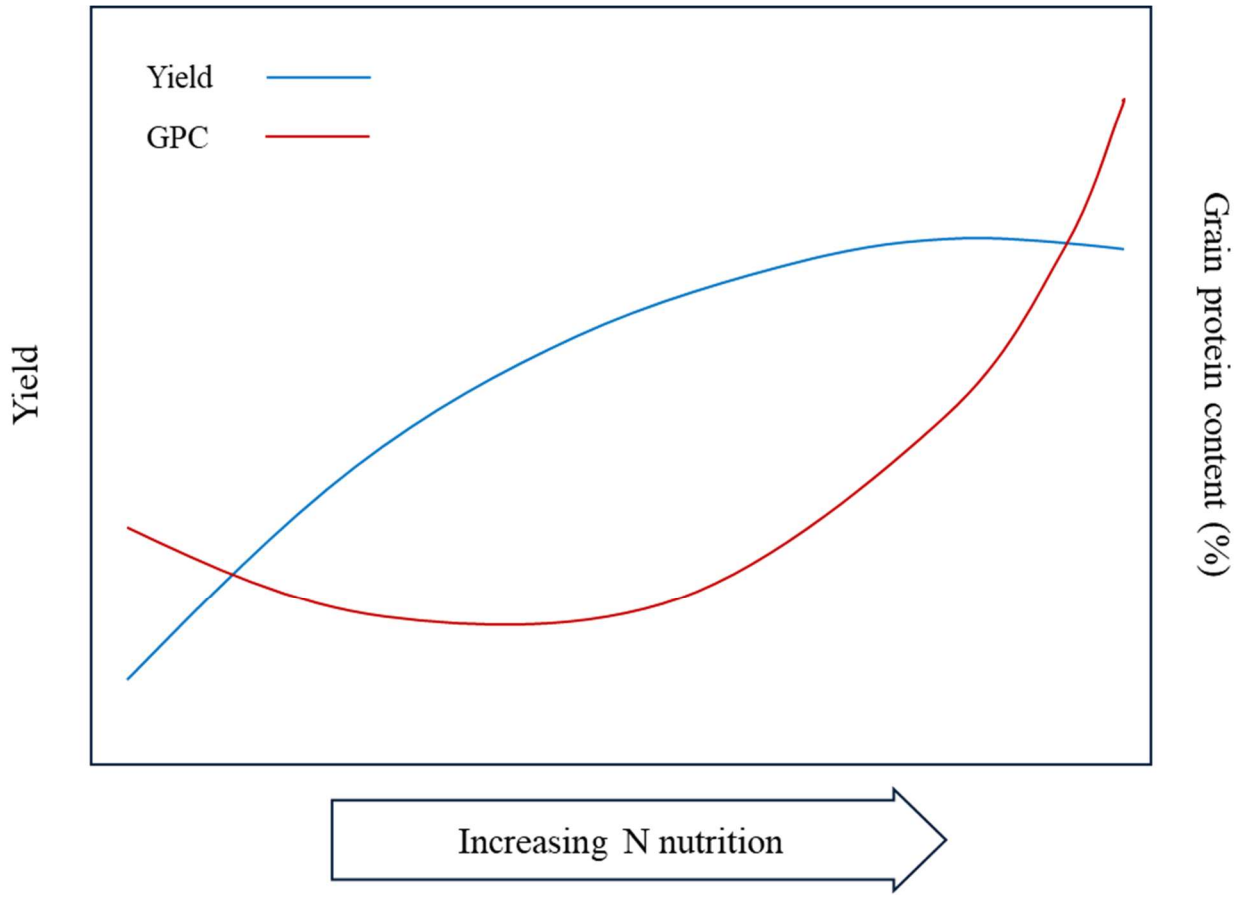

Fig. 2.1. Generalised yield and grain protein responses to additional nitrogen availability.

With regard to N management then, the so-called 'sweet spot' is where additional N does not increase yield but augments GPC and does not flow to the environment as waste. In general, for a given water supply, the relationship between yield, GPC and N follows a limited range of agronomically relevant patterns: low yield and low protein indicate limiting N; high yield, low protein may indicate that N was sufficient early, e.g., for tillering, but later became deficient; high yield with GPC in the target range indicate that a crop approached its potential. However, low yield with high protein can indicate either that growth was N-limited prior to stem elongation, limiting tillers and heads, hence GN, and thus irreversibly preventing both high yields and high canopy N and CHO content. Alternatively, excess N early can lead to excessive biomass, exhausting soil moisture and constraining grain filling;

this also produces high GPC and low yield (van Herwaarden *et al*., 1998). In severe cases, such 'haying off,' where late assimilation is restricted so much that grains finish with a characteristic 'pinched' shape because they fail to fill with CHO, can result in the complete loss of a grain crop, whereby as the name suggests it is cut for hay, but where less severe, grain could simply end up with high protein.

The timing of N fertilisation is also important. For example, in experimental conditions, N fertiliser applied after early stem elongation (Z30) was associated with increased GPC, while earlier applications elicited a strong yield response but no change to GPC (Fischer *et al*., 1993). 1993. Likewise in both bread (Lollato *et al*., 2021) and durum wheat (Ottman *et al*., 2000) GPC response to supplementary N was stronger late in seasons, although this has not been universal (Garrido-Lestache *et al*., 2005). A further proviso to this recipe is that supplementary N applied in solid form late in seasons needs to be washed into the root zone; this is not always operationally possible. Foliar applications of N, as liquid urea or ammonium nitrate may be made to lessen reliance on wash-in and circumvent uptake problems from dry, late-season soil. These can elicit GPC improvements far greater than broadcast fertiliser (Gooding and Davies, 1992) or earlier foliar applications (Bly and Woodard, 2003). However, Garrido-Lestache *et al*. (2005) noted only a minor, though statistically significant, GPC increase when foliar urea was applied at ear emergence; nevertheless, these findings support the notion that strategic fertiliser applications can affect the market value of grain.

2.1.3. Temperature stress

Like water stress, heat and cold can affect both yield and GPC. Exposure effects accumulate across seasons, but short periods of extreme ambient temperatures (T_{amb}) at crucial times can also cause severe damage. According to the dilution principle, such yield moderation should

increase GPC, but in the case of temperature extremes this is not universal. For example, even small T_{amb} increases can substantially increase respiration rates, promoting the loss of assimilated C (Heskel *et al*., 2016); this may reduce yields more than heat-inhibited photosynthesis (Li *et al*., 2021). Moreover, because of plants' reduced ability to cool themselves by evapotranspiration, heat damage is expected to be worse where soil moisture is low. Given the complex and interacting physiological pathways involved, it may be difficult to maintain cereal protein under increasingly variable temperature extremes (Kawasaki and Uchida, 2016).

Higher average T_{amb} accelerates phenology, shortening the time available for biomass accumulation in all vegetative growth phases (Altenbach *et al*., 2003; Dupont *et al*., 2006; Zhao *et al*., 2022), but particularly strongly during grain filling (Gooding *et al*., 2003). This has been demonstrated in bread and durum wheat, and in various experimental contexts. Phenological compression has compounding modes of action, some common to water / N limitation: lower aggregate CHO accumulation for later translocation; less biomass available for partitioning to photosynthetic structures, further reducing overall photosynthetic capacity, and lower root biomass, hence capacity to forage for water, N, and other nutrients. There is also a strong relationship between the stem elongation rate and grain N, as this phase is critical for accumulating canopy N, the majority of which is later transferred to the grain (Chapin and Wardlaw, 1988; Magney *et al*., 2016a; Miller *et al*., 1993). Moreover, slowed grain filling has been recorded at both low and high T_{amb} (Zhao *et al*., 2022).

As also found by Rharrabti *et al*. (2003), Flagella *et al*. (2010) saw higher GPC in durum when T_{amb} was elevated throughout the season, attributing this to lower total assimilation and hence dilution. Likewise, Dupont *et al*. (2006) saw higher bread wheat GPC, and reduced effect of N supplements on GPC, under elevated daytime T_{amb} of 37 °C applied throughout the season to irrigated wheat, also attributing the changes to accelerated phenology. These

findings are in common with Rao *et al*. (1993). High T_{amb} can also directly reduce both total protein accumulation and affect partitioning by protein type, affecting food quality (Panozzo *et al*., 2001; Triboï *et al*., 2003). For example, Flagella *et al*. (2010) observed increased gluten content correlated with the number of days > 30 °C during grain filling. The T_{amb} considered in the named studies are not unexpected for late-season winter wheat in Australian and other Mediterranean climes, and indeed are becoming more common (Asseng *et al*., 2015, 2011).

Extremes of T_{amb} around anthesis can strongly reduce grain number (GN) through damage to reproductive structures and kernel abortion (Fischer, 2011; Hays *et al*., 2007; Thakur *et al*., 2010); GN is the aspect of wheat plant development most susceptible to abiotic damage (Dolferus *et al*., 2011). Low spikelet counts (hence GN) have been observed at both high and low T_{amb}, and, at high T_{amb}, attributed to phenological compression (Halse and Weir, 1974; Rahman and Wilson, 1978). More recent work confirmed this mechanism and attribution, wherein GN-related yield reduction was proportional to a nocturnal-only T_{amb} increase (García *et al*., 2015). Although Ugarte *et al*. (2007) found that GN was restricted more by high T_{amb} during stem elongation than later, regardless of its cause and timing, GN shows a strong, positive association with yield via the overall CHO sink size (Dolferus *et al*., 2011; García *et al*., 2015; Thorne and Wood, 1987). However, the sink size effect is stronger on CHO than on proteins (Borghi *et al*., 1986), potentially exaggerating the inverse GN~GPC relationship.

Based on metabolic rates, cold stress should depress yield via a generalised reduction assimilation; this was supported in recent meta-analyses of the individual and combined effects of cold and drought on wheat (Ejaz *et al*., 2023) and by others (X. Li *et al*., 2015; Thakur *et al*., 2010). While early-season chilling can cause irreversible damage (Subedi *et al*., 1998), later cold stress can reduce GN through abortion or deformation of fruiting bodies,

impair pollen viability, decrease translocation and grain filling rates for both CHO and proteins, slow phenological progression (Subedi *et al*., 1998; Thakur *et al*., 2010; Zhao *et al*., 2022). Zhao *et al*. (2022) tested this experimentally in wheat, showing that cold stress during grain fill increased GPC compared with a control, but had a weaker effect than heat stress. Labuschagne *et al*. (2009) found similarly: flour protein increased under severe cold but this was cultivar-dependent, and the protein response to heat was stronger. However, others have found, despite lower CHO deposition in grains, no GPC difference between cold-treated and control wheat plots (X. Li *et al*., 2015). Cooling sufficient to damage plant tissues can vary spatially within fields because cold air pools in topographic depressions, effects relevant where this study was conducted.

2.2. Precision agriculture and on-combine monitors

To improve N efficiency and maximise returns on their cropping investments regarding GPC within seasons, dryland wheat growers can undertake site-specific nutrient management (SSNM) and / or strategic harvesting. In attempting to account for soil variability and variations in crop N demand, and to optimise GPC, SSNM has to account for both rainfall and phenology, and therefore has a temporal dimension. This fits within the '4R' principles of PA, which attempts to guide fertiliser application to the right time, place, type and amount (Gebbers and Adamchuk, 2010; Johnston and Bruulsema, 2014). To this end, RS-based GPC estimates could be used as a tool for targeting SSNM and/or for harvest planning. More comprehensive coverage of PA principles is beyond the scope of this thesis, but recent reviews cover numerous RS-relevant foci: thermal RS (Khanal *et al*., 2017); updated RS platforms, techniques, and applications, including big data and machine learning (ML; Sishodia *et al*., 2020); and Sentinel-2 (S2; Segarra *et al*., 2020).

Harvester-mounted devices capable of recording location-specific readings of yield and GPC have added to the data sources available for analysis but are also applied by growers pursuing

practical PA. Yield monitors have undergone substantial scientific research (Ping and Dobermann, 2005; Whelan and McBratney, 2003), including many studies integrating RS data (Campos *et al*., 2019; Dobermann and Ping, 2004; Stoy *et al*., 2022; Vallentin *et al*., 2020). Yield and GPC monitors are usually used in tandem; by combining their outputs, N harvest offtake can be mapped, encouraging engagement with N dynamics and loss pathways. Such mapping could be used to investigate N and moisture supply issues or identify underperforming areas that may be either remediable or more profitably used for other purposes (Whelan, 2019). However, while *post hoc* analyses improve knowledge of high- and low-performing areas of paddocks and farms and assists planning for PA in the following season, it cannot assist within seasons (Moffitt, 2020). Apart from Whelan *et al*. (2009) and Stoy *et al*. (2022), there is little reference to RS data used with on-combine GPC monitors in the peer-reviewed literature.

The monitor devices leveraged in the current studies temporarily remove grain samples from the combine's clean grain elevator approximately every 10 s (frequency of 0.1 Hz) during harvest operations and estimate GPC by NIR spectroscopy (720—1100 nm), with accurate geolocation by real-time kinetic GPS. The fidelity of similar units was good (R^2 = 0.86) when compared in field conditions against bench-top analysis equivalent to those used for grain pricing at receival (Long *et al*., 2005). Moreover, the devices are designed to satisfy statutory accuracy ± 0.4 % GPC mandated in Australia and the United States (Clancy and Heiken, n.d.).

2.3. Remote sensing data sources and retrieval methods

Remote sensing (RS) is a suite of methods used to obtain plant canopy information and investigate plant function from global to molecular scales, on the basis of radiation emitted from the surface under observation. For optical RS, this consists of the interactions of light with vegetated surfaces, which we observe as reflectance in the visible wavelengths, while

thermal RS relies on infrared emissions. It is well established that the biochemical and biophysical characteristics of the wheat crop canopy affect the associated reflectance spectra.

Nitrogen atoms are central to all amino acids and hence to the light harvesting protein complexes that contain C_{a+b}, C_{x+c}, and non-pigment polypeptides; thus, most of the N in leaves is associated with C_{a+b} and incorporated in such structures, themselves in the chloroplasts (Bassi *et al*., 1987; Yoder and Pettigrew-Crosby, 1995). Leaves, and collectively, canopies, are the site of most of the photosynthesis occurring in a wheat plant, and N and C_{a+b} are strongly and positively correlated at both leaf and canopy levels, (Evans, 1989), Plants with better access to N typically have higher concentrations of leaf N and C_{a+b}, and the reverse is also true (Feng *et al*., 2014; Xue *et al*., 2007; Zhao *et al*., 2005). Under severe N deficiency, chlorosis, the visible yellowing of leaves, becomes discernible; thus, the hue of reflected sunlight has long been used to assess N sufficiency in plants. However, less extreme N stress can also be detected by RS spectroscopy, many such methods have been developed and are surveyed below (§2.5.1). However, the strong N~C_{a+b} association has been shown to weaken at high N levels (Evans, 1983), many widely used N and/or C_{a+b} estimation methods also weaken at high N/C_{a+b} (Clevers and Kooistra, 2012; Flowers *et al*., 2003; Nguy-Robertson *et al*., 2012).

Through its centrality to proteins in leaf and grain, and its links to light harvesting and plant performance, then, N is a central theme of this work. However, because N atoms do not interact with light, they have no direct effect on reflectance and cannot be directly observed. As well as via C_{a+b} content, methods of N estimation therefore rely on other indicators of plant performance less directly related to N sufficiency, including biomass, light interception and the overall greenness of target areas. Nor is N the only, nor even the most influential plant trait in relation to wheat GPC; indeed, the relative importance of plant traits to GPC is dynamic across agronomic situations and seasons. As surveyed below, decades of research

have developed and refined methods of obtaining, through RS, insights into ever more complex aspects of plant performance, many of which are linked to GPC. These efforts take into consideration the contemporaneous advance of sensor technologies, from broadband through multispectral to hyperspectral.

Many VIs have been developed; their performance and specialisation have increased across decades of use and alongside increased sensor development; a class of VI, the narrow-band hyperspectral index (NBHI), has arisen to take advantage of ever-increasing sensor resolution. Many of these specialisations have been aimed at quantifying C_{a+b}, N or related vegetation traits, as surveyed in §0. However, VIs/NBHIs remain by definition empirical estimates of plant traits, in itself a deficit, and are often poorly adaptable across growing situations. Following, but now in parallel with VIs, the development of radiative transfer models (RTMs) has been based on mechanistic understandings of light~vegetation interactions: RTMs estimate the probabilistic fate of photons meeting a vegetated surface. Such models, often reported as more robust and flexible than indices (Jacquemoud and Baret, 1990; Wang *et al*., 2015), are also covered below (§2.3.5**)**, but it is useful to briefly encapsulate here their relevance to the current work. As described, N does not directly interact with incident radiation. However, many aspects of plant performance related to N- and water sufficiency, including LAI, dry matter, C_{a+b} and leaf water content, can be retrieved from multispectral data by RTM inversion (e.g. Baret *et al*., 2007; Delloye *et al*., 2018; Wang *et al*., 2022). Inversions have been instrumental in estimating specifically N-related parameters in higher plants (e.g. Wang *et al*., 2015; Zarco-Tejada *et al*., 2004) and in wheat (e.g. Camino *et al*., 2018; Raya-Sereno *et al*., 2022). Still more parameters can be inverted from HS reflectance, in particular given their potential relevance to GPC, the stress-associated auxiliary pigments Anth and C_{x+c}.

The current section presents retrieval methods relevant to the current PhD project without extended discussion of specific traits; these are treated in dedicated eponymous subsections in §2.4 and §2.5. Surveys of airborne HS and thermal, and Sentinel-2 (S2) multispectral RS, and their relative advantages and disadvantages are prioritised, as these are most relevant to the study. Broadband and active sensors, and retrieval methods relevant only to these, are avoided, as are references to proximal sensing, where sensors are hand-held or mounted on vehicles or towers. A focus on wheat, GPC, N, and plant stress is maintained. Estimation of biophysical variables from reflectance spectra generally takes one of four main approaches: 1) parametric regression; 2) nonparametric regression; 3) physically based, i.e., RTM inversion; and 4) hybrid methods in which elements of RTM inversion are combined with ML (Verrelst *et al*., 2015a). In this project, parametric regression was applied via vegetation indices (§0). Hybrid RTM inversion was also applied and is described in §2.3.5. Nonparametric methods were not used in this project for retrieval of plant biophysical traits, so are not further specifically considered.

2.3.1. Hyperspectral remote sensing

Specialised HS sensors typically record canopy radiance in the visual and near infrared (VNIR; 400 – 1000 nm) and near infrared (NIR; 800 – 1700 nm) domains. Spectral resolution < 10 nm is now typical for airborne HS instruments, and specialised nano-hyperspectral sensors have extended this into the sub-nanometre range (Belwalkar *et al*., 2022). This resolution provides very large amounts of information useful for assessing crop performance. In addition, HS data are often collected at very high spatial and radiometric resolution, while temporal resolution can be adjusted according to operational requirements or constraints. Piloted aircraft can carry multiple sensors concurrently and therefore allow deployment of HS and thermal sensors in equivalent light conditions. This allows efficient data collection over large extents, up to thousands of hectares on a single day as is demonstrated below (Longmire

et al., 2022). This extends the feasibility of airborne RS in large and spatially distributed study units, relevant to broadacre crops. Compared to piloted aircraft, unmanned aerial vehicle (UAV / drone) payloads and flight times are as yet insufficient for operations at such scales (e.g. Kanning *et al*., 2018). Such platforms may also be inherently unable to carry multiple instruments and are legally limited to operation within visual line-of-sight. While hand-held and vehicle-mounted HS instruments are also available and feature strongly in the literature, these are largely inappropriate to operations at commercial scale. Spaceborne HS sensors likewise played no role in this project.

2.3.2. Satellite multispectral remote sensing

While satellite platforms record data at lower spectral and spatial resolution than airborne sensors, for crop monitoring, they have an important advantage: repeated coverage of large areas. The studies below leverage data from the European Space Agency's Sentinel-2 (S2) satellites, whose frequent return times allow for timeseries observations, and whose images are cost-free. Like most satellites relevant to PA at within-field spatial scales, the S2 satellites are in polar orbits, meaning that they observe a swath that runs diagonally SW—NE across the earth's surface. A sun-synchronous orbit provides constant viewing conditions, and the paired platforms revisit any given location in the mid-latitudes at a delay of around < days. The multispectral instruments (MSI) aboard the S2 platforms capture a 290 km swath with 13 bands in the visible, red edge and near-infrared domains (Drusch *et al.*, 2012). This strong red edge focus, where three bands are recorded, is conceived for observing vegetation, particularly C_{a+b}, N content and LAI estimation (Frampton *et al.*, 2013; Herrmann *et al.*, 2011). The MSI measure reflectance in 13 spectral bands from visible and near-infrared (VNIR) to short-wave infrared (SWIR), with 12-bit radiometric resolution and a spatial resolution of 10—60 m (

Table 2.1). The spatial resolution, or ground sampling distance (GSD) is on the order of 10 m for the bands most relevant to vegetation, which also approximately coincides with that of GPC monitor data. For comparison, the contemporary iteration of Landsat, launched in 2013, offers nine bands in the visual, NIR and SWIR, with GSD = 30 m.

Table 2.1. Sentinel-2 band specifications.

Band number	Central wavelength (nm)	Band width (nm)	GSD (m)
1	443	20	60
2	490	65	10
3	560	35	10
4	665	30	10
5	705	15	20
6	740	15	20
7	783	20	20
8	842	115	10
8b	865	20	20
9	945	20	60
10	1380	30	60
11	1610	90	20
12	2190	180	20

The frequent S2 overpasses and its collection of multispectral data offer opportunities to monitor intra- inter-seasonal variation in leaf and canopy traits related to GPC, without the expense of piloted aircraft. However, the timing of satellite overpasses is inflexible and cannot account for cloud cover; while a substantial proportion of days are expected to be cloudy across a winter wheat growing season in southern Australia, in practice sufficient images were obtained to form adequate timeseries.

2.3.3. Thermal remote sensing

Thermal RS imaging, or infrared thermography, utilises passively recorded emissions in the thermal infrared domain (8—15 μm), exploiting the fact that all non-black body objects with temperature > 0 K emit radiation in this range. Airborne or satellite thermal RS may complement optical or other forms of RS. The thermal properties of plant leaves depend to a

large extent on their water content and capacity to release water to the atmosphere; when water is available, evapotranspiration cools plant surfaces and when it is not, plants accumulate heat. Stomatal closure is initiated when plant water potential Ψ reaches a critical level. This controls water loss but also restricts CO_2 diffusion into leaves, reducing the CO_2 concentration in leaf mesophyll and hence the substrate available for photosynthesis; it is one of plants' principal responses to low humidity, droughting and high temperature among other stimuli (Farquhar and Sharkey, 1982; von Caemmerer *et al*., 2004). Hence low stomatal conductance is associated with both increased leaf temperature and reduced photosynthetic rate; in dryland agriculture, these are often induced by water stress.

The physiological effects of heat accumulation, as relevant to GPC, were discussed in §2.1.3, while methods of assessing water stress from RS are detailed in §2.6. Thermal data can detect or assist in the detection of water deficits and are applied for irrigation scheduling, assessing PAWC variability, soil salinity and other osmotic constraints, phenotyping, pathogen detection and diagnosis of infestation or disease severity. Combined with hyperspectral indices, thermal information improved estimation of wheat N status (Pancorbo *et al*., 2021). Khanal *et al*. (2017) review the use of thermal RS in PA.

2.3.4. Reflectance indices

Vegetation indices (VIs) combine reflectance at two or more bands by arithmetical operations, transformation, or both, to enhance the influence of a desired plant trait, diminish the influence other plant traits, unwanted components such as soil, shadow, observational or atmospheric effects, or combinations of these (Fang and Liang, 2014, p. 2143). In general, VIs offer computational simplicity and efficiency, and have been refined over decades (Berger *et al*., 2018). The normalised difference VI (NDVI; (Rouse *et al.*, 1974)), a structural index sensitive to green cover, remains common in agriculture (Herrmann *et al.*, 2010), continues in use (e.g., Stoy *et al*., 2022; C.-W. Tan *et al*., 2020), and its basic form has been adopted by many other indices. NDVI correlates with wheat yield, especially at coarse spatial scales (e.g. Becker-Reshef *et al.*, 2010; Dempewolf *et al.*, 2014; Lai *et al.*, 2018; Toscano *et al.*, 2019), although correlations with yield are growth stage-dependent and reduce after anthesis (Magney *et al*., 2016a). NDVI and its variants also saturate at high biomass, which can blunt estimation of even fundamental measures of crop performance; for example, the green NDVI, (gNDVI; Gitelson *et al*. (1996)) can saturate as low as LAI ≤ 2.5 (Nguy-Robertson *et al*., 2014). As they do not explicitly account for observer viewing angle, solar zenith and azimuth angles, soil reflectance and other components, VIs in effect discard some of the information available in spectroscopic data (Bacour *et al*., 2006). This limits their transferability between locations, times and agronomic situations (Baret and Guyot, 1991; Eitel *et al*., 2008; Zarco-Tejada *et al*., 2005). Moreover, VIs provide only an empirical estimate of plant traits, as they are unitless, they can lack interpretability.

Sensor development from broadband through multi- and hyperspectral sensors has allowed the construction of many new indices with better sensitivity and specificity to an expanding range of detectable plant traits. Indices now account better for confounding from observational artifacts, overlapping absorbance among plant pigments, differing water

content, or differences in canopy structure. There has been much progress, and many VIs developed, since the NDVI. Broadband indices and early development were reviewed by Bannari *et al*. (1995), and later broadband and early hyperspectral progress by Dorigo *et al*. (2007). Hyperspectral VIs were evaluated by Li *et al*. (2010) with specific reference to wheat N status, and NBHIs were covered more recently by Camino *et al*. (2018) and Liang *et al*. (2018). All vegetation indices considered in the current thesis are detailed in Table A1 andTable A2, in Appendix A.

2.3.5. Radiative transfer models and their inversion

Radiative transfer models (RTMs) approximate the behaviour of photons incident on vegetation, modelling the complex physical interaction of solar radiation with leaves and canopies. They generate synthetic reflectance spectra based on the combined optical properties of leaf and canopy biochemical and biophysical variables, and observational factors. As detailed below, these variables include plant traits C_{a+b} and other pigment concentrations, leaf structural and moisture components, and LAI. However, environmental factors including observer- and sun angles and soil reflectance are also specifically enumerated in RTMs, meaning that their effect on plant trait estimation is considered. This specific inclusion of components which are either not captured or are discarded by empirical models makes physically-based RTMs more transferable across time, location, crop types, phenological stages and agronomic situations than the empirical models (Clevers and Kooistra, 2012; Dorigo *et al.*, 2007; Jacquemoud *et al.*, 1995). In further contrasts to empirical methods, field observations are used to validate RTM outputs, rather than for establishing relationships, and untransformed spectral data from the observing sensor are used to directly estimate plant parameters in their biologically relevant units. Inversion of RTMs allows accurate estimation of both structural and physiological traits at leaf and canopy levels from observed spectra (Bacour *et al*., 2002; Féret *et al*., 2008; Jacquemoud *et*

al., 2009, 1995; Z. Li *et al*., 2015). Concurrent retrieval of multiple traits is common (Camino *et al*., 2018; Poblete *et al*., 2020a; Zarco-Tejada *et al*., 2018). However, Schiefer *et al*. (2021) warn of that phenology-related effects on RTM retrievals; although they did not include wheat in their work, they caution that LAI and C_w are best estimated early in seasons, and C_{a+b}/C_{x+c} are best later.

Many RTMs have been developed since their initial inception in the 1960s (Allen *et al*., 1969; Duncan *et al*., 1967). This project employs the coupled models PROSPECT and SAIL, often referred to as PROSAIL, which are briefly described immediately below and dominate references to RTMs in this literature review. The coupled use of these models was comprehensively reviewed by Jacquemoud *et al*. (2009) and more recently, with a focus on their use with hyperspectral data, by Berger *et al*. (2018), who also provide a comprehensive overview of each model's input parameters. Verrelst *et al*. (2019) also thoroughly cover inversion and look-up table (LUT) methods.

The earliest version of the PROSPECT model of leaf optical properties accounted for reflectance and transmittance, between 400—2500 nm, as a function of three properties: absorption of light by C_{a+b} and water content (equivalent water thickness; C_w) as well as light scattering by refraction and leaf mesophyll structure (N), characterised as the number of tissue/air interfaces in leaf mesophyll (Jacquemoud and Baret 1990). Later incarnations included more leaf traits: PROSPECT 4-5 (Féret *et al*., 2008) added carotenoid pigments (C_{x+c}) and dry matter content (C_m), improving C_{a+b} estimation. The PROSPECT-D version (Féret *et al*., 2017a) improved the representation of seasonal dynamics by estimating anthocyanins (Anth). This improved pigment discrimination allows quantification of these auxiliary pigments and improves the retrieval of other parameters. The latest addition to the family, PROSPECT-PRO includes proteinaceous components and therefore improves retrieval of leaf biomass per unit area (Féret *et al*., 2021).

The Scattering by Arbitrarily Inclined Leaves (SAIL) canopy-level RTM is a representation of canopy reflectance as a function of canopy structural traits LAI and leaf inclination distribution function (LIDFa) to model light interception and scattering (Verhoef, 1984). SAIL has since been modified to include a hotspot parameter (Kuusk, 1991) and, as 4SAIL, extended into the thermal domain (Verhoef *et al*., 2007) and has inspired several other models (Jacquemoud *et al*., 2009). The contributions to reflectance of viewing geometry – azimuth, solar and observer zenith angles – are considered in SAIL, as are the proportion of diffuse radiation and soil reflectance. In the commonly used PROSAIL, SAIL adopts leaf reflectance and transmittance from the PROSPECT model; by linking the models, biochemical and biophysical characteristics at leaf and canopy levels are estimated concurrently. This reduces the number of unparametrised variables, facilitating model inversion, and improves retrievals, as the spectral signature of each component is accounted for in estimation of the others (Blackburn, 2007; Jacquemoud *et al*., 2009). PROSAIL is well suited to the relatively homogeneous canopies of agricultural crops, but is also applied in horticultural and forest contexts (Jacquemoud *et al*. 2009).

Although the PROSAIL models are used in forward mode, for example to identify and optimise reflectance bands for VI development (Cheng *et al*., 2014; Haboudane *et al*., 2002), most users invert RTMs in order to quantify plant traits listed among their input parameters (Berger *et al*., 2018; Jacquemoud *et al*., 2009). Since early demonstrations (e.g. Baret *et al*. (1992)), inversions have been widely implemented and good retrieval accuracies shown for inversions of PROSPECT, SAIL and their variants (Bacour *et al*., 2002; Féret *et al*., 2008; Jacquemoud *et al*., 1995; Z. Li *et al*., 2015; Ustin *et al*., 2009; Zarco-Tejada *et al*., 2021, 2018).

In RTM inversions, the models are run for permutations of vegetation attributes sampled from their biologically feasible ranges, while artefactual parameters such as azimuth, solar

angle, viewing angle and other physical characteristics at the time of image acquisition can be held at their measured value or similarly sampled. Thus, a database, or look-up table (LUT), typically of 100,000—200,000 simulations, is constructed wherein synthetic spectra are stored in association with the attributes used to generate them. In this way single or multiple plant attributes are commonly inverted from observed reflectance (Camino *et al*., 2018; Locherer *et al*., 2015). Observed reflectance spectra are then matched against the synthetically generated spectra, and the underlying variables of the best-fitting synthetic spectrum are assumed to represent the relevant leaf/canopy characteristics for the real canopy (Locherer *et al*., 2015).

Model inversion problems are ill-posed, meaning solutions are not unique and uncertainty in measured or modelled quantities can change the output (Combal *et al*., 2003; Verrelst *et al*., 2019). In RTM inversion, this means that the same or very similar spectra can be generated by different plausible combinations of biophysical parameters (Combal *et al*., 2003; Dorigo *et al*., 2007). Nevertheless, there are ways to limit the effects of the ill-posed problem. Prior knowledge of agronomic parameters or measured quantities can be used to reduce dimensionality, as widely described and implemented (e.g., Combal *et al*., 2003; Z. Li *et al*., 2015). Indeed, this approach is usual and is an extension of the logic that all canopy biophysical variables to be estimated should be constrained to their biologically realistic ranges. To further reduce the ill-posed problem, Wang *et al*. (2015) and Camino *et al*. (2018) each conducted stepwise inversions whereby specific input parameters were derived using the spectral domain in which they most affect light dynamics, and ordered by the effects' size. Each subsequent parameter was retrieved using as inputs the values of those established previously; such stepwise procedures can be better than inversions with concurrent estimation of several parameters (Camino *et al*., 2018). As detailed below, in sections named for each

parameter, PROSAIL inversion has often retrieved structural and physiological variables in wheat.

Hybrid inversion techniques have been shown effective (Verrelst *et al*., 2015a) and, due to shorter processing times, able to handle large volumes of data (Gómez-Dans *et al*., 2016). This can become important when multiple traits are retrieved, pixel-wise and/or timeseries retrievals are required or when these are combined, as is done in the current work. Several ML algorithms have been assessed, and performed well; the general approach is that RTM output spectra are used – as ML inputs – to train an algorithm, which is implemented to compare these with observed spectra; many such retrievals have been done (Verrelst *et al*., 2015a). In the component studies of this PhD project (Longmire *et al*., 2022), LUTs are interrogated by support vector machine (SVM), after the procedures of Poblete *et al*. (2021).

2.4. Retrieval of GPC-linked structural traits

Leaf area index (LAI) is the ratio of total upward-facing green leaf to ground surface area, so indicates the number of leaf layers between soil and sky, as $m^2.m^{-2}$. It is therefore a purely structural parameter that indicates canopy development and biomass and represents the surface available for sunlight capture and the exchange of mass and energy with the atmosphere. LAI is closely related to photosynthesis (Liu *et al*., 2006) and evapotranspiration (Nearing *et al*., 2012) and often linearly associated with N demand (Lemaire *et al*., 2008). As such, LAI is closely related to yield and, less directly, to GPC in wheat (Kanning *et al*., 2018; Liu *et al*., 2006). Through its correlation with biomass, LAI is a reliable indicator of crop CHO and protein sink sizes. For wheat, LAI as a function of days after sowing (DAS) generally follows an approximately symmetric curve, peaking before anthesis with maximum of 2—6 (Baret and Guyot, 1991; Waldner *et al*., 2019; Zhang *et al*., 2007) but may be as high as 8 (C.-W. Tan *et al*., 2020).

Despite the ubiquity of LAI in vegetation remote sensing, it is useful to define similar parameters. Green LAI is a refinement of LAI, measuring the leaf area involved in photosynthesis (Daughtry *et al*., 1992). The fraction of absorbed photosynthetically active radiation (fAPAR) quantifies the proportion of incident sunlight (400—700 nm) absorbed by a canopy, while fractional cover (fCover) gives the fraction of ground surface obscured by plant matter in the sensor's nadir view; as such, fCover saturates relatively early in seasons so is useful mainly for assessments of early vigour (Bacour *et al*. 2006). Like LAI, measures of leaf angle and orientation, such as leaf inclination distribution function (LIDFa) affect light distribution and use efficiency, affecting photosynthesis and reflectance.

The similar effects of LAI and C_{a+b} content on canopy reflectance complicated their separation in index-based retrieval from RS images, until the sensitivity of C_{a+b} retrieval to LAI (Haboudane *et al*., 2002), and vice-versa (Haboudane *et al*., 2004) were reduced, improving the estimation of each from hyperspectral data, including in wheat. LAI is highly correlated with leaf-level C_{a+b} and N, and can be used to estimate them at canopy level (Houborg and Boegh, 2008); the three quantities can also be retrieved together, for example, in wheat, by Kanning *et al*. (2018) and Li *et al*. (2015). Using VIs and nonparametric regression to estimate LAI from S2 data, predictors which include RE components have been shown best, although many VIs have been shown as broadly similar in accuracy, and few have maintained this accuracy along seasons (Herrmann *et al*., 2011; Kamenova and Dimitrov, 2021).

Wheat LAI retrievals by RTM inversion have also obtained moderate results from airborne hyperspectral (Z. Li *et al*., 2015; Richter *et al*., 2011) and good results from S2 data (Upreti *et al*., 2019), although saturation at high LAI values may remain problematic (Lunagaria and Patel, 2019; Weiss and Baret, 1999). PROSAIL LAI retrieval within commercial fields was recently demonstrated across diverse wheat cropping operations in Italy and China, with

strong correlation (R^2 = 0.78) and low error (RMSE ≈ 0.69) as validated by ground measurements propagating into good estimations of CCC also (Upreti *et al*., 2019). This work, which used a range of ML-assisted hybrid retrieval approaches to estimate crop traits from S2 images, reported that a least-squares linear regression also overcame the tendency of LAI estimates to saturate at high values. Likewise, Jiao *et al*. (2022) used prior knowledge of average leaf angle to improve LAI and C_{a+b} retrieval by PROSAIL-D, and showed moderately good CCC estimations. Pan *et al*. (2019) also retrieved LAI and C_w from S2 by PROSAIL and were able to show LAI progression and later decline through a season, also with low error, in commercial wheat fields although with a relatively low sampling density.

2.5. Retrieval of GPC-linked physiological traits

Chlorophylls *a* and *b* (C_{a+b}) influence GPC because they are both a substantial proportion of plant N stores, and an indication of photosynthetic capacity. However, pigments other than C_{a+b} also offer potential as diagnostic of GPC, given their links with stress; anthocyanins (Anth) and carotenoids C_{x+c} are the accessory pigments considered below. Solar induced fluorescence (SIF) is an indicator of photosynthetic rate, hence both early-season plant vigour and later protein dilution, and indirectly of C_{a+b} concentration and abiotic stress. The subsections immediately below summarise retrievals of these quantities from RS through both VIs and RTM inversions.

2.5.1. Chlorophyll and N content

Concentrations of C_{a+b} and N in leaves are closely correlated (Evans, 1983) because N atoms make up ~ 6% by mass of C_{a+b}**,** and most leaf N is contained in C_{a+b} molecules (Yoder and Pettigrew-Crosby, 1995); moreover, a large majority of plant N is contained in chloroplasts. Canopy N is strongly correlated with GPC (Feng *et al*., 2014; Xue *et al*., 2007; Zhao *et al*., 2005). VIs, many designed specifically for C_{a+b} or N content retrieval, have disentangled

canopy biochemistry from structural factors. Chlorophyll *a* absorbance peaks at 430 and 662 nm and chlorophyll *b* at 453 and 642 nm. The red edge (RE), a domain of distinct reflectance increase where pigments' influence on canopy reflectance weakens and that of cell walls, mesophyll and water strengthens (Hatfield *et al*., 2008), and therefore gives insight into both pigments and structure. The RE lies between 680 and 750 nm, in the NIR, and shows a close relationship to C_{a+b} concentration as well as LAI and water status (Filella and Peñuelas, 1994; Horler *et al*., 1983). Because it involves a relatively narrow range, RE analysis required the development of narrow-band sensors; thenceforth many VIs exploit the links between the RE and C_{a+b} to retrieve this or N at leaf or canopy level (Clevers and Gitelson, 2013; Haboudane *et al*., 2008; Prey and Schmidhalter, 2019). For example, the normalised difference red edge index (NDRE) was devised to monitor and separate N from water status effects, by comparing a RE band (R_{720}) with a reference in the NIR (R_{790}; Barnes *et al*., 2000), and was used to retrieve C_{a+b} in wheat (Z. Li *et al*., 2015). Also relying on the RE, the Modified Chlorophyll Absorption in Reflectance Index (MCARI; Daughtry *et al*., 2000) comprises C_{a+b} absorption at 670 nm with bands at 550 nm 700 nm, and is sensitive to C_{a+b}. Combined with the second Modified Triangular Vegetation Index as MCARI/MTVI2, this approach has retrieved wheat canopy N from proximal sensing, robust to LAI variations, in the original work (Eitel *et al*., 2007) and subsequently (Eitel *et al*., 2008; Z. Li *et al*., 2018), and from RS (Li *et al*., 2019). Other NBHIs of relevance below, and which leverage C_{a+b}~RE relationship, are the Zarco-Tejada and Miller index (ZMI; R_{750}/R_{710}), Vogelmann index 1 (R_{740}/R_{720}) and the RE chlorophyll index (CI_{re}; (R_{750}/R_{700})-1) (Gitelson et al., 2003; Vogelmann et al., 1993; P.J. Zarco-Tejada et al., 2001). By contrast, the Normalized Pigment Chlorophyll ratio Index (NPCI; Peñuelas *et al*., 1994), calculated as $(R_{680} - R_{430}) / (R_{680} + R_{430})$, was conceived to describe the ratio of C_{x+c} to C_{a+b} and is now categorised among C_{a+b} indices.

Other approaches to C_{a+b} estimation combine multiple VIs: The Transformed Chlorophyll Absorption in Reflectance Index (TCARI; Haboudane *et al.*, 2002) and Optimized Soil-Adjusted Vegetation Index (OSAVI; Rondeaux *et al.*, 1996), combined as TCARI/OSAVI are sensitive to C_{a+b} despite LAI variability, soil background and shadow effects (Haboudane *et al*. 2002). Further, the canopy chlorophyll content index (CCCI; Barnes *et al*. 2000), derived by plotting NDRE as a function of NDVI, was used to detect wheat N deficiency (Rodriguez *et al*., 2006), then adapted to estimate canopy N under field conditions in the Wimmera (Cammarano *et al*., 2011); each of these studies found good results before Z35, emphasising their potential for practical implementation. In a further example, of 14 standard structural and chlorophyll VIs recently tested in Australian commercial wheat fields, the best yield predictions were obtained from seasonal peaks of CI_{re} and NDRE, while NDVI and OSAVI were the best structural VIs (Zhao *et al.*, 2020). TCARI/OSAVI have been widely applied in wheat, estimating, for example, C_{a+b} (Gonzalez-Dugo *et al.*, 2015) and total N (Klem *et al.*, 2018).

Among those described here and other indices, Chen *et al*. (2010) found that VIs incorporating hyperspectral reflectance at three or more wavelengths, including their own double-peak canopy N index (DCNI) were generally better than those relying on two bands, a conclusion supported by Wang *et al*. (2012). Li *et al.* (2014) also reported that no two-band VI offered transferability between locations, phenological stages and soil conditions for wheat canopy N estimation, attributing this to confounding by canopy structural components. This confounding, likely arising from saturation at high biomass was, however, not problematic for more complex VIs.

Reliable retrieval of wheat C_{a+b} by RTM inversion is well demonstrated. From airborne RS in experimental plots, Camino *et al*. (2018) found PROSAIL C_{a+b} was strongly correlated with leaf-level C_{a+b} from spectroscopy (R^2 = 0.81), canopy N from destructive sampling (R^2 =

0.71), and with CHO assimilation (R^2 = 0.59). In commercial wheat crops, both Delloye *et al*. (2018) and Upreti *et al*. (2019) recently retrieved leaf and canopy C_{a+b} content (CCC; LAI (or GAI) x C_{a+b}; Clevers *et al.*, 2017; Houlès *et al.*, 2007) from S2 data by PROSAIL inversion. Delloye *et al*. (2018) argued that the relatively low accuracy of their leaf C_{a+b} estimate related to the vertical pigment distribution gradient in the plant, but showed a substantially more accurate CCC estimate, particularly when this was based solely on the S2 RE bands. Upreti *et al*. (2019) also had better results for at canopy than leaf level; each of these studies had R^2 for wheat CCC in the range 0.6—0.7, as did Jiao *et al*. (2022), also from S2 and PROSAIL. Such differentials in estimation accuracy may arise from inhibited propagation of the reflectance signal from leaf to canopy scale (Asner, 1998), but such an effect is not always evident: Sehgal *et al*. (2016) achieved lower error, assessed by nRMSE, for C_{a+b} than for CCC from multispectral satellite RS by PROSAIL. Despite differences in absolute fidelity as described, PROSAIL inversion is hence sensitive to C_{a+b} variability across wheat canopies in both experimental and commercial contexts, so offers potential as a predictor of GPC.

2.5.2. Carotenoids

The carotenoids (C_{x+c}), comprising carotenes and xanthophylls, are substantial contributors to the red-yellows of nature (Cazzonelli and Pogson, 2010); these pigments are an important aspect of grain quality, especially in durum wheat (Ficco *et al*., 2014). The C_{x+c} are auxiliary photosynthetic pigments, extending plants' ability to absorb energy across the EM spectrum; β-carotene has absorbance peaks at 451 and 470 nm. By dissipating excess radiation and scavenging ROS, they moderate damage from heat and photooxidative stress (Demmig-Adams and Adams, 2006; Strzałka *et al*., 2003); moreover, compounds produced in the ensuing reversible oxidation of C_{x+c} to xanthophylls promote stress-adaptive responses (Havaux, 2014). Carotenoids are upregulated under photooxidative, water, frost and heat stress in wheat (Borrelli *et al*., 2011; Fratianni *et al*., 2013; Groth *et al*., 2020; Janeczko *et al*.,

2018), and in response to higher N availability (Shah *et al*., 2017). Increases in the ratio of C_{x+c} to C_{a+b} are related to reduced photosynthetic activity (Fréchette *et al*., 2016; Sonobe *et al*., 2018).

Although the overlap between C_{x+c}, and C_{a+b} absorbance made C_{x+c} estimation challenging, many index-based methods have been developed since the advent of hyperspectral sensing (Blackburn, 1998; Chappelle *et al*., 1992; Féret *et al*., 2011; Hernández-Clemente *et al*., 2012; Ustin *et al*., 2009). Of the NBHIs developed for C_{x+c} estimation, most are based on reflectance in the green at 470–530 nm (Blackburn, 1998; Chappelle *et al*., 1992; Fassnacht *et al*., 2015; Gitelson *et al*., 2002b, 2006; Hernández-Clemente *et al*., 2012; Sonobe and Wang, 2018). In addition, NBHI have been developed to assess the epoxidation state of xanthophylls, including the photochemical reflectance index (PRI; Gamon *et al*., 1992), several modifications of the PRI which better account for canopy structure, stomatal conductance and other factors (Hernández-Clemente *et al*., 2011). The changing ratio of C_{x+c} to C_{a+b} is the basis of the carotenoid/chlorophyll index (CCI), which can be used to track photosynthesis in evergreen plants (Gamon *et al*., 2016), and the carotenoid/chlorophyll ratio index (CCRI), calculated as their own carotenoid index (CARI; Zhou *et al*., 2017) divided by CI_{re}. The CCRI has delivered moderately good estimates of the C_{x+c} : C_{a+b} in wheat canopies (Zhou *et al*., 2019). Also sensitive to C_{x+c}, the photochemical reflectance index (PRI; (Gamon *et al.*, 1992)) indicates instantaneous changes in water stress and photosynthetic rate (Feng *et al*., 2017; Magney *et al*., 2016b). The PRI is related to the C_{x+c}/C_{a+b} ratio, and hence to photosynthetic efficiency (Filella *et al*., 2009; Sims and Gamon, 2002), and has been shown to indicate wheat plant performance related to WS and N supply (Magney *et al*., 2016b). Characterised as the chlorophyll/carotenoid index (CCI), this ratio indicates photosynthetic activity (Gamon *et al*., 2016); PRI and CCI have been used to track gross primary productivity (GPP), and hence phenology, in evergreen forests (Wong *et al*., 2022). The

hyperspectral carotenoid index (CAR; Zarco-Tejada *et al*., 2013b) has shown moderate, significant relationships with yield in rainfed and irrigated bread and durum wheat, as did the PRI (Gonzalez-Dugo *et al*., 2015).

Féret *et al*. (2008) specifically included C_{x+c} as an input to PROSPECT-5, meaning they can also be estimated by inversion of this and later PROSPECT variants. However while RTM inversions have been conducted specifically for C_{x+c} retrieval in wine grapes (*Vitis vinifera* L.; (Zarco-Tejada *et al*., 2013b), tea (*Camellia sinensis* (L.) Kuntze; Sonobe *et al*., 2018) and savanna woodlands (Miraglio *et al*., 2020), so far as the current thesis could establish, they have not in wheat. Findings that leaf C_{x+c} concentrations correlate inversely with photosynthesis, and that they respond to stress in wheat, allow the hypothesis of their association with GPC via reduced assimilation. This hypothesis is supported by findings of positive correlations between genes associated with C_{x+c} production and GPC in durum wheat (Colasuonno *et al*., 2019). Nevertheless, publications linking C_{x+c} or PRI in the living wheat plant with GPC appear absent.

2.5.3. Anthocyanins

Anthocyanins (Anth) are water-soluble pink, purple, and red non-photosynthetic cytoplasmic pigments with many roles in the leaf and broader plant structures; they number over 500 (Féret *et al*., 2017a). Showing osmoregulatory function, they are involved in protection against water stress (Chalker-Scott, 1999; Naing and Kim, 2021; Shoeva *et al*., 2017), and by scavenging free radicals and reactive oxygen species (ROS), they reduce photooxidation under excessive light (Gould, 2004). They play defensive roles against biotic stresses including pathogen attack and herbivory (Gould, 2004) and contribute to tolerance of both cold and heat in developing wheat plants (Calderon Flores *et al*., 2021). Anth accumulate in stressed plants (Chalker-Scott, 2002, 1999; Naing and Kim, 2021) and senescing leaves

(Sims and Gamon, 2002), and water stress in post-anthesis growth phases increases their concentration in wheat grains (X. Li *et al*., 2018). In rice (*Oryza sativa* L.), increased leaf, but not grain Anth concentration has also been observed in response to increased N availability (Yamuangmorn *et al*., 2018).

As they convey useful information about plants' physiological status, Anth have been targeted for RS stress detection; due to their strong absorbance in the green, retrieval methods have commonly included spectral components around 550 nm (Féret et al., 2017a; A. A. Gitelson et al., 2001; Gitelson et al., 2006, 2009). While their overlapping absorption features made difficult the separation of Anth, C_{a+b} and C_{x+c} with index-based methods, particularly at canopy scale, this has improved with increasing spectral resolution of sensors (Gitelson *et al*., 2006; Ustin *et al*., 2009) and more recently with the development of the PROSPECT-D RTM, conceived particularly to follow the evolution of foliar pigments across plant phenological advance (Féret *et al*., 2017a). In ML-based studies, and concordant with Gitelson *et al*. (2006) for maize (*Zea mays* L.) and soy (*Glycine max* (L.) Merr.) crops, Shah *et al*. (2019) and Odilbekov *et al*. (2018) each found that Anth concentration was strongly predictive of wheat leaf C_{a+b}; like C_{x+c} and PRI, though, investigations of links between Anth and GPC are lacking from the literature.

2.5.4. SIF

Solar induced fluorescence (SIF) is emitted from the photosystems at 1—5% of incident radiation and in proportion to instantaneous photosynthetic rate; despite its relatively small flux from the canopy, its origin within the photosynthetic apparatus and close links to physiological processes make SIF highly relevant to remote sensing (Damm *et al*., 2014; Verrelst *et al*., 2015b). SIF is dependent on leaf C_{a+b} concentration (Houborg *et al*., 2013; Walker *et al*., 2014), indicates plants' general functional status (Meroni *et al*., 2009) and is a

proxy for carbon assimilation (Genty *et al.*, 1989; Mohammed *et al.*, 2019; Porcar-Castell *et al.*, 2014).

A broadband emission from 650—800 nm, SIF has maxima at 690 nm and, more strongly, at 740 nm (Meroni *et al*., 2009; Verrelst *et al*., 2015b). Because of absorption by atmospheric O_2 molecules, solar radiation at ground level is attenuated by up to 90 %, compared with top-of-atmosphere (TOA) irradiance, in the telluric O_2 absorption (O_2-A) bands (Meroni *et al*., 2009). This allows passive SIF quantification by Fraunhofer line depth (FLD) methods, usually focused on the 740 nm peak, which overlaps the O_2-A line centred at 760.4 nm (Verrelst *et al*., 2015b; Zarco-Tejada *et al*., 2016, 2013a). The contribution of SIF to total radiance is estimated by comparing irradiance (E) and radiance (L) measured within the Fraunhofer line with E and L measured at a close wavelength or wavelengths outside the line (Plascyk, 1975; Plascyk and Gabriel, 1975).

In agriculture, spaceborne SIF has been shown to improve crop productivity monitoring on regional to global scales (Guan *et al*., 2016; Guanter *et al*., 2014; Miao *et al*., 2018; Y. Zhang *et al*., 2018), and yield estimation in maize and soy (Peng *et al.*, 2020). By detecting the depression of SIF emissions associated with accumulated heat stress, Song *et al*. (2020) improved regional scale wheat yield prediction, demonstrating also the usefulness of SIF in heat stress monitoring. However, SIF is also less effective as a yield predictor during heatwaves (Wohlfahrt *et al.*, 2018); this is likely attributable either to lower net assimilation when respiratory loads are high (Heskel *et al.*, 2016), heat damage to reproductive organs, undetected by RS (Sloat *et al.*, 2021), or a combination of these. However, spaceborne SIF products have had mixed outcomes in crop assessments. For example, SCIAMACHY/GOME-2 SIF products were no better than the structural enhanced VI (EVI) in estimates of Australian wheat yield at regional scale, due to the SIF products' low spatiotemporal resolution and relatively poor signal-to-noise ratio (SNR; Cai *et al.*, 2019).

Similarly, SIF from the higher-resolution OCO-2 and TROPOMI sensors was inferior to both VI-based and NIR reflectance methods for maize and soy yield estimation (Peng *et al.*, 2020). Although recent sensors have improved results at regional scales (Sloat *et al.*, 2021), access to spaceborne SIF at higher spatial resolution resides in the future, and the benefits of using it for predicting harvest variables remain unclear (Peng *et al.*, 2020). On smaller spatial scales, by contrast, airborne SIF has contributed to WS, nutrient deficiency and biotic stress diagnoses (Poblete *et al*., 2020a; Zarco-Tejada *et al*., 2018), including in wheat (Camino *et al*., 2019, 2018). In wheat, airborne SIF strongly improved leaf N estimation over C_{a+b} alone, providing far more information than C_m or C_w (Camino *et al*., 2018), and was sensitive to small differences in soil N sufficiency and C_{a+b} (Raya-Sereno *et al*., 2022). From proximal sensing, within-field SIF variability has been associated with GPC, albeit relatively weakly (Song *et al*., 2017); SIF~GPC correlations were negative, and at their weakest, until heading, then consistently positive.

Based on the principles laid out above, SIF could influence GPC through various pathways. Spatial variability in SIF should reflect the relative availability of new assimilates, at least during grain filling and under WS, as follows: low SIF > low dilution > higher GPC. Absent abiotic stress, high SIF may imply more N / protein for translocation via the relationship of SIF with C_{a+b}; this may be more subtle. Apart from work associated with this thesis (Longmire *et al*. 2022), that of Song *et al*. (2017) appears to be the only research into the SIF~GPC relationship within fields.

2.6. Retrieval of water stress indicators

Under water stress, stomatal gas exchange is reduced to limit leaf water loss. Low stomatal conductance in turn reduces evaporative cooling of the leaf surface, allowing heat to accumulate, while the photosynthetic apparatus is starved of atmospheric CO_2, constraining photosynthesis. This means that thermal data can be used to detect WS (Grant *et al*., 2007;

Idso, 1982), and because of the effect of stress on assimilation, assist GPC estimation (Zhao *et al*., 2005).

2.6.1. Water stress indicators – thermal

In the crop water stress index (CWSI), which estimates instantaneous evapotranspiration against its minimum and maximum potential rates, the difference between canopy temperature (T_c) and air temperature (T_a) is normalised against the vapour pressure deficit multiplied by a crop-specific factor which represent a non-water stressed baseline. As they are inversely proportional to stomatal conductance, CWSI and the normalised relative canopy temperature (NRCT) can be used to diagnose sub-optimal performance (Gonzalez-Dugo *et al*., 2015; Idso, 1982; Jackson *et al*., 1981), although both require knowledge of T_c. Added to multivariate regressions, CWSI has consistently improved wheat yield estimates (Gonzalez-Dugo *et al*., 2015; Klem *et al*., 2018); the fate of N in the wheat shoot, and whether it is translocated to the grain as protein, has also been linked to CWSI (Klem *et al*., 2018). However, thermal data are not always available; nor is it possible to accurately derive CWSI in canopies with substantially incomplete canopy closure, because of the often large difference between the thermal emissions of foliage and bare soil, which can exaggerate T_c in mixed vegetation/soil pixels (Jackson *et al*., 1981).

2.6.2. Water stress indicators – non-thermal

Some limitations of CWSI were alleviated by Moran et. al (1994), whose water deficit index (WDI) does not require measurement of T_c and instead uses a VI/temperature trapezoid in which fractional vegetation cover is estimated with the soil-adjusted VI (SAVI; Huete, 1988). Included with the CCCI (Fitzgerald *et al*., 2010) in multivariate regressions, the WDI has been used to overcome confounding of N- and water stress in wheat (Pancorbo *et al*., 2021), but it was also shown to correlate less well with yield than CWSI (Gonzalez-Dugo *et al*.,

2015). The normalised difference water index (NDWI; Gao, 1996), which estimates canopy water content based on NIR reflectance, consistently improved wheat leaf C_{a+b} estimation (Shah *et al*., 2019); Wolanin *et al*. (2020) found similarly for several ML models predicting wheat yield. Nevertheless, the NDWI has also shown only a moderate, though significant, relationship with wheat yield and very weak correlation with GPC (Liu *et al*., 2006). Yield estimates have also been improved by including CWSI alternatives, measures of relative canopy temperature which do not require measurement of T_{amb} (Elsayed *et al*., 2017). By including an evapotranspiration index modelled from met statistics (Fitzpatrick and Nix, 1969), Zhao *et al*. (2020) substantially improved their field-level predictions; this method of incorporating WS improved on the use of Landsat thermal data at far higher spatial resolution.

The PRI, discussed above as an estimator of C_{x+c}, also indicates pre-visual water stress and recovery therefrom, hence and is proposed as an alternative to thermal RS for water stress detection (Kohzuma *et al*., 2021; Suárez *et al*., 2008; Thenot *et al*., 2002). Indeed, the PRI has been shown to correlate with leaf stomatal conductance and T_c-T_a (Calderón *et al*., 2013). In wheat, PRI has been shown to improve yield estimation and to improve discrimination between water and disease stresses (Feng *et al.*, 2017; Magney *et al.*, 2014) but prior to the studies comprising this thesis, no reference was made to PRI-assisted GPC estimation. RTM inversions have shown strong correlations between the PROSAIL C_w parameter and NDWI in wheat, somewhat better at leaf than canopy level but susceptible to saturation at high C_w (C. Zhang *et al*., 2018). These authors also showed a moderately strong relationship of yield to C_w at jointing but did not consider GPC.

2.7. Grain protein content from remote sensing

The focus on GPC estimation in published works has been relatively minor compared with that on yield but has nevertheless advanced in recent years. The following emphasises GPC

estimates in commercial settings, especially from S2 data. Because it lacks sensitivity to C_{a+b}/N content, NDVI alone is a relatively poor predictor of GPC (Bonfil, 2017; Magney *et al*., 2016a; Wang *et al*., 2004). Despite this, Feng *et al.* (2014) estimated GPC in commercial dryland wheat, by combining satellite NDVI from two growth stages (R^2 = 0.567—0.632, rRMSE = 0.141—0.144), albeit at field level and hence not applicable to PA approaches. Wright *et al.* (2004) estimated leaf N and GPC in irrigated wheat crops, finding the gNDVI (Gitelson and Merzlyak, 1998) best for GPC, airborne more accurate than satellite and acceptable results for both (airborne R^2 = 0.53**,** satellite R^2 = 0.48). Recently, this link with gNDVI was confirmed by Pancorbo *et al*. (2023), who used it in multivariate linear regressions with a C_{a+b}-sensitive hyperspectral VI and SWIR information to estimate GPC in plots with R^2 = 0.73. The same study estimated GPC from S2 data by similar means with R^2 = 0.69. From Landsat, Zhao *et al.*, (2005) reported relationships between the VI_{green} (Gitelson *et al*., 2002a) and GPC (R^2 = 0.46).

While interest in direct estimation of GPC has accelerated, small number of studies focused on commercial fields remains small; in these, more specialised VIs have shown success. For example, Raya-Sereno *et al.*, (2021) found that RE indices consistently correlated better with GPC than did other VIs, and that the NDRE was a relatively stable indicator of GNC over four years. Li *et al.* (2020) obtained the best of their moderate results with the RE-focused DCNI and CI, and Rodrigues *et al.* (2018) found combinations of green- and RE bands most effective for GPC estimation. By contrast, greater importance was attributed to the red and NIR domains, than to the RE for GPC estimation from airborne multispectral data (Zhou *et al*., 2021). While the relevance of the RE is driven by C_{a+b}, perhaps weighted toward its influence on the overall N/protein source size, the NIR remains in focus because of its relationship with water content and therefore the duration and extent of photosynthesis. Moreover, N uptake, translocation and utilisation for GPC is strongly dependent on water

stress (Klem *et al*., 2018). Ma *et al*. (2022) review the use of hyperspectral RS for GPC estimation in wheat but substantially omit RTM inversion; this is justified by the fact that no prior work appears to have directly investigated GPC estimation using RTM-inverted parameters. This is despite quantities directly associated with GPC, including C_{a+b}, LAI, and others that respond instantaneously to changes in stress such as Anth and C_{x+c}, and are hence likely to influence GPC, having been widely quantified by RTM inversion. Moreover, while PROSAIL-inverted C_w was used by C. Zhang *et al*. (2018) to refine wheat yield estimates, and should be directly related to GPC via photosynthetic rate to date its use for GPC estimation appears lacking from the literature.

Wheat GPC is a complex trait which has proven difficult to predict in field conditions (Zhao *et al*., 2019). For example, while Zhou *et al*. (2021) found mostly moderately strong relationships (R^2 = 0.50—0.62) between predicted and observed yield across linear regressions based on single VIs, metrics for GPC prediction with equivalent methods were far weaker (R^2 = 0.12—0.34); in each case, the EVI performed best. However, the same study found that GPC prediction metrics (R^2 = 0.55—0.63) were similar to those for yield (R^2 = 0.56—0.62) across several ML-based models, emphasising the appropriateness of analysis by ML to the complex physiology of GPC; such matters are considered below.

2.7.1. Timeseries remote sensing for GPC estimation

In general, estimates of wheat GPC, and the specific predictors most suited to doing so, have lacked consistency, but many studies have noted stages ≥ Z65 as best for estimating yield (e.g. Gómez *et al*., 2021; Liao *et al*., 2022), GPC (e.g. Apan *et al*., 2006; Zhao *et al*., 2005, 2019), or both (e.g., Jensen *et al*., 2007; Liu *et al*., 2006; Lopez-Bellido *et al*., 2004; Magney *et al*., 2016a; Raya-Sereno *et al*., 2021; Sun *et al*., 2022; Wang *et al*., 2014). Estimating wheat yield from S2, Hunt *et al.* (2019) saw a large improvement on adding a second image,

diminishing returns for further images but better performance closer to harvest, as did (Wolanin *et al.*, 2020). In a novel use of NDVI, using daily observations to define the duration of phenological stages, Magney *et al*. (2016a) improved GPC prediction vastly over raw NDVI, showing that faster heading and slower ripening phases each had moderately strong associations with higher GPC.

While additive use of timeseries VIs from proximal sensing, including ratio vegetation index (RVI), NDVI, and gNDVI brought substantial improvement to yield estimates over single-epoch use, the same was not seen for GPC estimation (Xue *et al.*, 2007). These authors attributed the difference between results for yield and GPC to a failure of their methods to account for protein translocation, which indeed seems a likely difficulty given the range of VIs employed. A similar deficit with respect to GPC was also seen by Rodrigues *et al*. (2018), based on airborne hyperspectral imagery. By contrast, Wang *et al.* (2014) reported that stacked TS images were best for GPC modelling from satellite RS via standard VIs, despite poor correlations early in seasons, improving later. Performance improvements with TS data integration were seen for both yield and GPC by Sun *et al*. (2022), who demonstrated a DL-based method of estimating the harvest parameters concurrently, albeit from proximal sensing. The best successes in the studies cited here, which as far as could be ascertained covers all publications concerning GPC estimation from TS imaging, arise from proximal sensing. While such methods are poorly applicable in commercial contexts, they suggest that TS analysis of airborne and satellite RS images could be fruitful.

2.8. Use of machine learning algorithms in RS

Machine learning regression algorithms (MLRAs) find wide application in analysis of RS data in studies of cereal phenotype, yield, and quality, and are generally suited to crop trait prediction because of their capacity to identify latent relationships between observed and target variables, and hence build effective predictive models in large datasets (Ma *et al*.,

2019; van Klompenburg *et al*., 2020). In agriculture and other land use analyses, MLRAs are applied to both classification (e.g. Abdi, 2020; Poblete *et al.*, 2021) and regression tasks (e.g. Wolanin *et al.*, 2020). Techniques include partial least squares regression (PLSR) and its derivatives (Hansen *et al*., 2002; Jin *et al*., 2014; Kanning *et al*., 2018; Øvergaard *et al*., 2013; Yendrek *et al*., 2017), artificial, convolutional and recurrent neural networks (ANN, CNN, RNN; Atzberger, 2010; Wolanin *et al*., 2020, 2019; Yao *et al*., 2015), support vector machines / regression (SVM/R; Poblete *et al*., 2020; Rivera-Caicedo *et al*., 2014; Yao *et al*., 2015) and random forests (RF; Liang *et al*., 2018; Shah *et al*., 2019). MLRAs have been used to retrieve traits directly from hyperspectral reflectance spectra (Shah *et al*., 2019; Yao *et al*., 2015), via NBHIs (Hansen *et al*., 2002; Yang *et al*., 2021), RTM inversions (Atzberger, 2010; Liang *et al*., 2016; Zhang *et al*., 2013) and as part of hybrid retrieval techniques (§2.3.5). Alongside these more traditional, study-scale applications of ML, the prodigious amounts of inhomogeneous RS data now collected puts out of reach analysis by humans alone (Chlingaryan *et al*., 2018), and ML is applied ever more often. As such, recent reviews have considered its use for classification tasks with RS (Maxwell *et al*., 2018), RS for yield and N status prediction in PA (Chlingaryan *et al*., 2018), and yield prediction with ML and deep learning (van Klompenburg *et al*., 2020).

Recently, data from diverse sources including socio-economic, meteorological and crop records have been analysed through ML for their influence on grain yield and quality (Cao *et al*., 2020; Gómez *et al*., 2021; Li *et al*., 2020; Wolanin *et al*., 2020). From airborne hyperspectral images, Kanning *et al*. (2018) first derived LAI and leaf C_{a+b} by PLSR, then predicted wheat yield within paddocks with good accuracy, proposing that such mapping could be used to reduce N applications. From S2 data, and via PROSAIL inversion, LAI and CCC were estimated in wheat (Upreti *et al.*, 2019). Zhou *et al*. (2021), instead showed RF as marginally better than SVR, and ANNs and linear models in both yield and GPC prediction

from multispectral images across two years in commercial wheat fields. Despite this increasing use, several authors point out that while ML approaches are well-suited, they are underutilised for the investigation of N-related plant traits, including GPC (Prey and Schmidhalter, 2019; Raya-Sereno *et al*., 2021).

Successes demonstrate the potential of MLRAs in elucidating real-world variables from large and complex datasets, and suggest they are well suited to the complex and nonlinear relationships between canopy traits and GPC (Zhou *et al*., 2021). However, MLRAs also require caution and do not always bring a distinct advantage. For example, while both Zhou *et al*. (2021) and Gómez *et al*. (2021) reported superior results from ML as compared with the linear regression techniques they tested, the gap was sometimes small, while in contrast Upreti *et al*. (2019) reported better performance from least squares and PLSR methods than from boosting- and RF-based algorithms. Pancorbo *et al*. (2023) also obtained better wheat GPC estimates from multiple linear regressions than from ML applications, across both airborne HS and S2 analyses. Yao *et al*. (2015) also showed that ANN and SVM were subject to overfitting in wheat canopy N estimation, and MLRAs are criticised as lacking transparency, either because information is commonly reduced to latent variables or because relationships of interest are obscured behind complex functions (Wolanin *et al*. 2020). This interpretability issue leaves a small subset of candidate MLRAs capable of providing insight into relative input feature importance while also coping with large datasets.

2.8.1. Gradient boosting

Gradient boosting/boosted machines (GBM) are a MLRA from work by Friedman (2001), owing heritage to (Freund and Schapire, 1997). Boosting is an *"... effective method of producing a very accurate prediction rule by combining rough and moderately inaccurate rules of thumb…*" (Freund and Schapire, 1999). Hence many sub-models, each with weak

predictive skill but better than chance, are added sequentially, incrementally 'boosting' estimates of the response variable. Thus, later trees learn from their predecessors through the weighting of observations based on their error, the difference between observed and predicted values for the target variable. Observations with higher error are hence more likely to appear in later iterations, and each subsequent iteration trains a new model with respect to the error of the ensemble of previous trees. As GBMs use decision trees as sub-models, or base learners, they are also known as 'boosted regression trees', 'gradient boosting decision tree regression' and similar names.

Like other regression models, GBMs are explicitly unable to extrapolate beyond the limits of observed data when predicting but offer good predictive skill and can efficiently deal with large numbers of both input features and observations, but also work well with smaller sets (Ruan *et al*., 2022). GBMs are able to handle explicitly NA values in features, while NAs in the response simply mean that nothing is learned from that observation. GBMs can handle both continuous and categorical input data and have been reported as relatively unaffected by feature reduction. Moreover, inputs can be strongly collinear, up to perfect correlation, without affecting predictive skill (Elith *et al*., 2008; Gómez *et al*., 2021), although such multicollinearity will affect analyses of feature importance. Finally, GBM input features do not require scaling or normalisation, which improves the interpretability of model outputs.

Many GBM hyperparameters can be tuned; those relevant to the current studies, and their effects, are summarised in Table 3.8. Overfitting is limited through direct control of model complexity, via the maximum tree depth, minimum node size and γ hyperparameters; and via stochastic gradient descent (SGD). In SGD, a randomly-sampled subset of the training set is passed forward as the training data in each epoch (Friedman, 2002). This can be done column- or row-wise, adding randomness to improve robustness to noise, and can allow the algorithm to avoid local minima in the loss function gradient. However, because removal of

columns causes gain redistribution, SGD should be used with caution where feature importance is a focus. The learning rate, also known as step size or shrinkage, by which residuals are multiplied at each iteration, scaling the contribution of each tree, can also be adjusted. Lower learning rates mean that smaller steps are taken towards the optimal solution and may increase the required number of trees (Elith *et al*., 2008). In addition, k-fold cross-validation (CV) can be applied, in which folds are randomised so as not to reflect any systematic ordering and should remain representative of the whole dataset. Predictions are made against these CV samples, and error is assessed; in the current studies, by minimising the root mean square error (RMSE) of cross validation ($RMSE_{cv}$).

Until recently, GBMs have had relatively little use in agriculture (van Klompenburg *et al.*, 2020), perhaps because of the good reportability of, and hence strong focus on, good predictive results. Indeed, GBMs have been shown among the most accurate algorithms in agricultural and land use contexts (Abdi, 2020; Ruan *et al*., 2022; Zhang *et al*., 2020). In the works of Ruan *et al*. (2022), GBMs were a top performer among 11 algorithms tested in wheat yield prediction based on proximal sensing and met data and offered a large reduction in computational overhead relative RF, while Gómez *et al*. (2021) likewise ranked the GBM highly. For leaf N estimation in wheat, gradient boosting again outperformed other algorithms (Yang *et al*., 2021). Nevertheless, GPC estimation with GBM has to date been based only on genotype data, where it also outperformed other algorithms especially in more complex situations (Grinberg *et al*., 2020). However, like RFs, GBMs assess the relative contribution of input features to target variable estimation, termed importance or gain, and this is encouraging uptake (Abdi, 2020; Grinberg *et al*., 2020; Hunt *et al*., 2019). Recent GBM use with RS includes regional yield estimation in wheat, where feature importance was a strong focus (Cao *et al*., 2020; Gómez *et al*., 2021; Ruan *et al*., 2022). This interpretability is a major asset to the current study, which seeks to identify traits linked to GPC.

2.9. Multicollinearity reduction

Multicollinearity describes a situation in which two or more input features (explanatory variables) are linearly associated between themselves and with the response variable (Akinwande *et al*., 2015). Such highly-correlated predictors supply the model with redundant information, so their removal does not make a large impact on predictive skill (Akinwande *et al*., 2015). For prediction purposes, tree-based ML algorithms including random forests and gradient boosting are generally robust to collinearity among input features (Elith *et al*., 2008; Gómez *et al*., 2021). However, a substantial focus of this project is to quantify the relative importance to GPC estimation of input features; indeed, this is one of the reasons for choosing a GBM. This objective parallels fundamental early applications of RTMs to identify and optimise reflectance bands for VI development (e.g. Dorigo *et al*., 2007; Haboudane *et al*., 2002). Because gain assessment is crucial to the greater project, it was important in this case to exclude multicollinear input features to prevent gain leakage between them.

Variance inflation factor (VIF) analysis is a method of identifying collinear coefficients in a multiple regression dataset; as an example, at VIF = 10, the variance of a coefficient *i* is ten times higher than it would be if the *i*th independent variable were linearly independent of other variables (Dormann *et al*., 2013; O'Brien, 2007). VIF functions, therefore, as an effective, data-driven dimensionality reduction method, which also shows the relative magnitude of collinearity (Poblete *et al*., 2021). Learned opinion contends that there is no definitive VIF threshold above which multicollinearity is excessive, and that, while it should be given to readers, the threshold depends on the user's objective (Craney and Surles, 2002). For example, while O'Brien (2007) argues that there is no definitive VIF value at which a variable must be excluded, Dormann *et al*. (2013) consider VIF = 10 to be a critical value. Magney *et al*. (2016a) use VIF = 10, as do Zarco-Tejada *et al*. (2018), while Akinwande *et*

al. (2015) and Poblete *et al*. (2021) adopt VIF = 5. For the reasons laid out above, this project adopts these latter, lower values.

2.10. Summary of knowledge gaps

Most studies estimating GPC from RS spectroscopy have used plot experiments (e.g. Pancorbo *et al*., 2023; Raya-Sereno *et al*., 2021; Walsh *et al*., 2023; Wang *et al*., 2004; Zhao *et al*., 2005, 2019). Overall, analyses that include commercial growing conditions, and airborne or satellite images, are both relatively few and are mostly conducted at the scale of whole fields (e.g. C. Tan *et al*., 2020; Wang *et al*., 2014; Zhao *et al*., 2019). Research works at within-field scales in commercial crops appear to be limited to Rodrigues *et al.* (2018) and Stoy *et al*., (2022); GPC variability driven by natural soil variability under normal farming conditions has thus had scant attention. This gap should be addressed because it is at this scale that practical GPC-oriented PA interventions would be applied, and incomplete knowledge currently impedes progress. While Rodrigues *et al.* (2018) and Stoy *et al*., (2022) do consider soil variability in real crops, only the latter base their work on on-combine GPC monitor data.

Moreover, according to comprehensive searches of the academic databases for this PhD project, and as cited in the current literature review, works published to date linking RS with GPC appear exclusively to have used VIs, with no reference to RTM inversions. This research gap also demands attention, because of the poor consistency among VIs revealed herein and noted by Raya-Sereno *et al*. (2021), and because the potential contribution of RTM-inverted parameters to GPC estimation itself merits thorough investigation.

Although TS have been investigated for yield estimation, their additive use for GPC estimation is far more limited. This work has seen limited success and, due to its exclusive reliance on VIs, is likely poorly transferable between agronomic situations; indeed this

reliance is identified as a potential source of inconsistency across TS within such studies (Feng *et al*., 2014; Rodrigues *et al*., 2018; Xue *et al*., 2007). Notwithstanding their occasional instability at extreme values, RTM-inverted parameters have potential to improve both overall predictive skill in models based on them, and such models' interpretability along TS.

Yet another knowledge gap occurs in relation to the plant traits used for GPC estimation. Most work has focused on the obvious physiological precursors N and C_{a+b}, and on water stress; this is reasonable given the direct and unambiguous relationships of these quantities with GPC. However, as discussed above, GPC is strongly mediated by plant stress, and the detection of pre-visual stress is relatively little covered in the existing literature. Despite their proven physiological links to such stress, and some investigation of their association with wheat yield, Anth, C_{x+c}, and PRI have not been directly tested as GPC predictors.

Finally, apart from Stoy *et al*., (2022), few recent studies leverage on-combine GPC data as a source of ground truth for RS studies, so there has been scant testing of the potentially substantial scientific value of these data. This is important because such instruments are becoming more common in farm operations; moreover, they are equivalent to those used to grade wheat and set its price at receival depots. To gain traction, any proposed PA solution should be based on technology and methods already used in industry.

3. Materials and Methods

This chapter contains included material from:

Longmire, A.R., Poblete, T., Hunt, J.R., Chen, D., Zarco-Tejada, P.J., 2022. Assessment of crop traits retrieved from airborne hyperspectral and thermal remote sensing imagery to predict wheat grain protein content. *ISPRS Journal of Photogrammetry and Remote Sensing 193, 284–298*. https://doi.org/10.1016/j.isprsjprs.2022.09.015

Longmire, A.R., Poblete, T., Hornero, A., Chen, D., Zarco-Tejada, P.J., (accepted 2023-10-27). Estimation of grain protein content in commercial bread and durum wheat fields via traits inverted by radiative transfer modelling from Sentinel-2 timeseries. *ISPRS Journal of Photogrammetry and Remote Sensing*.

—

To robustly characterise plant traits that are both retrievable from airborne hyperspectral RS data, and closely associated with GPC, it was necessary to consider wheat grown under semi-controlled conditions in which soil N supply could be manipulated and data collected at ground level, but under open skies, and preferably with divergent soil and/or rainfall conditions. I therefore struck collaborations with two agricultural research and extension organisations with the capacity to establish plot trials in 2019. The recent advent of on-combine GPC monitor data collected during harvests, and a desire to test its scientific value, motivated me to include commercial crops in the project. The use of airborne hyperspectral and thermal sensors mounted in tandem on a piloted aircraft allowed data capture in light and meteorological conditions identical between the sensors and permitted several thousand hectares' data to be acquired in a single day, contingent on just two factors: cloud-free conditions over the target, and feasible flying distance from the aircraft's home base. The selection of both plot and commercial sites satisfied these conditions while also providing for wide soil variability both between sites and across small spatial scales within them. Further, the scale of dryland wheat cropping is appropriate to the areal coverage and regular return time of spaceborne sensors and offers an opportunity both to test the utility of plant traits

retrieved from satellite data and to assess the contribution of image timeseries. Lastly, to more fully assess traits' and indices' relationship with GPC, it was desirable to analyse data from at least two years.

3.1. Study sites

3.1.1. Plot experiments

This study considers both experimental plots and commercial crops, grown in rainfed conditions across two meteorologically diverse seasons and in contrasting soils at four locations in the southern Australian wheat belt. The plot trials were located at Birchip (experimental site 1 (ES1); average annual rainfall (AAR) = 353 mm) and Yarrawonga (experimental site 2 (ES2); AAR = 470 mm; Fig. 3.1) in the southern Australian wheatbelt. Each plot layout was planted in a randomised complete block design with hard white bread wheat (cv. Scepter) for N fertiliser treatment trials in 2019. Site ES1 takes the Köppen-Geiger climate classification Bsk (arid, dry summer, cold) and ES2, Cfa (temperate, without dry season, hot summer; Peel *et al.*, 2007). In the Australian Soil Classification (ASC; Isbell, (2002)), soils at ES1 are classified as calcarosols, while sodosols predominate at ES2. Pre-season soil mineral N was equivalent at the two sites.

The ES1 plots were sown on 2019-05-16 and fertiliser was applied on 2019-08-10 (Zadoks growth stage Z31). N application rates at ES1 only were adjusted according to season-to-date rainfall accumulation, targeting yield decile predictions modelled with the Yield Prophet® fertiliser decision support tool (Hochman *et al.*, 2009; Hunt *et al.*, 2006), plus rates based on grower advice, the maximum yield formulae of French and Schultz (1984) and those of Sadras and Angus (2006). Some of these N rate increments were grouped for analysis (

Table 3.1). ES2 was sown on 2019-05-09 and fertiliser was applied in equally split doses at tillering (Z23; 2019-06-22) and at stem elongation (Z31; 2019-08-06), producing five N treatment levels. Plots at both sites were approximately 12 x 2 m, treatments were replicated four times and agronomic procedures were equivalent. Each layout was located within fields regularly used for commercial cropping and with zero slope. As per normal cropping practices for Australian wheat, no irrigation was applied.

Table 3.1 Location and fertiliser N applied to plots sown to wheat (cv. Scepter) at ES1 (Birchip) and ES2 (Yarrawonga). Pre-sowing soil N to 120 cm (ES1) and 100 cm (ES2).

Location details	Soil N (mg kg^{-1})	Total fertiliser (kg N ha^{-1})	Treatment (aggregated)	Plots (n)
Site ES1	46.8	0	B0	4
Birchip		30	B1	8
35.969° S		37		
142.822° E		98	B2	8
102 m ASL		104		
		162	B3	12
		167		
		171		
Site ES2	46.1	0	Y1	4
Yarrawonga		46	Y2	
36.050° S		92	Y3	
145.983° E		138	Y4	
129 m ASL		184	Y5	

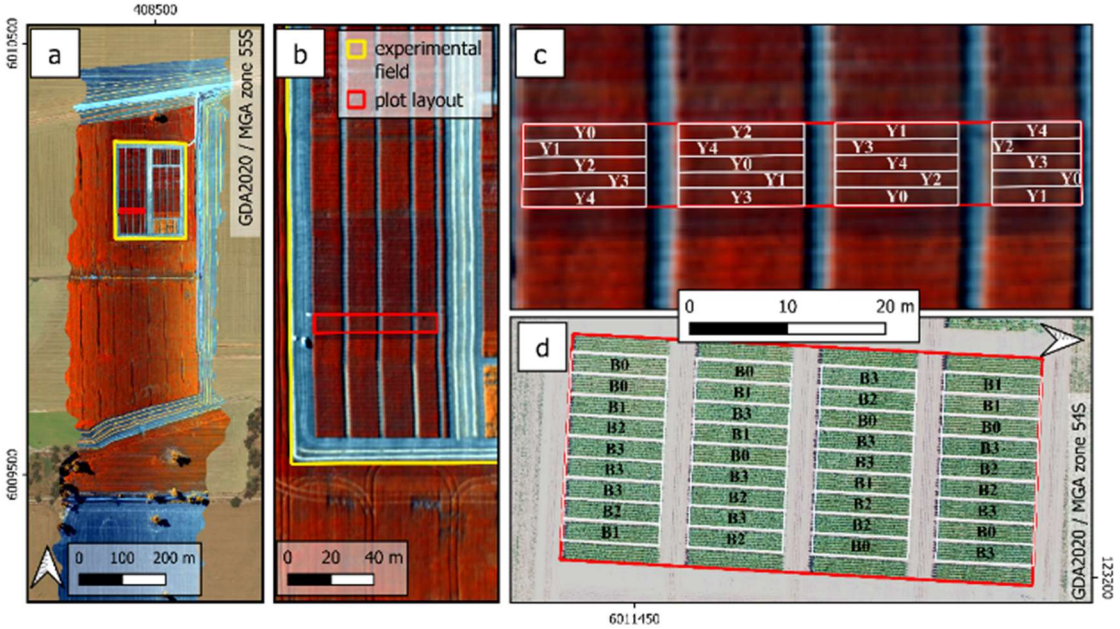

Fig. 3.1. Experimental site 2 (ES2; a); ES2 layout on false colour hyperspectral image (R = 749 m, G = 710 nm, B = 678 nm; b); ES2 (c) and ES1 (d) plots with treatment levels (kg.N/ha): Y0 = 0, Y1 = 46, Y2 = 92, Y3 = 138, Y4 = 184; B0 = 0, B1 = 30 – 37, B2 = 98 – 104, B3 = 162 – 171.

3.1.2. Field transect

In 2020 only, a transect was laid out in one field (M01; 36.30° S 141.35° E), part of commercial cropping zone 1 (CZ1), described below. To facilitate later spatial joining of transect waypoints to GPC and airborne data the transect, comprising 20 points approximately 50 m apart, was laid out parallel to the direction of travel for combine harvesters. The points were GPS marked at the time of field work / hyperspectral data collection, 2020-10-28.

3.1.3. Commercial crops hyperspectral

This study considers 6355 ha of rainfed commercial hard white bread and durum wheat crops grown in two areas of the southern Australian wheat belt. In CZ1, around Kaniva (36.37° S, 141.24° E; altitude 142 m; AAR = 451 mm; Fig. 3.2a); the subject bread and durum crops were spread across 30 km of latitude and 15 km of longitude. For CZ1, growers supplied on-combine protein monitor data from 662 ha (seven paddocks) for the 2019 harvest and 858 ha (10 fields) in 2020. Bread wheat only was sown around Manangatang (CZ2; 35.05° S, 142.88° E; altitude 55 m; AAR = 316 mm; Fig. 3.2b). For CZ2, data were available from 12 fields in 2019 and 26 in 2020, totalling > 2000 ha in each year and distributed across a similar extent as at CZ1. In each location, these winter crops were sown between 2019-05-15—2019-06-03 and 2020-05-12—2020-06-05. Fertiliser was applied as urea 1—3 times each season, according to commercial priorities and on growers' assessment of likely yield. The Köppen-Geiger classification for CZ1 is Cfb, indicating a temperate climate, without dry season, and with a warm summer, while CZ2, like ES2, is classified Bsk. Location and rainfall details are provided in Table 3.2.

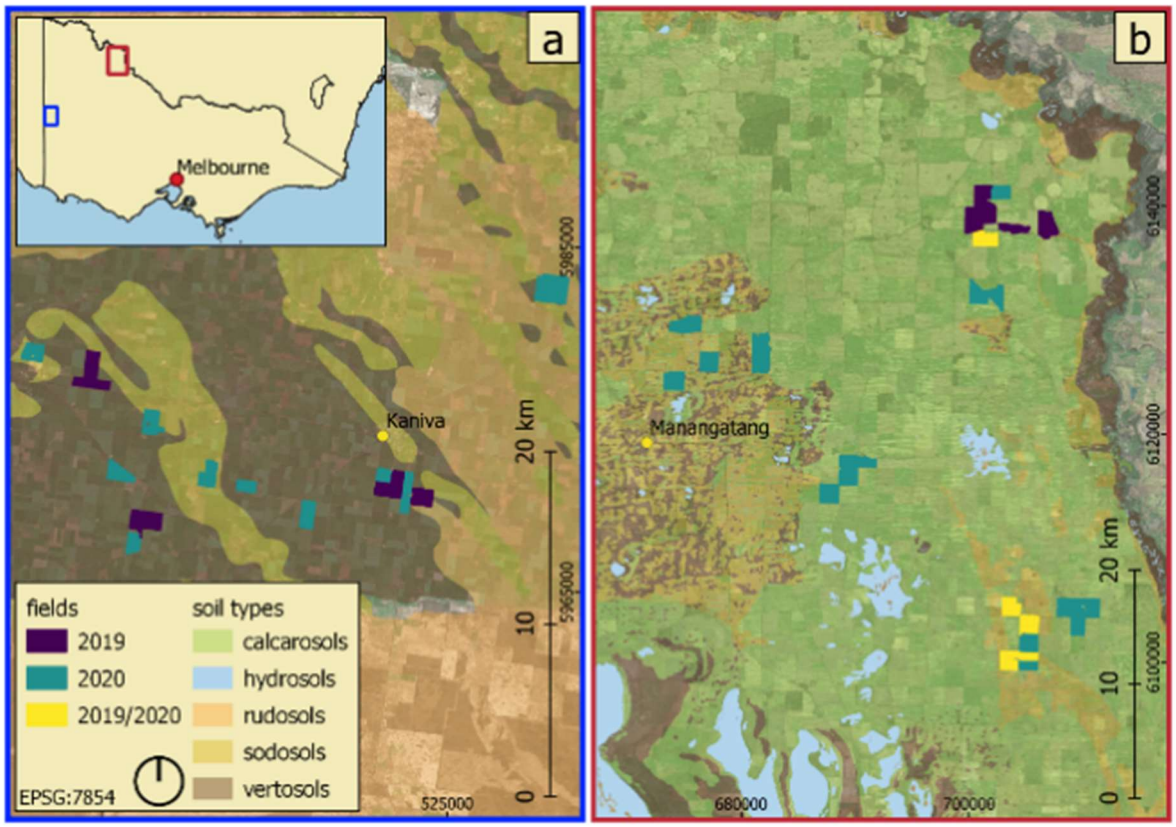

Fig. 3.2. Commercial cropping zones 1 (CZ1; a) and 2 (CZ2; b) regions, Victoria, Australia, with Australian Soil Classification to soil type and wheat fields for which grain protein content observations were available in 2019 and 2020.

The dominant cropping soils in CZ1 are vertosols; the equivalent in the World reference base for soil resources (WRB) are vertisols; (FAO, 2015), but both sodosols (WRB: solonetz / abruptic luvisols / planosols) and chromosols (abruptic luvisols / lixisols) are also common. Landforms are typically level to gently undulating. In CZ2 calcarosols (calcisols) predominate, while vertosols and sodosols are also present. The remnant dune fields, on which agricultural fields in CZ2 are overlaid, formed through aeolian deposition of extensively re-worked material of marine origin (Pell *et al*., 2001; Sadras *et al*., 2002). The dunes are 2–6 m high and 200–1200 m apart, an amplitude less than the dimensions of most of the fields considered, so the dune-swale morphology induces great within-field soil variability. Due to their effects on plant-available water content (PAWC), differences in soil and subsoil properties at sub-paddock scale may drive yield variability as much as inter-annual rainfall differentials (Whitbread *et al*., 2008). Dunes are composed predominantly of

deep, loose sand and aggregates in which rainfall infiltrates to a depth where it is protected from evaporation and contributes to PAWC (Llewellyn *et al*., 2008; Sadras *et al*., 2002). Swale soils are heavier, their shallow hardpans may limit infiltration, and boron toxicity may reduce root penetration (Nuttall et al., 2003). Sodicity and higher salt loads also reduce PAWC, through their osmotic effects, in swale soils.

Table 3.2. Location, rainfall, climate zone and total cropped areas included for CZ1 (Kaniva) and CZ2 (Manangatang), 2019–2020.

Zone	Lat.	Lon.	Alt. (m)	AAR* (mm)	Clim. zone†	Year	Rain* (mm)	GSR* (mm)	Area (Ha)
CZ1	-36.37°	141.24°	142	449	Cfb	2019	288	238	662
						2020	444	291	858
CZ2	-35.05°	142.88°	55	316	Bsk	2019	194	135	2341
						2020	342	277	2494

*AAR = long term average annual rainfall; GSR = growing season rainfall (Apr – Oct); rain = annual total. Rainfall data are from the nearest Bureau of Meteorology station with complete records for the year in question: CZ1 Nhill Aerodrome (#78015), CZ2 Ouyen (#076047). † Climate zone as per Köppen-Geiger climate classification (Peel *et al*., 2007).

3.2. On-ground and lab processes

To validate estimates of plant variables obtained through remote sensing and image analysis, data were collected by three general methods in the plot experiments and field transect:

- Destructive sampling, where removal of plant and / or soil material and laboratory assays produced ‘ground truth’ data;
- Leaf level spectroscopy, where non-destructive, optical leaf-clip instruments were used to estimate pigments and other quantities; and
- Canopy level work, where NBHIs and inverted parameters were retrieved from airborne hyperspectral and thermal images

3.2.1. Plot experiments

Leaf measurement and tissue sampling at each experimental site targeted the adaxial surface of uppermost sun-adapted leaves central in plots and were conducted quasi-concurrently with hyperspectral flight campaigns, in equivalent light and meteorological conditions. Ten leaf measurements in each plot were taken with Dualex (FORCE-A, Orsay, France) and SpectraPen (PSI, Drasov, Czech Republic). Approximately 80 g of entire flag or other upper, sun-exposed leaves were cut from the central rows of plots, sealed in plastic, refrigerated in transit then kept at -20 °C until processing. Ten subsamples were taken from the central area of each cut leaf with punches of known diameter for calculation of areal N content; these discs were weighed before and after drying. Leaf tissue was dried at 65 °C for 48 h, then ground to a uniform powder in a ball mill and analysed by Dumas combustion for total N (mass %). Plots were machine harvested and the grain assessed for protein content by near infrared (NIR) spectroscopy (CropScan 3000B Grain Analyser, Next Instruments, Sydney Australia). Ambient temperature, barometric pressure and incoming shortwave and longwave radiation were recorded on the ground using a portable weather station (model WXT510, Vaisala, Helsinki, Finland).

3.2.2. Field transect

Dualex and SpectraPen measurements were also made during the transect of field M01 at Miram. Further, three soil samples were taken with a hand auger from the top 15 cm and mixed (total approx. 50 g) to represent each waypoint, which was marked with a hand-held GPS (Garmin, Olathe, Kansas, USA). Leaf and soil samples were oven-dried at 65 °C for 48 h and mineral N extracted as per Rayment and Lyons (2010) on a Skalar San++ SFA (FlowAccess V 3.2). Absent differential fertiliser treatments in the crops by which to group observations for analysis, K-means clustering (Hartigan and Wong, 1979) was used to divide the transect data into three levels by GPC.

3.2.3. Protein monitor data

Data analysis for commercial fields was based on GPC records collected during harvest. These data were collected with NIR spectrometers (CropScan 3000/3300H, Next Instruments, Sydney Australia) mounted on the clean grain elevator of combine harvesters with geolocation from real-time kinetic GPS receivers at positional accuracy of ± 1 cm. GPC point data was first cleaned according to principles detailed for yield data by Ping and Dobermann, (2005): those with outlier GPC values, in slivers or isolated from neighbouring observations, within 30 m of perimeter fences or headlands and within 20 m of trees, dams, fences and, or in other harvester turn zones were discarded. By buffering the point location of each remaining GPC record to a 5 m radius in a GIS, then drawing bounding geometries for these new circular polygons, square 100 m^2 regions of interest (ROI) were established (Fig. 3.3). The 5 m radius was chosen so that ROI width was less than typical harvester swath width (12 m). Each ROI had a GPC point as its centroid and took the associated grain protein value. Where overlap between ROIs occurred, sufficient of them were removed to ensure spatial independence. Finally, areas of hyperspectral and satellite image rasters affected by cloud shadow were identified, then intersecting ROIs again removed to exclude erroneous reflectance spectra.

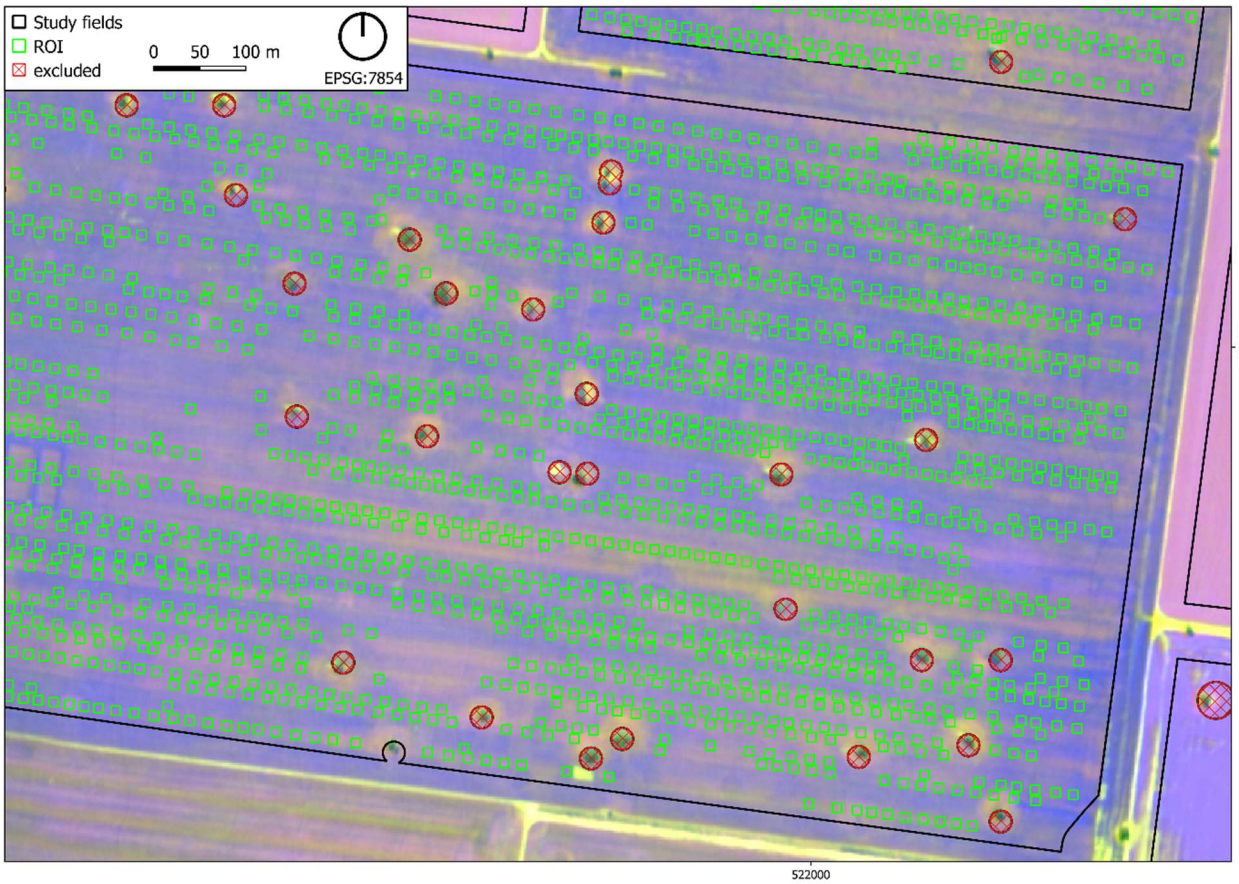

Fig. 3.3. Detail of commercial wheat field showing regions of interest (ROI) established with grain protein observation points as their centroids, buffering of field boundaries and non-crop vegetation excluded from analyses.

A Wilcoxon test (Bauer, 1972) was applied a) across zones within year / wheat type permutations and b) across wheat types within zone / year permutations to test the significance and effect size of differences in median GPC.

3.3. Remote-sensed data: acquisition and pre-processing

3.3.1. Airborne data

In each year, logistical constraints dictated a later start to remote sensing campaigns than desired. Nevertheless, airborne hyperspectral and thermal images were collected by sensors flown in tandem on a Cessna-172 light aircraft operated by the Hypersens Laboratory, University of Melbourne. Missions were flown over plots at site 1 on 2019-10-03 (1409 growing degree days after sowing (GDDAS)), site ES2 on 2019-10-09 (1559 GDDAS), and

commercial fields near Kaniva (CZ1) on 2019-10-22 (bread wheat 1514 GDDAS, durum 1736 GDDAS) and 2020-10-28 (bread 1592 GDDAS, durum 1742 GDDAS). Thermal time calculations, after Baskerville and Emin (1969), were based on data from the Australian Bureau of Meteorology's station #78015 (Bureau of Meteorology, 2021). Flights at 350 m and 400 m above ground level (AGL) gave ground sampling distances (GSD) of 0.15 m (hyperspectral) and 0.25 m (thermal) at experimental site ES1 and GSD ≈ 0.2 m (hyperspectral) and 0.35 m (thermal) at site ES2, respectively, while flights over commercial fields at ≈ 2000 m AGL yielded GSD ≈ 1.0 m (hyperspectral) and GSD ≈ 1.7 m (thermal).

Hyperspectral radiance data were collected in the visible and near infrared (VNIR) domains with a hyperspectral VNIR sensor (VNIR E-Series model; Headwall Photonics, Fitchburg, MA, USA), capturing 371 bands from 400-1001 nm at 8 nm per pixel, yielding 7 nm FWHM with a 25 μm slit. At 12-bit radiometric resolution, storage rate was 50 frames per second with exposure time of 18 ms and an 8 mm focal length. Radiometric and spectral calibration was completed in the laboratory prior to flights: The hyperspectral imager was calibrated using an integrating sphere (Labsphere XTH2000C, Labsphere Inc., North Sutton, NH, USA), deriving coefficients at four illumination levels. Atmospheric correction of radiance was applied with the Simple Model of Atmospheric Radiative Transfer of Sunshine (SMARTS) model (Gueymard, 1995), using aerosol optical depth (AOD) observed at 550 nm at the time of flight with a Micro-Tops II sunphotometer (Solar LIGHT Co., Philadelphia, PA, USA). This method has previously been used for hyperspectral data (Calderón *et al*., 2015; Poblete *et al*., 2020a; Zarco-Tejada *et al*., 2018). Thermal images were collected in the 7.5—14 μm region with an A655c camera (FLIR systems, Wilsonville, Oregon, USA), a scientific-grade instrument radiometrically calibrated by the manufacturer; nevertheless, a further indirect calibration was carried out during flights using ground observations from a handheld infrared thermometer (LaserSight from Optris GmbH, Berlin, Germany), after

Calderón *et al*. (2015). Orthorectification was performed using Parametric Geocoding and Orthorectification for Airborne Optical Scanner Data (PARGE; ReSe applications GmbH, Wil, Switzerland) using an inertial measurement unit and GPS data from a VN-300 (VectorNav Technologies LLC, Dallas, TX, USA). These methods follow the detailed proposals in Zarco-Tejada *et al*., (2016). The bandset was convolved to 79 bands on conversion to reflectance.

3.3.2. Spaceborne data

A multi-temporal dataset of Sentinel-2 rasters, orthorectified and expressed in top-of-atmosphere reflectance (Level-1C product) was downloaded from the Copernicus Open Access Hub (https://scihub.copernicus.eu/). The target areas during the 2019 and 2020 growing seasons intersected with tiles 54HWE (CZ1), 54HXG and 54HYG (CZ2). Correction for atmospheric effects, converting images from Level-1C to Level-2A surface reflectance was done using Sen2Cor (version 2.3.1). All images between 1 July and 31 October in each season were visually assessed, and those in which any subject field was cloud affected were discarded. This left 23 images in which all fields were cloud free, five or six in each combination of year and zone (Table 3.3). Bands of GSD = 20 m were resampled to 10 m prior to stacking as a multiband raster.

Table 3.3. Cloud-free Sentinel-2 images available in CZ1 and CZ2, 2019–2020 with associated growing degree days after sowing (GDDAS; °C day) and Zadoks (Z) stage/name. Bold entries are images from which traits were retrieved for comparison against indicators from airborne hyperspectral analysis.

	Image		Bread			Durum		
Zone	Year	Date	GDDAS	Z stage	Z name	GDDAS	Z stage	Z name
1	2019	18-Jul	481	15	seedling	684	16	seedling
		23-Jul	537	15		739	17	
		17-Aug	770	17	stem elong.	972	31	stem elong.
		11-Sep	1008	32		1211	37	
		1-Oct	1237	43	booting	1439	52	ear emerg.
		21-Oct	**1508**	**69**	**anthesis**	**1710**	**74**	**grain fill**
	2020	17-Jul	452	14	seedling	560	15	seedling
		1-Aug	581	15		689	17	
		26-Aug	803	31	stem elong.	910	31	stem elong.
		10-Sep	979	32		1087	33	
		10-Oct	**1375**	**51**	**ear emerg.**	**1483**	**67**	**anthesis**
2	2019	17-Jul	904	17	seedling			
		28-Jul	1028	31	stem elong.			
		12-Aug	1169	31				
		17-Aug	1220	32				
		27-Aug	1325	32				
		1-Oct	**1778**	**54**	**ear emerg.**			
	2020	17-Jul	719	17	seedling			
		27-Jul	803	17				
		26-Aug	1099	32	stem elong.			
		31-Aug	1161	32				
		10-Sep	1307	34				
		15-Sep	**1385**	**42**	**booting**			

3.3.3. Modelling of phenological advance

To facilitate comparison across years, zones and crop types, we calculated growing degree days after sowing (GDDAS; Asseng *et al.* (2010)), based on daily temperatures for each cropping zone taken from the SILO dataset ("https://www.longpaddock.qld.gov.au/," n.d.). To allow for potential differences in rates of phenological advance among zones and cultivars, phenology in Zadoks stages was modelled in the Agricultural Production Systems sIMulator (APSIM) Next Generation (Holzworth *et al.*, 2018) biophysical systems model (wheat module; Brown *et al*., 2014), using the same SILO temperature data.

3.3.4. Extraction of raster values to study units

Radiance (L), reflectance (R) and canopy temperature (T_c; e.g. Fig. 3.6c) values contained in hyperspectral and thermal rasters were aggregated to mean pixel value per plot and ROI. After exclusion of edges and end zones, and areas where whole shoot biomass cuts had been taken, plots contained 100—250 pure vegetation pixels each, from which mean reflectance spectra and canopy temperature (T_c) were calculated. Commercial crop ROIs contained for L and R, $n \approx 100$ pixels, and for T_c $n \approx 36$ pixels. Mean hyperspectral L and R spectra by fertiliser treatment are shown for both experimental sites (Fig. 3.4).

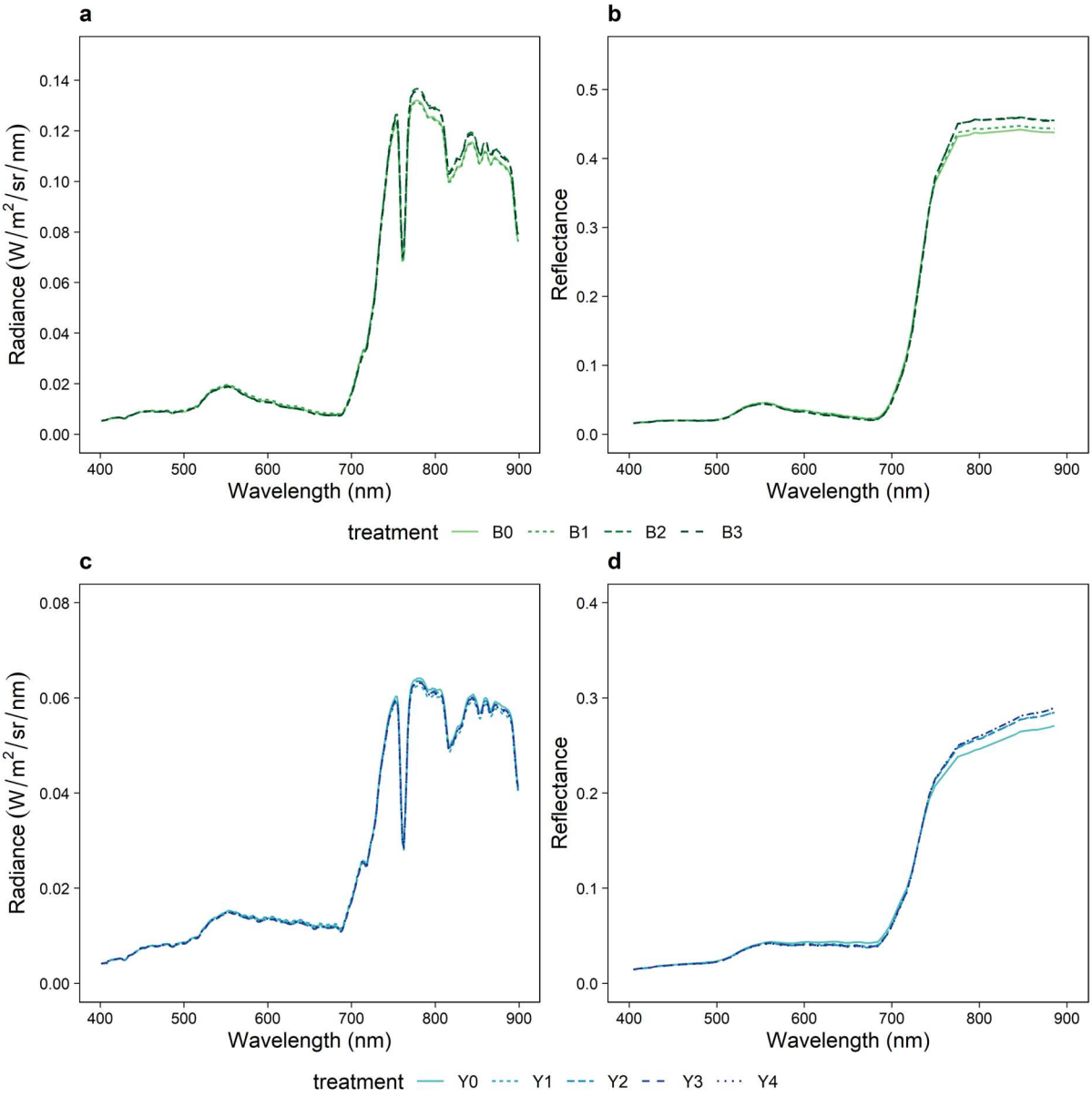

Fig. 3.4. Airborne hyperspectral radiance (W/m^2/sr/nm) and reflectance spectra captured at ES1 (a, b) and ES2 (c, d). Plot-wise mean by treatment.

Pixel values of stacked S2 band rasters were extracted to corresponding GPC points. Spatial analyses were done in QGIS (QGIS Development Team, 2020) and R (R Core Team, 2020).

3.4. Data processing – spectroscopic data

Vegetation indices and inverted leaf and canopy traits were retrieved from both airborne hyperspectral and S2 multispectral data. Solar-induced fluorescence (SIF), from hyperspectral L and R, and the crop water stress index (CWSI), from thermal T_c data, were also retrieved for the plots at both experimental sites and for ROIs in CZ1. Narrow-band

hyperspectral indices were also calculated from SpectraPen reflectance at plot and transect scales. Dualex observations ≥ 2 standard deviations from the plot or waypoint mean, and erroneous SpectraPen spectra were removed, then the clean data were aggregated to mean values and spectra per plot or transect waypoint. Where transect waypoints intersected with ROIs, variables retrieved from airborne data and associated with the ROI, plus GPC values, were assigned to waypoints by spatial joining. Data collection, extraction, handling and filtering processes are summarised in Fig. 3.5.

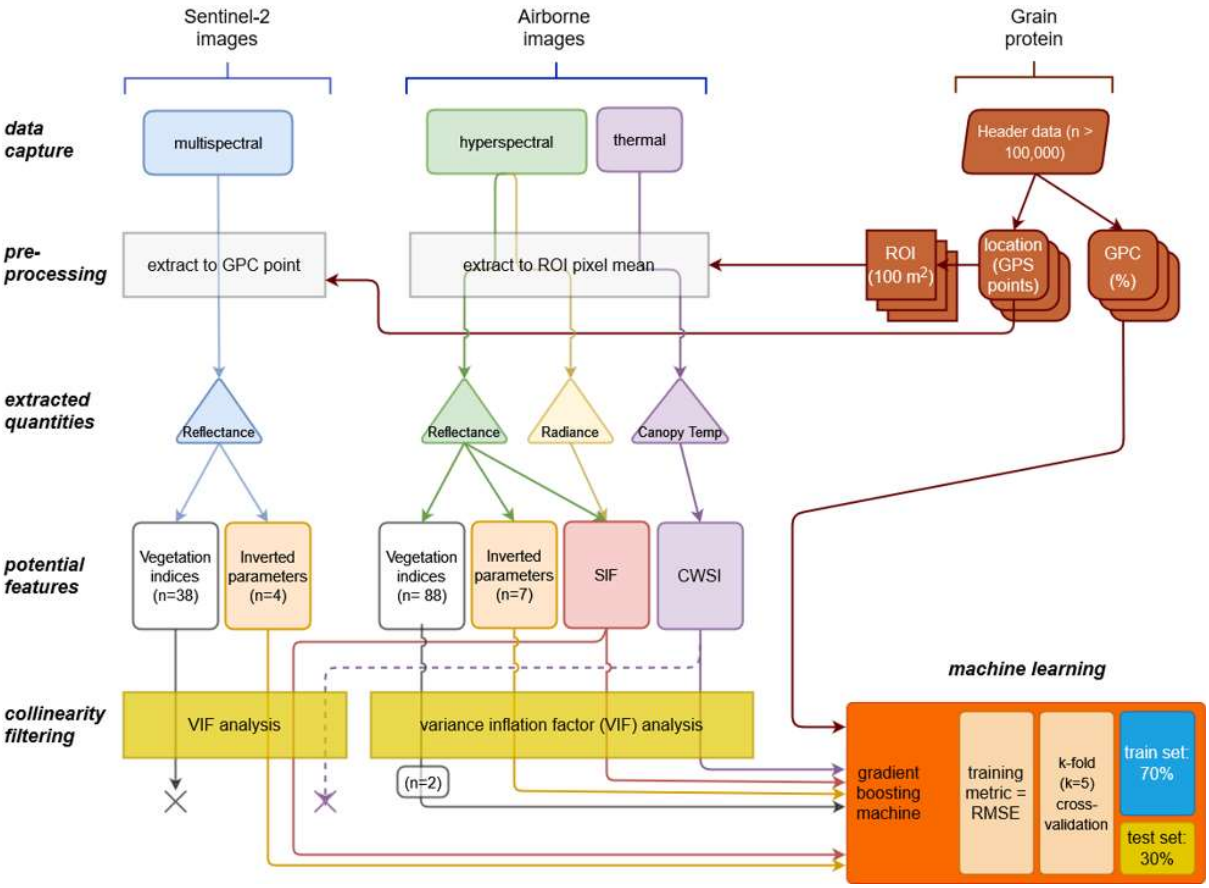

Fig. 3.5. Schematic summary of data handling and machine learning processes for estimation of wheat grain protein content in commercial crops from plant traits retrieved from airborne hyperspectral and thermal and Sentinel-2 multispectral remote sensed images.

3.4.1. Radiative transfer modelling

Leaf and canopy variables were retrieved with the PRO4SAIL radiative transfer model (RTM), which comprises PROSPECT-D (Féret *et al*., 2017b) and 4SAILH (Verhoef *et al.*, 2007). The PROSPECT-D component was used to retrieve leaf traits including the pigments xanthophylls / carotene (C_{x+c}), C_{a+b} and anthocyanins (Anth), plus leaf dry matter (C_m) and water content (C_w) from hyperspectral data, while canopy structural traits leaf area index (LAI) and leaf angle distribution function (LIDFa) were estimated with the coupled 4SAILH were outputs of the 4SAIL component. From multispectral S2 data, only C_{a+b}, C_m, C_w and LAI were retrieved through a similar PRO4SAIL inversion procedure wherein simulated reflectance was convolved to S2 specifications and solar zenith angle was varied to account for changing image dates.

In each such hybrid procedure, after Xu *et al*. (2019), separate look-up tables (LUT), each of 200,000 simulated reflectance spectra with their associated leaf and canopy values, were generated for hyperspectral and multispectral data. This number of permutations, and the associated procedures have previously been shown sufficient (Poblete *et al*., 2021; Xu *et al*., 2019; Zarco-Tejada *et al*., 2018). To build each LUT, model parameters were randomly sampled from uniform distributions in the ranges given in Table 3.4. To compare synthetic spectra against those observed in the study units, the LUTs were interrogated using a support vector machine (SVM) machine learning algorithm (MATLAB; Statistics and Machine Learning, Deep Learning and Parallel Computing toolboxes; Mathworks Inc., Natick, MA, USA), with simulated reflectance spectra as model inputs and plant/canopy traits as outputs to retrieve each trait independently. The SVM algorithm applied a radial basis function and hyperparameters were optimised for each target variable during training. Hybrid retrieval methods of this type effectively address the ill-posed problem (Verrelst *et al*., 2015a). For S2 timeseries, inversions were conducted for each separate image date in each location.

Table 3.4. Values and ranges of leaf and canopy traits used to invert the PRO4SAIL radiative transfer model inversion to retrieve plant and canopy traits from hyperspectral data only and from both hyperspectral and Sentinel-2 data (italics).

Parameter	Abbreviation	Value/range
Anthocyanin content [μg/cm^2]	Anth	1–10
Carotenoid content [μg/cm^2]	C_{x+c}	1–20
Chlorophyll a + b content [μg/cm^2]	*C_{a+b}*	*3–70*
Dry matter content [g/cm^2]	*C_m*	*0.001–0.035*
Hot spot parameter	h	0.01
Leaf area index [m^2/m^2]	*LAI*	*1–5*
Leaf Inclination Dist. Func. [°]	$LIDF_a$	0–90
Mesophyll struct. Coef.	N	0.5–3.0
Observer angle [deg.]	tt_o	0
Relative azimuth angle [deg.]	ψ	0
Sun zenith angle [deg.]	tt_s	47.7
Water content [g/cm^2]	*C_w*	*0.001–0.035*

3.4.2. Reflectance indices

In total, 88 NBHIs were calculated from hyperspectral reflectance for each study unit; those relevant (section 3.4.5) to later procedures are detailed in Table 3.5, while a full list appears in Table A1, Appendix A.

Table 3.5. Selected narrow-band hyperspectral indices, calculated from SpectraPen leaf clip spectra in plots at ES1 and ES2 (2019), in a field transect (2020), and from airborne hyperspectral imagery captured over the same plots and commercial wheat fields in CZ1, relevant to downstream processes.

Type	Name	Formula	Reference
Chlorophyll *a+b*	Red edge chlorophyll index	$CI = (R_{750}/R_{700})-1$	Gitelson *et al.* (2003)
	Normalised Pigments Index	$NPCI = (R_{680} - R_{430})/(R_{680} + R_{430})$	Peñuelas *et al.*, (1994)
	Zarco-Tejada and Miller index	$ZMI = R_{750} / R_{710}$	Zarco-Tejada *et al.* (2001)
	Vogelmann index 1	$VOG1 = R_{740}/R_{720}$	Vogelmann *et al.* (1993)
Other pigments	Photochemical Refectance Index	$PRI = (R_{531} - R_{570})/(R_{531} + R_{570})$	Gamon *et al.* (1992)
	Photochemical Refl.Index (670)	$PRI_{m3} = (R_{670} - R_{531})/(R_{670}+R_{531})$	Hernández-Clemente *et al.* (2011)
	Photochemical Refl.Index (670 and 570)	$PRI_{m4} = (R_{570}-R_{531}-R_{670})/(R_{570}+R_{531}+R_{670})$	Hernández-Clemente *et al.* (2011)
Structure	Enhanced VI (NIR)	$EVI_{NIR} = 2.5*(R_{800}-R_{670})/(R_{800}+6*R_{670}-7.5*R_{800})$	Longmire *et al.*, (2022)
	Normalised Difference VI	$NDVI = (R_{840}-R_{670})/(R_{840}+R_{670})$	Rouse *et al.*, (1974)
Structure / water	R_{920}/R_{729}	R_{920}/R_{729}	L. Suárez (personal communication)

From each image in the multispectral satellite timeseries, 38 VIs compatible with the S2 spectral specifications and associated with canopy structure or pigment concentration were calculated for each GPC point in the commercial cropping zones. Indices found relevant (§ 3.5) to downstream processes are given in Table 3.6, while a complete list is in Table A2, Appendix A.

Table 3.6. Selected vegetation indices calculated from Sentinel-2 (S2) reflectance observed over commercial wheat fields in CZ1 and CZ2, and found to be linearly independent of S2 inverted parameters from the same data.

Name	Formula	Reference
Global Environment Monitoring Index	GEMI = (n(1−0.25n)−RED−0.125/1−RED) where n = (2 * (NIR^2 - RED^2) + 1.5 * NIR + 0.5 * RED) / (NIR + RED + 0.5)	Pinty and Verstraete 1992
Maccioni Index	Macc = $(R_{780}-R_{710})/(R_{780}-R_{680})$	Maccioni *et al.* 2001
MERIS Terrestrial Chlorophyll Index	MTCI = $(R_{754}-R_{709})/(R_{709}-R_{681})$	Dash and Curran 2007
TCARI/OSAVI	$TCARI/OSAVI = 3\cdot[(R_{700}-R_{670})-0.2\cdot(R_{700}-R_{550})\cdot(R_{700}/R_{670})]/((1+0.16)\cdot(R_{800}-R_{670})/(R_{800}+R_{670}+0.16))$	Haboudane *et al.* (2002)
Transformed Chlorophyll Absorption in Reflectance Index 1610	$TCARI_{1510} = 3\cdot[(R_{700}-R_{1510})-0.2\cdot(R_{700}-R_{550})\cdot(R_{700}/R_{1510})]$	Herrmann *et al.* 2010

3.4.3. Solar induced fluorescence

From plot- and ROI mean R and L values, Fluorescence was retrieved using bands inside and outside of the telluric O_2-A Fraunhofer line (FLD2; Plascyk and Gabriel, 1975) and on the basis of radiance estimated with the SMARTS model (Gueymard, 1995) then convolved to the FWHM and spectral sampling interval (SSI) of the hyperspectral sensor. The band inside the Fraunhofer line corresponds to the local minimum incoming irradiance (E) at the relevant SSI (λ_{in}; 762 nm) while the outside band (λ_{out}; 750 nm) is on the shoulder of the O_2-A line at a shorter wavelength. SIF was calculated for each plot and ROI for that study unit's mean L and R, derived as above, according to:

Eqn. 1

$$SIF = d - Rb$$

where

$$R = \frac{(c - d)}{(a - b)}$$

and

$a = E_{750}$, $b = E_{762}$, $c = L_{750}$, $d = L_{762}$

3.4.4. Crop water stress index

Using the mean T_c for each study unit, the crop water stress index (CWSI) was calculated after the proposal of Idso *et al.* (1981), which normalises T_c with T_a and vapour pressure deficit (VPD). Air temperatures and relative humidity were from portable weather stations at the plot sites and from BoM recording station 78015 (Nhill Airport) for CZ1 at the time of flights (Bureau of Meteorology, 2021).

Eqn. 2

$$CWSI = \frac{(Tc - Ta) - (Tc - Ta)_{LL}}{(Tc - Ta)_{UL} - (Tc - Ta)_{LL}}$$

The canopy / air temperature differential in a canopy transpiring at its maximum rate for a given VPD is given by the lower limit $(Tc - Ta)_{LL}$, whereas the upper limit $(Tc - Ta)_{UL}$ represents the same for a canopy where transpiration is zero. Here I adopted $(Tc - Ta)_{LL}$ defined by Idso (1982) and used by Gonzalez-Dugo *et al.* (2015) in studies of wheat in arid conditions.

Eqn. 3

$$(Tc - Ta)_{LL} = -3.25 \cdot \text{VPD} + 3.38$$

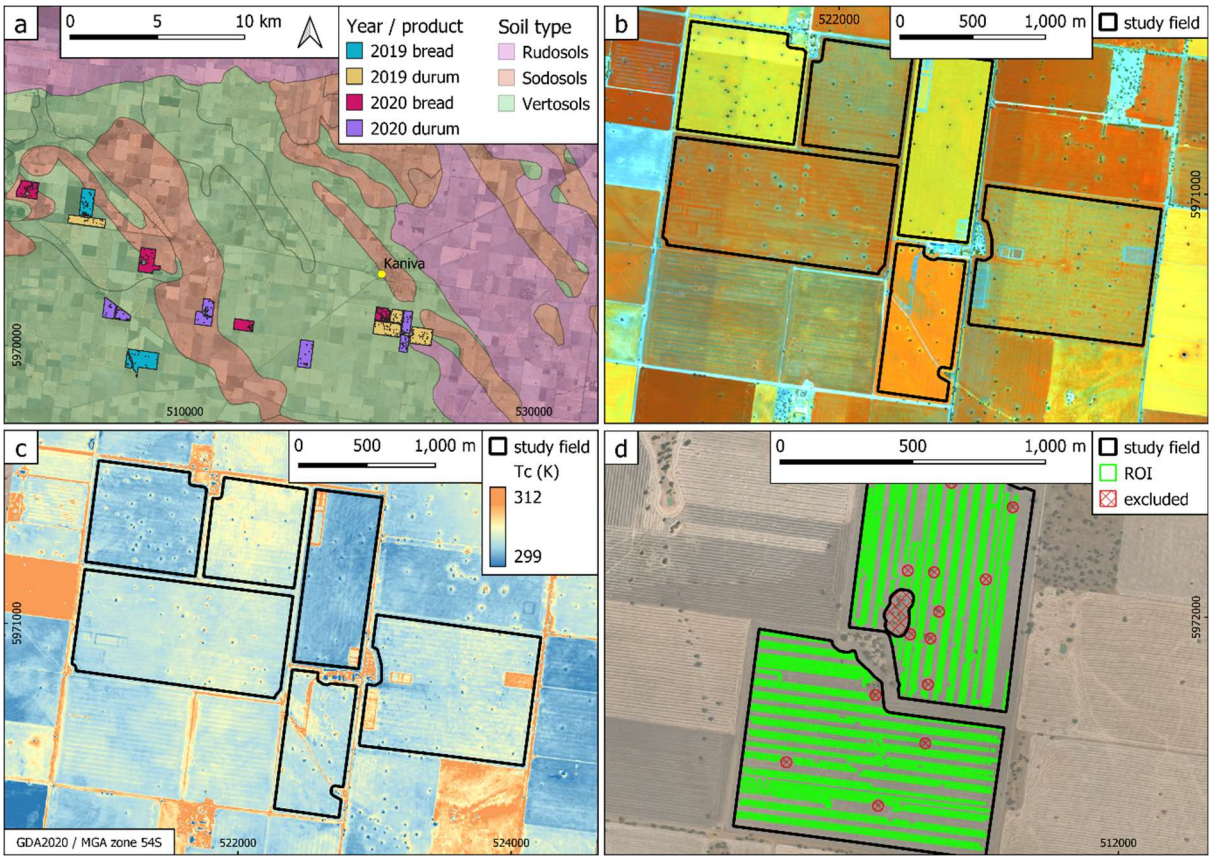

Fig. 3.6. Wheat fields in CZ1, 2019–2020, with soil types a), false-colour hyperspectral image (R = 775 nm, G = 710 nm, B = 678 nm; b), canopy temperature in Kelvin (c) and fields showing regions of interest (ROI) and areas excluded from analyses (d).

3.4.5. Performance metrics

Coefficient of determination (R^2) and root mean square error (RMSE) were used as primary model performance metrics. R^2 expresses the proportion of change in one variable is explained by change in another; a measure of correlation, in the current case between predicted and observed values of GPC. RMSE, a measure of model error, was calculated according to equation 4:

$$RMSE = \sqrt{\frac{\sum(P_i - O_i)}{n}}$$

where P_i is the predicted value of the i^{th} observation and O_i is the observed value for the i^{th} observation. Both were calculated using the package *caret* (Kuhn, 2020) in R (R Core Team, 2020), and RMSE was used to tune ML models. Relative RMSE (rRMSE) is also given for

GPC estimates and was calculated by dividing the RMSE by the range of the target variable. However, RMSE is chosen for emphasis below; this facilitates comparison with other studies because RMSE is in the units of the target variable, and because with respect to GPC, these are the units used by growers and at grain receival depots. Where 'model skill' is referred to below, this is considered to summarise both R^2 and RMSE.

3.5. Variance inflation factor analysis

To exclude collinearity between potential machine learning (ML) model inputs and to reduce data dimensionality, each year's airborne hyperspectral data were assessed by variance inflation factor analysis (VIF; R package *fmsb*; (Nakazawa, 2022)). Like other recent work, this study used VIF thresholds (t) of 5—10 (Akinwande *et al*., 2015; Magney *et al*., 2016a; Poblete *et al*., 2021; Zarco-Tejada *et al*., 2018). Multicollinear features were first excluded among the 88 NBHIs; those surviving (VIF < 5) then underwent a second VIF analysis, again at t = 5, in which all the inverted parameters Anth, C_{a+b}, C_{x+c}, C_m, C_w, LAI and LIDFa, plus CWSI and SIF, were forced inclusions. Indices retained after these analyses were then tested for their contribution to the ML model; those which did not improve GPC prediction in a small number of testing runs were discarded, and only EVI_{NIR} (hereafter 'EVI') and PRI were retained. These final, linearly independent features were categorised into ML input layers: physiological (Anth, C_{a+b}, C_{x+c}, SIF, PRI); structural (LAI, LIDFa); and thermal (CWSI).

The 38 VIs calculated from S2 reflectance were also subjected to VIF analysis. Initially considering only data from the last image in each timeseries for each permutation of site, year and wheat type, VIF was applied first among all VIs, then with forced inclusion of inverted parameters plus, for CZ1 only where these were available, CWSI and SIF. No VI was kept at t = 5 in the second step; those kept at t = 10 (

Table 3.6) were added stepwise as ML input features, to test for model skill improvements against otherwise equivalent models built only with inverted parameters. After this step, all S2 VIs were discarded because none substantially improved performance (R^2 contributions of ≤ 0.03 units). CWSI was excluded as collinear with the inverted parameter C_w. In contrast, SIF (CZ1 only) was independent of the inverted parameters and improved prediction, so was included. Inverted parameters only were then tested for multicollinearity among themselves, both along the timeseries and between features within dates. Minor collinearity was observed at t = 5 between like inverted traits from images captured less than 14 days apart but was allowed to persist so that feature importance dynamics along timeseries could be assessed.

3.6. Machine learning procedures

The gradient boosted machine (GBM) algorithm is asserted in the literature to accurately model complex, nonlinear relationships between numerous input features and a single target variable, notably with regard to wheat yield (Cao et al., 2020; Cheng et al., 2022; Ruan et al., 2022) but also in other fields of research. To test such claims in the current context, the GBM was assessed for its performance estimating GPC from plant traits: 12 different MLRAs were tested in parallel using the R package *tidymodels* (Kuhn and Wickham, 2020). For these tests, a diverse range of ML techniques (Table 3.7) was applied to subsets of GPC observations from commercial crops.

Table 3.7. Machine learning regression algorithms tested to establish relative performance in estimation of grain protein content from airborne hyperspectral and thermal plant traits in bread wheat crops in CZ1, 2020.

algorithm name	abbreviation	gain assessment	R package	reference
artificial neural network	ANN	no	*nnet*	Ripley (2023)
classification and regression trees	CART	no	*rpart*	Therneau and Atkinson (2023)
bagged classification and regression trees	CART (bagged)	no		
cubist	cubist	no	*Cubist*	Kuhn *et al*. (2023)
gradient boosting machine	GBM	yes	*xgboost*	Chen *et al*. (2021)
K nearest neighbours	KNN	no	*kknn*	Schliep *et al*. (2016)
full quadratic K nearest neighbours	quad. KNN	no		
quadratic linear regression model	quad. LM	no	*glmnet*	Friedman *et al*. (2010)
multivariate adaptive regression splines	MARS	no	*earth*	Milborrow (2023)
random forest	RF	yes	*ranger*	Wright and Ziegler (2017)
polynomial support vector machine	SVM (poly)	no	*kernlab*	Karatzoglou *et al*. (2004)
radial support vector machine	SVM (radial)	no		

Hyperspectral and thermal traits found through VIF analysis to be both linearly independent and robust across years were used as input features. The GBM was found to be consistently among the top-performing candidate algorithms as assessed by R^2 and RMSE (Fig. 3.7). Overall, its performance across combinations site, year and wheat types was stable and sufficient to justify its choice as the single MLRA on which to focus.

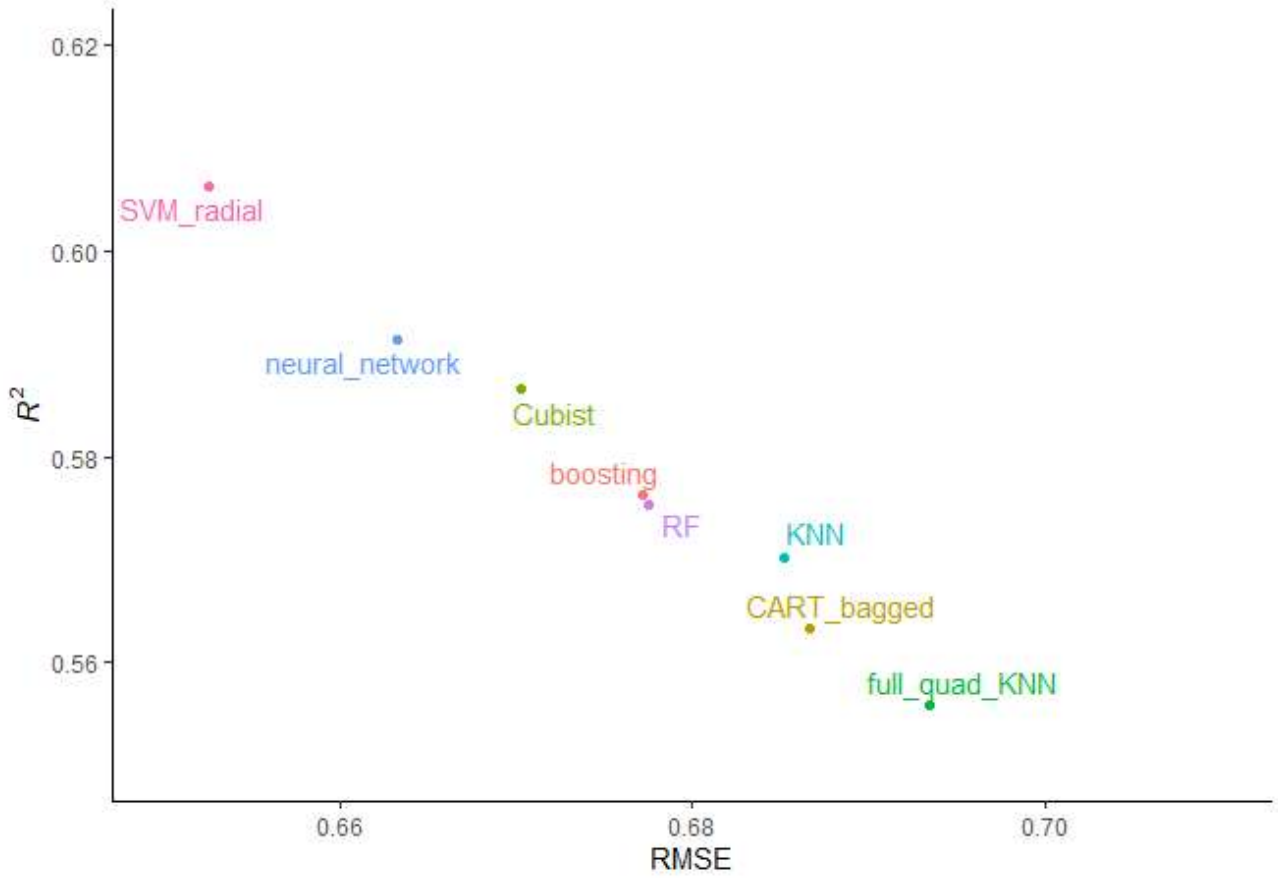

Fig. 3.7. Performance of machine learning algorithms in estimation of grain protein content from airborne hyperspectral and thermal plant traits in bread wheat crops at CZ1 in 2020. Several lower-performing algorithms are removed to allow zooming to more relevant ones, and to distinguish between the gradient boosting machine ("boosting") and RF.

In view of these results, and the capacity of the GBM to estimate the model gain contributions of its input features, this algorithm was applied to estimate, through supervised learning, relationships between leaf and canopy traits retrieved from hyperspectral and multispectral reflectance and the target variable, GPC. The primary focus was to assess input features' relative importance in GPC prediction (Chen and Guestrin, 2016), and secondarily to assess prediction accuracy in samples unseen during algorithm training. For the latter purpose, data were randomly split 70:30 into training and test sets for each run (Hunt *et al.*, 2019; Wu *et al.*, 2021).

Stochastic gradient descent (SGD) reduces the likelihood of overfitting by training models on random subsets of observations, introducing noise (Friedman, 2002). Here, SGD was implemented by training models on either 65% or 85% of rows, whereas all columns were considered in every run to prevent gain leakage. In addition to SGD, learning rate, tree depth and minimum node size were varied, with full factorial hyperparameter searching. Randomised K-fold cross-validation (K = 5) was also implemented as further protection against overfitting. In the tuning phase, the hyperparameter combination that minimised RMSE of prediction was adopted as the top model for each training set, then applied to predict test set GPC one final time. Data analysis and ML were done in R (R Core Team, 2020) using the *xgboost* (Chen *et al.*, 2021) and *caret* (Kuhn, 2020) packages for gradient boosting and model tuning.

Table 3.8. *xgboost* hyperparameters (Chen *et al.*, 2021) and values used to model grain protein content from canopy traits retrieved from hyperspectral and Sentinel-2 imagery in a gradient boosted machine. Values in italics were used for tuning and testing only.

hyperparameter	Name	Description/behaviour	Values used
`eta`	learning rate / shrinkage	Lower values increase robustness, increase run time	0.01, 0.025, *0.1, 0.25*
`max_depth`	maximum tree depth	Higher values increase tree complexity, the likelihood of overfitting, and the run time	*3, 4,* 5, 6
`min_child_weight`	minimum child weight / node size	Minimum sum of observation weights in a child. Lower values produce leaf nodes with fewer children; higher values reduce overfitting and increase run time but can also cause underfitting if too high.	*1, 3,* 5, 7
`subsample`	subsample ratio of instances	Proportion of training data (observations) sampled. Lower values reduce model complexity and introduce stochastic variability, reducing the risk of overfitting.	0.65, 0.8, 1
`colsample_bytree`	subsample ratio of columns	Proportion of training data (features) included. Lower values reduce model complexity and introduce stochastic variability but can affect feature importances.	*0.8, 0.9,* 1

Within each of hyperspectral and S2 datasets, the algorithm was run for each combination of zone, year and wheat type. The three hyperspectral models, comprising physiological +

structural + thermal (Model 1); physiological + structural (Model 2); and physiological inputs only (Model 3) were assessed separately for each of these combinations.

Regarding S2 traits, the GBM was first run for each site/year/product combination, using as inputs only the plant traits inverted from the image closest to the hyperspectral flight date, in each case the last of the timeseries. In the case of CZ1, airborne SIF was available for the end of the season and was added facultatively to assess its potential to improve model performance. Next, traits from each image capture date were analysed in turn, whereby the input feature set contained only C_{a+b}, C_m, C_w and LAI retrieved from the single relevant image. Finally, we combined the timeseries such that each permutation of inverted parameter and image date was used as an individual input feature, disregarding minor collinearity in the temporal dimension between inverted parameters. Thus, trait importance dynamics and skill were comprehensively assessed through seasons, first with TS elements as separate models, then additively stacking them within site-years to form single predictive models.

4. Results

This chapter contains included material from:

Longmire, A.R., Poblete, T., Hunt, J.R., Chen, D., Zarco-Tejada, P.J., 2022. Assessment of crop traits retrieved from airborne hyperspectral and thermal remote sensing imagery to predict wheat grain protein content. *ISPRS Journal of Photogrammetry and Remote Sensing 193, 284–298*. https://doi.org/10.1016/j.isprsjprs.2022.09.015

Longmire, A.R., Poblete, T., Hornero, A., Chen, D., Zarco-Tejada, P.J., (accepted 2023-10-27). Estimation of grain protein content in commercial bread and durum wheat fields via traits inverted by radiative transfer modelling from Sentinel-2 timeseries. *ISPRS Journal of Photogrammetry and Remote Sensing*.

—

4.1. Plot experiments

Both experimental sites had rainfall substantially below average in 2019, and this was seen in crop performance and GPC. Although growing season rainfall (GSR) was only 65% of long-term AAR at experimental site 1 (ES1), these plots started with good soil moisture remaining from heavy falls in the summer of 2018-2019 and were in good condition at the time of concurrent field work and remote sensing. Experimental site 2 (ES2) was more severely droughted: GSR was only 54% of AAR and starting soil moisture was very low. By the time of data collection, ES2 plots were under severe water stress, with visible signs of premature senescence. Combined with soil moisture conditions, the treatments applied at the two experimental sites produced strong GPC gradients, parallel with fertiliser dosing. The two sites were sown to the same cultivar, but GPC was in a higher range at ES2, reaching 14.9 % while the maximum at ES1 was < 12 % (Fig. 4.1, Fig. 4.2c-f). The ES2 GPC response also saturated at heavy N applications, and at the highest level of 184 kg.N ha^{-1} was lower than the next highest N treatment (Fig. 4.1, Fig. 4.2c, f).

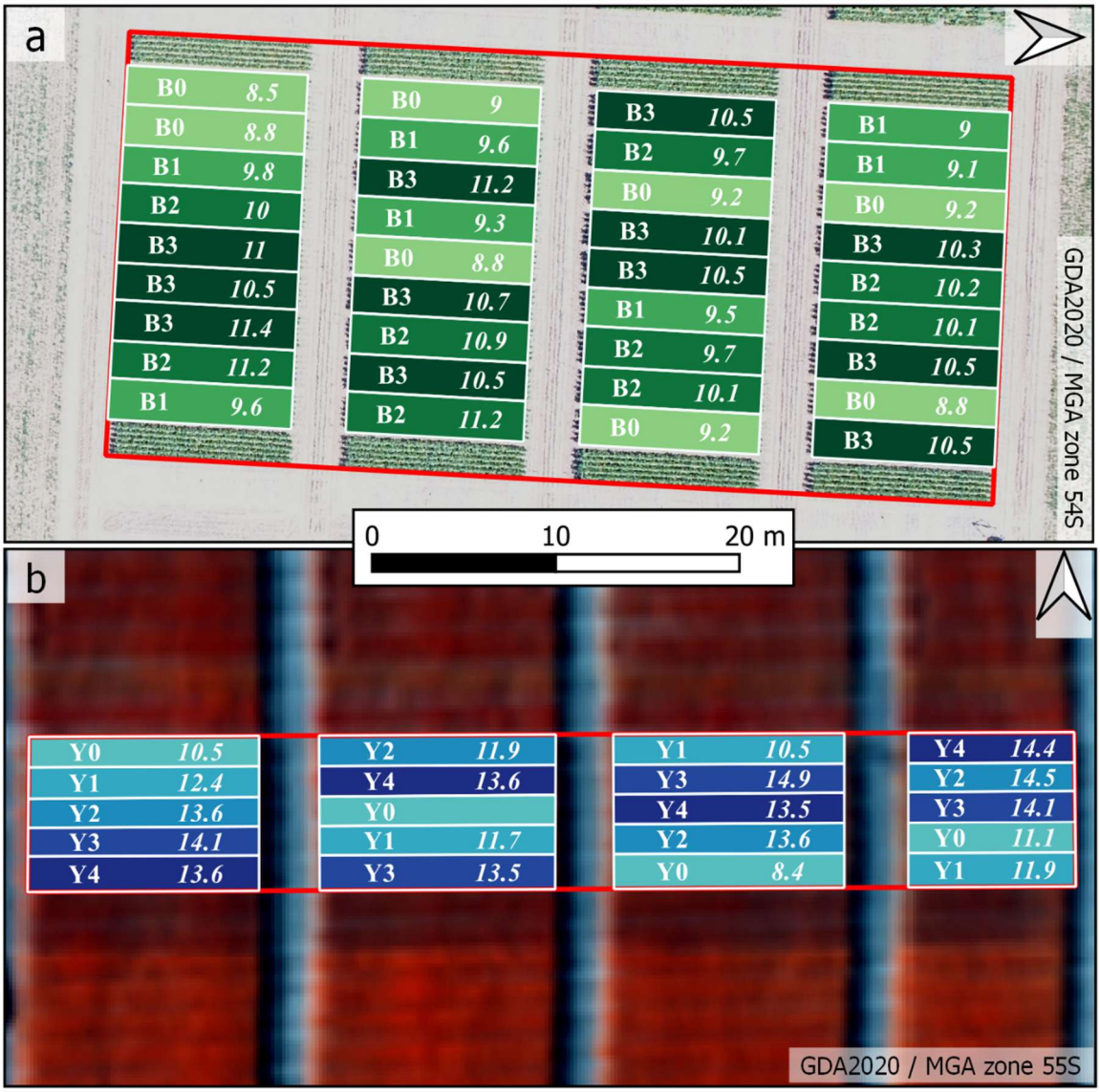

Fig. 4.1. Plots at ES1 (a) and ES2 (b) with treatment (left of plots) and GPC (%, in italics; right of plots) over false-colour hyperspectral image (R = 749 nm, G = 710 nm, B = 678 nm. Colours represent treatments (kg.N/ ha): B0 = 0, B1 = 30 – 37, B2 = 98 – 104, B3 = 162 – 171; Y0 = 0, Y1 = 46, Y2 = 92, Y3 = 138, Y4 = 184.

The leaf-level reflectance indices NPCI, VOG1, ZMI, R_{920}/R_{729} and PRI_{m3} increased with fertiliser N availability at both sites (Fig. 4.2a, d). This trend, parallel with GPC, was distinct at ES1 but was less distinct and showed saturation at ES2. Indices derived from airborne images also showed strong responses to fertiliser rate at ES1, but weaker responses at the water stressed ES2 (Fig. 4.2b, e). The CI, PRI_{m4} and, as at leaf level, R920/R729 responded to fertiliser N more strongly at the unstressed ES1. The responses of structural indices NDVI

and EVI also followed the N gradient but saturated after the first treatment level at ES2 and after the second at ES1. Dualex C_{a+b} and Nitrogen Balance Index (NBI) were similar between ES1 and ES2, though with higher absolute readings at ES1. In each case, these were positively associated with GPC; Dualex Anth was instead inverted against GPC. Leaf N from Dumas combustion generally increased parallel with N treatment at both sites but, in contrast to C_{a+b} and NBI at the same site, declined in treatments Y4 and Y5. Across sites and observation levels, there was close agreement between leaf level PRI_{m3} and canopy PRI_{m4}, with leaf level readings lower at both sites (Fig. 4.2a, d.). Such similarities were also seen between leaf- and canopy levels, and across sites, in R_{920}/R_{729}. Each version of the PRI_{m*} followed GPC and leaf N from both optical and destructive methods.

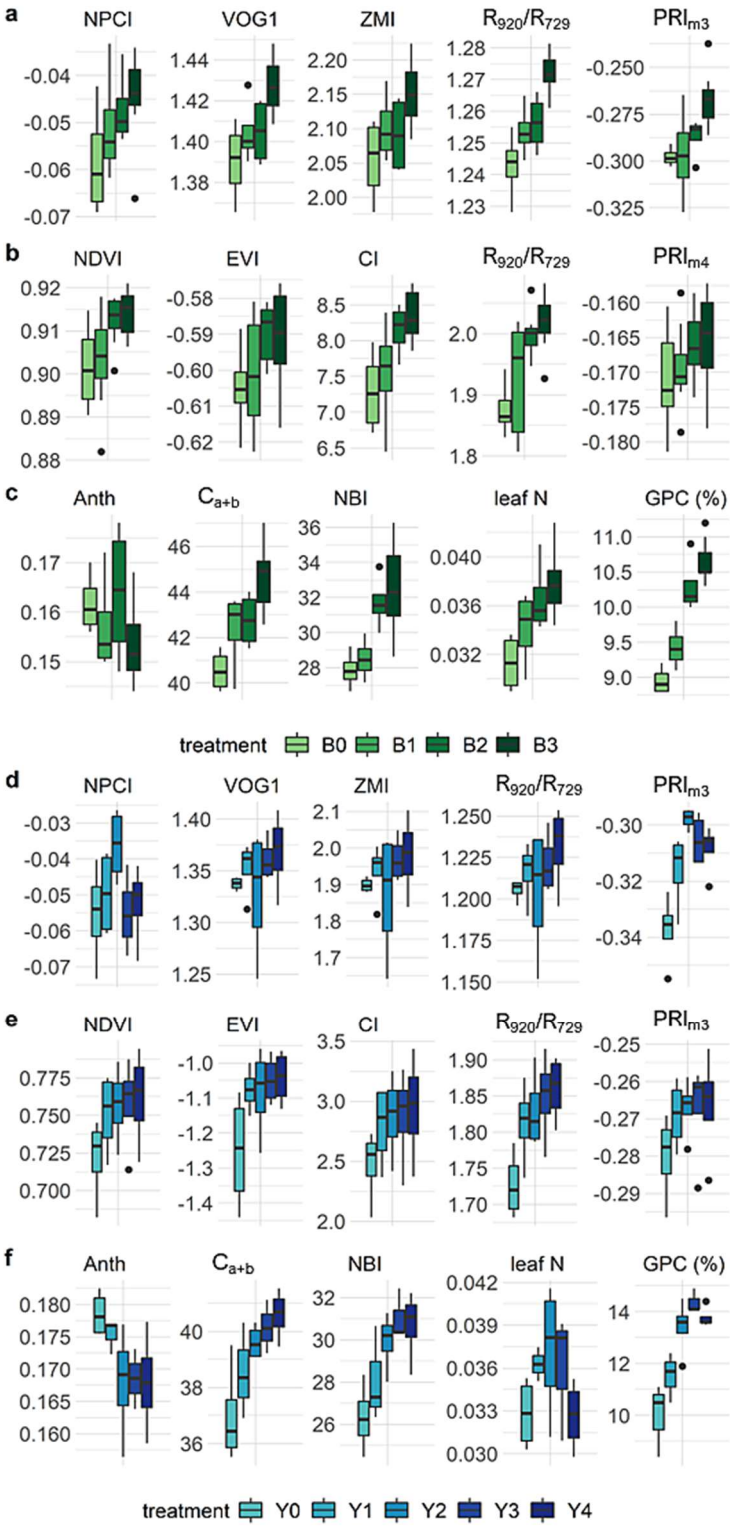

Fig. 4.2. Vegetation indices at leaf (a, d) and canopy level (b, e) and ground-truth indicators (c, f) by treatment at ES1 (upper) and ES2 (lower). At ES1, n = 20 for ground observations (a) and n = 36 for airborne indices (c); at ES2, n = 19. Anth, C_{a+b} and NBI in Dualex proprietary units; leaf N in mg.N/cm^2.

Parameters retrieved by PRO4SAIL inversion from airborne HS data were generally aligned with those from proximal spectroscopy and destructive sampling (Fig. 4.3). Increases along the nutrient gradient in inverted HS C_{a+b} aligned with those seen in Dualex C_{a+b}, and these two measures were significantly correlated ($R^2 = 0.61$, $p < 0.0001$). C_{a+b} from both sources was also parallel with GPC across both sites, though with some saturation at higher N rates at ES2, where C_{a+b} concentrations were also in a lower range (Fig. 4.3a, b). Despite saturating after one treatment level, C_{x+c} also followed N rate, parallel to GPC at ES2 but not at ES1. Airborne inverted Anth trended higher with N treatment only at ES1, where they were also in higher concentration overall, but like Dualex Anth at both sites (Fig. 4.2c, f), the Anth~GPC relationship was reversed at ES2. Inverted LAI was similar to inverted Anth, in that it increased with N treatment at ES1, but declined at ES2, where LAI values were also very high and showed saturation. This very high foliage density was also visually evident. No treatment response was evident in SIF at ES1, while at ES2 SIF reduced, against GPC, with higher N. The range of CWSI was higher at the severely stressed ES2 site. At ES1, the following relationships were statistically significant: Anth ($R = 0.61$, $p < 0.01$), C_{a+b} ($R = 0.63$, $p < 0.01$) and CWSI ($R = -0.52$, $p < 0.05$). At ES2 these were: Anth ($R = -0.62$, $p < 0.01$) and LAI ($R = -0.55$, $p < 0.05$).

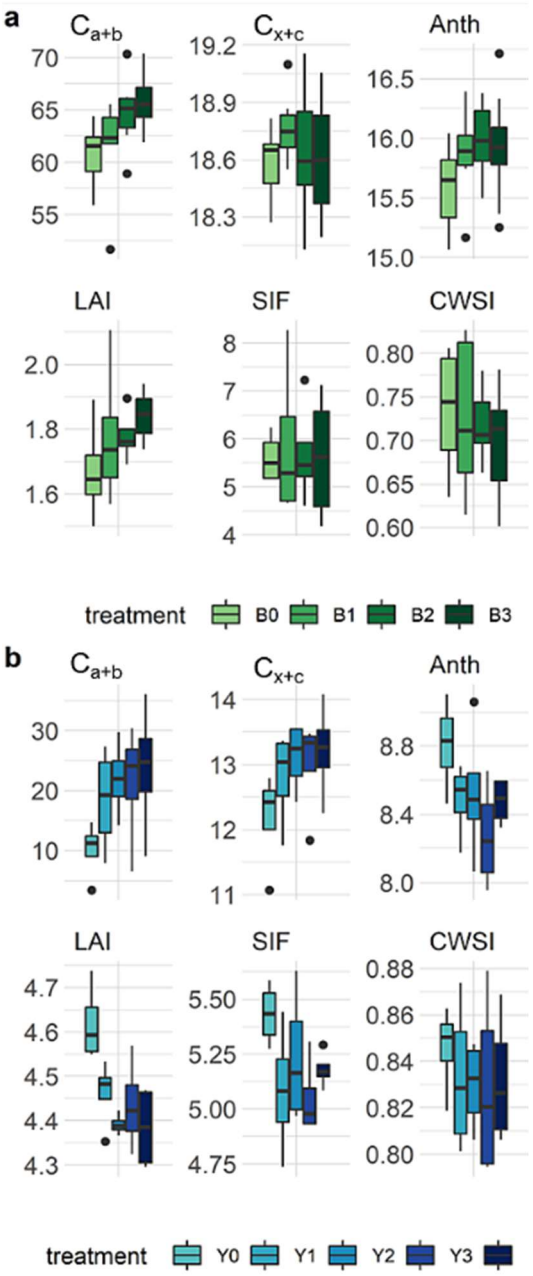

Fig. 4.3. Chlorophyll, Carotenes, Anthocyanins and LAI (C_{a+b}, C_{x+c}, Anth; µg/cm^2, LAI; m^2/m^2), retrieved by radiative transfer model inversion, plus solar-induced fluorescence (SIF; mW/m^2/ sr/nm) and crop water stress index (CWSI) from airborne hyperspectral and thermal data at ES1 (a; n = 36) ES2 (b; n = 19).

4.2. Field transect

The M01 crop retained soil moisture during RS campaigns, conducted 1592 GDDAS, but despite this later time in the crop cycle, plants had not senesced as much as those in the ES2 plots (images captured 1409 GDDAS at ES1 and 1559 GDDAS at ES2). GPC at the transect was in a range intermediate between ES1 and ES2, and K-means clustering (K = 3) based on GPC produced class separation adequate for analysis of relationships between retrieved leaf and canopy quantities (Fig. 4.4). Dualex NBI showed a markedly similar trend to GPC, as did soil mineral N from Dumas combustion of sampled topsoil (Fig. 4.4d). At leaf level, VOG1, ZMI, R_{920}/R_{729} and PRI_{m3} trended with GPC, showing some saturation, while NPCI showed little association (Fig. 4.4a). Canopy level NDVI, EVI, CI, R_{920}/R_{729} and PRI_{m4} calculated from airborne data all trended with GPC (Fig. 4.4b). As in the plot studies, the two PRI_{m*} versions agreed across observational scales, while the structural NDVI and EVI were also aligned with GPC. Regarding terminal drought, conditions in the sample field appeared intermediate between the plot sites, and this was confirmed by C_w and CWSI values. Among parameters inverted from airborne HS data, C_{a+b} were negatively associated with GPC, while C_{x+c} and Anth had positive relationships with GPC (Fig. 4.4c). In common with plot studies, CWSI, C_w and SIF (not shown) showed no relationship with GPC, but the ranges of both CWSI and C_w were intermediate between the plot sites.

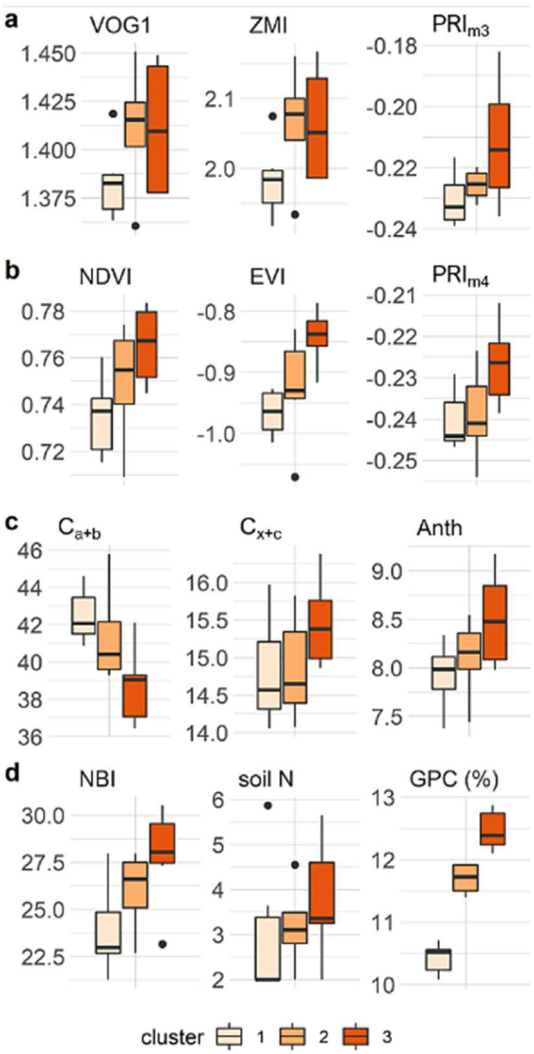

Fig. 4.4. Leaf-level (a) and airborne (b) reflectance indices; Chlorophyll, Carotenes, and Anthocyanins (C_{a+b}, C_{x+c}, Anth; µg/cm^2), retrieved by radiative transfer model inversion (c) and ground-truth indicators Dualex Nitrogen Balance Index (NBI), soil N (mg/kg; d), and grain protein content (GPC; %; d) by GPC-based k-means clusters in a transect of commercial wheat field M01 (n = 20).

4.3. Commercial crops

4.3.1. Growing conditions and GPC

In copping zone 1 (CZ1), this study incorporates both durum and bread wheat cultivars, while only bread wheat was sown in CZ2 during the period of study. Hyperspectral data were collected in CZ1 only. Growing conditions differed substantially between the two years, strongly affecting the commercial crops at both locations. Both total and growing season rainfall were extremely low in 2019, in CZ1-19, in the lowest decile of long-term AAR records (280 mm) while CZ1-20 was at the long-term average (444 mm) and GSR reached almost 300 mm. Moreover, rainfall accumulated in the relevant pre-season (December 2018—May 2019 was also low (87 mm) compared to the subsequent equivalent period (164 mm) in CZ1. In CZ2 also, both total and growing season rainfall were extremely low in 2019 while in 2020 they were close to long-term averages (

Table 2.1). Hyperspectral RS campaigns took place 1514 GDDAS (bread wheat) and 1736 GDDAS (durum) in 2019 when crops were under severe terminal water stress. In 2020, HS data capture was done 1592 GDDAS (bread wheat) and 1742 GDDAS (durum). Around anthesis 2019 in CZ1, conditions changed suddenly: frost (≈- 4 °C) was observed on 9 October, and maxima of > 35°C were recorded on several days later in the same month. As detailed in the theoretical basis to this work, such a contrast in moisture conditions can have large and varied effects on grain protein. Descriptive statistics for GPC across all combinations of years, zones and wheat types are given in Table 4.1.

Table 4.1. Statistical summary of commercial zone / year / wheat type combinations.

			GPC				
Zone	**Year**	**Wheat type**	**Obs (n)**	**Mean**	**SD**	**Median**	**IQR**
1	2019	bread	7114	11.63	1.52	11.20	2.66
		durum	5013	11.85	1.18	11.84	1.46
	2020	bread	8075	11.90	0.97	11.91	1.26
		durum	9541	12.81	0.93	12.84	1.22
2	2020	bread	13976	10.59	2.49	10.70	3.67
		bread	18650	11.54	1.05	11.46	1.27

Wilcoxon tests (Bauer, 1972) showed differences in GPC between all zone/year/product combinations; while at the scale of cropping zones, effect sizes were small to moderate, differences were highly statistically significant (Table 4.2). Significant differences were also seen when wheat products were compared between seasons. Bread wheat GPC was higher in 2019 (mean = 11.6, SD = 1.52) than 2020 (mean = 11.3, SD = 1.05; Wilcoxon's $p < 0.0001$, effect size $r = 0.489$). Mean durum wheat GPC was higher in 2020 (mean = 12.7, SD = 0.94) than in 2019 (mean = 11.9, SD = 1.18; Wilcoxon's $p < 0.0001$, $r = 0.360$). GPC was significantly higher in durum than bread wheat in each year also (2019: $p < 0.0001$, $r = 0.112$; 2020: $p < 0.0001$, $r = 0.564$). In a subset of the data, hyperspectral PRO4SAIL C_{a+b} and CI were strongly and significantly correlated ($R^2 = 0.86$, $p < 0.0001$).

Table 4.2. Wilcoxon test statistics for all commercial zone / year / wheat type combinations.

comparator 1				comparator 2				effect			
zone	**year**	**wheat**	**obs (n)**	**zone**	**year**	**wheat**	**obs (n)**	**size**	**desc.**	**p-value**	**sig.**
1	2019	bread	7114	1	2019	durum	5013	0.111	small	1.19E-34	****
					2020	bread	8075	0.136	small	1.76E-63	****
				2	2019		13976	0.200	small	2.22E-185	****
					2020		18650	0.021	small	6.01E-04	***
		durum	5013	1	2020	durum	9541	0.406	mod.	0	****
	2020	bread	8075					0.438	mod.	0	****
				2	2019	bread	13976	0.289	small	0	****
					2020		18650	0.183	small	1.56E-197	****
2	2019	bread	13976					0.207	small	1.28E-305	****

Substantial differences in GPC were also seen both within and between fields for each site, zone and wheat type, often over short distances (Fig. 4.5, Fig. 4.6). Dune-swale morphology in CZ2 is also evident in the Sentinel-2 RGB sections of Fig. 4.6b.

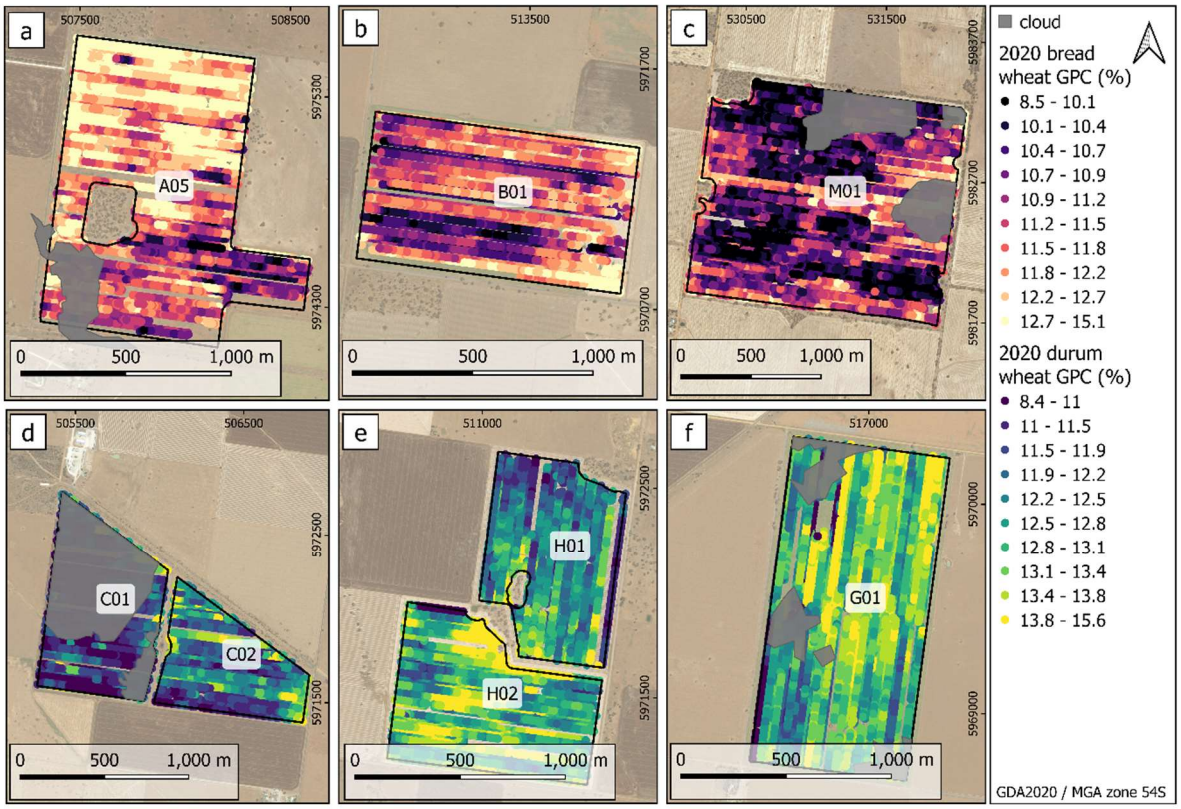

Fig. 4.5 Spatial variability in grain protein content (GPC; %) in bread (a—c) and durum wheat (d—f) fields in CZ1, 2020.

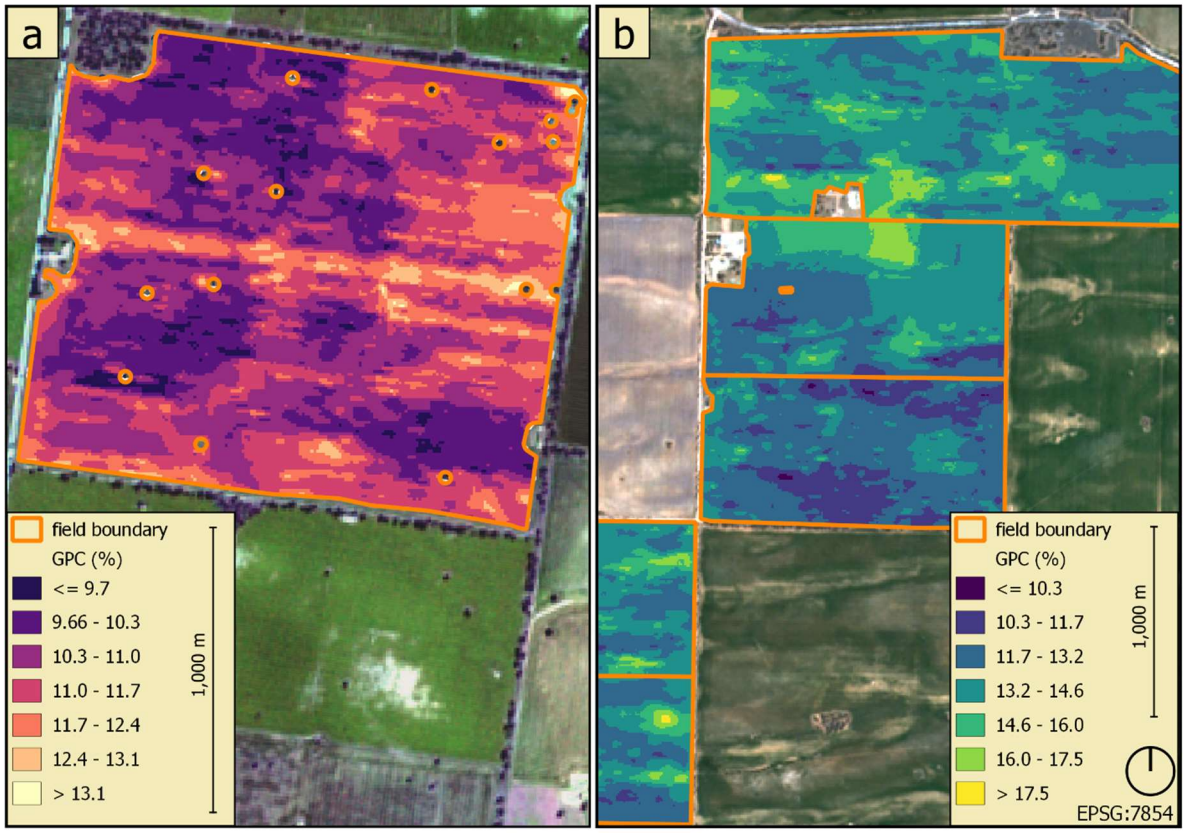

Fig. 4.6. Spatial variability in grain protein content (GPC; %) in wheat fields in CZ1 (a) and CZ2 (b) displayed on Sentinel-2 images (R = band 2, G = band 3, B = band 4).

4.3.2. Variance inflation factor analysis

Of the 88 HS reflectance indices tested, only EVI and PRI were selected after VIF and proved robust across both years of the study; they were therefore retained as ML input features. EVI and PRI were both also correlated with GPC at plot scale. When VIF was applied to S2 features, within the respective single final image capture dates and with inverted parameters forcibly retained, all 38 VIs were excluded in every site/year/wheat type permutation, either because of multicollinearity with at least one inverted parameter or because they did not add model skill. VIF, therefore, removed only redundant features and with only minor effects on model performance; Sentinel-2 VIs were hence not further considered in analyses. Neither CWSI nor SIF can be calculated from S2 data; for CZ1 these were taken from airborne RS. CWSI was also excluded, but SIF retained, at VIF threshold

t = 10, so was included in later analyses. Regarding timeseries, S2 inverted plant traits from each date per season were treated as separate entities, and the entire set subjected to independent VIF tests in two modes: i) across traits, within dates and ii) across dates, within traits. In these analyses, no feature was removed by test ii) that wasn't also removed by test i), so test 1 outputs were retained. Minor multicollinearity (t = 10) between like plant traits in consecutive images was tolerated in order to inspect trait dynamics through time.

4.3.3. Crop-scale airborne hyperspectral indices, CWSI and SIF

Thermal and HS traits estimated on a single date for each GPC point demonstrated distinct and sometimes wide variability at within-field scale for both bread and durum wheat, spatial associations between traits, and variability in these spatial associations. Two study fields from images captured in CZ1-19 are shown in Fig. 4.7. In bread wheat, high Anth, SIF, C_{x+c} and, to a lesser extent, PRI were associated with relatively low CWSI, while C_{a+b} appeared relatively weakly associated with areas of greatest stress (Fig. 4.7a). Durum wheat underwent less severe water stress; higher C_{x+c}, PRI and SIF were associated with areas of low water stress, while Anth and C_{a+b} showed a positive spatial association with stress; again, the C_{a+b}~CWSI association was relatively weak (Fig. 4.7b).

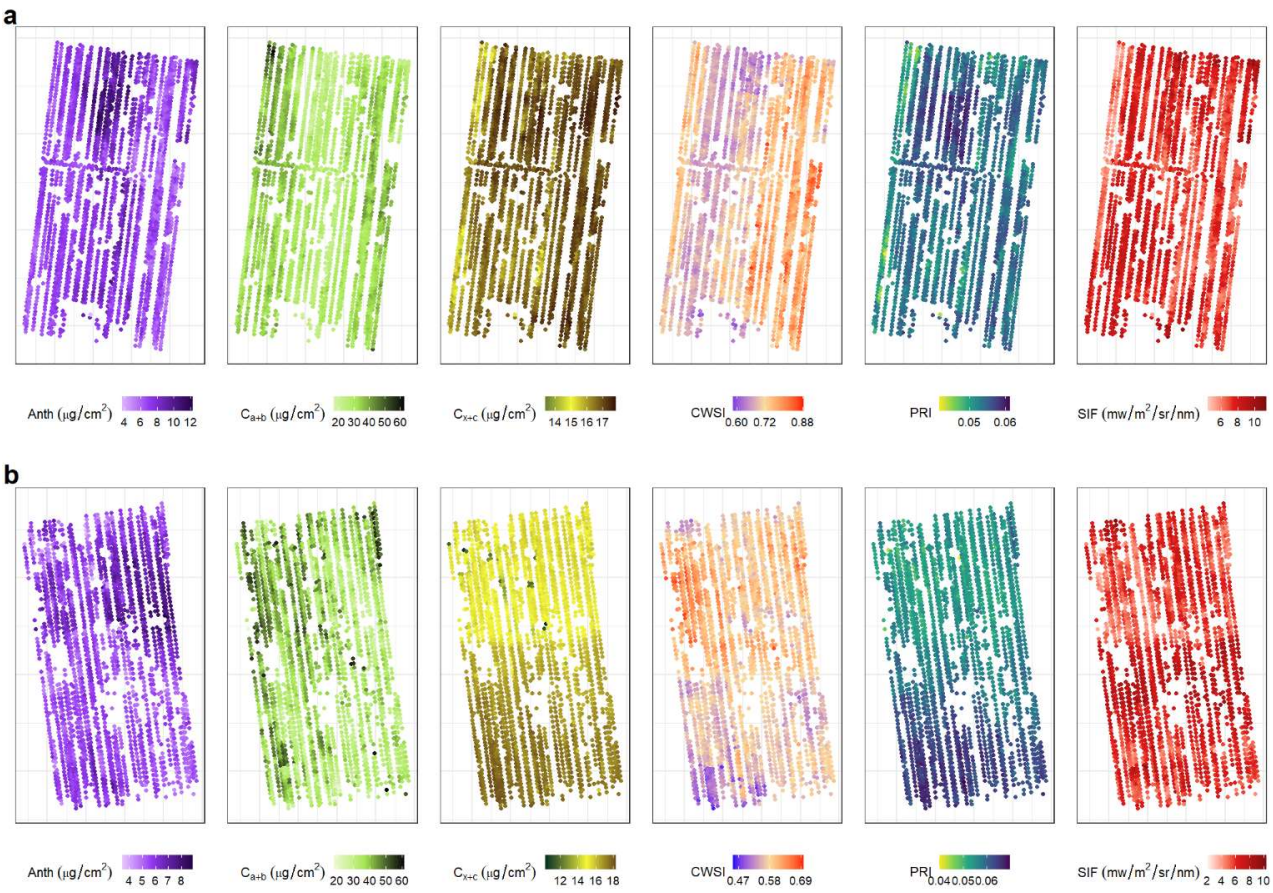

Fig. 4.7. Spatial variability in anthocyanins (Anth), chlorophyll (C_{a+b}), carotenes (C_{x+c}), crop water stress index (CWSI), photochemical reflectance index (PRI), and solar-induced fluorescence (SIF) retrieved from hyperspectral and thermal images captured over fields sown to bread (a; n = 2242) and durum (b; n = 2017) wheat in CZ1, 2019.

Trait density distributions (Fig. 4.8) further clarify the differences in water stress undergone at the same two fields, one bread and one durum wheat, in the same year (2019) and despite their proximity (~10 km) and similar planting dates. Bimodal distributions were common in retrieved hyperspectral traits across the study and were seen in CWSI for bread wheat and in C_{x+c} and PRI for durum wheat in the current example (Fig. 4.8).

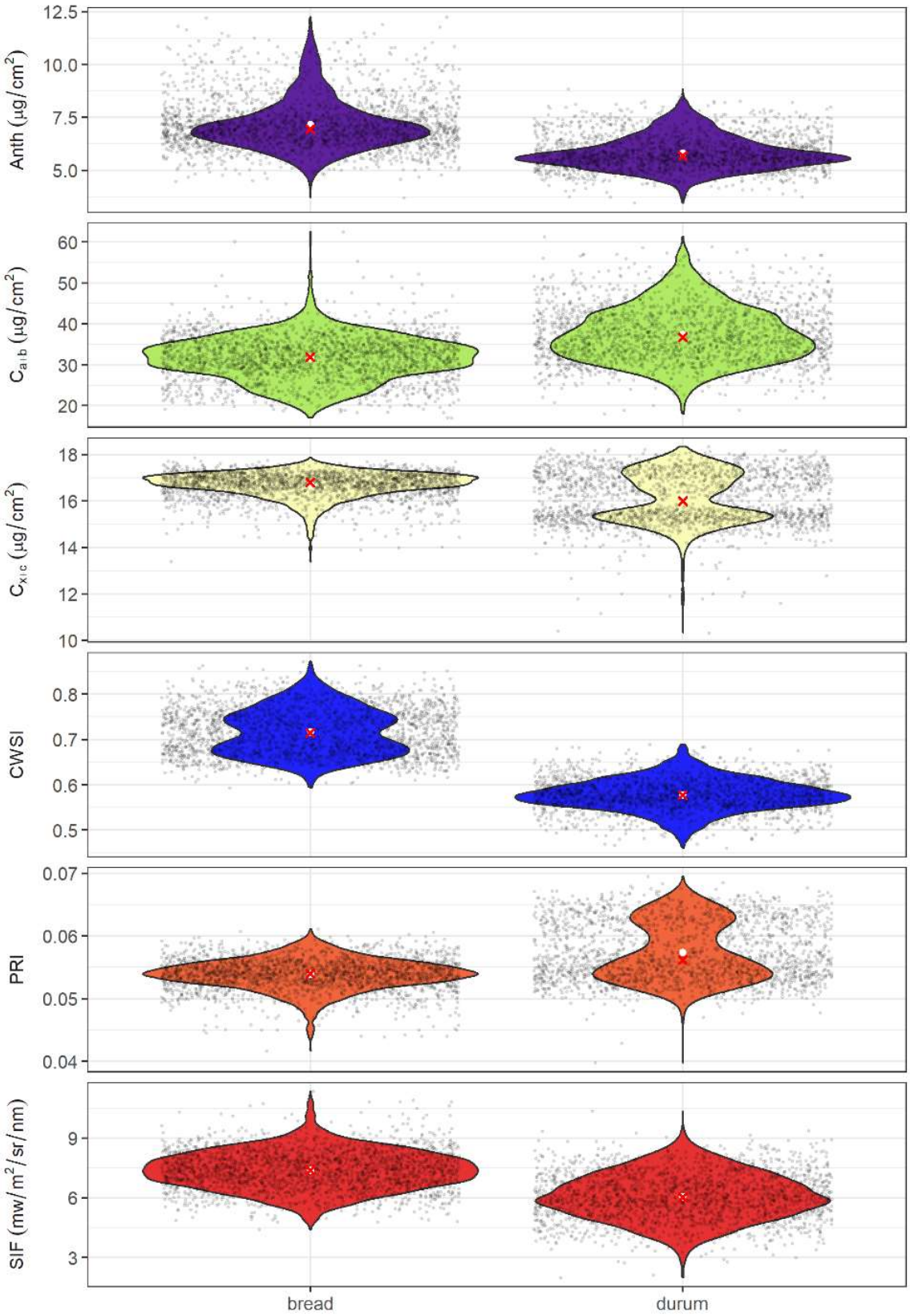

Fig. 4.8. Frequency distribution of anthocyanins (Anth), chlorophyll (C_{a+b}), carotenes (C_{x+c}), crop water stress index (CWSI), photochemical reflectance index (PRI), and solar-induced fluorescence (SIF) retrieved from hyperspectral and thermal images captured over example fields sown to bread (n = 2242) and durum (n = 2017) wheat in CZ1, 2019. Violin plots show data distribution, white circle = mean, red cross = median.

4.3.4. Crop-scale Sentinel-2 radiative transfer modelling

Trait estimations for each GPC point across seasonal timeseries showed seasonal progression and spatial heterogeneity across site/year/wheat type combinations. Example fields from CZ1-19 are plotted to map traits over time for bread wheat (Fig. 4.9) and to show progression with phenological advance, relationships between traits and density distributions within trait and stage combinations in durum (Fig. 4.10).

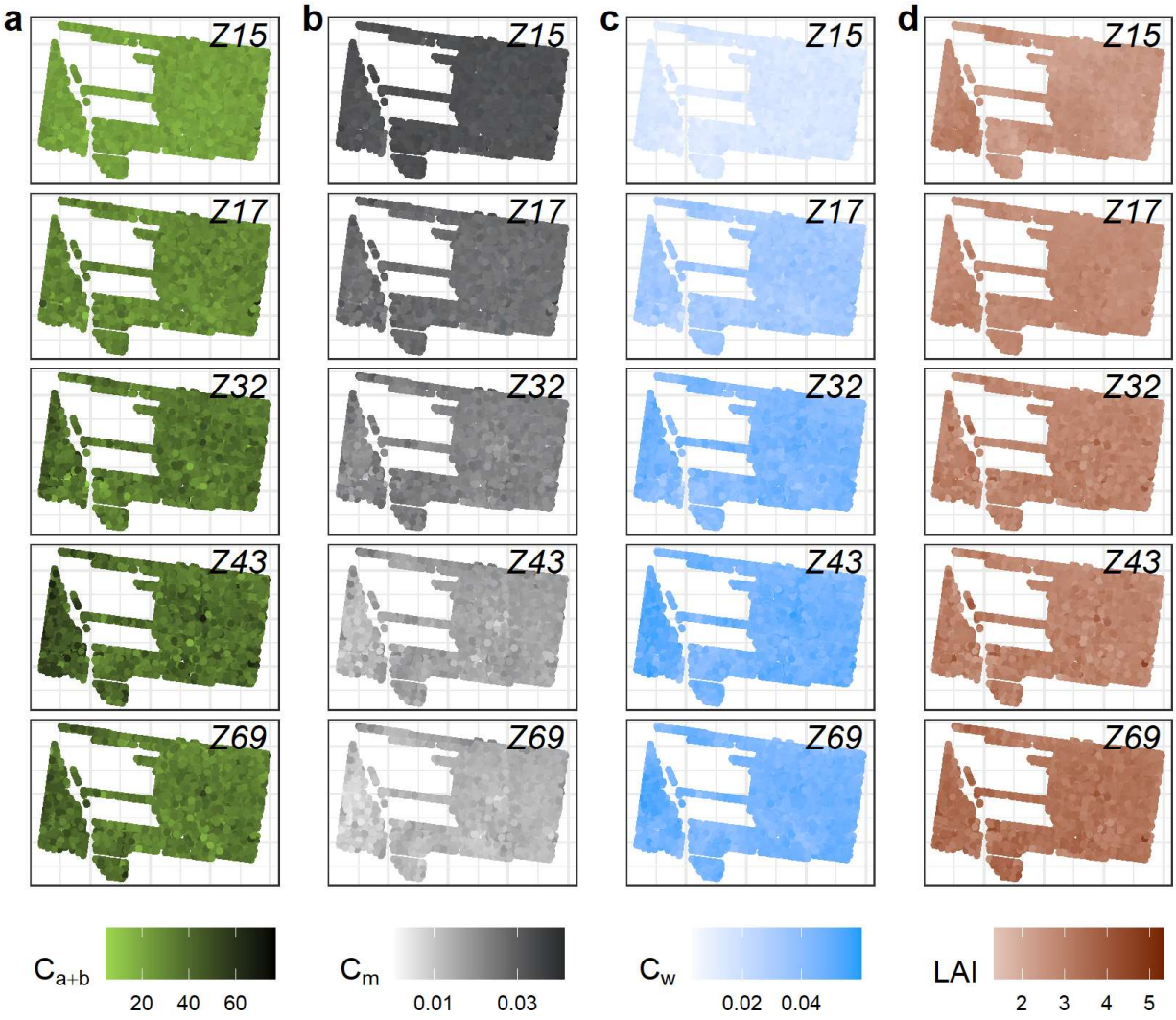

Fig. 4.9. Spatial variability in chlorophyll (C_{a+b}; µg/cm²; a), leaf dry matter (C_m; g/cm²; b), leaf water content (C_w; g/cm²; c) and leaf area index (LAI; m²/m²; d) retrieved by radiative transfer model inversion from Sentinel-2 image timeseries captured over a field sown to bread wheat in CZ1, 2019 (n = 4897). Plots arranged by Zadoks (Z) growth stage from top down.

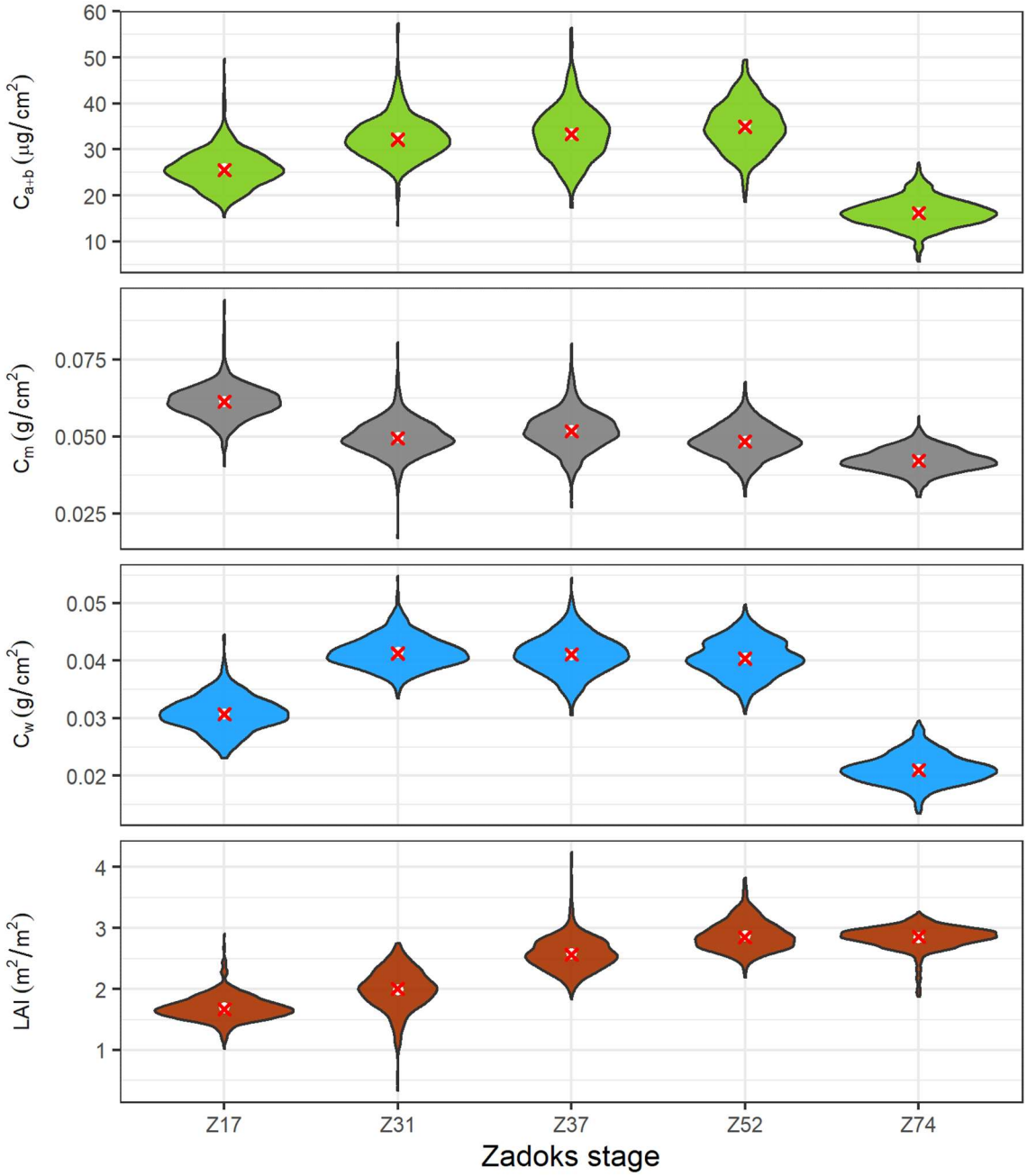

Fig. 4.10. Frequency distribution of chlorophyll (C_{a+b}), leaf dry matter (C_m), leaf water content (C_w) and leaf area index (LAI) retrieved by radiative transfer model inversion from Sentinel-2 image timeseries captured over a field sown to durum wheat in CZ1, 2019 (n = 782). Violin plots show data distribution within Zadoks (Z) growth stage; white circle = mean, red cross = median.

4.3.5. Hyperspectral feature importance

The relative importance to GPC estimation of plant traits retrieved from airborne HS data was quantified for input layers comprising: (i) physiological indicators Anth, C_{a+b} and C_{x+c} from

model inversion, SIF and the photochemical reflectance index (PRI); (ii) structural indicators EVI, LAI and LIDFa; and (iii) the thermal-based CWSI. These inputs were supplied to the GBM algorithm as separate models: physiological + structural + CWSI (Model 1); physiological + structural (Model 2); physiological (Model 3).

The thermal CWSI contributed 69% of total information to model 1 in moisture-stressed bread wheat in 2019. The structural reflectance index EVI contributed 13% and the rest was split among other physiological and structural features. Chlorophyll was of minor importance (Fig. 4.11a). In unstressed 2020 bread wheat crops, the same model showed more even gain distribution, in which carotenoids (C_{x+c}) brought > 20% of total gain and CWSI was second but contributed only marginally more than any other single component. In the stressed 2019 bread crops, LAI then C_{x+c} contributed most to model 2, which lacked CWSI, while C_{x+c} maintained high importance in model 3, below C_{a+b} under stress but substantially more important in milder conditions. In bread wheat, 2020, C_{x+c} was the most important component regardless of model, accumulating proportional importance as thermal then structural features were removed (Fig. 4.11a-c). Anth were not prominent components of model 1 but were grouped with the remaining physiological features in both wheat types.

Where model 1 was used for durum wheat GPC, CWSI was consistently important across the two study years, although all nine inputs shared importance in a range closely clustered around 9—14 % (Fig. 4.11d). Also in durum, SIF ranked highly in both years and all models and where CWSI was absent, SIF was the largest contributor while all components were again in a tight range. In 2020 durum, SIF was quasi-equivalent to PRI and C_{x+c} and was top ranked in durum models 2 and 3. The structural EVI, LAI and LIDFa were grouped in the lower ranks of importance to 2019 durum. In model 2, importance was evenly distributed and ordinal rankings were largely stable across inputs and years (Fig. 4.11e): SIF consistently contributed the most gain and the structural components lowest. Anth and C_{a+b} ranked highly

in 2019 but somewhat lower in 2020, when PRI and C_{x+c} were relatively higher; nevertheless, these physiological elements remained closely grouped in all models (Fig. 4.11d-f).

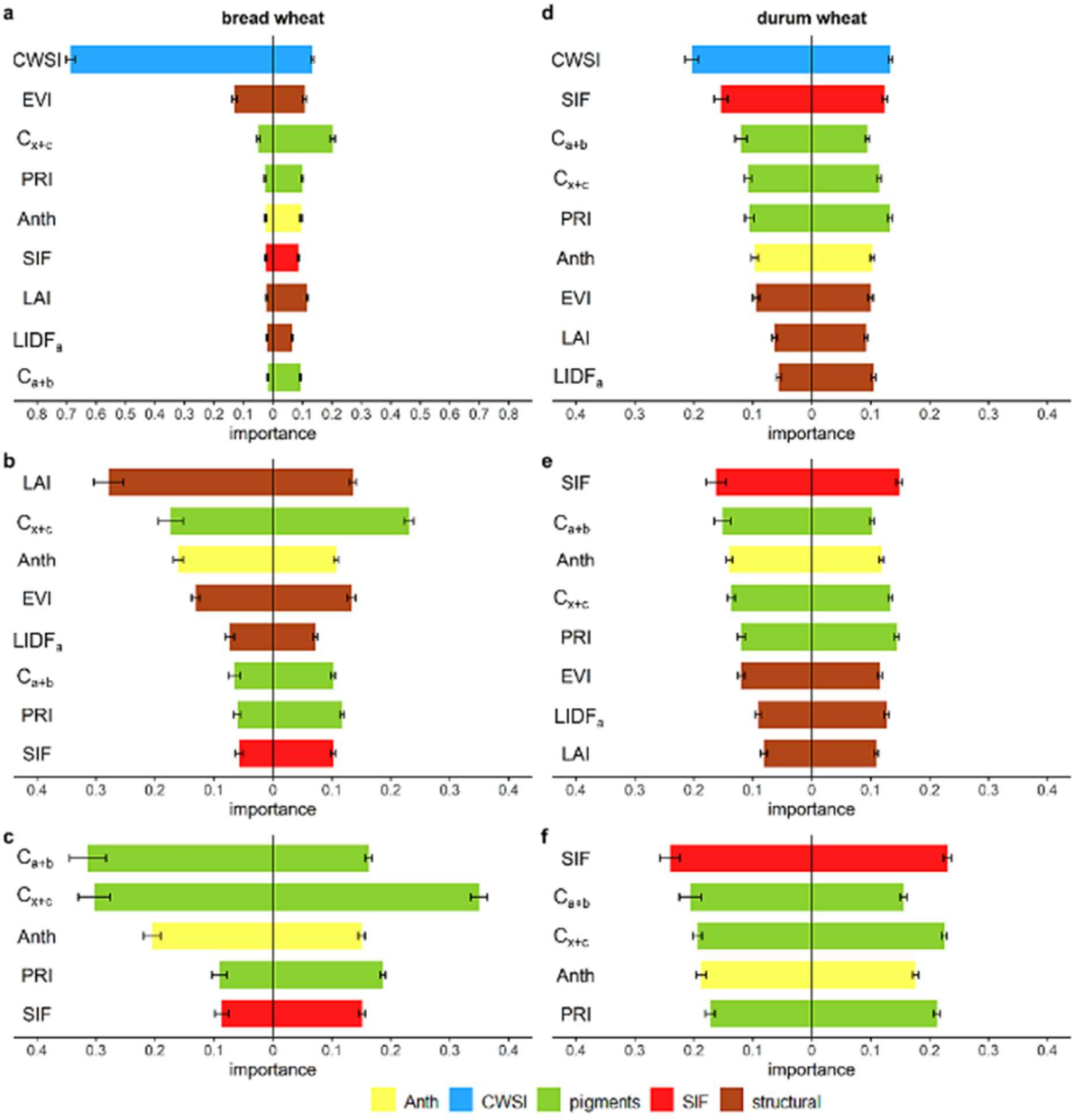

Fig. 4.11. Relative importance (proportion) of input features to a gradient boosting machine estimating grain protein content in commercial bread (a, b, c) and durum wheat (d, e, f) fields in CZ1, 2019–2020. Three models are shown: physiological + structural + CWSI (model 1; a, d); physiological + structural (model 2; b, e); physiological (model 3; c, f). Each sub-figure represents 2019 (left; bread n = 7213, durum n = 5030) and 2020 (right; bread n = 11060, durum n = 17310). Error bars show standard deviation of the mean proportional importance over 80 runs.

4.3.6. Sentinel-2 feature importance

Sentinel-2 plant trait importance results are presented in separate sections, first with a survey of S2 traits and comparisons against hyperspectral traits, then models constructed with plant traits from individual timeseries images, and finally models containing traits from the entire TS. Models were first built with traits from single cloud-free S2 images acquired on dates as close as possible to the airborne HS missions; for 2019, this was the day after the HS mission, while in 2020 the closest available was 18 days earlier. Fluorescence was facultatively added to each CZ1 model. Overall, relative S2 feature importance mirrored hyperspectral importance across conditions and wheat types and remained stable when SIF was added. The S2~hyperspectral similarity is marked when the traits are grouped by indicator type: S2-derived C_{a+b} occupied similar importance ranks as hyperspectral C_{a+b}, C_{x+c} and PRI; the structural indicators S2 C_m and hyperspectral EVI are well-aligned, as are LAI (S2) with hyperspectral LAI and $LIDF_a$ and C_w with CWSI (Fig. 4.12).

Leaf water thickness (C_w) dominated model gain in the severe water stress of 2019 CZ1 bread wheat. By contrast in 2020 importance was spread evenly in a tight range (26—27 %) for each of C_{a+b}, C_w and C_m, while LAI was less important (Fig. 4.12a) in each year. In durum wheat, (CZ1-19), importance was evenly spread between plant traits, with marginal differences in the order $C_w > C_{a+b} > C_m$, and in 2020 durum, feature importance was shared very evenly (Fig. 4.12b). In CZ2, where only bread wheat was grown, C_{a+b} was most important in 2019, with C_m and C_w intermediate and LAI lowest. In CZ2-20, C_w took > 40% of importance, with the remainder spread evenly between C_{a+b}, LAI and C_m. Airborne SIF added to S2 traits in 2019 bread wheat took low importance and did not disrupt feature order relative to the model without SIF, while in 2020 SIF provided gain approximately equal to C_w and other traits and greater than LAI, but importance remained distributed as it was without SIF, with no change to ranking (Fig. 4.12c). For 2019 durum wheat, SIF contributed

marginally more gain than C_w but, similarly, features' relative magnitude was otherwise unchanged, while SIF became the most important in 2020 durum while other features were approximately equal (Fig. 4.12d).

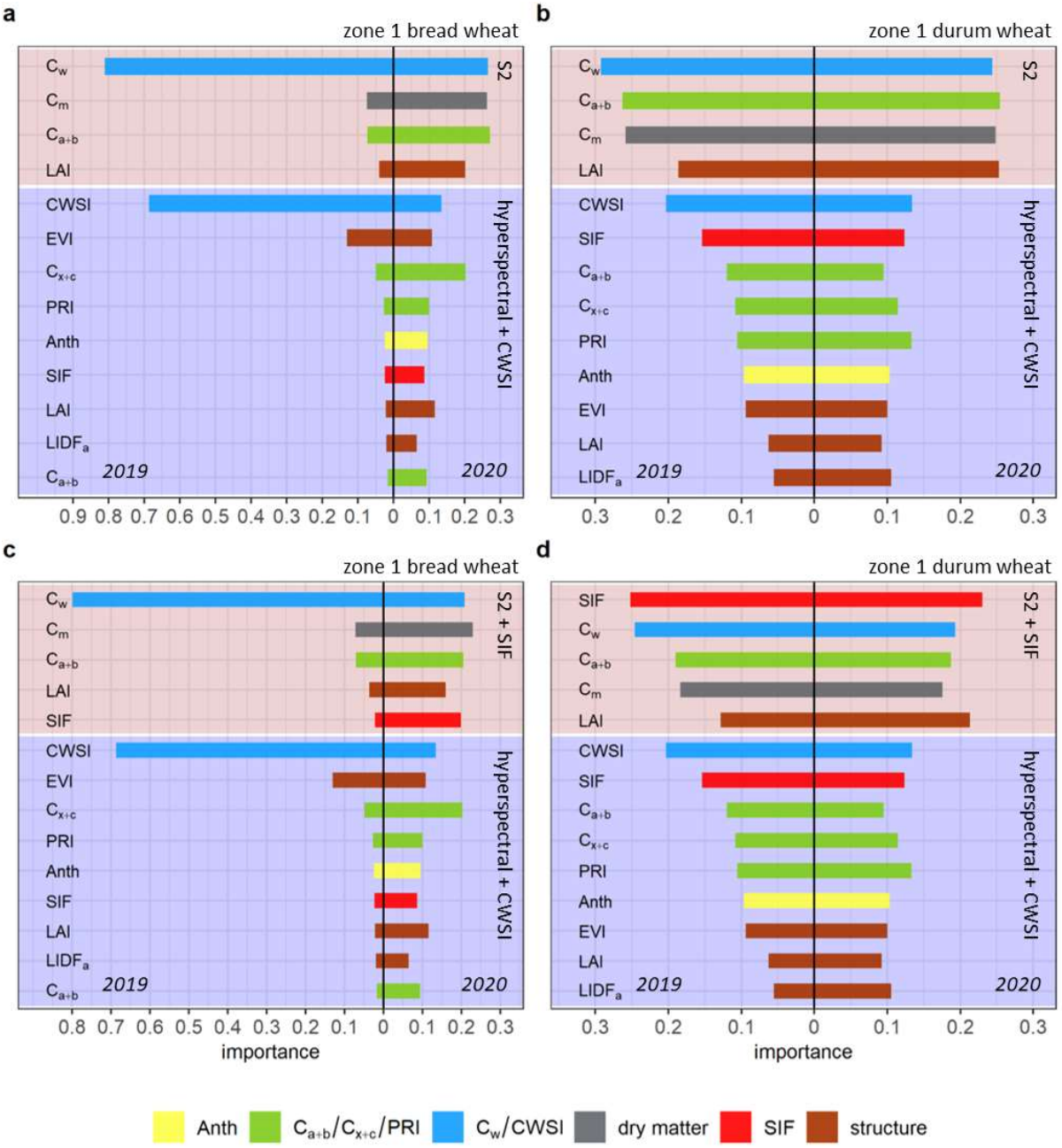

Fig. 4.12. Relative importance (proportion) of input features to a gradient boosting machine estimating grain protein content in commercial bread (a, c) and durum wheat (b, d) grown in CZ1, 2019–2020. On brown background: traits retrieved by radiative transfer model inversion from single Sentinel-2 images: chlorophyll (C_{a+b}), leaf dry matter (C_m), leaf water content (C_w) and leaf area index (LAI), ± solar induced fluorescence (SIF; airborne). On blue background: vegetation indices EVI and PRI, inverted anthocyanins (Anth), C_{a+b}, carotenes (C_{x+c}), LAI and leaf inclination ($LIDF_a$) from airborne hyperspectral images, plus airborne crop water stress index (CWSI).

For each combination of site, year and wheat type, separate ML models were built with traits retrieved from each image of the S2 timeseries. Within each model, represented by a separate horizontal sub-panel and marked with its Zadoks growth stage (Z) in Fig. 4.13, feature importance sums to 1. Empty sub-panels represent Z stages where cloud-free images were available in only one of the years. LAI was the dominant feature in each model from Z15 (five-leaf seedling) until Z43 (booting) in the 2019 CZ1 bread wheat crops. While C_w became predominant in the model based on images from near anthesis (Z69), neither C_m nor C_{a+b} contributed significant gain at any time (Fig. 4.13). In the benign 2020 season, and in both years' durum crops, gain was distributed among traits in each consecutive model, changing little and without a clear pattern through the season. In CZ2-19, where images were available from Z17 to Z54 (mid-ear emergence), C_{a+b} was consistently though marginally the more important feature, and LAI was consistently the least important, while in 2020 the gain contributed by C_w increased from ~35 % early in the season to > 50 % at Z42, the last image capture date.

Fig. 4.13. Relative importance (proportion) of input features to a gradient boosting machine (GBM) estimating grain protein content in commercial bread (a) and durum wheat (b), CZ1, 2019–2020. Chlorophyll (C_{a+b}), leaf dry matter (C_m), leaf water content (C_w) and leaf area index (LAI) inverted from Sentinel-2 image timeseries were used as separate GBM training feature sets. Importance sums to 1 within year/stage combinations. Image dates by Zadoks (Z) growth stage from top down.

In addition to treating plant traits from each image in a given combination of site, year and wheat type as input features to separate models, ML models were built with all traits retrieved from each entire S2 timeseries. In this instance, each sub-panel in Fig. 4.14 represents a trait in timeseries with the relevant Z stages on the y-axis and feature importance summing to 1 through the site / year / wheat type. Absent bars again represent an absence of cloud-free images. This modelling approach revealed patterns of feature importance consistent with those in individual image models: for bread wheat, high LAI importance early in 2019, switching to C_w after anthesis, and a minor emphasis on early- to mid-season LAI in 2020, and importance evenly distributed between traits and stages in other situations (Fig. 4.14). In CZ2, 2019, C_{a+b} estimated at early stem elongation (Z31) contributed 17 %, and C_m from a

similar stage 14 %, of total gain. The 2020 season in CZ2 had C_w at Z42 as the sole major gain contributor for GPC estimation; no other feature reached importance > 10 %.

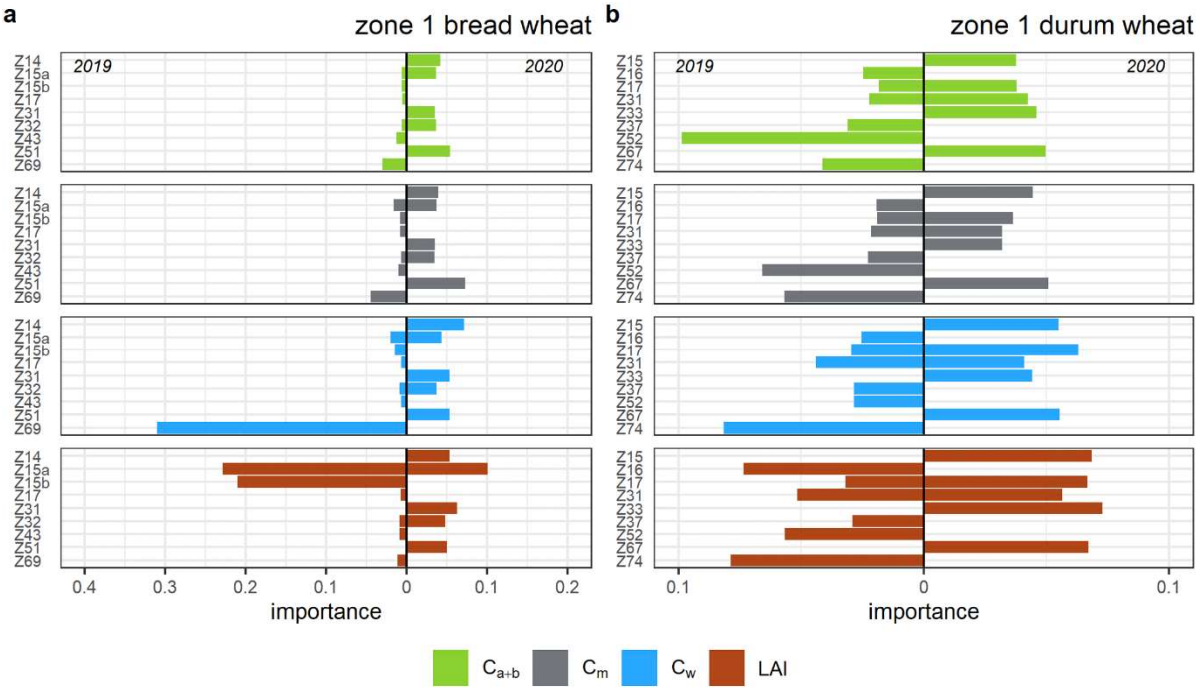

Fig. 4.14. Relative importance (proportion) of input features to a gradient boosting machine (GBM) estimating grain protein content in commercial bread (a) and durum wheat (b), CZ1, 2019–2020. Chlorophyll (C_{a+b}), leaf dry matter (C_m), leaf water content (C_w) and leaf area index (LAI) inverted from Sentinel-2 image timeseries were used together as a single GBM training feature set. Importance sums to 1 within years. Features arranged by Zadoks (Z) growth stage from top down, within plant trait category.

4.4. Model predictive skill

4.4.1. Hyperspectral models

For models built with airborne hyperspectral and thermal traits, mean skill in GPC prediction over 80 runs per model was assessed for each combination of site, year, product and input layer. Predictions of GPC were made for a randomly sampled 30% of observations not used to train the algorithm (the 'unseen' sample). Coefficient of determination (R^2) was used to assess pred~obs correlations, and RMSE (% GPC); Table 4.3) to assess model error. In all cases, the more information was provided to the model, the better the predictions: Model 1,

with physiological, structural and thermal inputs, performed better than models 2 and 3. The best predictions were in the 2019 bread wheat crop (R^2 = 0.80, RMSE = 0.62), when CWSI contributed a mean 69% of total predictive power; this year and crop also had the best predictions under models 2 and 3. In lower water stress, the proportional contribution of CWSI to the accuracy of models in which it was included was lower, and that of the tested physiological variables higher (Table 4.3). Although the individual importance of the structural components LAI, LIDFa and EVI was generally low, when added to the physiological layer they collectively increased model skill by between 11—21% and reduced error more than CWSI in any year/product/model combination.

Table 4.3. Predictive skill (R^2, RMSE; %) for Model 1, built with physiological + structural + CWSI layers, Model 2 (physiological + structural) and Model 3 (physiological only) across bread and durum wheat. Each model / product / year combination was run 80 times.

Model	Bread wheat				Durum wheat			
	2019		2020		2019		2020	
	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
1	0.8	0.62	0.54	0.66	0.54	0.81	0.49	0.67
2	0.7	0.76	0.5	0.69	0.49	0.84	0.46	0.69
3	0.57	0.91	0.39	0.76	0.43	0.9	0.37	0.75

4.4.2. Sentinel-2 models

Predictive skill results for models based on Sentinel-2 plant traits are presented in separate sections, dealing first with models built from single end-of-season images and their comparison with hyperspectral results, then those constructed with plant traits from individual images from the timeseries, and lastly models with traits from full timeseries.

In parallel with results for HS ± CWSI detailed above, predictive skill for single-image S2 models, with or without SIF, was best in CZ1-19 bread wheat. The S2 models predicted GPC better than HS models only in this combination of site, year, and product, and performed

substantially worse than HS in 2020 when conditions were more benign (Fig. 4.15). Augmentation of S2 traits with SIF, which was available only in CZ1, made only a minor difference in the dry conditions but improved prediction relatively more, from a lower base, in each other site/year/product combination (Fig. 4.15). Airborne SIF improved model skill most strongly in 2020 durum crops, when prediction based solely on S2 traits was poor.

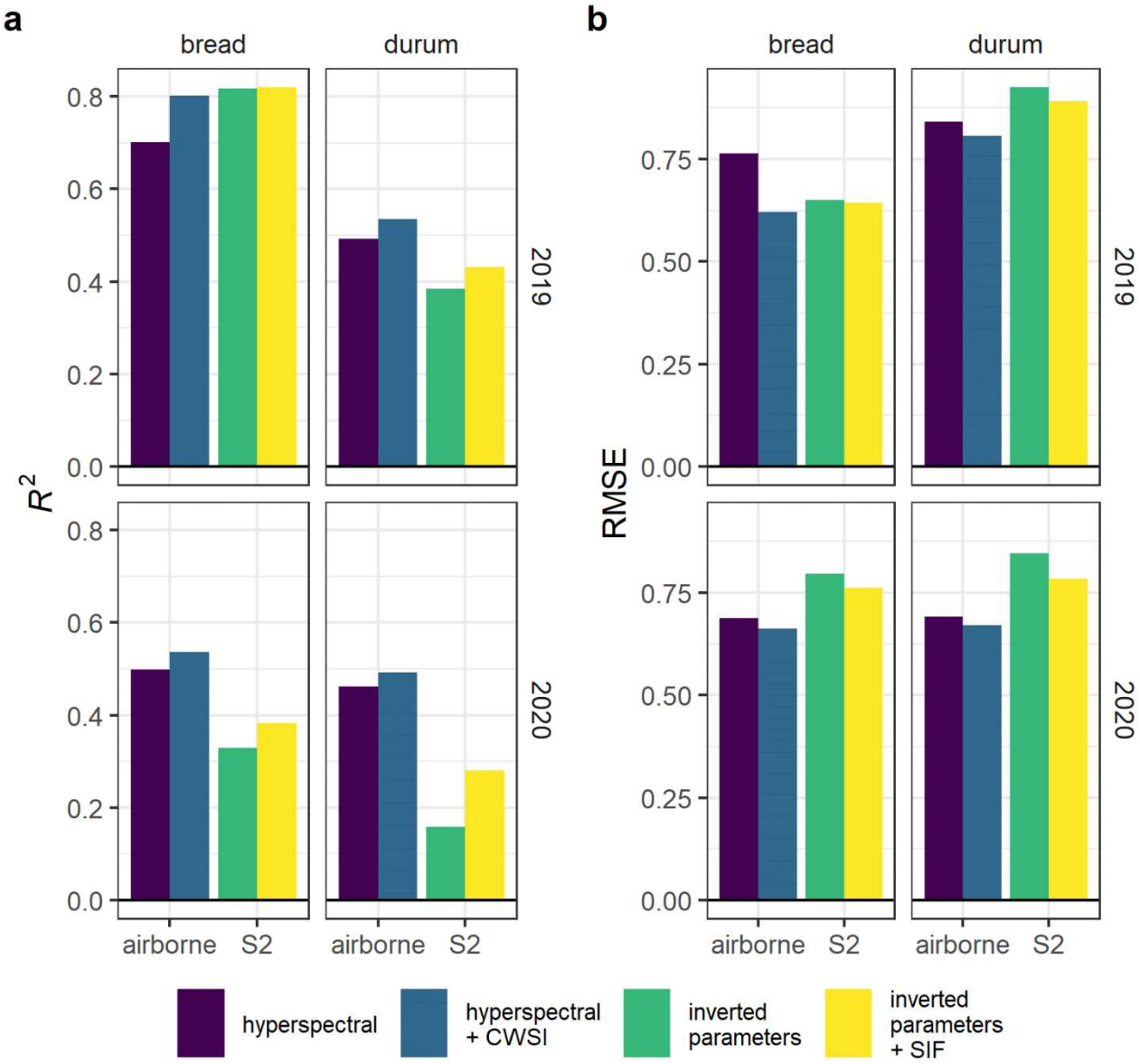

Fig. 4.15. Mean skill in wheat grain protein content estimation in commercial bread and durum wheat crops, CZ1, 2019–2020. Models comprise chlorophyll (C_{a+b}), leaf dry matter, leaf water content and leaf area index (LAI) inverted from single, late-season Sentinel-2 images, ± airborne fluorescence (SIF), and models comprising inverted anthocyanins, C_{a+b}, carotenes, LAI, and leaf angle plus vegetation indices EVI and PRI and SIF from airborne hyperspectral images, ± airborne crop water stress index.

The predictive skill of models built with traits from single successive images from the collected timeseries varied substantially in some seasons. Performance was high early in the CZ1-19 bread wheat season but diminished from late seedling stage (Z17) until Z69 when the last usable image was captured (Fig. 4.16). Although differences were relatively small, models based on traits from early phenology generally performed better than those built with mid-season traits, usually with a subsequent increase late in seasons. The 2020 season in CZ2 lacked images past Z42, while images were obtained until at least Z51 in all other seasons.

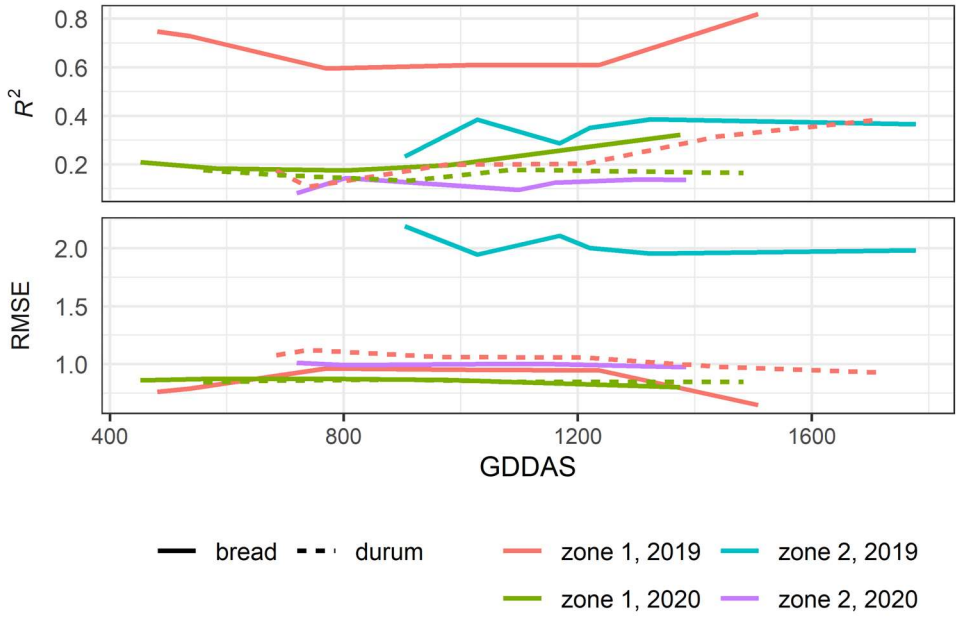

Fig. 4.16. Evolution through seasons of mean skill in wheat grain protein content estimation in commercial bread and durum wheat crops, CZ1, 2019–2020. Models comprise chlorophyll, leaf dry matter, leaf water content and leaf area index inverted from single Sentinel-2 images in a timeseries. Metrics are shown as a function of growing degree days after sowing (GDDAS) at image capture date.

When predictive skill was compared between late-season single-image (base) models and those built from timeseries, the timeseries brought improvements in all site / year / wheat type combinations (Fig. 4.17). This included 2019 bread wheat, where base performance was already strong, but was emphatic in cases where the base model skill was lower. In two cases, CZ1-20 durum, and CZ2-20 bread wheat, the R^2 of prediction was more than doubled with TS addition, although error was not similarly improved.

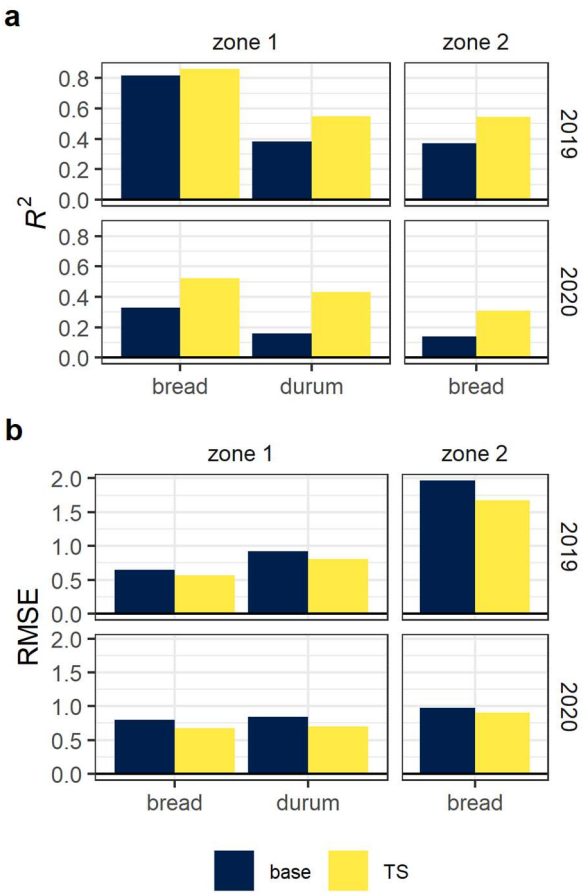

Fig. 4.17. Mean skill in wheat grain protein content estimation in commercial bread and durum wheat crops, CZ1 and CZ2, 2019–2020. Base models comprise chlorophyll, leaf dry matter, leaf water content and leaf area index inverted from single late-season Sentinel-2 images; timeseries (TS) models comprise the same traits retrieved from 5-6 S2 images across the same season.

5. Discussion

This chapter contains included material from:

Longmire, A.R., Poblete, T., Hunt, J.R., Chen, D., Zarco-Tejada, P.J., 2022. Assessment of crop traits retrieved from airborne hyperspectral and thermal remote sensing imagery to predict wheat grain protein content. *ISPRS Journal of Photogrammetry and Remote Sensing 193, 284–298*. https://doi.org/10.1016/j.isprsjprs.2022.09.015

Longmire, A.R., Poblete, T., Hornero, A., Chen, D., Zarco-Tejada, P.J., (accepted 2023-10-27). Estimation of grain protein content in commercial bread and durum wheat fields via traits inverted by radiative transfer modelling from Sentinel-2 timeseries. *ISPRS Journal of Photogrammetry and Remote Sensing*.

—

Grain protein content is a complex plant trait under genetic, environmental and management control (Zhao *et al.*, 2019). Once a grower has selected a cultivar to be sown, when and with what initial fertiliser rates, decisions typically made in view of growing- and market conditions, and based on a mixture of experience, expert guidance and sometimes an algorithmic decision support, they may have little further possibility of influencing crop outcomes. This scenario is normal for wheat cropping in the areas considered in this study, where irrigation is never available, and rainfall is the predominant influence on yield and quality; such conditions represent a large majority of wheat cropping in Australia and are transferable to many locations worldwide. Hence for many wheat growers, possibilities of influencing harvest outcomes, and therefore optimising financial returns, through in-season interventions are limited to strategic N fertiliser applications, which will require calibration from year to year according to rainfall both to date and forecast. However, because of the inherent complexity, growers may have relatively little insight into soil variability within their fields and its likely effects on GPC. Moreover, the soil conditions with the largest potential to affect GPC, soil water and N supply may vary substantially on a scale of meters, interact, and are likely to grade into one another along a continuum from repleteness to stress

without distinct delineation. The fact that these relationships can themselves alter along seasons according to changing plant demands and soil conditions, and that they may not be consistent from year to year either, adds further complexity.

The effects of substantial soil variability on GPC were observed in both the experimental plots and in each of the fields considered (e.g. Fig. 4.5, Fig. 4.6). For example, within each plot study, and therefore with minimal influence from natural soil conditions, GPC varied by 3—4 %, sufficient to increase the quality grading of the wheat by at least two classes of the Australian Wheat Standards (GrainCorp Ltd., 2021). In single fields, GPC from < 9 % to > 16 % was observed, a range greater than that between the minimum for Australian (soft) white noodle wheat and the minimum for Australian Premium Hard 1. At the date of writing this represented a range of AUD322—AUD385 per tonne paid at receival depots (Cargill Australia Ltd., 2023); at the yields attained in the same fields, a difference of > AUD150 / ha.

These challenges present motives and opportunities to calibrate N inputs: strategic increases where a positive GPC response is likely, with reductions where none is expected. Moreover, it is known that GPC is more likely to respond to supplementary N at certain growth stages than at others. This adds to the challenge of GPC prediction but also motivates investigation of the temporal dynamics of plant trait effects on GPC. With better in-season DSS, tactical N applications could be calibrated in space, quantity and time to optimise GPC. Hence the objectives of this thesis: to make progress toward implementable PA algorithms capable of estimating GPC prior to harvest, by identifying plant traits and indicators related to GPC in bread and durum wheat, quantifying these traits' and indicators' relative importance to GPC estimation, and testing the fidelity of estimated against measured GPC.

5.1. Plot experiments

Overall, GPC is influenced by the amount of protein available for translocation, the size of the protein sink, via grain count, and by late-season conditions affecting photosynthesis and phloem transport. Late in seasons, conditions unfavourable for CHO production and translocation tend to promote high GPC. From the plot studies, it was evident that relatively small increases in soil N can elicit a strong GPC signal, even where there was zero imposed difference in moisture levels and any natural difference in soil properties was accounted for in experimental design. In each location, the second and third N fertiliser increments/classes sat well within the normal range of full-season commercial applications for the relevant regions, and these increments produced distinct differences in GPC. This was likely compounded at ES2 by soil drying from high early biomass, which would have contributed to the observed saturation and reversal of the GPC response. These phenomena, and the large GPC signal from realistic increments in N nutrition, emphasise the potential for accurate fertiliser calibration to both increase and moderate GPC, depending on conditions. Moreover, they confirm that adjustment of fertiliser dosing can help to optimise GPC and reduce N losses.

Overall, the results for experimental plots showed that associations between leaf C_{a+b}/N content, plant stress indicators and GPC were consistently stronger than those between structural measures and GPC. The close alignment of CI, VOG1 and ZMI with each of Dualex C_{a+b} and NBI at leaf level, inverted canopy C_{a+b}, and GPC is related to these indicators' links to leaf N, as found previously (Ustin et al., 2009; Vogelmann et al., 1993; P.J. Zarco-Tejada et al., 2001), and to GPC (Li *et al.*, 2020). While present at both locations, these associations were more pronounced at ES1 than at ES2. For example, VOG1 was strongly correlated with leaf N at site 1 and with Dualex C_{a+b}, NBI and GPC at both sites. This, and close associations between NPCI and leaf N, support other findings in wheat

(Ranjan *et al.*, 2012). The NPCI was more closely associated with leaf N/C_{a+b}, for which it was conceived, than with GPC. Further, the combination of higher ranges of CI, VOG1, ZMI and airborne canopy C_{a+b} at site 1, suggest higher C_{a+b} accumulation there, a hypothesis supported by higher Dualex C_{a+b} but not by the lower leaf N from destructive sampling. In contrast, higher GPC at site 2, with lower CI, VOG1, ZMI and airborne C_{a+b}, may indicate that translocation was more advanced at ES2, driven by senescence. This latter hypothesis is also supported by the reduction in LAI along the N gradient at site 2. At site 1 only, and from airborne RS, the C_{a+b}- and pigment-focused VIs MTCI, SRPI and TGI were also moderately, though significantly, correlated with GPC, as were the structural or RGB-based MTCI, NGRDI, BRI2, VARI, and each of the R, G, and B indices (Table A1).

At ES2, N supply at the upper treatment levels, imposed for experimental reasons but in anticipation of higher rainfall, manifestly exceeded plant requirements given the low starting soil moisture and very low eventual GSR. This excessive N nutrition is shown by the high LAI, which was extreme for Australian wheat (Waldner *et al.*, 2019) and far exceeded the LAI recorded at ES1 (Fig. 4.3). Very high GPC in all treatments and declining LAI and leaf N, especially at high fertiliser rates, are further signs of excessive N at ES2 but were not seen at ES1 (Fig. 4.2, Fig. 4.3). While a greater GPC response to N fertiliser is expected when grain filling is water limited (Angus and Fischer, 1991; Holford *et al.*, 1992), and this was seen in the comparison of the two sites, the GPC response at ES2 saturated at heavy N applications, and at the highest level was below the next highest N treatment. Site 2 was also more advanced in phenology at the time of flights and was visibly stressed; it is probable that little extractable soil moisture remained. This is witness to the 'haying off' phenomenon. The unusually high biomass, seen in the LAI values, is likely to have worsened water stress, further restricting late season plant N uptake, especially at the higher treatments, evidence for which is seen in declining leaf N in treatments Y3 and Y4 but conversely also contributed to

high GPC through its function as a large protein source. These phenomena were not seen in the more moderate moisture and calibrated N dosing of ES1. Besides causing actual reductions in leaf C_{a+b}/N, and hence indicators associated therewith, the extreme water stress at ES2 may also have accelerated senescence, further obscuring the spectroscopic signature of differential N rates and, by disrupting translocation, breaking the physiological link between leaf- and grain protein contents. Such earlier senescence and translocation of leaf proteins than at lower, and hence less water-stressed, treatment levels is a further explanation for lower leaf N, which was destructively sampled, in the site 2 high N treatments. Although no conclusive trend was seen in either inverted C_w or thermal CWSI at either site, the lower C_w and higher CWSI ranges at site 2 were in agreement; hence both leaf water content and canopy temperature broadly confirmed the soil moisture contrast between sites. While these observations were inconclusive regarding increased soil drying at higher N supply, they were also supported by the lower site 2 SIF, discussed below.

The traits and indices discussed above, linked to GPC via leaf N / C_{a+b} content, are stable over days to weeks, because it is over such periods that C_{a+b} is stable in wheat (Hamblin *et al*., 2014), but PRI_{m3} and PRI_{m4} can show changes on much shorter timescales (Magney *et al*., 2016b). Given the stable weather, however, here the PRI likely shows a more stable condition of stress caused by droughting; again, as assessed by PRI, site 2 shows greater stress, with distinct saturation at high N treatments. Evidence for a physiological link between PRI and photosynthesis was seen, especially at site 2, in the marked alignment between PRI and GPC, via a clear inverse relationship between airborne SIF and GPC, and in lower overall SIF. SIF, in turn, declined at higher N, which adds evidence for greater drought stress in those treatments. Such findings accord with inverse PRI~SIF relationships reported previously (Magney *et al.*, 2016; Peñuelas *et al.*, 2011; Suárez *et al.*, 2009) and show the

substantial potential of PRI to improve GPC estimation, as it provides a signal of constrained assimilation.

Taken together, high GPC, the distinct alignment between PRI_{m3} and GPC, CWSI and SIF responses show strongly constrained assimilation at ES2 but not at ES1. This was also seen in the C_{x+c} response, which despite saturation was also associated with GPC at ES2 under severe stress, though not at site 1. However C_{x+c} also parallel leaf N and C_{a+b}, so their response may simply reflect N nutrition, as seen by Shah *et al.* (2017) in wheat. C_{x+c} should be higher under stress, but at the time of image capture, were lower at ES2. These lower values may indicate prior C_{x+c} remobilisation related to senescence: Low C_{a+b} values at site 2, compared with site 1 and other work (Hamblin *et al.*, 2014), and which, given the high LAI and recognised C_{a+b}:LAI relationship, would have been higher earlier in the year at that location, suggest that C_{a+b} declined before C_{x+c} (Gitelson and Merzlyak, 1994a, 1994b). Nevertheless, on evidence from the current studies, C_{x+c} are indicative of potential GPC, in severely droughted conditions as reliably as C_{a+b} and PRI.

Given that Anth are upregulated under stress (Naing and Kim, 2021; Shoeva *et al*., 2017), and are known to accumulate in the wheat grain, increasingly from post-anthesis drought (X. Li *et al*., 2018), one may expect high Anth to correspond to high GPC. Some evidence was seen for this mechanism: at ES1, where inverted canopy Anth were aligned with both leaf N and GPC; in parallel, the airborne anthocyanin reflectance index ARI1 was also significantly correlated with GPC, at ES1 only. However, if C_{x+c} and Anth concentrations are understood to indicate stress, site 1 appears more stressed, but by all other measures it was not. As an alternative to presuming that elevated Anth in this case relate solely to elevated stress, and via this to GPC, which was not observed, it is possible that leaf Anth content instead shows only a positive response to N fertilisation, as recorded by Yamuangmorn *et al.* (2018) in rice.

Indeed, such hypothetical positive responses of both C_{x+c} and Anth simply to high N supply suggest more accessible soil N at site 1 due to higher soil moisture, while at ES2, strong declines in Anth at high N appear to confirm lower soil N accessibility at high N treatments and/or earlier senescence. Moreover despite showing some, though limited correlation with C_{a+b} in other work (Shah *et al*., 2019), here the range of each was higher at ES1 but showed no such association along treatment levels within either site. At site 2, similar Anth responses at leaf and canopy levels, and their reversal with respect to N treatment and GPC, suggest that like C_{a+b} and C_{x+c}, Anth translocation had begun before data collection, especially at high N and attributable to accelerated senescence. Based on these plot experiments, Anth show some association with GPC under mild water stress, but there is no evidence of such under severe stress.

In toto, evidence from the current plot studies and relevant literature as cited allows the conclusion that Anth, C_{x+c}, and PRI are each associated with GPC via moderate to severe stress, although because of their intimate connection it remains difficult to separate the effects of N and moisture limitation. In some contrast to most of the physiological indicators discussed above, structural VIs NDVI and EVI_{NIR} showed generally inferior performance as GPC indicators; their relatively poor performance in stress assessment has been seen previously (Gamon *et al.*, 1992; Peñuelas *et al.*, 1994). Predictably, however, both saturated at higher N treatments, particularly under water stress, and showed relatively limited association with GPC. A weak relationship with GPC was also common among other structural and greenness VIs tested, including EVI2, gNDVI, the green-blue NDVI, MSAVI and RDVI. The consistency of the R_{920}/R_{729} index with C_{a+b}, NBI, and GPC, likely due to its component bands' sensitivity to water, structure and C_{a+b}, suggests it could be applied in GPC estimation.

5.2. Field transect

The field transect was conceived in order to directly compare destructively sampled soil and crop quantities generated under commercial cropping conditions, and ground-level spectroscopic measures, with VIs and retrieved parameters from airborne RS. While the transect sample size was limited by practicalities of data collection, this segment of the project tested the degree to which results from the controlled environments of plot studies would translate to commercial growing environment. Unlike the plot experiments, the transect was not subject to an imposed N gradient, so its classification into three classes was done *post hoc* by clustering based on GPC. Wheat in the transect field was about as advanced in phenology (≥ Z75; mid-milk development), at the time of field work and RS campaigns, as the plot sites. The transect's substantially more moderate C_w and CWSI therefore appear to reflect a better match of N supply to soil moisture, with less soil drying by high early biomass, than at either plot site, and highlight the contrast between experimental and commercial farmer's objectives. Overall, the transect results showed that some insight into relative soil N levels could be gained via spectroscopic estimation of leaf and canopy quantities. This is an important finding given that the thesis objective is to contribute information about plant performance which arises from soil conditions but is estimated from RS observations of the canopy.

Given the lack of imposed soil N gradient, it was expected that differences in N-related plant traits would be more subtle, as seen in VOG1, ZMI and airborne inverted C_{a+b}. Indeed C_{a+b} trended against GPC and soil N this probably reflects the progress of protein translocation, whereby crop areas with higher remaining C_{a+b} continued to accumulate photosynthate, diluting GPC. Airborne NDVI and EVI were relatively consistent with soil N, Dualex NBI and GPC; overall, their performance was quite stable across experimental plots and transect. While these structural measures were not strongly indicative of GPC in either context, their

consistency does demonstrate a link with GPC. This is consistent with other studies which have generally integrated NDVI across time or with other indicators to achieve acceptable GPC estimates (Feng *et al*., 2014; Stoy *et al*., 2022). In the transect, as at ES1, inverted canopy level Anth increased with GPC. Also as at both plot sites, C_{x+c} was parallel with GPC, and with soil N, supporting an effect of N nutrition, stress or both on GPC. Moreover, the association of PRI with lower assimilation is evident through its alignment with GPC in the transect, a dynamic also seen in the plots. These highly consistent relationships support the proposition of Anth, C_{x+c} and PRI as GPC predictors in commercial wheatfields as in the experimental context.

5.3. Commercial fields

As established by the many previous studies covered in the literature review above and confirmed by the plot- and transect-scale components of this study, various plant and canopy traits important to the biophysical determination of GPC can be retrieved using RS spectroscopy. To advance predictive estimation of GPC for eventual application in PA, a strong focus on commercial fields was considered crucial, because while crop genetics are invariant within fields and rainfall relatively stable across regions, soil conditions can show wide spatial variability within fields. Soil attributes including clay and organic matter fractions, native N, and subsoil constraints affecting water and root penetration including B and Al, calcrete, and hardpans can vary in severity across tens to hundreds of metres in the regions considered here, exerting strong influence over plant-available water capacity (PAWC) and N (Nuttall *et al.*, 2003; Sadras *et al.*, 2002) and therefore also over GPC. These edaphic realities would be difficult or impossible to replicate in experimental plots.

Again, the influence of soil physicochemical properties on wheat crops, in particular on GPC precursors such as leaf N, C_{a+b} and LAI and their modes of action, as well as on yield and GPC directly are well covered above; these physiological relationships are often both

complex and interactive. Further, although soil conditions are relatively stable across time, wide variability within and between seasons in the amount and timing of rainfall is a fact of life in the cropping environments considered here. Such variability has clear influences on plant performance, and therefore on the spectroscopic traits antecedent to GPC, but also strongly affects relationships between soil conditions and harvest variables. Indeed, even for a single location within a field, the relative effect of subsoil constraints, N supply and PAWC on plant growth is so strongly mediated by seasonal rainfall that patterns of variation seen at harvest can vary to the point of reversal from year to year (Armstrong *et al*., 2009; Sadras, 2003).

For plant traits to be important in GPC estimation, in particular to be assessed as contributing model gain, they should vary spatially across a crop in some relationship with GPC. Such spatial variability is visible both in RGB images of the subject fields and in GPC readings (Fig. 4.6). The choice of the GBM algorithm was made largely because of its capacity to attribute relative gain to model input features, hence both estimating the relative importance of plant traits to GPC determination and predicting GPC itself. The chosen ML, and others, are designed to handle multiple complex and nonlinear relationships between independent and dependent variables, so nonlinearity in precursor~GPC relationships should not of itself hinder prediction. Inconsistencies in these relationships, mediated by phenological advance, soil and moisture variability complicate GPC prediction, as detailed in the next sections. These biophysical realities are revealed in traits' relative importance, which varies substantially across time, both within and between seasons.

To avoid leakage of importance between ML input features – the plant traits – multicollinearity among these was reduced through variance inflation factor (VIF) analyses. These were largely data-driven, iterative processes, commencing with the full suite of HS ($n > 90$) or S2 ($n > 40$) features, in which all VIF values changed at each successive calculation

because variables were successively discarded. However, the procedures were forced to retain RTM-inverted parameters because these offer better transferability across agronomic situations, locations and phenological stages, and hence were considered more likely to be more stable than VIs. Based on HS imaging, the narrow-band VIs EVI_{NIR}, a novel index assessed during the current research, and PRI, sensitive to water stress and photosynthetic rate, were linearly independent of and complemented inverted parameters across both years' data, so were included in analyses. In contrast to the HS analysis, none of the S2-compatible, multispectral VIs tested was both a) linearly independent of all inverted parameters and b) contributed substantially to model skill. While the lack of linear independence among VIs contrasts with the findings of equivalent tests in the current HS analyses, it is as expected for data of lower spectral resolution.

Given that no Sentinel-2 VI made a unique contribution to model skill, it can be argued that the S2 band set, and the retrieval methods adopted, adequately captured the spectral information available and needed for GPC estimation. This in itself is a finding of minor importance, in accordance with the findings of Wolanin and colleagues (2019) that S2 bands were sufficient to accurately estimate gross primary productivity and the fraction of incident radiation captured by C_{a+b}, without recourse to VIs or non-RS data, via RTM inversion and ML.

5.3.1. Sites and conditions

Differences in PAW influence GPC via lowered stomatal conductance and assimilation, hence lower dilution of protein with new photosynthate, especially late in the season when water is limiting. These effects were, in particular in CZ1, 2019, detected in measures of crop water sufficiency C_w and CWSI. Constrained carbohydrate (CHO) additions during grain filling can therefore increase GPC and reduce yield. However, grain number (GN) is the

primary determinant of sink size for both proteins and photosynthates and, therefore, also a strong driver of eventual GPC. While early vigour tends to increase GN, stress conditions around anthesis can sharply reduce GN. However, strong early vigour can also result in excessive shoot biomass, in turn causing soil drying sufficient to severely restrict later assimilation. This effect may be sufficient to increase GPC but in severe cases can cause complete crop loss (van Herwaarden *et al.*, 1998). These dynamics help to illustrate the large swings in GPC that may be precipitated by sub-optimal calibration of N nutrition made in incomplete knowledge of PAW changes.

Besides being very dry, CZ1 saw other weather extremes during critical periods of 2019. From mid-September into early October, frosts below -4°C affected the durum crop at crucial stages (Z42—69; booting, ear emergence and anthesis). Bread wheat fields were less affected by frosts due to their location higher in the landscape, a matter of farmer choice, and somewhat earlier, less susceptible growth stages (Z33—51; stem elongation to incipient ear emergence), related to genetics and later sowing. Frost around anthesis severely reduces grain count; durum is more susceptible to both frost and heat damage than bread wheat and other cereals (Beres *et al.*, 2020; McCallum *et al.*, 2019). Immediately after these frosts, four days had maxima in the range 35—38 °C. Such high temperatures impose high respiratory loads, consuming CHO (Heskel *et al.*, 2016), and have the potential to cause permanent damage, reduce total N uptake (van Ittersum *et al.*, 2003) and/or simply exceed the optimum for photosynthesis (Asseng *et al.*, 2011; Lobell and Gourdji, 2012). Heat stress can also severely alter phenology: At T > 34 °C, senescence is accelerated by a factor of three, and sixfold at T ≥ 36 °C, foreshortening grain filling (Asseng *et al.*, 2011; Porter and Gawith, 1999). In addition, droughted plants have less capacity for cooling through evapotranspiration, so accumulate more heat. Each of these effects reduce yield, especially if cumulative (Asseng *et al.*, 2011); each should therefore affect GPC. Heat stress may increase mean GPC, while

reducing GPC variability, and can reduce total N uptake (van Ittersum *et al.*, 2003) through a generalised reduction of assimilation. Foreshortened CHO accumulation during grain fill (CZ1-19) would also be associated with lower C_{a+b}, C_m, LAI and SIF variability, diminishing these as GPC estimators, while likely enhancing the importance of moisture sufficiency measures, as observed in both CWSI and C_w. By interfering with grain development and filling, these large variations in weather conditions, seen at crucial times in crop development, likely had substantial effects on GPC, without similar effects on canopy reflectance characteristics. This would constitute a large source of error, reducing the ability of the current methods to predict GPC, and likely would also affect bread and durum wheat differently because of their genetic and phenological differences. Such challenges emphasise the importance and difficulty of conducting studies in uncontrolled, real-world conditions.

In CZ2, where rainfall is already substantially lower and crops are grown on former dune fields, phenomena influencing PAW and N, including both osmotic and physical impediments to root growth, arise from the dune-swale morphology and can be severe (Nuttall *et al.*, 2003; Sadras *et al.*, 2002). Though high-resolution soil data were unavailable, the influence of the CZ2 dune-swale systems is clearly seen in GPC (Fig. 4.6b); the dune effect on biomass can also be seen in the Sentinel-2 RGB areas of the same figure. In these areas, the effect of height within the dunes is often strongly discernible at harvest and plays a large role in wheat growers' cropping decisions. Crop responses include complete reversal between wet and dry years of yield response on dunes and in swales; mid-slope areas, in which soil conditions grade between light-textured soils on one extreme and heavy clays on the other, can occupy a high proportion of fields (Armstrong *et al.*, 2009; Hoffmann *et al.*, 2016). The severe soil variability, known yield response differences, and the proportion of fields affected, are also reflected both in less definitive feature importance dynamics, and lower GPC estimation skill, seen in CZ2. This difficulty is analysed in depth by Rab *et al*.

(2009): a large majority of the ~ 90 ha they studied across four years in the same region showed great wheat yield variability, and this was substantially explained by soil water dynamics which they linked in turn to dune-swale topography.

5.3.2. Single-image analyses

Especially in severe water stress, but also more generally, CWSI was an important indicator of GPC, in hyperspectral analyses. In the S2 analyses, CWSI was excluded from ML analyses due to collinearity with at least one S2 inverted parameter in each iteration of VIF analysis, but in lieu of CWSI, S2-inverted C_w assumed high importance in the very dry conditions of CZ1, 2019 (Fig. 4.12a, c) and throughout the analyses was parallel with CWSI. These observations reflect the well-recognised relationship of water stress with assimilation rate, hence the protein dilution effect. In addition, the finding that C_w supplanted CWSI when the latter was excluded allows the assertion that C_w retrieved from S2 data can, in some situations, reduce the requirement for thermal imaging. This is supported by findings of SWIR information as important for GPC estimation via its link with plant water relations (Pancorbo *et al*., 2023; Rodrigues *et al*., 2018), despite findings that C_w was retrieved relatively poorly, and substantially less well than LAI, in wheat (Sehgal *et al*., 2016). However, this latter study would have included relatively little C_w variability as it considered only two replicates in a plot layout.

The elevated importance of both water stress indicators in severe conditions, however, comes not only of their inherently large contribution to model gain in those conditions, but also of their complementarity to the physiological features, which show less spatial variability when fields are so dry as to universally depress photosynthesis, and hence contribute little model gain. Under such dry conditions, it is apparent that C_w retains more variability than other S2 traits, especially C_{a+b} and C_m, whose low importance in dry conditions highlights their lack of

spatial variability and hence small contribution to GPC differences. This low variability in dry conditions is physiologically supported because they were established early in the season when water and nutrients were less limiting, and therefore remained relatively little affected by soil variability, while C_w shows what was happening at the end of the season, after differential soil drying. C_{a+b}, C_m, and shoot biomass, the latter indicated by LAI, contribute to the pools of both N and CHO available for translocation. It follows that, as the translocation process itself is also impeded by water stress, with protein transport more strongly affected than CHO transport (Giuliani *et al*., 2011; Wardlaw, 1967), these quantities' influence on GPC is more likely to be seen under low water stress. Conversely under high stress the combined influence of C_{a+b}, C_m, and LAI would be both smaller than that of the ongoing photosynthetic rate and obscured by inefficient translocation; this comparative reduction of their signal likely contributes to the dominance of C_w under severe stress.

By classifying input features into thermal, structural and physiological layers, and sequentially removing these from the HS model, each input's contribution to GPC estimation and each layer's influence on predictive skill were assessed. When CWSI was excluded as a model input for severely stressed crops, LAI importance was high, but in benign conditions, LAI and other indicators of canopy structure were not similarly promoted in the absence of CWSI. This shows that structural indicators have a greater relative influence on ML outputs in terminal drought than in lower stress. Also in drought, the displacement of EVI by LAI in model gain on CWSI removal suggests some redundancy between them, and that each was less affected by drought than were the physiological traits. These statistical phenomena also share a basis in plant development, in that both LAI and EVI indicate canopy structure variability established prior to the onset of water stress but which persisted into the late season. In benign conditions, structural indicators LAI and LIDFa generally showed low to moderate importance in HS models, as did LAI in single-image S2 analyses. This alignment

between HS and S2 provides cross-validation of each result. S2-derived C_m, defined as leaf dry matter mass per unit area, hence indicative of leaf biomass and likely to correlate with other leaf components, crucially C_{a+b}, and to proxy the protein source size, was also in all analyses of equal or greater importance than S2 LAI. The consistently high importance of C_m across weather conditions and wheat products appeared to confirm the sensitivity of this variable to changes in water status that in HS analyses was adopted by HS stress indicators Anth, C_{x+c} and PRI, treated below, but which are unavailable with the S2 band set.

Physiological features also showed commonalities across moisture conditions, within both HS and S2 analyses and when these were compared. In severe stress, in CZ1-19 bread wheat, physiological components contributed little gain. With stomatal conductance and photosynthesis universally depressed, the physiological links of HS C_{a+b}, SIF, and PRI with GPC were less prominent as predictive features for ML, showing their lack of power to indicate GPC. In these conditions, C_{x+c} was highest among the generally low importances for physiological indicators, but provided more gain on removal of thermal and structural layers, as did Anth and C_{a+b}.

In more moderate conditions, spatial variability in photosynthetic rate and hence in physiological indicators, driven by soil heterogeneity across and between fields, is greater. This greater variability allows physiological model features to convey more information, conferring to them higher importance. At moderate stress, in the CZ1-19 durum crop, airborne SIF approached CWSI in importance and all remaining HS physiological features were closely associated below SIF, none individually prominent but each contributing more gain than the structural traits. The order of physiological features changed little on removal of thermal and structural layers, showing their robustness and utility as GPC predictors in such conditions. Meanwhile in CZ1-20 durum, when water stress was lowest, CWSI was either demoted, below C_{x+c} in bread wheat, or varied with PRI between first and second rank. In

these crops, importance was more evenly spread between physiological, thermal and structural components. In bread wheat, C_{x+c} was the best indicator overall, and in durum was below only CWSI, SIF and PRI, likely showing a water stress differential between the two wheat types.

At the time of the attendant studies, the availability of spaceborne SIF was limited and its spatial resolution was unsuitable to the spatial scales considered. Given this shortcoming, this project combined airborne SIF with S2 data by sampling each at the GSD of the GPC data. In contrast to all tested S2 VIs and CWSI, and despite its close links to C_{a+b} and C_w, SIF was linearly independent of all other HS and S2 parameters and was included in CZ1 models. This linear independence, and the substantial extra skill it conferred to our predictions, are important findings in themselves. SIF is a proxy for instantaneous photosynthetic rate; as such, its inclusion can greatly improve estimates of other complex biophysical quantities, including N status and assimilation (Camino *et al*., 2019, 2018; Raya-Sereno *et al*., 2022). It follows that, as an estimate of instantaneous assimilation and hence protein dilution, SIF should improve GPC estimation.

Like C_{a+b} and PRI from HS and S2-derived C_{a+b}/C_m, SIF was of minor importance to GPC estimation in very dry conditions. The SIF contribution to predictive skill and accuracy was also minimal in these cases. Though this was assessed only in S2 analyses, as SIF was included and removed from models alongside other HS physiological components, it is generalisable to HS because of its basis in a lack of variability in biophysical activity across crops and the dominance of water stress indicators. In support of these findings, low or ambiguous sensitivity of SIF to wheat performance has also been observed by others: at regional scales and over multiple years and growing conditions, Cai *et al.* (2019) found that the addition of satellite SIF brought no improvement in wheat yield estimation over EVI-based measures. Also in regional-scale assessments from satellite RS, Sloat *et al*. (2021)

found that SIF products were better correlated with biological productivity than similar measures based on NDVI in moderate growing conditions, but that seasonal SIF-based estimates were inferior under water stress. Relevant to the current project, both Cai *et al.* (2019) and Sloat *et al*. (2021) reported better model skill when SIF was combined with reflectance-based input(s), and proposed that estimates might be improved in future by higher quality spaceborne SIF. In milder growing conditions, where base model accuracy was low (Fig. 4.17), the inclusion of SIF in S2 analyses brought larger estimation improvements. This is parallel to and arises for the same reasons as the higher importance of other physiological indicators under benign conditions, as discussed above. However, a smaller skill improvement was seen on adding SIF to S2 features in CZ1-19 durum, despite the elevation of SIF importance to the top rank, above even C_w. This may relate to weather damage to reproductive tissues not picked up in the SIF signal, but which is crucial to GPC.

The current observations at moderate and low water stress appear to confirm that C_{x+c}, Anth, PRI and SIF are sensitive to stress that is yet insufficient to strongly diminish photosynthesis or cause visible symptoms, a sensitivity recognised elsewhere (Borrelli *et al*., 2011; Fratianni *et al*., 2013; Suárez *et al*., 2008). Moreover, in low stress, each of C_{x+c}, Anth, PRI and SIF consistently had equal or higher importance than C_{a+b}. This demonstrates that short-term stress indicators also may have a more substantial role than C_{a+b} as precursors to GPC. In addition to their detection of mild water stress, this may arise from detection of differing degrees of senescence, particularly regarding C_{x+c}. Overall, physiological features were stable across conditions, and consistently supplied more information than structural features. This concurs with findings at plot scale in the present work and those of others that Anth, C_{x+c}, PRI and SIF, or combinations thereof, were crucial in stress diagnosis (Poblete *et al.*, 2021; Suárez *et al.*, 2008; Zarco-Tejada *et al.*, 2018). The PRI~SIF alignment also mirrors the current plot and transect studies and commends PRI as a proxy for photosynthesis in wheat,

as proposed by Magney *et al.* (2016b). Anth importance was moderate in all conditions and stable between wheat types in low stress; this shows a differential Anth response under variable water stress, supported both at plot scale and in the literature (Chalker-Scott, 2002; X. Li *et al*., 2018).

While model predictive skill is treated in detail below, it is valuable to consider here the apparent effects of the three HS layers. Although GPC estimations produced by the physiological layer alone (Model 3) were substantially less accurate than those from physiological and other information combined (Models 1 & 2), Model 3 accuracy and error metrics were nevertheless acceptable (Table 4.3). This skill stems from the layer's comprehensive coverage of GPC-relevant traits, from instantaneous SIF and PRI to the relatively stable C_{a+b} and suggests that progress can be made without recourse to thermal or structural features. In addition, such results align with findings of SIF and C_{a+b} as far better than structural measures in estimating wheat leaf N (Camino *et al*., 2019), and of PRI as a proxy for water stress (Suárez *et al*., 2008). Except in extreme stress, unstable importance in EVI, LAI and LIDFa limited their value for GPC estimation; this is in accordance with recent findings that LAI was unrelated to GPC (Pancorbo *et al*., 2023). Nevertheless, together they contributed to model performance and generalisability: the use of structural in addition to physiological information (Model 2) brought a substantial improvement in prediction metrics across all growing conditions, with the largest impact under water stress, and demonstrates the strong influence of biophysical state variables that change gradually through seasons. Such consistent influence of structure also provides a counterpoint to claims of leaf N measures as uniquely powerful in GPC estimation (Fu *et al*., 2022; Xue *et al*., 2007; Zhao *et al*., 2005, 2019).

5.3.3. Model skill

In summary, for CZ1 bread and durum wheat, and regarding both relative trait importance and predictive skill, the results obtained from single S2 images captured at or near the end of the growing season largely mirrored those from airborne HS (Fig. 4.12**,** Fig. 4.15). The combined relationships of S2 C_{a+b}, C_m, LAI and SIF with GPC were sufficient to produce GPC predictions of broadly similar accuracy to those provided by the full suite of thermal and HS indices and retrieved parameters, including stress-specific measures and SIF. S2 predictions were, however, less accurate than those from HS. This is explained by the lower spectral resolution of S2 data, which contains 10 bands intended for land observations (Drusch *et al*., 2012; Wolanin *et al*., 2019) as compared to the 75 bands of HS reflectance used here: Anth, C_{x+c} and PRI, linked to GPC through their detection of mild stress, cannot be retrieved from S2 data so are absent from those analyses, but made substantial contributions to HS predictions. Meanwhile, an exception to the rule of better results from HS than S2 occurred in the very dry conditions of CZ1-19. In that case, S2 was marginally better than HS + CWSI and substantially better than HS without thermal; this confirms C_w as a good proxy for thermal information in broadacre crops under differential water stress, while also showing the diminished influence of stress-based HS indicators.

Regardless of the data source, under mild stress the combination of structural and physiological inputs was insufficient to bring model performance to the level provided, under severe stress, by the strong dynamic linking C_w/CWSI to GPC via stomatal conductance, photosynthesis and assimilatory dilution of protein content. Because PAWC variability is reduced in mild growing conditions, soil properties have less influence over crop outcomes than they do under water stress and hence produce a smaller range of GPC; this narrower range itself makes prediction less accurate. However, as established above, PAW variability affects many facets of plant performance, from nutrient uptake, through leaf physiology and

canopy structure, to rate and duration of photosynthesis and assimilate translocation. While the signatures of such differences are detectable in canopy reflectance-based measures, their underlying biological phenomena are less directly connected with GPC than those caused by severe stress and therefore convey less information for prediction.

Moreover, in mild growing conditions the influence of CZ2 dune-swale systems on PAW is likely to be less definitive; less drying should occur in the heavy-textured but shallow swale soils, reducing the PAW advantage conferred in drier seasons by light soils higher on the dunes. This dynamic appears verified in CZ2-20, where model accuracy was worst, but may also have affected CZ2-19, where error was high despite acceptable pred~obs correlations. Assessed by correlation and error metrics, then, CZ2 was the hardest to estimate. The lack of HS data for CZ2 hampered more accurate GPC estimation but more importantly also the assessment of HS physiological indicators. In this situation, where the mid-slope areas, themselves a large proportion of fields, are neither severely droughted nor well-watered and grade between these extremes (Rab *et al*., 2009), the increased power of HS traits, particularly C_{x+c} PRI and SIF, to differentiate moderate levels of water stress may improve estimation.

However, with regard to predictive skill and error metrics, the question arises: how good is good enough? As a matter of pure science, the pursuit of maximum fidelity and minimum error is valid; this is how methods are refined and compared. With a view to practical implementation of PA, where science meets the pragmatism of farming; growers are accustomed to uncertainty, neither seek nor expect an exact prognosis, requesting instead estimates that can be combined with other information for decision support. Hence because PA practices will always be aimed at increasing the likelihood of a desired outcome rather than at precise predictions, the answer to “how good?” is more flexible. Of direct, practical relevance here is the error associated with modelling, here stated as RMSE, which in all cases

except CZ2-19 was < 1 % GPC, within the threshold ranges (1—1.5% GPC) used to define wheat classes in Australia. The errors implicit in GPC estimation by the methods and in the regions explored in this project can therefore be considered acceptable and are likely accurate enough to motivate growers in some situations to take strategic actions. The current project's results in severe water stress are encouraging and taken in isolation may be actionable in terms of agronomic interventions. However, it is unlikely that management action would be taken in such severe conditions, especially if the intervention itself incurred risks or costs. In such cases a cost-neutral intervention would likely be preferred, such as strategic harvesting for grain blending or segregation. In moderate growing conditions, strategic fertiliser applications may be considered, which although they entail a greater risk of misdirecting resources may be worthwhile if parts of a crop can be lifted into a higher GPC class.

Grain protein estimation based on crop information derived from RS, as presented in the attendant studies, is acknowledged as considerably more complex than estimation of intermediate quantities, or indeed of yield (Pancorbo *et al*., 2023; Zhao *et al*., 2019). Arising from an intermittent and relatively small but growing body of work, the published literature contains a wide variety of results. Although direct comparisons, particularly of error metrics, are not always possible, particularly when metrics are neither consistently computed nor supplied, the current studies' GPC predictions from single airborne HS images captured late in seasons were within the wide range of previous results and improved on some. For example, Øvergaard *et al.* (2013) obtained pred~obs R^2 = 0.16—0.68, while Hansen *et al.* (2002) achieved R^2 = 0.56 / RMSE 0.4—0.79 % and Apan *et al.* (2006) R^2 = 0.92 with RMSE 0.65—1.18 %. With the addition of met data to an interannual study based on VIs, Li *et al.* (2020) obtained R^2 = 0.13—0.85 / 1.02 ≤ RMSE ≤ 3 %, testing on 33% of their observations. However, each of Hansen *et al.* (2002), Apan *et al.* (2006), Øvergaard *et al.* (2013), and Li *et*

al. (2020), like many others, relied on proximal HS sensing of experimental plots, and hence a relatively small n. In plots,

Some more recent research has, validly with respect to relevance in commercial settings as they therefore account for natural soil variability, considered commercial wheatfields: among these, Zhou *et al.* (2021) realised validation R^2 = 0.55—0.63 and RMSE = 1.07—1.18 % GPC in an unseen 31% of their full (n = 327) dataset, from airborne multispectral RS. These authors tested both single VIs and combinations of several VIs via machine learning, finding the combinations performed better. Also from airborne multispectral RS, Stoy *et al*. (2022) succeeded in explaining 40% of GPC variability in a single 24 ha field with highly variable soils, using only NDVI, via a model that accounted for NDVI changes across a seasonal TS. Touching on methods used in this project, Stoy *et al*. (2022) also used GPC data from a combine harvester and satellite data, but found Landsat NDVI ineffective for GPC estimation. Their poor result from satellite data was perhaps related to their use of only a single measure – peak seasonal NDVI – and the relatively coarse Landsat GSD (30 m). Also working under commercial cropping conditions, with VIs from airborne HS in timeseries, Rodrigues *et al.* (2018) obtained a best correlation between any index and GPC of R^2 = 0.21; while their prediction errors were small (RMSE < 0.5 % GPC), these results were obtained in a single field with low GPC variability.

The estimation accuracies obtained here from single-image S2 analyses, $0.14 \leq R^2 \leq 0.82$ / $0.65 \leq$ RMSE ≤ 1.97, and with full-season TS $0.31 \leq R^2 \leq 0.86$ / $0.56 \leq$ RMSE ≤ 1.68 (Fig. 4.17) also largely sit within a comparable range to other recent results from satellite RS. Zhao *et al.* (2019) recorded $0.428 \leq R^2 \leq 0.467$ and $0.45 \leq$ RMSE ≤ 1.01 for wheat GPC from S2 via standard VIs, while Tan *et al.* (2020) had best verification metrics of R^2 = 0.81 and RMSE = 0.54, also from the four-band multispectral HJ-CCD sensors, although these were achieved at field level and not within fields. In large experimental plots, and using only

timeseries RS data from two multispectral satellite platforms, Wang *et al*. (2014) achieved moderately strong correlations ($R^2 \approx 0.6$) but relatively large RMSE = 1.28—2.86 % for various VIs with GPC.

As in the work of Li *et al.* (2020) and Zhou *et al.* (2021), predictive skill was tested on a substantial unseen hold-out of observations across all fields, more robust than the leave-one-out (LOO) cross validation metrics often presented as proof of model performance and here used only to tune ML models. However, to progress toward a flexible algorithm able to estimate GPC from RS data only, validated against but independent of ground-based GPC data for training. Successful demonstration of such out-of-sample (OOS) estimation would add substantially to the potential for practical application of the wider methods. To this end, this project attempted to predict GPC in entire paddocks unseen by the ML algorithm in training; predictions were tested in a field-wise LOO method, such that each field's data were successively used as the unseen test set for a model calibrated on all other fields. With all data from remaining fields included for model training, the amount of data from the LOO field available for training was reduced stepwise from 70 % to 10 %. Predictive skill declined, though not dramatically, at each step down to 10 %, but models failed when zero data from the LOO field were included in training. This result is to be expected: Like other regression-based estimates, a GBM prediction can perform adequately only within the bounds of the data on which it was trained; extrapolations are not reliable. In practice, for the present case, this means that a training set should incorporate representative coverage of the diverse plant trait ~ GPC relationships sufficient to allow the algorithm to learn the patterns inherent in them. This is problematic, because as established above, these relationships' relative weight of influence and even the direction of their effects are subject to drastic change according to soil properties, weather and phenology.

Other recent studies have also shown good results for OOS analyses while training and test datasets were consistent, but reported model failure where this was not true. Exemplifying the first case, Gómez *et al*. (2021) achieved good regional wheat yield prediction results (R^2 > 0.8) for both random 20 % holdouts and for a single year held out of ML training by adding met and historical yield data as GBM input features to MODIS seasonal NDVI. For wheat yield at regional scales in Australia, Cai *et al*. (2019) did similarly, adding non-RS meteorological inputs to satellite EVI and SIF, holding out single years and training ML on the remainder of their sample. They obtained very similar results to Gómez *et al*. (2021) but saw a sharp reduction in model skill for a single year of very different weather, emphasising the importance of training ML models on representative data. While the focus of these works on regional-scale yield diminishes their relevance to work on within-field variability, each demonstrated both the value of adding information from sources other than RS, i.e., met and production history, and the centrality of RS through its in-season currency. RS assessment of plant performance is indispensable to prospective in-season fertiliser DSS at scales useful for PA and aimed at improving GPC and/or N efficiency. However, to improve OOS results, RS could be augmented with year-to-date met or PAW assessments and historical information; while the former are difficult to obtain at spatial scales similar to GPC variability, the collection of high-resolution historical GPC data should be increasingly feasible as monitors become more common and may be a valuable model input.

Satellite TS are likely to be very relevant to any operational implementation of the methods described here, if only because repeated airborne HS campaigns would be prohibitive in terms of cost and complexity, while satellite TS are easily available at low cost. Besides, the ultra-high spectral and spatial resolution of the airborne RS used in this project, while crucial to scientific advance, may not be strictly necessary for GPC estimation in commercial contexts. This provides a counterpoint to recent recommendations of HS data as

indispensable for such uses (Pancorbo *et al*., 2023; Raya-Sereno *et al*., 2021); while such appears true for optimal scientific results, and HS sensing is certainly better when only single images are available, multispectral satellite TS is able to compensate at least in some situations.

Despite the negative effects of lower spectral resolution, and the associated loss of physiological indicators found highly relevant to HS analyses, particularly in moderate growing conditions, the advantage of satellite RS becomes clear through TS. Prediction metrics from TS were, overall, similar to and in severe water stress, better than those based on HS (Fig. 4.17), demonstrating that S2 TS can be used to predict GPC with moderately good fidelity even where prediction was at its most difficult. Bread wheat from CZ1-20 is a case in point: while HS+CWSI gave R^2 = 0.54 / RMSE = 0.66 %, and accuracy from single-image S2, without SIF, diminished to R^2 = 0.33 / RMSE = 0.80 %, addition of the TS reinstated performance back to equivalence with HS+CWSI, at R^2 = 0.52 and RMSE = 0.67 %. In the same site/year, a parallel result was seen for durum, albeit with a more severe drop in performance from single-image S2. The improvement over single-date S2 estimates is also large in CZ2, where they come off a particularly low base (Fig. 4.17).

Such improvements are owed to the incorporation of early-season structural information. Along TS and within site/year combinations in all situations, predictive performance was better during early development than in the mid-season, showing both that establishment and early vigour are important to eventual GPC outcomes, and that our ML approach is sensitive to the same biological reality. Canopy structure (LAI) played a far greater role in S2 TS than in single-image S2 analyses because it encompasses differences in vigour during the early part of seasons. For example, when TS images were considered separately for CZ1-19 bread, LAI took 60—80% of total importance at each stage up to Z65, whereafter importance switched entirely to C_w (Fig. 4.13a). This may be linked to the acknowledged strong role for

early vigour as a predictor of crop performance more generally, including yield prediction, for which satellite EVI has been found informative in the early- and mid-season but of little value later (Cai *et al*., 2019). Similarly, early season airborne NDVI was found more closely related to GPC than later NDVI (Stoy *et al*., 2022). It is also likely, however, that the importance of LAI stems from its function as a proxy for biomass, hence soil drying, the degree of later drought and relative amount and duration of photosynthesis, which are then shown by high C_w importance in the final image of the same CZ1-19 bread wheat season. This site/year/product combination also demonstrates the ability of TS to diagnose GPC-related crop performance changes in response to severe stress, which in some circumstances could inform strategic agronomic interventions. Under lower water stress, an even spread of importance across all four inverted parameters and throughout seasons shows relatively even plant performance across time and within fields. In turn, trait ~ GPC associations are weaker, and a smaller degree of PAW change through time is evident; this dynamic supplies less intelligence for potential agronomic interventions.

When all dates were pooled for CZ1-19 bread, Z15 (seedlings, pre-tillering) LAI took > 40% of importance, and C_w at anthesis took > 30% (Fig. 4.14a), similarly as described above for image-by-image modelling. This sudden change was seen only under drought, but a gradual switch from LAI to C_w was seen in CZ1-19 durum also, despite a likely reduced crop water demand after frosting and noise from weather damage more broadly. In other situations, the distribution of importance between dates and traits was quite even whether image dates were separate or pooled but always included substantial contributions from LAI. This, and the decline of LAI from a mid-season peak when all date / feature combinations were pooled shows the high relevance to GPC of above-ground biomass, as a source of both proteins and CHO. Again, PAW variability, shown by C_w importance, is low early in the season partly because crops have had less time – and biomass – to withdraw water. Increasing C_w

importance through seasons likely reflects increasingly variable soil moisture as plants differentially accumulate biomass, hence capacity to dry the soil, allowing soil properties in turn to exert more influence. Of the inverted traits retrievable from S2, C_{a+b}, and C_m were consistently the least influential but peaked during mid-late vegetative growth (Z50—70) in both image-by-image and pooled models. This reflects the physiological links of biochemistry and leaf biomass content to GPC, most prominently in mild conditions, as sources for translocation. In pooled CZ2-19 models, C_{a+b} and C_m at early stem elongation (Z31) together contributed 31 % of total gain, a very substantial contribution where this was spread across 20 features. This high importance of combined leaf thickness and C_{a+b} in dry conditions also appears to indicate a strong role for early N content in GPC determination.

Cai *et al.* (2019) have argued that the optimal timing for Australian wheat yield prediction accuracy is before October; although this assertion comes from a national-scale study, where definition of phenological stage was not possible, in the context of the current work it is reasonable, corresponding to Z43—70. Others report that the addition of extra S2 TS data after June, in a study considering a large number of fields across a region, gave little improvement for yield estimation due to incipient senescence (Hunt *et al.*, 2019). Similar timings may hold for GPC prediction also, given that the two quantities co-vary in opposition; though little research has addressed this directly, some contend that Z65 is effective for GPC (C. Tan *et al*., 2020; Wang *et al*., 2014; Zhao *et al*., 2019), and at plot scale, Raya-Sereno *et al*. (2021) had best results for grain N content at early milk (during grain filling; ~Z73). Rodrigues *et al.* (2018) contend instead that early- to mid-season RS (Z31—Z65) is best for hyperspectral VI-based GPC assessment under commercial conditions, but this assertion is based on very weak coefficients of determination between selected VIs and GPC.

Evidence from the current TS analyses did not contradict the fairly diverse findings above, but also supported them only relatively weakly. Although the paucity of cloud free images after ~Z50 makes it difficult to support suggestions of Z65 and later, hence from around flowering and into grain filling, as best for prediction, the attendant studies do confirm that late-season images tend to bring higher accuracy. The lack of late images leaves an undesirable gap, most notably in the milder, wetter 2020 when no image was available after Z50/1400 GDDAS in either zone (Table 3.3, Fig. 4.16). According to the survey of RS timing for GPC estimation immediately above, this missingness excluded perhaps the most crucial crop information; this is reflected in the lower skill of single-image analyses from 2020. While performance may improve further with later cloud-free images, the likelihood of obtaining these would not be substantially higher in any similarly mild season, a familiar problem in satellite RS research. TS stacking, however, partially compensated for these issues in 2020, and improved performance in every case. Moreover, despite the many factors that can intervene between early potential and harvest reality, depending on conditions specific S2 inverted parameters can contribute to GPC prediction as early as Z15, and predictions this early can give reasonable results.

Overall, the predictive skill improvements gained from adding airborne SIF to S2 analyses suggest that if it were available at higher spatial and temporal resolution, TS SIF may strongly improve GPC prediction. Nevertheless, it is probable that a strong role for SIF, with substantial improvement of predictive performance and associated high gain, would be weighted toward the grain filling period, when spatial variability in soil moisture is most pronounced and influential, at least in dryland, semi-arid environments, and to relatively benign moisture conditions, because if soil is either very wet or very dry there should be less variability in SIF. These hypotheses also accord with the findings of Sloat *et al*. (2021); in their work, early-season SIF contributed to overestimation of seasonal productivity,

particularly under drought. Given that terminal drought is expected in virtually all Australian wheat cropping, such relative unreliability of early-season SIF in harvest prediction will likely affect future within-field TS analyses.

Again, from the perspective of practical application of such methods into agronomic interventions, viz, strategic N applications, a question arises regarding how early in seasons actionable information would be useful. A recent meta-analysis showed that later (≥ Z55) foliar N applications were more effective in increasing GPC than earlier applications, and that higher ratios of late to total N applied also increased GPC relatively more (Giordano *et al*., 2023). This implies that, as long as crops are not senescent or water stressed to the point of impeded translocation, supplementary N applications can potentially be delayed until very late in seasons. However, a strong caveat also exists, at least for Australia where foliar fertiliser applications are rare: broadcast urea, here the most common source of synthetic N, was among the least likely forms to elicit a positive GPC response (Giordano *et al*., 2023). Though broadcast N has also been associated with GPC increases (Zebarth *et al*., 2007; Zhang *et al*., 2022), solid forms require rain for wash-in, so any such strategy would require quick reactions to weather forecasts.

6. Conclusions

This chapter contains included material from:

Longmire, A.R., Poblete, T., Hunt, J.R., Chen, D., Zarco-Tejada, P.J., 2022. Assessment of crop traits retrieved from airborne hyperspectral and thermal remote sensing imagery to predict wheat grain protein content. *ISPRS Journal of Photogrammetry and Remote Sensing 193, 284–298*. https://doi.org/10.1016/j.isprsjprs.2022.09.015

Longmire, A.R., Poblete, T., Hornero, A., Chen, D., Zarco-Tejada, P.J., (accepted 2023-10-27). Estimation of grain protein content in commercial bread and durum wheat fields via traits inverted by radiative transfer modelling from Sentinel-2 timeseries. *ISPRS Journal of Photogrammetry and Remote Sensing*.

—

A comprehensive literature review highlighted that despite solid understandings in crop science of the physiological links between plant stress and wheat GPC, and substantial work on RS estimation of plant traits associated with GPC, published investigation into GPC estimation directly from those or other RS traits was sparse. This situation persists despite the crucial importance of GPC on farms and in diets and the prodigious amount of research into grain yield prediction. Hence while techniques for estimating relevant parameters including LAI, leaf- and canopy C_{a+b} and N content are advanced, comparatively little research linked these with harvest GPC. A further omission concerned plant stress indicators and their associations with GPC.

To address these gaps, airborne hyperspectral and thermal data were collected above experimental plots and commercial wheatfields, each in two seasons of diverse precipitation and across two locations. Plant traits for each GPC location were retrieved by recognised means: vegetation indices, radiative transfer model inversions, solar-induced fluorescence and water stress. The project used a machine learning algorithm to concurrently estimate GPC and quantify the relative importance of retrieved traits to estimations. Similar retrieval

and ML methods were adapted to the reduced spectral but higher temporal resolution of Sentinel-2 satellite images: again, suitable VIs were calculated and RTM inversions conducted for each GPC location, and for each image capture date in season timeseries. Solar-induced fluorescence and CWSI, unavailable from S2 data, were used to supply supplementary information to satellite traits. These parallel procedures allowed comparison of trait importances and predictive skill between hyperspectral and multispectral data sources, evaluating the importance of thermal and hyperspectral data to account for water stress and fluorescence emission, respectively. The S2 timeseries data were then applied in two ways: first using features from different dates to build separate models, then stacking all features/dates into a single model. These approaches allowed trait importance and predictive skill to be tracked through seasons, and for the potential of stacked TS images to improve model skill. At each step, multicollinearity among plant traits was reduced using data-driven methods to avoid diluting their relative importance as ML input features. Overall, these procedures permitted a clear view of the relative magnitude of plant traits as GPC predictors in commercial conditions.

Hence, the first study identified the most important hyperspectral-based plant traits related to GPC in rainfed wheat under variable stress levels. In experimental plots, two variants of the PRI, related to the xanthophyll pigment cycle, showed consistent trends very similar to GPC along the induced N nutrition gradient, and in this respect performed better than any other spectral trait. Implementation of a gradient boosted machine (GBM) elucidated relationships between HS- derived input features and GPC in commercial crops. The thermal CWSI, a proxy indicator of canopy transpiration and hence of water stress, contributed strongly to GPC modelling under water stress conditions, while Anth, C_{x+c}, PRI and SIF consistently showed high importance in moderate conditions. Structural indicators LAI or any other structural proxy such as NDVI contributed substantially less. Using a model built with

thermal and physiological traits quantified by RTM inversion, the GBM predicted GPC with R^2 = 0.80 and RMSE = 0.62 % between predicted and observed GPC in bread wheat crops.

These methods were then adapted to the reduced spectral but higher temporal resolution of satellite images. Adding further wheat fields to those considered in the HS analyses, and therefore more diverse soil and weather conditions, plant and canopy traits inverted from Sentinel-2 images were assessed for GPC prediction. Using equivalent modelling methods, accuracy from single images was generally lower when based on S2 than on airborne HS traits. However, in very dry conditions, the best model using a single S2 image and made with PRO4SAIL-inverted C_{a+b}, C_m, C_w and LAI outperformed the HS+CWSI model in the same environment. Adding timeseries S2 inverted traits substantially improved all models over single-image versions; improvement was very strong in benign conditions and compensated for the accuracy reduction caused by switching from HS to S2. The best predictive performance was achieved by stacking retrieved parameters from all dates as inputs to a single model (R^2 = 0.86, RMSE = 0.56 %). The order and relative importance of S2 plant traits were similar to airborne HS: S2 importance was dominated by C_w in drought but evenly spread between structural and physiological features in benign conditions.

6.1. Major findings

As detailed in Chapter 2, this thesis and the broader research project were structured around the below research questions; each is now addressed:

1) *Which imaging spectroscopy-based plant physiological traits, retrievable from hyperspectral and thermal airborne remote sensing, are most closely associated with GPC?*
2) *Are these traits stable across experimental and commercial contexts, different locations and contrasting seasonal weather conditions?*

Very dry conditions increased the importance of water stress measures relative to other plant traits, while in milder conditions physiological traits were emphasised. Airborne CWSI was associated with higher GPC in plot trials and was the dominant predictor of GPC in dry cropping conditions and remained important in moderate conditions. Canopy structural components were of relatively low value for GPC prediction, except under drought where CWSI was excluded, and showed limited association with GPC in plot trials. In addition to C_{a+b} and SIF, physiological indicators of moderate plant stress C_{x+c} and PRI were closely associated with GPC in experimental and commercial settings, and their contribution to GPC estimation was substantial in moderate growing conditions. These conclusions are, in general, valid across locations and for both experimental and commercial contexts.

3) *Do plant traits retrieved from Sentinel-2 satellite images show similarities to those from hyperspectral images as predictors of GPC?*

With respect to their categories, whether physiological, structural or related to water stress, and their importance for GPC estimation, single-image S2 traits were markedly similar to airborne thermal and HS traits: S2-inverted C_w assumed the importance and predictive power of CWSI in dry conditions, and indeed was excluded due to its collinearity with the S2 parameters. In moderate conditions, across both years, and in both HS and S2 analyses, model gain was more evenly spread between the relevant measure of water sufficiency and other features. C_{a+b}, as the only purely physiological parameter retrievable from S2, was consistently of strong secondary importance as were the HS physiological measures in moderate conditions. In each analysis of moderate conditions, SIF where included was quasi-equivalent with C_w/CWSI as contributing most gain.

4) *What is the effect on model predictive skill of bandset reduction from hyperspectral to Sentinel-2 when traits from similar time points in each season are retrieved from the respective datasets?*

In very dry conditions this change in data source is associated with a marginal decline in coefficient of determination; in moderate conditions the decline is worse but not dramatic. Relative error increases under such a change, but in all analyses RMSE remains below 1 % GPC.

5) *Crop water stress index is a proxy for transpiration and therefore atmospheric gas exchange, while SIF is proportional to photosynthetic rate. Each should therefore add predictive power for GPC estimation via the carbohydrate dilution principle, but neither is available from Sentinel-2. What are the contributions of CWSI and/or SIF to model predictive skill when added to S2 traits?*

Especially in very dry conditions, CWSI substantially improved GPC estimation relative to HS traits without CWSI. However, CWSI was not added to S2 models due to collinearity with S2 inverted trait C_w; because this absence did not reduce model skill, at least in terminal drought, it is asserted that C_w can be an effective proxy for CWSI. SIF improved S2 models' performance metrics, most strongly under moderate cropping conditions.

6) *What plant trait dynamics are evident within and across seasons?*

With all features pooled, in dry conditions, gain transitions from LAI early to C_w late in the season; in moderate conditions, gain is distributed evenly between traits and image dates, with some weighting toward late images.

7) *What is the effect on model predictive skill of including TS elements in predictive models?*

When TS images are treated as input features for separate models, in coefficient of determination generally increases and error decreases as seasons proceed. In very dry conditions, skill declined in the midseason, after the first image, and recovered later. Stacking of trait/date combinations to form a single model generated the best results of any modelling, including HS, in very dry conditions, and reached similar levels to HS otherwise. Stacked, seasonal models were always better than the single-image S2 models tested.

6.2. Challenges and limitations

Remote sensing and its associated retrieval methods are well developed and amply able to capture many aspects of plants' complex interactions with their environments; this is demonstrated in the detailed retrievals of plant traits, and in their evolution through time, presented above. Machine learning algorithms are also powerful tools, capable of interpreting multiple nonlinear relationships between predictor and target variables to estimate the latter. However, given the inherent complexity, ML is not expected to fully capture or precisely interpret the influence of RS traits on GPC.

- The failure of the predictive model to estimate GPC for entire fields unseen the algorithm during the training phase exposed a substantial limitation of the study. Although this failure is easily explained as the inability of regression-based approaches to reliably extrapolate beyond the bounds of observed data, it indicates a need for further work. Moreover, this is a limitation of the ML methods, not of the remote sensing procedures, and is likely to be remediable with refinements. Data augmentation approaches may prevent model failure in fields contributing sparse or zero data to training and facilitate OOS prediction methods for entire fields where GPC is not known *ab initio*. Although this would require large datasets and exhaustive testing, it warrants pursuit.

- The short season of S2 images in some cases terminated data collection well before anthesis. Given that plant development and conditions around and after anthesis, particularly during grain filling, are known to strongly influence GPC, this shortcoming was significant.
- With specific relevance to the regions considered here, there is a need for more information, for example regarding soil variability, to improve GPC predictions in the difficult edaphic conditions of CZ2.

6.3. Future steps

The author hypothesises that the methods proposed, developed, and tested in this project, or similar methods, could function as well in other continents and agronomic contexts and proposes that this be tested. Indeed, with regard to RS-based GPC estimation in wheat, many refinements are possible and many avenues remain open for investigation. For example, larger collections of GPC and RS data should be sought, so that a wider range of seasonal and agronomic conditions can be incorporated as training data for predictive algorithms. As protein monitors become more commonplace, GPC data will be more readily available for this purpose. Regarding temporal aspects, later TS images, extending at least past the primary vegetative growth phases and preferably anthesis, are also likely to improve GPC prediction where these are available. Spaceborne SIF timeseries, when in future they become available at within-field resolution, should be tested for their impact on model skill. Assessed by the results obtained with airborne SIF, such impacts should be positive, particularly late in seasons. The effectiveness of C_w as a replacement for CWSI should be tested in a broader range of agricultural and horticultural applications.

Although pure RS approaches beguile, with distinct practical advantages, many other sources of ground-based information are available and would likely improve GPC prediction. It would be valuable to assess inputs including year-to-date rainfall or real-time soil moisture

information from monitoring probes. In regions affected by dune-swale morphology, relative elevation may also add predictive power. From the management perspective, agronomic information such as cropping history and past N mass balance or training on multi-year databases incorporating historic GPC responses may also be beneficial. Furthermore, RS may also be used to generate additional inputs, such as information from soil moisture monitoring. The studies presented here provide a basis for such future investigations.

6.4. Scientific contribution and implications

Specifically, the current thesis contributes to remote sensing and crop science because it:

- Demonstrates robust new RS-based methods for GPC estimation in commercial wheat crops, identifying plant traits retrievable from airborne hyperspectral and Sentinel-2 data that are physiologically relevant to GPC.
- Proposes and advances GPC estimation via plant performance parameters retrieved by RTM inversion. The component studies are the first published works to use inverted parameters for this purpose.
- Demonstrates that HS indicators of photosynthetic activity PRI and SIF are consistently associated with GPC in plot studies and make robust contributions to GPC prediction in crops. C_{x+c}, indicators of pre-visual physiological stress, are also shown as linked to harvest GPC. Further investigation of the physiological stress indicators found here to be associated with GPC could advance knowledge of stress diagnosis and trait~quality relationships in other food crops.
- Demonstrates the sufficiency of S2 data, via timeseries analysis of PROSAIL-inverted parameters, for GPC estimation with low error. Because S2 data are available across large spatial extents at spatial resolution relevant to PA applications, associated research has potential to advance practical DSS and interventions to improve farm incomes and N efficiency.

- Confirms the high importance of water status indicators in GPC estimation in crops under terminal drought and shows S2 inverted C_w to be an effective proxy for thermal CWSI. As assessed in the conditions and with the methods of this study, C_w may obviate the need for thermal data at all. This has potential to simplify and expedite RS-based assessments of water stress without recourse to thermal sensors or ground data collection.
- Confirms the usefulness of on-combine protein monitor data and machine learning as tools for predictive GPC modelling.

The approaches and analyses developed in this project constitute novel contributions toward GPC estimation from RS data sources. As wheat is one of the world's staples and primary sources of protein, and while access to RS, GPC monitoring and PA techniques become more common, wheat GPC prediction will be increasingly relevant. It is hoped that the methods presented herein can assist in maintaining the profitability and minimising the environmental impact of cropping industries.

***

7. Bibliography

Abdi, A.M., 2020. Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data. GIScience Remote Sens. 57, 1–20. https://doi.org/10.1080/15481603.2019.1650447

Adamchuk, V.I., 2013. Theoretical basis for sensor-based in-season nitrogen management, in: Stafford, J.V. (Ed.), Precision Agriculture '13. Academic Publishers, Wageningen, pp. 403–410. https://doi.org/10.3920/978-90-8686-778-3_49

Akinwande, M.O., Dikko, H.G., Samson, A., 2015. Variance Inflation Factor: As a Condition for the Inclusion of Suppressor Variable(s) in Regression Analysis. Open J. Stat. 05, 754. https://doi.org/10.4236/ojs.2015.57075

Allen, W.A., Gausman, H.W., Richardson, A.J., Thomas, J.R., 1969. Interaction of Isotropic Light with a Compact Plant Leaf. JOSA 59, 1376–1379. https://doi.org/10.1364/JOSA.59.001376

Altenbach, S.B., DuPont, F.M., Kothari, K.M., Chan, R., Johnson, E.L., Lieu, D., 2003. Temperature, Water and Fertilizer Influence the Timing of Key Events During Grain Development in a US Spring Wheat. J. Cereal Sci. 37, 9–20. https://doi.org/10.1006/jcrs.2002.0483

Andrews, M., Lea, P.J., 2013. Our nitrogen 'footprint': the need for increased crop nitrogen use efficiency. Ann. Appl. Biol. 163, 165–169. https://doi.org/10.1111/aab.12052

Angus, J.F., Fischer, R.A., 1991. Grain protein responses to nitrogen applied to wheat growing on a red earth. Aust. J. Agric. Res. 42, 735–746. https://doi.org/10.1071/ar9910735

Angus, J.F., Grace, P.R., 2017. Nitrogen balance in Australia and nitrogen use efficiency on Australian farms. Soil Res. 55, 435–450. https://doi.org/10.1071/SR16325

Apan, A., Kelly, R., Phinn, S., Strong, W., Lester, D., Butler, D., Robson, A., 2006. Predicting grain protein content in wheat using hyperspectral sensing of in-season crop canopies and partial least squares regression. Int. J. Geoinformatics 2, 93–108.

Armstrong, R.D., Fitzpatrick, J., Rab, M.A., Abuzar, M., Fisher, P.D., O'Leary, G.J., 2009. Advances in precision agriculture in south-eastern Australia. III. Interactions between soil properties and water use help explain spatial variability of crop production in the Victorian Mallee. Crop Pasture Sci. 60, 870–884. https://doi.org/10.1071/CP08349

Asner, G.P., 1998. Biophysical and Biochemical Sources of Variability in Canopy Reflectance. Remote Sens. Environ. 64, 234–253. https://doi.org/10.1016/S0034-4257(98)00014-5

Asseng, S., Ewert, F., Martre, P., Rötter, R.P., Lobell, D.B., Cammarano, D., Kimball, B.A., Ottman, M.J., Wall, G.W., White, J.W., Reynolds, M.P., Alderman, P.D., Prasad, P.V.V., Aggarwal, P.K., Anothai, J., Basso, B., Biernath, C., Challinor, A.J., De Sanctis, G., Doltra, J., Fereres, E., Garcia-Vila, M., Gayler, S., Hoogenboom, G., Hunt, L.A., Izaurralde, R.C., Jabloun, M., Jones, C.D., Kersebaum, K.C., Koehler, A.-K., Müller, C., Naresh Kumar, S., Nendel, C., O'Leary, G., Olesen, J.E., Palosuo, T., Priesack, E., Eyshi Rezaei, E., Ruane, A.C., Semenov, M.A., Shcherbak, I., Stöckle, C., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Thorburn, P.J., Waha, K., Wang, E., Wallach, D., Wolf, J., Zhao, Z., Zhu, Y., 2015. Rising temperatures reduce global wheat production. Nat. Clim. Change 5, 143–147. https://doi.org/10.1038/nclimate2470

Asseng, S., Foster, I., Turner, N.C., 2011. The impact of temperature variability on wheat yields. Glob. Change Biol. 17, 997–1012. https://doi.org/10.1111/j.1365-2486.2010.02262.x

Atzberger, C., 2010. Inverting the PROSAIL canopy reflectance model using neural nets trained on streamlined databases. J. Spectr. Imaging 1. https://doi.org/10.1255/jsi.2010.a2

Bacour, C., Baret, F., Béal, D., Weiss, M., Pavageau, K., 2006. Neural network estimation of LAI, fAPAR, fCover and LAI×Cab, from top of canopy MERIS reflectance data: Principles and validation. Remote Sens. Environ. 105, 313–325. https://doi.org/10.1016/j.rse.2006.07.014

Bacour, C., Jacquemoud, S., Leroy, M., Hautecœur, O., Weiss, M., Prévot, L., Bruguier, N., Chauki, H., 2002. Reliability of the estimation of vegetation characteristics by inversion of three canopy reflectance models on airborne POLDER data. Agronomie 22, 555–565. https://doi.org/10.1051/agro:2002039

Baffes, J., Koh, W.C., 2022. Fertilizer prices expected to remain higher for longer. World Bank Blogs. URL https://blogs.worldbank.org/opendata/fertilizer-prices-expected-remain-higher-longer (accessed 5.26.22).

Bannari, A., Morin, D., Bonn, F., Huete, A.R., 1995. A review of vegetation indices. Remote Sens. Rev. 13, 95–120. https://doi.org/10.1080/02757259509532298

Baret, F., Guyot, G., 1991. Potentials and limits of vegetation indices for LAI and APAR assessment. Remote Sens. Environ. 35, 161–173. https://doi.org/10.1016/0034-4257(91)90009-U

Baret, F., Houlès, V., Guérif, M., 2007. Quantification of plant stress using remote sensing observations and crop models: the case of nitrogen management. J. Exp. Bot. 58, 869–880. https://doi.org/10.1093/jxb/erl231

Baret, F., Jacquemoud, S., Guyot, G., Leprieur, C., 1992. Modeled analysis of the biophysical nature of spectral shifts and comparison with information content of broad bands. Remote Sens. Environ. 41, 133–142. https://doi.org/10.1016/0034-4257(92)90073-S

Barnes, E.M., Clarke, T.R., Richards, S.E., Colaizzi, P.D., Haberland, J., Kostrzewski, M., Waller, P., Choi, C., Riley, E., Thompson, T., Lascano, R.J., Li, H., Moran, M.S., 2000. Coincident detection of crop water stress, nitrogen status and canopy density using ground-based multispectral data, in: Proceedings of the Fifth International Conference on Precision Agriculture. p. 15.

Barnes, J.D., Balaguer, L., Manrique, E., Elvira, S., Davison, A.W., 1992. A reappraisal of the use of DMSO for the extraction and determination of chlorophylls a and b in lichens and higher plants. Environ. Exp. Bot. 32, 85–100. https://doi.org/10.1016/0098-8472(92)90034-Y

Barraclough, P.B., Howarth, J.R., Jones, J., Lopez-Bellido, R., Parmar, S., Shepherd, C.E., Hawkesford, M.J., 2010. Nitrogen efficiency of wheat: Genotypic and environmental variation and prospects for improvement. Eur. J. Agron. 33, 1–11. https://doi.org/10.1016/j.eja.2010.01.005

Baskerville, G.L., Emin, P., 1969. Rapid Estimation of Heat Accumulation from Maximum and Minimum Temperatures. Ecology 50, 514–517. https://doi.org/10.2307/1933912

Bassi, R., Høyer-Hansen, G., Barbato, R., Giacometti, G.M., Simpson, D.J., 1987. Chlorophyll-proteins of the photosystem II antenna system. J. Biol. Chem. 262, 13333–13341. https://doi.org/10.1016/S0021-9258(18)45205-2

Bauer, D.F., 1972. Constructing Confidence Sets Using Rank Statistics. J. Am. Stat. Assoc. 67, 687–690. https://doi.org/10.1080/01621459.1972.10481279

Becker-Reshef, I., Vermote, E., Lindeman, M., Justice, C., 2010. A generalized regression-based model for forecasting winter wheat yields in Kansas and Ukraine using MODIS data. Remote Sens. Environ. 114, 1312–1323. https://doi.org/10.1016/j.rse.2010.01.010

Belwalkar, A., Poblete, T., Longmire, A., Hornero, A., Hernandez-Clemente, R., Zarco-Tejada, P.J., 2022. Evaluation of SIF retrievals from narrow-band and sub-nanometer airborne hyperspectral imagers flown in tandem: Modelling and validation in the context of plant phenotyping. Remote Sens. Environ. 273, 112986. https://doi.org/10.1016/j.rse.2022.112986

Beres, B.L., Rahmani, E., Clarke, J.M., Grassini, P., Pozniak, C.J., Geddes, C.M., Porker, K.D., May, W.E., Ransom, J.K., 2020. A Systematic Review of Durum Wheat: Enhancing Production Systems by Exploring Genotype, Environment, and Management (G × E × M) Synergies. Front. Plant Sci. 11. https://doi.org/10.3389/fpls.2020.568657

Berger, K., Atzberger, C., Danner, M., D'Urso, G., Mauser, W., Vuolo, F., Hank, T., 2018. Evaluation of the PROSAIL Model Capabilities for Future Hyperspectral Model Environments: A Review Study. Remote Sens. 10, 85. https://doi.org/10.3390/rs10010085

Blackburn, G.A., 2007. Hyperspectral remote sensing of plant pigments. J. Exp. Bot. 58, 855–867. https://doi.org/10.1093/jxb/erl123

Blackburn, G.A., 1998. Quantifying Chlorophylls and Carotenoids at Leaf and Canopy Scales: An Evaluation of Some Hyperspectral Approaches. Remote Sens. Environ. 66, 273–285. https://doi.org/10.1016/S0034-4257(98)00059-5

Bly, A.G., Woodard, H.J., 2003. Foliar Nitrogen Application Timing Influence on Grain Yield and Protein Concentration of Hard Red Winter and Spring Wheat. Agron. J. 95, 335–338. https://doi.org/10.2134/agronj2003.3350

Bogard, M., Allard, V., Brancourt-Hulmel, M., Heumez, E., Machet, J.-M., Jeuffroy, M.-H., Gate, P., Martre, P., Le Gouis, J., 2010. Deviation from the grain protein concentration–grain yield negative relationship is highly correlated to post-anthesis N uptake in winter wheat. J. Exp. Bot. 61, 4303–4312. https://doi.org/10.1093/jxb/erq238

Bonfil, D.J., 2017. Wheat phenomics in the field by RapidScan: NDVI vs. NDRE. Isr. J. Plant Sci. 64, 41–54. https://doi.org/10.1080/07929978.2016.1249135

Borghi, B., Corbellini, M., Cattaneo, M., Fornasari, M.E., Zucchelli, L., 1986. Modification of the Sink/Source Relationships in Bread Wheat and its Influence on Grain Yield and Grain Protein Content*. J. Agron. Crop Sci. 157, 245–254. https://doi.org/10.1111/j.1439-037X.1986.tb00073.x

Borrelli, G.M., Ficco, D.B.M., Giuzio, L., Pompa, M., Cattivelli, L., Flagella, Z., 2011. Durum Wheat Salt Tolerance in Relation to Physiological, Yield and Quality Characters. Cereal Res. Commun. 39, 525–534. https://doi.org/10.1556/CRC.39.2011.4.7

Broge, N.H., Leblanc, E., 2001. Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. Remote Sens. Environ. 76, 156–172. https://doi.org/10.1016/S0034-4257(00)00197-8

Brooks, A., Jenner, C.F., Aspinall, D., 1982. Effects of Water Deficit on Endosperm Starch Granules and on Grain Physiology of Wheat and Barley. Funct. Plant Biol. 9, 423–436. https://doi.org/10.1071/pp9820423

Brown, H.E., Huth, N.I., Holzworth, D.P., Teixeira, E.I., Zyskowski, R.F., Hargreaves, J.N.G., Moot, D.J., 2014. Plant Modelling Framework: Software for building and running crop models on the APSIM platform. Environ. Model. Softw. 62, 385–398. https://doi.org/10.1016/j.envsoft.2014.09.005

Bureau of Meteorology, 2021. Climate Data Online [WWW Document]. URL http://www.bom.gov.au/climate/data/ (accessed 9.24.21).

Cai, Y., Guan, K., Lobell, D., Potgieter, A.B., Wang, S., Peng, J., Xu, T., Asseng, S., Zhang, Y., You, L., Peng, B., 2019. Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches. Agric. For. Meteorol. 274, 144–159. https://doi.org/10.1016/j.agrformet.2019.03.010

Calderon Flores, P., Yoon, J.S., Kim, D.Y., Seo, Y.W., 2021. Effect of chilling acclimation on germination and seedlings response to cold in different seed coat colored wheat (Triticum aestivum L.). BMC Plant Biol. 21, 252. https://doi.org/10.1186/s12870-021-03036-z

Calderón, R., Navas-Cortés, J.A., Lucena, C., Zarco-Tejada, P.J., 2013. High-resolution airborne hyperspectral and thermal imagery for early detection of Verticillium wilt of olive using fluorescence, temperature and narrow-band spectral indices. Remote Sens. Environ. 139, 231–245. https://doi.org/10.1016/j.rse.2013.07.031

Calderón, R., Navas-Cortés, J.A., Zarco-Tejada, P.J., 2015. Early Detection and Quantification of Verticillium Wilt in Olive Using Hyperspectral and Thermal Imagery over Large Areas. Remote Sens. 7, 5584–5610. https://doi.org/10.3390/rs70505584

Camino, C., González-Dugo, V., Hernández, P., Sillero, J.C., Zarco-Tejada, P.J., 2018. Improved nitrogen retrievals with airborne-derived fluorescence and plant traits quantified from VNIR-SWIR hyperspectral imagery in the context of precision agriculture. Int. J. Appl. Earth Obs. Geoinformation 70, 105–117. https://doi.org/10.1016/j.jag.2018.04.013

Camino, C., Gonzalez-Dugo, V., Hernandez, P., Zarco-Tejada, P.J., 2019. Radiative transfer Vcmax estimation from hyperspectral imagery and SIF retrievals to assess photosynthetic performance in rainfed and irrigated plant phenotyping trials. Remote Sens. Environ. 231, 111186. https://doi.org/10.1016/j.rse.2019.05.005

Cammarano, D., Fitzgerald, G., Basso, B., O'Leary, G., Chen, D., Grace, P., Fiorentino, C., 2011. Use of the Canopy Chlorophyl Content Index (CCCI) for Remote Estimation of Wheat Nitrogen Content in Rainfed Environments. Agron. J. 103, 1597–1603.

Campbell, C.A., Davidson, H.R., Winkelman, G.E., 1981. Effect of nitrogen, temperature, growth stage and duration of moisture stress on yield components and protein content of manitou spring wheat. Can. J. Plant Sci. 61, 549–563. https://doi.org/10.4141/cjps81-078

Campos, I., González-Gómez, L., Villodre, J., Calera, M., Campoy, J., Jiménez, N., Plaza, C., Sánchez-Prieto, S., Calera, A., 2019. Mapping within-field variability in wheat yield and biomass using remote sensing vegetation indices. Precis. Agric. 20, 214–236. https://doi.org/10.1007/s11119-018-9596-z

Cao, J., Zhang, Z., Tao, F., Zhang, L., Luo, Y., Han, J., Li, Z., 2020. Identifying the Contributions of Multi-Source Data for Winter Wheat Yield Prediction in China. Remote Sens. 12, 750. https://doi.org/10.3390/rs12050750

Cargill Australia Ltd., 2023. Cargill Pricing Hub [WWW Document]. URL https://au.mycargill.com/PricingHub/Bid (accessed 7.27.23).

Carter, G.A., 1994. Ratios of leaf reflectances in narrow wavebands as indicators of plant stress. Int. J. Remote Sens. 15, 697–703. https://doi.org/10.1080/01431169408954109

Cassman, K.G., Dobermann, A., Walters, D.T., 2002. Agroecosystems, Nitrogen-use Efficiency, and Nitrogen Management. AMBIO J. Hum. Environ. 31, 132–140. https://doi.org/10.1579/0044-7447-31.2.132

Cazzonelli, C.I., Pogson, B.J., 2010. Source to sink: regulation of carotenoid biosynthesis in plants. Trends Plant Sci. 15, 266–274. https://doi.org/10.1016/j.tplants.2010.02.003

Ceccato, P., Flasse, S., Tarantola, S., Jacquemoud, S., Grégoire, J.-M., 2001. Detecting vegetation leaf water content using reflectance in the optical domain. Remote Sens. Environ. 77, 22–33. https://doi.org/10.1016/S0034-4257(01)00191-2

Chalker-Scott, L., 2002. Do anthocyanins function as osmoregulators in leaf tissues? Adv. Bot. Res. 37, 103–127. https://doi.org/10.1016/S0065-2296(02)37046-0

Chalker-Scott, L., 1999. Environmental Significance of Anthocyanins in Plant Stress Responses. Photochem. Photobiol. 70, 1–9. https://doi.org/10.1111/j.1751-1097.1999.tb01944.x

Chapin, F.S., Wardlaw, I.F., 1988. Effect of Phosphorus Deficiency on Source-Sink Interactions Between the Flag Leaf and Developing Grain in Barley. J. Exp. Bot. 39, 165–177. https://doi.org/10.1093/jxb/39.2.165

Chappelle, E.W., Kim, M.S., McMurtrey, J.E., 1992. Ratio analysis of reflectance spectra (RARS): An algorithm for the remote estimation of the concentrations of chlorophyll A, chlorophyll B, and carotenoids in soybean leaves. Remote Sens. Environ. 39, 239–247. https://doi.org/10.1016/0034-4257(92)90089-3

Chen, D., Lam, S.K., Mosier, A.R., Eckard, R., Vitousek, P., 2016. Should soil nitrogen be mined?, in: Proceedings of the 2016 International Nitrogen Initiative Conference. Presented at the 2016 International Nitrogen Initiative Conference, Melbourne, Australia.

Chen, D., Suter, H., Islam, A., Edis, R., Freney, J.R., Walker, C.N., 2008. Prospects of improving efficiency of fertiliser nitrogen in Australian agriculture: a review of enhanced efficiency fertilisers. Soil Res. 46, 289–301.

Chen, J.M., 1996. Evaluation of Vegetation Indices and a Modified Simple Ratio for Boreal Applications. Can. J. Remote Sens. 22, 229–242. https://doi.org/10.1080/07038992.1996.10855178

Chen, P., Haboudane, D., Tremblay, N., Wang, J., Vigneault, P., Li, B., 2010. New spectral indicator assessing the efficiency of crop nitrogen treatment in corn and wheat. Remote Sens. Environ. 114, 1987–1997. https://doi.org/10.1016/j.rse.2010.04.006

Chen, T., Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16. Association for Computing Machinery, New York, NY, USA, pp. 785–794. https://doi.org/10.1145/2939672.2939785

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., Zhou, T., Li, M., Xie, J., Lin, M., Geng, Y., Li, Y., 2021. xgboost: Extreme Gradient Boosting.

Chen, X., Cui, Z., Fan, M., Vitousek, P., Zhao, M., Ma, W., Wang, Zhenlin, Zhang, Weijian, Yan, X., Yang, J., Deng, X., Gao, Q., Zhang, Q., Guo, S., Ren, J., Li, S., Ye, Y., Wang, Zhaohui, Huang, J., Tang, Q., Sun, Y., Peng, X., Zhang, J., He, M., Zhu, Y., Xue, J., Wang, G., Wu, Liang, An, N., Wu, Liangquan, Ma, L., Zhang, Weifeng, Zhang, F., 2014. Producing more grain with lower environmental costs. Nature 514, 486–489. https://doi.org/10.1038/nature13609

Chen, X.-P., Cui, Z.-L., Vitousek, P.M., Cassman, K.G., Matson, P.A., Bai, J.-S., Meng, Q.-F., Hou, P., Yue, S.-C., Römheld, V., Zhang, F.-S., 2011. Integrated soil–crop system management for food security. Proc. Natl. Acad. Sci. 108, 6399–6404. https://doi.org/10.1073/pnas.1101419108

Cheng, E., Zhang, B., Peng, D., Zhong, L., Yu, L., Liu, Y., Xiao, C., Li, C., Li, X., Chen, Y., Ye, H., Wang, H., Yu, R., Hu, J., Yang, S., 2022. Wheat yield estimation using remote sensing data based on machine learning approaches. Front. Plant Sci. 13. https://doi.org/10.3389/fpls.2022.1090970

Cheng, X., Yang, G., Xu, X., Chen, T., Li, Z., Feng, H., Wang, D., 2014. [Estimating canopy water content in wheat based on new vegetation water index]. Guang Pu Xue Yu Guang Pu Fen Xi Guang Pu 34, 3391–3396.

Chlingaryan, A., Sukkarieh, S., Whelan, B., 2018. Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. Comput. Electron. Agric. 151, 61–69. https://doi.org/10.1016/j.compag.2018.05.012

Clancy, P., Heiken, D., n.d. Montana Field Trials for On Combine NIR Analyzer.

Clevers, J.G.P.W., Gitelson, A.A., 2013. Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on Sentinel-2 and -3. Int. J. Appl. Earth Obs. Geoinformation 23, 344–351. https://doi.org/10.1016/j.jag.2012.10.008

Clevers, J.G.P.W., Kooistra, L., 2012. Using Hyperspectral Remote Sensing Data for Retrieving Canopy Chlorophyll and Nitrogen Content. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 5, 574–583. https://doi.org/10.1109/JSTARS.2011.2176468

Clevers, J.G.P.W., Kooistra, L., Van den Brande, M.M.M., 2017. Using Sentinel-2 Data for Retrieving LAI and Leaf and Canopy Chlorophyll Content of a Potato Crop. Remote Sens. 9, 405. https://doi.org/10.3390/rs9050405

Colaço, A.F., Bramley, R.G.V., 2018. Do crop sensors promote improved nitrogen management in grain crops? Field Crops Res. 218, 126–140. https://doi.org/10.1016/j.fcr.2018.01.007

Colaço, A.F., Povh, F.P., Molin, J.P., Romanelli, T.L., 2012. Energy assessment for variable rate nitrogen application. Agric. Eng. Int. CIGR J. 14, 85–90.

Colaço, A.F., Richetti, J., Bramley, R.G.V., Lawes, R.A., 2021. How will the next-generation of sensor-based decision systems look in the context of intelligent agriculture? A case-study. Field Crops Res. 270, 108205. https://doi.org/10.1016/j.fcr.2021.108205

Colasuonno, P., Marcotuli, I., Blanco, A., Maccaferri, M., Condorelli, G.E., Tuberosa, R., Parada, R., de Camargo, A.C., Schwember, A.R., Gadaleta, A., 2019. Carotenoid Pigment Content in Durum Wheat (Triticum turgidum L. var durum): An Overview of Quantitative Trait Loci and Candidate Genes. Front. Plant Sci. 10.

Combal, B., Baret, F., Weiss, M., Trubuil, A., Macé, D., Pragnère, A., Myneni, R., Knyazikhin, Y., Wang, L., 2003. Retrieval of canopy biophysical variables from bidirectional reflectance: Using prior information to solve the ill-posed inverse problem. Remote Sens. Environ. 84, 1–15. https://doi.org/10.1016/S0034-4257(02)00035-4

Congreves, K.A., Otchere, O., Ferland, D., Farzadfar, S., Williams, S., Arcand, M.M., 2021. Nitrogen Use Efficiency Definitions of Today and Tomorrow. Front. Plant Sci. 12.

Craney, T.A., Surles, J.G., 2002. Model-Dependent Variance Inflation Factor Cutoff Values. Qual. Eng. 14, 391–403. https://doi.org/10.1081/QEN-120001878

Cross, L., Gruère, A., 2022. International Fertilizer Association Public Summary Medium-Term Fertilizer Outlook 2022 – 2026. International Fertilizer Association, Paris, France.

Cubadda, R.E., Carcea, M., Marconi, E., Trivisonno, M.C., 2007. Influence of Gluten Proteins and Drying Temperature on the Cooking Quality of Durum Wheat Pasta. Cereal Chem. 84, 48–55. https://doi.org/10.1094/CCHEM-84-1-0048

Damm, A., Guanter, L., Laurent, V.C.E., Schaepman, M.E., Schickling, A., Rascher, U., 2014. FLD-based retrieval of sun-induced chlorophyll fluorescence from medium spectral resolution airborne spectroscopy data. Remote Sens. Environ. 147, 256–266. https://doi.org/10.1016/j.rse.2014.03.009

Daniel, C., Triboï, E., 2002. Changes in wheat protein aggregation during grain development: effects of temperatures and water stress. Eur. J. Agron. 16, 1–12. https://doi.org/10.1016/S1161-0301(01)00114-9

Dash, J., Curran, P.J., 2004. The MERIS terrestrial chlorophyll index. Int. J. Remote Sens. 25, 5403–5413. https://doi.org/10.1080/0143116042000274015

Datt, B., 1998. Remote Sensing of Chlorophyll a, Chlorophyll b, Chlorophyll a+b, and Total Carotenoid Content in Eucalyptus Leaves. Remote Sens. Environ. 66, 111–121. https://doi.org/10.1016/S0034-4257(98)00046-7

Daughtry, C.S.T., Gallo, K.P., Goward, S.N., Prince, S.D., Kustas, W.P., 1992. Spectral estimates of absorbed radiation and phytomass production in corn and soybean canopies. Remote Sens. Environ. 39, 141–152. https://doi.org/10.1016/0034-4257(92)90132-4

Daughtry, C.S.T., Walthall, C.L., Kim, M.S., de Colstoun, E.B., McMurtrey, J.E., 2000. Estimating Corn Leaf Chlorophyll Concentration from Leaf and Canopy Reflectance. Remote Sens. Environ. 74, 229–239. https://doi.org/10.1016/S0034-4257(00)00113-9

De'ath, G., Fabricius, K.E., Sweatman, H., Puotinen, M., 2012. The 27–year decline of coral cover on the Great Barrier Reef and its causes. Proc. Natl. Acad. Sci. 109, 17995–17999. https://doi.org/10.1073/pnas.1208909109

Delegido, J., Verrelst, J., Alonso, L., Moreno, J., 2011. Evaluation of Sentinel-2 Red-Edge Bands for Empirical Estimation of Green LAI and Chlorophyll Content. Sensors 11, 7063–7081. https://doi.org/10.3390/s110707063

Delegido, J., Verrelst, J., Meza, C.M., Rivera, J.P., Alonso, L., Moreno, J., 2013. A red-edge spectral index for remote sensing estimation of green LAI over agroecosystems. Eur. J. Agron. 46, 42–52. https://doi.org/10.1016/j.eja.2012.12.001

Delloye, C., Weiss, M., Defourny, P., 2018. Retrieval of the canopy chlorophyll content from Sentinel-2 spectral bands to estimate nitrogen uptake in intensive winter wheat cropping systems. Remote Sens. Environ. 216, 245–261. https://doi.org/10.1016/j.rse.2018.06.037

Demmig-Adams, B., Adams, W.W., 2006. Photoprotection in an ecological context: the remarkable complexity of thermal energy dissipation. New Phytol. 172, 11–21. https://doi.org/10.1111/j.1469-8137.2006.01835.x

Dempewolf, J., Adusei, B., Becker-Reshef, I., Hansen, M., Potapov, P., Khan, A., Barker, B., 2014. Wheat Yield Forecasting for Punjab Province from Vegetation Index Time Series and Historic Crop Statistics. Remote Sens. 6, 9653–9675. https://doi.org/10.3390/rs6109653

Dobermann, A., Ping, J.L., 2004. Geostatistical Integration of Yield Monitor Data and Remote Sensing Improves Yield Maps. Agron. J. 96, 285–297. https://doi.org/10.2134/agronj2004.2850

Dolferus, R., Ji, X., Richards, R.A., 2011. Abiotic stress and control of grain number in cereals. Plant Sci. 181, 331–341. https://doi.org/10.1016/j.plantsci.2011.05.015

Dorigo, W.A., Zurita-Milla, R., de Wit, A.J.W., Brazile, J., Singh, R., Schaepman, M.E., 2007. A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. Int. J. Appl. Earth Obs. Geoinformation, Advances in airborne electromagnetics and remote sensing of agro-ecosystems 9, 165–193. https://doi.org/10.1016/j.jag.2006.05.003

Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitão, P.J., Münkemüller, T., McClean, C., Osborne, P.E., Reineking, B., Schröder, B., Skidmore, A.K., Zurell, D., Lautenbach, S., 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 36, 27–46. https://doi.org/10.1111/j.1600-0587.2012.07348.x

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., Bargellini, P., 2012. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. Remote Sens. Environ., The Sentinel Missions - New Opportunities for Science 120, 25–36. https://doi.org/10.1016/j.rse.2011.11.026

Duncan, W., Loomis, R., Williams, W., Hanau, R., 1967. A model for simulating photosynthesis in plant communities. Hilgardia 38, 181–205.

Dupont, F.M., Hurkman, W.J., Vensel, W.H., Tanaka, C., Kothari, K.M., Chung, O.K., Altenbach, S.B., 2006. Protein accumulation and composition in wheat grains: Effects of mineral nutrients and high temperature. Eur. J. Agron., Modelling Quality Traits and Their Genetic Variability for Wheat 25, 96–107. https://doi.org/10.1016/j.eja.2006.04.003

Eitel, J.U.H., Long, D.S., Gessler, P.E., Hunt, E.R., 2008. Combined Spectral Index to Improve Ground-Based Estimates of Nitrogen Status in Dryland Wheat. Agron. J. 100, 1694–1702. https://doi.org/10.2134/agronj2007.0362

Eitel, J.U.H., Long, D.S., Gessler, P.E., Smith, A.M.S., 2007. Using in-situ measurements to evaluate the new RapidEye™ satellite series for prediction of wheat nitrogen status. Int. J. Remote Sens. 28, 4183–4190. https://doi.org/10.1080/01431160701422213

Ejaz, I., Pu, X., Naseer, M.A., Bohoussou, Y.N. 'Dri, Liu, Y., Farooq, M., Zhang, J., Zhang, Y., Wang, Z., Sun, Z., 2023. Cold and Drought Stresses in Wheat: A Global Meta-analysis of 21st Century. J. Plant Growth Regul. https://doi.org/10.1007/s00344-023-10960-x

Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. J. Anim. Ecol. 77, 802–813. https://doi.org/10.1111/j.1365-2656.2008.01390.x

Elsayed, S., Elhoweity, M., Ibrahim, H.H., Dewir, Y.H., Migdadi, H.M., Schmidhalter, U., 2017. Thermal imaging and passive reflectance sensing to estimate the water status and grain yield of wheat under different irrigation regimes. Agric. Water Manag. 189, 98–110. https://doi.org/10.1016/j.agwat.2017.05.001

Erenstein, O., Jaleta, M., Mottaleb, K.A., Sonder, K., Donovan, J., Braun, H.-J., 2022. Global Trends in Wheat Production, Consumption and Trade, in: Reynolds, M.P., Braun, H.-J. (Eds.), Wheat Improvement: Food Security in a Changing Climate. Springer International Publishing, Cham, pp. 47–66. https://doi.org/10.1007/978-3-030-90673-3_4

Evans, J.R., 1989. Photosynthesis and Nitrogen Relationships in Leaves of C_3 Plants. Oecologia 78, 9–19.

Evans, J.R., 1983. Nitrogen and Photosynthesis in the Flag Leaf of Wheat (Triticum aestivum L.). Plant Physiol. 72, 297–302. https://doi.org/10.1104/pp.72.2.297

Fang, H., Liang, S., 2014. Leaf Area Index Models☆, in: Reference Module in Earth Systems and Environmental Sciences. Elsevier. https://doi.org/10.1016/B978-0-12-409548-9.09076-X

FAO, 2022. FAOSTAT (production statistics), Crops and livestock products. Food and Agriculture Organization of the United Nations Statistics Division, Rome, Italy.

FAO, 2015. World reference base for soil resources 2014 (World Soil Resources Reports No. 106). FAO, Rome.

Farquhar, G.D., Sharkey, T.D., 1982. Stomatal Conductance and Photosynthesis. Annu. Rev. Plant Physiol. 33, 317–345. https://doi.org/10.1146/annurev.pp.33.060182.001533

Fassnacht, F.E., Stenzel, S., Gitelson, A.A., 2015. Non-destructive estimation of foliar carotenoid content of tree species using merged vegetation indices. J. Plant Physiol. 176, 210–217. https://doi.org/10.1016/j.jplph.2014.11.003

Feng, M., Xiao, L., Zhang, M., Yang, W., Ding, G., 2014. Integrating Remote Sensing and GIS for Prediction of Winter Wheat (Triticum aestivum) Protein Contents in Linfen (Shanxi), China. PLOS ONE 9, e80989. https://doi.org/10.1371/journal.pone.0080989

Feng, W., Qi, S., Heng, Y., Zhou, Y., Wu, Y., Liu, W., He, L., Li, X., 2017. Canopy Vegetation Indices from In situ Hyperspectral Data to Assess Plant Water Status of Winter Wheat under Powdery Mildew Stress. Front. Plant Sci. 0. https://doi.org/10.3389/fpls.2017.01219

Féret, J.-B., Berger, K., de Boissieu, F., Malenovský, Z., 2021. PROSPECT-PRO for estimating content of nitrogen-containing leaf proteins and other carbon-based constituents. Remote Sens. Environ. 252, 112173. https://doi.org/10.1016/j.rse.2020.112173

Féret, J.-B., François, C., Asner, G.P., Gitelson, A.A., Martin, R.E., Bidel, L.P.R., Ustin, S.L., le Maire, G., Jacquemoud, S., 2008. PROSPECT-4 and 5: Advances in the leaf optical properties model separating photosynthetic pigments. Remote Sens. Environ. 112, 3030–3043. https://doi.org/10.1016/j.rse.2008.02.012

Féret, J.-B., François, C., Gitelson, A., Asner, G.P., Barry, K.M., Panigada, C., Richardson, A.D., Jacquemoud, S., 2011. Optimizing spectral indices and chemometric analysis of leaf chemical properties using radiative transfer modeling. Remote Sens. Environ. 115, 2742–2750. https://doi.org/10.1016/j.rse.2011.06.016

Féret, J.-B., Gitelson, A.A., Noble, S.D., Jacquemoud, S., 2017a. PROSPECT-D: Towards modeling leaf optical properties through a complete lifecycle. Remote Sens. Environ. 193, 204–215. https://doi.org/10.1016/j.rse.2017.03.004

Féret, J.-B., Gitelson, A.A., Noble, S.D., Jacquemoud, S., 2017b. PROSPECT-D: Towards modeling leaf optical properties through a complete lifecycle. Remote Sens. Environ. 193, 204–215. https://doi.org/10.1016/j.rse.2017.03.004

Ficco, D.B.M., Mastrangelo, A.M., Trono, D., Borrelli, G.M., Vita, P.D., Fares, C., Beleggia, R., Platani, C., Papa, R., 2014. The colours of durum wheat: a review. Crop Pasture Sci. 65, 1–15. https://doi.org/10.1071/CP13293

Filella, I., Peñuelas, J., 1994. The red edge position and shape as indicators of plant chlorophyll content, biomass and hydric status. Int. J. Remote Sens. 15, 1459–1470. https://doi.org/10.1080/01431169408954177

Filella, I., Porcar-Castell, A., Munné-Bosch, S., Bäck, J., Garbulsky, M.F., Peñuelas, J., 2009. PRI assessment of long-term changes in carotenoids/chlorophyll ratio and short-term changes in de-epoxidation state of the xanthophyll cycle. Int. J. Remote Sens. 30, 4443–4455. https://doi.org/10.1080/01431160802575661

Fischer, R.A., 2011. Wheat physiology: a review of recent developments. Crop Pasture Sci. 62, 95–114. https://doi.org/10.1071/CP10344

Fischer, R.A., Howe, G.N., Ibrahim, Z., 1993. Irrigated spring wheat and timing and amount of nitrogen fertilizer. I. Grain yield and protein content. Field Crops Res. 33, 37–56. https://doi.org/10.1016/0378-4290(93)90093-3

Fitzgerald, G., Rodriguez, D., O'Leary, G., 2010. Measuring and predicting canopy nitrogen nutrition in wheat using a spectral index—The canopy chlorophyll content index (CCCI). Field Crops Res. 116, 318–324. https://doi.org/10.1016/j.fcr.2010.01.010

Fitzpatrick, E.A., Nix, H.A., 1969. A model for simulating soil water regime in alternating fallow-crop systems. Agric. Meteorol. 6, 303–319. https://doi.org/10.1016/0002-1571(69)90023-5

Flagella, Z., Giuliani, M.M., Giuzio, L., Volpi, C., Masci, S., 2010. Influence of water deficit on durum wheat storage protein composition and technological quality. Eur. J. Agron. 33, 197–207. https://doi.org/10.1016/j.eja.2010.05.006

Flexas, J., Escalona, J.M., Evain, S., Gulías, J., Moya, I., Osmond, C.B., Medrano, H., 2002. Steady-state chlorophyll fluorescence (Fs) measurements as a tool to follow variations of net CO2 assimilation and stomatal conductance during water-stress in C3 plants. Physiol. Plant. 114, 231–240. https://doi.org/10.1034/j.1399-3054.2002.1140209.x

Flowers, M., Weisz, R., Heiniger, R., 2003. Quantitative Approaches for Using Color Infrared Photography for Assessing In-Season Nitrogen Status in Winter Wheat. Agron. J. 95, 1189–1200. https://doi.org/10.2134/agronj2003.1189

Fowler, D., Steadman, C.E., Stevenson, D., Coyle, M., Rees, R.M., Skiba, U.M., Sutton, M.A., Cape, J.N., Dore, A.J., Vieno, M., Simpson, D., Zaehle, S., Stocker, B.D., Rinaldi, M., Facchini, M.C., Flechard, C.R., Nemitz, E., Twigg, M., Erisman, J.W., Butterbach-Bahl, K., Galloway, J.N., 2015. Effects of global change during the 21st century on the nitrogen cycle. Atmospheric Chem. Phys. 15, 13849–13893. https://doi.org/10.5194/acp-15-13849-2015

Fowler, D.B., 2003. Crop Nitrogen Demand and Grain Protein Concentration of Spring and Winter Wheat. Agron. J. 95, 260–265. https://doi.org/10.2134/agronj2003.2600

Frampton, W.J., Dash, J., Watmough, G., Milton, E.J., 2013. Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. ISPRS J. Photogramm. Remote Sens. 82, 83–92. https://doi.org/10.1016/j.isprsjprs.2013.04.007

Fratianni, A., Giuzio, L., Di Criscio, T., Zina, F., Panfili, G., 2013. Response of Carotenoids and Tocols of Durum Wheat in Relation to Water Stress and Sulfur Fertilization. J. Agric. Food Chem. 61, 2583–2590. https://doi.org/10.1021/jf304168r

Fréchette, E., Chang, C.Y.-Y., Ensminger, I., 2016. Photoperiod and temperature constraints on the relationship between the photochemical reflectance index and the light use efficiency of photosynthesis in Pinus strobus. Tree Physiol. 36, 311–324. https://doi.org/10.1093/treephys/tpv143

Freeman, K.W., Raun, W.R., Johnson, G.V., Mullen, R.W., Stone, M.L., Solie, J.B., 2003. Late-season Prediction Of Wheat Grain Yield And Grain Protein. Commun. Soil Sci. Plant Anal. 34, 1837–1852. https://doi.org/10.1081/CSS-120023219

French, R.J., Schultz, J.E., 1984. Water use efficiency of wheat in a Mediterranean-type environment. I. The relation between yield, water use and climate. Aust. J. Agric. Res. 35, 743–764. https://doi.org/10.1071/ar9840743

Freund, Y., Schapire, R.E., 1999. A Short Introduction to Boosting. J. Jpn. Soc. Artif. Intell. 14, 771–780.

Freund, Y., Schapire, R.E., 1997. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. J. Comput. Syst. Sci. 55, 119–139. https://doi.org/10.1006/jcss.1997.1504

Friedman, J., Tibshirani, R., Hastie, T., 2010. Regularization paths for generalized linear models via coordinate descent. J. Stat. Softw. 33, 1–22. https://doi.org/10.18637/jss.v033.i01

Friedman, J.H., 2002. Stochastic gradient boosting. Comput. Stat. Data Anal., Nonlinear Methods and Data Mining 38, 367–378. https://doi.org/10.1016/S0167-9473(01)00065-2

Friedman, J.H., 2001. Greedy Function Approximation: A Gradient Boosting Machine. Ann. Stat. 29, 1189–1232.

Fu, Z., Yu, S., Zhang, J., Xi, H., Gao, Y., Lu, R., Zheng, H., Zhu, Y., Cao, W., Liu, X., 2022. Combining UAV multispectral imagery and ecological factors to estimate leaf nitrogen and grain protein content of wheat. Eur. J. Agron. 132, 126405. https://doi.org/10.1016/j.eja.2021.126405

Galloway, J.N., Aber, J.D., Erisman, J.W., Seitzinger, S.P., Howarth, R.W., Cowling, E.B., Cosby, B.J., 2003. The Nitrogen Cascade. BioScience 53, 341–356. https://doi.org/10.1641/0006-3568(2003)053[0341:TNC]2.0.CO;2

Galloway, J.N., Leach, A.M., Erisman, J.W., Bleeker, A., 2017. Nitrogen: the historical progression from ignorance to knowledge, with a view to future solutions. Soil Res. 55, 417–424. https://doi.org/10.1071/SR16334

Gamon, J.A., Huemmrich, K.F., Wong, C.Y.S., Ensminger, I., Garrity, S., Hollinger, D.Y., Noormets, A., Peñuelas, J., 2016. A remotely sensed pigment index reveals photosynthetic phenology in evergreen conifers. Proc. Natl. Acad. Sci. 113, 13087–13092. https://doi.org/10.1073/pnas.1606162113

Gamon, J.A., Peñuelas, J., Field, C.B., 1992. A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. Remote Sens. Environ. 41, 35–44. https://doi.org/10.1016/0034-4257(92)90059-S

Gamon, J.A., Surfus, J.S., 1999. Assessing leaf pigment content and activity with a reflectometer. New Phytol. 143, 105–117. https://doi.org/10.1046/j.1469-8137.1999.00424.x

Gao, B., 1996. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sens. Environ. 58, 257–266. https://doi.org/10.1016/S0034-4257(96)00067-3

Gao, C., Zhang, T., 2010. Eutrophication in a Chinese Context: Understanding Various Physical and Socio-Economic Aspects. Ambio 39, 385–393. https://doi.org/10.1007/s13280-010-0040-5

García, G.A., Dreccer, M.F., Miralles, D.J., Serrago, R.A., 2015. High night temperatures during grain number determination reduce wheat and barley grain yield: a field study. Glob. Change Biol. 21, 4153–4164. https://doi.org/10.1111/gcb.13009

Garrido-Lestache, E., López-Bellido, R.J., López-Bellido, L., 2005. Durum wheat quality under Mediterranean conditions as affected by N rate, timing and splitting, N form and S fertilization. Eur. J. Agron. 23, 265–278. https://doi.org/10.1016/j.eja.2004.12.001

Garrity, S.R., Eitel, J.U.H., Vierling, L.A., 2011. Disentangling the relationships between plant pigments and the photochemical reflectance index reveals a new approach for remote estimation of carotenoid content. Remote Sens. Environ. 115, 628–635. https://doi.org/10.1016/j.rse.2010.10.007

Gebbers, R., Adamchuk, V.I., 2010. Precision Agriculture and Food Security. Science 327, 828–831. https://doi.org/10.1126/science.1183899

Genty, B., Briantais, J.-M., Baker, N.R., 1989. The relationship between the quantum yield of photosynthetic electron transport and quenching of chlorophyll fluorescence. Biochim. Biophys. Acta BBA - Gen. Subj. 990, 87–92. https://doi.org/10.1016/S0304-4165(89)80016-9

Giordano, N., Sadras, V.O., Lollato, R.P., 2023. Late-season nitrogen application increases grain protein concentration and is neutral for yield in wheat. A global meta-analysis. Field Crops Res. 290, 108740. https://doi.org/10.1016/j.fcr.2022.108740

Gitelson, A., Merzlyak, M., Zur, Y., Stark, R., Gritz, U., 2001. Non-destructive and remote sensing techniques for estimation of vegetation status, in: Proceedings of the 3rd European Conference on Precision Agriculture. Presented at the Third European Conference on Precision Agriculture, Montpelier, France.

Gitelson, A., Merzlyak, M.N., 1994a. Quantitative estimation of chlorophyll-a using reflectance spectra: Experiments with autumn chestnut and maple leaves. J. Photochem. Photobiol. B 22, 247–252. https://doi.org/10.1016/1011-1344(93)06963-4

Gitelson, A., Merzlyak, M.N., 1994b. Spectral Reflectance Changes Associated with Autumn Senescence of Aesculus hippocastanum L. and Acer platanoides L. Leaves. Spectral Features and Relation to Chlorophyll Estimation. J. Plant Physiol. 143, 286–292. https://doi.org/10.1016/S0176-1617(11)81633-0

Gitelson, A.A., Chivkunova, O.B., Merzlyak, M.N., 2009. Nondestructive estimation of anthocyanins and chlorophylls in anthocyanic leaves. Am. J. Bot. 96, 1861–1868. https://doi.org/10.3732/ajb.0800395

Gitelson, A.A., Gritz †, Y., Merzlyak, M.N., 2003. Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. J. Plant Physiol. 160, 271–282. https://doi.org/10.1078/0176-1617-00887

Gitelson, A.A., Kaufman, Y.J., Merzlyak, M.N., 1996. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. Remote Sens. Environ. 58, 289–298. https://doi.org/10.1016/S0034-4257(96)00072-7

Gitelson, A.A., Kaufman, Y.J., Stark, R., Rundquist, D., 2002a. Novel algorithms for remote estimation of vegetation fraction. Remote Sens. Environ. 80, 76–87. https://doi.org/10.1016/S0034-4257(01)00289-9

Gitelson, A.A., Keydan, G.P., Merzlyak, M.N., 2006. Three-band model for noninvasive estimation of chlorophyll, carotenoids, and anthocyanin contents in higher plant leaves. Geophys. Res. Lett. 33. https://doi.org/10.1029/2006GL026457

Gitelson, A.A., Merzlyak, M.N., 1998. Remote sensing of chlorophyll concentration in higher plant leaves. Adv. Space Res., Synergistic Use of Multisensor Data for Land Processes 22, 689–692. https://doi.org/10.1016/S0273-1177(97)01133-2

Gitelson, A.A., Merzlyak, M.N., 1996. Signature Analysis of Leaf Reflectance Spectra: Algorithm Development for Remote Sensing of Chlorophyll. J. Plant Physiol. 148, 494–500. https://doi.org/10.1016/S0176-1617(96)80284-7

Gitelson, A.A., Merzlyak, M.N., Chivkunova, O.B., 2001. Optical Properties and Nondestructive Estimation of Anthocyanin Content in Plant Leaves. Photochem. Photobiol. 74, 38–45. https://doi.org/10.1562/0031-8655(2001)0740038OPANEO2.0.CO2

Gitelson, A.A., Zur, Y., Chivkunova, O.B., Merzlyak, M.N., 2002b. Assessing Carotenoid Content in Plant Leaves with Reflectance Spectroscopy. Photochem. Photobiol. 75, 272–281. https://doi.org/10.1562/0031-8655(2002)0750272ACCIPL2.0.CO2

Giuliani, M.M., Giuzio, L., Caro, A.D., Flagella, Z., 2011. Relationships between Nitrogen Utilization and Grain Technological Quality in Durum Wheat: I. Nitrogen Translocation and Nitrogen Use Efficiency for Protein. Agron. J. 103, 1487–1494. https://doi.org/10.2134/agronj2011.0153

Gómez, D., Salvador, P., Sanz, J., Casanova, J.L., 2021. Modelling wheat yield with antecedent information, satellite and climate data using machine learning methods in Mexico. Agric. For. Meteorol. 300, 108317. https://doi.org/10.1016/j.agrformet.2020.108317

Gómez-Dans, J.L., Lewis, P.E., Disney, M., 2016. Efficient Emulation of Radiative Transfer Codes Using Gaussian Processes and Application to Land Surface Parameter Inferences. Remote Sens. 8, 119. https://doi.org/10.3390/rs8020119

Gonzalez-Dugo, V., Hernandez, P., Solis, I., Zarco-Tejada, P.J., 2015. Using High-Resolution Hyperspectral and Thermal Airborne Imagery to Assess Physiological Condition in the Context of Wheat Phenotyping. Remote Sens. 7, 13586–13605. http://dx.doi.org/10.3390/rs71013586

Gooding, M.J., Davies, W.P., 1992. Foliar urea fertilization of cereals: A review. Fertil. Res. 32, 209–222. https://doi.org/10.1007/BF01048783

Gooding, M.J., Ellis, R.H., Shewry, P.R., Schofield, J.D., 2003. Effects of Restricted Water Availability and Increased Temperature on the Grain Filling, Drying and Quality of Winter Wheat. J. Cereal Sci. 37, 295–309. https://doi.org/10.1006/jcrs.2002.0501

Gooding, M.J., Gregory, P.J., Ford, K.E., Ruske, R.E., 2007. Recovery of nitrogen from different sources following applications to winter wheat at and after anthesis. Field Crops Res. 100, 143–154. https://doi.org/10.1016/j.fcr.2006.06.002

Gould, K.S., 2004. Nature's Swiss Army Knife: The Diverse Protective Roles of Anthocyanins in Leaves. J. Biomed. Biotechnol. 2004, 314–320. https://doi.org/10.1155/S1110724304406147

Grace, J., Nichol, C., Disney, M., Lewis, P., Quaife, T., Bowyer, P., 2007. Can we measure terrestrial photosynthesis from space directly, using spectral reflectance and fluorescence? Glob. Change Biol. 13, 1484–1497. https://doi.org/10.1111/j.1365-2486.2007.01352.x

GrainCorp Ltd., 2021. Wheat-Standards-2021-2022.pdf [WWW Document]. URL https://grains.graincorp.com.au/wp-content/uploads/2021/08/Wheat-Standards-2021-2022.pdf (accessed 7.25.23).

Grant, O.M., Tronina, Ł., Jones, H.G., Chaves, M.M., 2007. Exploring thermal imaging variables for the detection of stress responses in grapevine under different irrigation regimes. J. Exp. Bot. 58, 815–825. https://doi.org/10.1093/jxb/erl153

Grinberg, N.F., Orhobor, O.I., King, R.D., 2020. An evaluation of machine-learning for predicting phenotype: studies in yeast, rice, and wheat. Mach. Learn. 109, 251–277. https://doi.org/10.1007/s10994-019-05848-5

Groth, S., Wittmann, R., Longin, C.F.H., Böhm, V., 2020. Influence of variety and growing location on carotenoid and vitamin E contents of 184 different durum wheat varieties (Triticum turgidum ssp. durum) in Germany. Eur. Food Res. Technol. 246, 2079–2092. https://doi.org/10.1007/s00217-020-03557-1

Gu, B., Zhang, X., Lam, S.K., Yu, Y., van Grinsven, H.J.M., Zhang, S., Wang, X., Bodirsky, B.L., Wang, S., Duan, J., Ren, C., Bouwman, L., de Vries, W., Xu, J., Sutton, M.A., Chen, D., 2023. Cost-effective mitigation of nitrogen pollution from global croplands. Nature 613, 77–84. https://doi.org/10.1038/s41586-022-05481-8

Guan, K., Berry, J.A., Zhang, Y., Joiner, J., Guanter, L., Badgley, G., Lobell, D.B., 2016. Improving the monitoring of crop productivity using spaceborne solar-induced fluorescence. Glob. Change Biol. 22, 716–726. https://doi.org/10.1111/gcb.13136

Guanter, L., Zhang, Y., Jung, M., Joiner, J., Voigt, M., Berry, J.A., Frankenberg, C., Huete, A.R., Zarco-Tejada, P., Lee, J.-E., Moran, M.S., Ponce-Campos, G., Beer, C., Camps-Valls, G., Buchmann, N., Gianelle, D., Klumpp, K., Cescatti, A., Baker, J.M., Griffis, T.J., 2014. Global and time-resolved monitoring of crop photosynthesis with chlorophyll fluorescence. Proc. Natl. Acad. Sci. U. S. A. 111, E1327–E1333. https://doi.org/10.1073/pnas.1320008111

Gueymard, C., 1995. Simple Model for the Atmospheric Radiative Transfer of Sunshine (SMARTS2) Algorithms and performance assessment. Florida Solar Energy Center, University of Central Florida.

Guttieri, M.J., Ahmad, R., Stark, J.C., Souza, E., 2000. End-Use Quality of Six Hard Red Spring Wheat Cultivars at Different Irrigation Levels. Crop Sci. 40, 631–635. https://doi.org/10.2135/cropsci2000.403631x

Haboudane, D., Miller, J.R., Pattey, E., Zarco-Tejada, P.J., Strachan, I.B., 2004. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. Remote Sens. Environ. 90, 337–352. https://doi.org/10.1016/j.rse.2003.12.013

Haboudane, D., Miller, J.R., Tremblay, N., Zarco-Tejada, P.J., Dextraze, L., 2002. Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. Remote Sens. Environ. 81, 416–426. https://doi.org/10.1016/S0034-4257(02)00018-4

Haboudane, D., Tremblay, N., Miller, J.R., Vigneault, P., 2008. Remote Estimation of Crop Chlorophyll Content Using Spectral Indices Derived From Hyperspectral Data. IEEE Trans. Geosci. Remote Sens. 46, 423–437. https://doi.org/10.1109/TGRS.2007.904836

Halse, N.J., Weir, R.N., 1974. Effects of temperature on spikelet number of wheat. Aust. J. Agric. Res. 25, 687–695. https://doi.org/10.1071/ar9740687

Hamblin, J., Stefanova, K., Angessa, T.T., 2014. Variation in Chlorophyll Content per Unit Leaf Area in Spring Wheat and Implications for Selection in Segregating Material. PLOS ONE 9, e92529. https://doi.org/10.1371/journal.pone.0092529

Hansen, P.M., Jørgensen, J.R., Thomsen, A., 2002. Predicting grain yield and protein content in winter wheat and spring barley using repeated canopy reflectance measurements and partial least squares regression. J. Agric. Sci. 139, 307–318. https://doi.org/10.1017/S0021859602002320

Harries, M., Flower, K.C., Scanlan, C.A., Harries, M., Flower, K.C., Scanlan, C.A., 2021. Sustainability of nutrient management in grain production systems of south-west Australia. Crop Pasture Sci. 72, 197–212. https://doi.org/10.1071/CP20403

Hartigan, J.A., Wong, M.A., 1979. Algorithm AS 136: A K-Means Clustering Algorithm. Appl. Stat. 28, 100. https://doi.org/10.2307/2346830

Hatfield, J.L., Gitelson, A.A., Schepers, J.S., Walthall, C.L., 2008. Application of Spectral Remote Sensing for Agronomic Decisions. Agron. J. Madison 100, S117–S131.

Hatfield, J.L., Walthall, C.L., 2015. Meeting Global Food Needs: Realizing the Potential via Genetics × Environment × Management Interactions. Agron. J. 107, 1215–1226. https://doi.org/10.2134/agronj15.0076

Havaux, M., 2014. Carotenoid oxidation products as stress signals in plants. Plant J. 79, 597–606. https://doi.org/10.1111/tpj.12386

Hays, D.B., Do, J.H., Mason, R.E., Morgan, G., Finlayson, S.A., 2007. Heat stress induced ethylene production in developing wheat grains induces kernel abortion and increased maturation in a susceptible cultivar. Plant Sci. 172, 1113–1123. https://doi.org/10.1016/j.plantsci.2007.03.004

Heffer, P., Prud'homme, M., 2020. Global nitrogen fertiliser demand and supply: trend, current level and outlook. Presented at the 7th International Nitrogen Initiative Conference (INI2016), Melbourne, Australia.

Henzell, T., 2007. Australian Agriculture: Its History and Challenges. CSIRO Publishing.

Hernández-Clemente, R., Navarro-Cerrillo, R.M., Suárez, L., Morales, F., Zarco-Tejada, P.J., 2011. Assessing structural effects on PRI for stress detection in conifer forests. Remote Sens. Environ. 115, 2360–2375. https://doi.org/10.1016/j.rse.2011.04.036

Hernández-Clemente, R., Navarro-Cerrillo, R.M., Zarco-Tejada, P.J., 2012. Carotenoid content estimation in a heterogeneous conifer forest using narrow-band indices and PROSPECT+DART simulations. Remote Sens. Environ. 127, 298–315. https://doi.org/10.1016/j.rse.2012.09.014

Herrmann, I., Karnieli, A., Bonfil, D.J., Cohen, Y., Alchanatis, V., 2010. SWIR-based spectral indices for assessing nitrogen content in potato fields. Int. J. Remote Sens. 31, 5127–5143. https://doi.org/10.1080/01431160903283892

Herrmann, I., Pimstein, A., Karnieli, A., Cohen, Y., Alchanatis, V., Bonfil, D.J., 2011. LAI assessment of wheat and potato crops by VENμS and Sentinel-2 bands. Remote Sens. Environ. 115, 2141–2151. https://doi.org/10.1016/j.rse.2011.04.018

Heskel, M.A., O'Sullivan, O.S., Reich, P.B., Tjoelker, M.G., Weerasinghe, L.K., Penillard, A., Egerton, J.J.G., Creek, D., Bloomfield, K.J., Xiang, J., Sinca, F., Stangl, Z.R., Martinez-de la Torre, A., Griffin, K.L., Huntingford, C., Hurry, V., Meir, P., Turnbull, M.H., Atkin, O.K., 2016. Convergence in the temperature response of leaf respiration across biomes and plant functional types. Proc. Natl. Acad. Sci. 113, 3832–3837. https://doi.org/10.1073/pnas.1520282113

Hikosaka, K., Noda, H.M., 2019. Modeling leaf CO2 assimilation and Photosystem II photochemistry from chlorophyll fluorescence and the photochemical reflectance index. Plant Cell Environ. 42, 730–739. https://doi.org/10.1111/pce.13461

Hochman, Z., Rees, H. van, Carberry, P.S., Hunt, J.R., McCown, R.L., Gartmann, A., Holzworth, D., Rees, S. van, Dalgliesh, N.P., Long, W., Peake, A.S., Poulton, P.L., McClelland, T., 2009. Re-inventing model-based decision support with Australian dryland farmers. 4. Yield Prophet® helps farmers monitor and manage crops in a variable climate. Crop Pasture Sci. 60, 1057–1070. https://doi.org/10.1071/CP09020

Hoffmann, M.P., Llewellyn, R.S., Davoren, C.W., Whitbread, A.M., 2016. Assessing the Potential for Zone-Specific Management of Cereals in Low-Rainfall South-Eastern Australia: Combining On-Farm Results and Simulation Analysis. J. Agron. Crop Sci. n/a-n/a. https://doi.org/10.1111/jac.12159

Högy, P., Fangmeier, A., 2008. Effects of elevated atmospheric CO2 on grain quality of wheat. J. Cereal Sci. 48, 580–591. https://doi.org/10.1016/j.jcs.2008.01.006

Holford, I.C.R., Doyle, A.D., Leckie, C.C., 1992. Nitrogen response characteristics of wheat protein in relation to yield responses and their interactions with phosphorus. Aust. J. Agric. Res. 43, 969–986. https://doi.org/10.1071/ar9920969

Holzworth, D., Huth, N.I., Fainges, J., Brown, H., Zurcher, E., Cichota, R., Verrall, S., Herrmann, N.I., Zheng, B., Snow, V., 2018. APSIM Next Generation: Overcoming challenges in modernising a farming systems model. Environ. Model. Softw. 103, 43–51. https://doi.org/10.1016/j.envsoft.2018.02.002

Horler, D.N.H., Dockray, M., Barber, J., 1983. The red edge of plant leaf reflectance. Int. J. Remote Sens. 4, 273–288. https://doi.org/10.1080/01431168308948546

Houborg, R., Boegh, E., 2008. Mapping leaf chlorophyll and leaf area index using inverse and forward canopy reflectance modeling and SPOT reflectance data. Remote Sens. Environ. 112, 186–202. https://doi.org/10.1016/j.rse.2007.04.012

Houborg, R., Cescatti, A., Migliavacca, M., Kustas, W.P., 2013. Satellite retrievals of leaf chlorophyll and photosynthetic capacity for improved modeling of GPP. Agric. For. Meteorol. 177, 10–23. https://doi.org/10.1016/j.agrformet.2013.04.006

Houlès, V., Guérif, M., Mary, B., 2007. Elaboration of a nitrogen nutrition indicator for winter wheat based on leaf area index and chlorophyll content for making nitrogen recommendations. Eur. J. Agron. 27, 1–11. https://doi.org/10.1016/j.eja.2006.10.001

https://www.longpaddock.qld.gov.au/, n.d.

Huete, A.R., 1988. A soil-adjusted vegetation index (SAVI). Remote Sens. Environ. 25, 295–309. https://doi.org/10.1016/0034-4257(88)90106-X

Hunt, J., van Rees, H., Hochman, Z., Carberry, P.S., Holzworth, D., Dalgliesh, N., Brennan, L., Poulton, P., van Rees, S., Huth, N.I., 2006. Yield Prophet®: An online crop simulation service, in: Proceedings of the 13th Australian Agronomy Conference. pp. 10–14.

Hunt, M.L., Blackburn, G.A., Carrasco, L., Redhead, J.W., Rowland, C.S., 2019. High resolution wheat yield mapping using Sentinel-2. Remote Sens. Environ. 233, 111410. https://doi.org/10.1016/j.rse.2019.111410

Idso, S.B., 1982. Non-water-stressed baselines: A key to measuring and interpreting plant water stress. Agric. Meteorol. 27, 59–70. https://doi.org/10.1016/0002-1571(82)90020-6

Idso, S.B., Jackson, R.D., Pinter, P.J., Reginato, R.J., Hatfield, J.L., 1981. Normalizing the stress-degree-day parameter for environmental variability. Agric. Meteorol. 24, 45–55. https://doi.org/10.1016/0002-1571(81)90032-7

Isbell, R., 2002. The Australian soil classification. CSIRO publishing, Melbourne, Australia.

Jackson, R.D., Idso, S.B., Reginato, R.J., Pinter, P.J., 1981. Canopy temperature as a crop water stress indicator. Water Resour. Res. 17, 1133–1138. https://doi.org/10.1029/WR017i004p01133

Jacquemoud, S., Baret, F., 1990. PROSPECT: A model of leaf optical properties spectra. Remote Sens. Environ. 34, 75–91. https://doi.org/10.1016/0034-4257(90)90100-Z

Jacquemoud, S., Baret, F., Andrieu, B., Danson, F.M., Jaggard, K., 1995. Extraction of vegetation biophysical parameters by inversion of the PROSPECT + SAIL models on sugar beet canopy reflectance data. Application to TM and AVIRIS sensors. Remote Sens. Environ. 52, 163–172. https://doi.org/10.1016/0034-4257(95)00018-V

Jacquemoud, S., Verhoef, W., Baret, F., Bacour, C., Zarco-Tejada, P.J., Asner, G.P., François, C., Ustin, S.L., 2009. PROSPECT+SAIL models: A review of use for vegetation characterization. Remote Sens. Environ., Imaging Spectroscopy Special Issue 113, S56–S66. https://doi.org/10.1016/j.rse.2008.01.026

Jamieson, P.D., Semenov, M.A., 2000. Modelling nitrogen uptake and redistribution in wheat. Field Crops Res. 68, 21–29. https://doi.org/10.1016/S0378-4290(00)00103-9

Janeczko, A., Dziurka, M., Pociecha, E., 2018. Increased leaf tocopherol and β-carotene content is associated with the tolerance of winter wheat cultivars to frost. J. Agron. Crop Sci. 204, 594–602. https://doi.org/10.1111/jac.12287

Jensen, T., Apan, A., Young, F., Zeller, L., 2007. Detecting the attributes of a wheat crop using digital imagery acquired from a low-altitude platform. Comput. Electron. Agric. 59, 66–77. https://doi.org/10.1016/j.compag.2007.05.004

Jia, M., Colombo, R., Rossini, M., Celesti, M., Zhu, J., Cogliati, S., Cheng, T., Tian, Y., Zhu, Y., Cao, W., Yao, X., 2021. Estimation of leaf nitrogen content and photosynthetic nitrogen use efficiency in wheat using sun-induced chlorophyll fluorescence at the leaf and canopy scales. Eur. J. Agron. 122, 126192. https://doi.org/10.1016/j.eja.2020.126192

Jiao, Q., Sun, Q., Zhang, B., Huang, W., Ye, H., Zhang, Z., Zhang, X., Qian, B., 2022. A Random Forest Algorithm for Retrieving Canopy Chlorophyll Content of Wheat and Soybean Trained with PROSAIL Simulations Using Adjusted Average Leaf Angle. Remote Sens. 14, 98. https://doi.org/10.3390/rs14010098

Jin, X., Xu, X., Feng, H., Song, X., Wang, Q., Wang, J., 2014. Estimation of Grain Protein Content in Winter Wheat by Using Three Methods with Hyperspectral Data. Int J Agric Biol 16, 7.

Johnston, A.M., Bruulsema, T.W., 2014. 4R Nutrient Stewardship for Improved Nutrient Use Efficiency. Procedia Eng., SYMPHOS 2013 - 2nd International Symposium on Innovation and Technology in the Phosphate Industry 83, 365–370. https://doi.org/10.1016/j.proeng.2014.09.029

Ju, X., Zhang, C., 2017. Nitrogen cycling and environmental impacts in upland agricultural soils in North China: A review. J. Integr. Agric. 16, 2848–2862. https://doi.org/10.1016/S2095-3119(17)61743-X

Kamenova, I., Dimitrov, P., 2021. Evaluation of Sentinel-2 vegetation indices for prediction of LAI, fAPAR and fCover of winter wheat in Bulgaria. Eur. J. Remote Sens. 54, 89–108. https://doi.org/10.1080/22797254.2020.1839359

Kanning, M., Kuehling, I., Trautz, D., Jarmer, T., 2018. High-Resolution UAV-Based Hyperspectral Imagery for LAI and Chlorophyll Estimations from Wheat for Yield Prediction. Remote Sens. 10, 2000. https://doi.org/10.3390/rs10122000

Kant, S., Bi, Y.-M., Rothstein, S.J., 2011. Understanding plant response to nitrogen limitation for the improvement of crop nitrogen use efficiency. J. Exp. Bot. 62, 1499–1509. https://doi.org/10.1093/jxb/erq297

Karatzoglou, A., Smola, A., Hornik, K., Zeileis, A., 2004. kernlab – an S4 package for kernel methods in R. J. Stat. Softw. 11, 1–20. https://doi.org/10.18637/jss.v011.i09

Karnieli, A., Kaufman, Y.J., Remer, L., Wald, A., 2001. AFRI — aerosol free vegetation index. Remote Sens. Environ. 77, 10–21. https://doi.org/10.1016/S0034-4257(01)00190-0

Kawasaki, K., Uchida, S., 2016. Quality Matters More Than Quantity: Asymmetric Temperature Effects on Crop Yield and Quality Grade. Am. J. Agric. Econ. 98, 1195–1209. https://doi.org/10.1093/ajae/aaw036

Khanal, S., Fulton, J., Shearer, S., 2017. An overview of current and potential applications of thermal remote sensing in precision agriculture. Comput. Electron. Agric. 139, 22–32. https://doi.org/10.1016/j.compag.2017.05.001

Kibite, S., Evans, L.E., 1984. Causes of negative correlations between grain yield and grain protein concentration in common wheat. Euphytica 33, 801–810. https://doi.org/10.1007/BF00021906

Kirkegaard, J.A., Lilley, J.M., 2007. Root penetration rate – a benchmark to identify soil and plant limitations to rooting depth in wheat. Aust. J. Exp. Agric. 47, 590. https://doi.org/10.1071/EA06071

Klem, K., Záhora, J., Zemek, F., Trunda, P., Tůma, I., Novotná, K., Hodaňová, P., Rapantová, B., Hanuš, J., Vavříková, J., Holub, P., 2018. Interactive effects of water deficit and nitrogen nutrition on winter wheat. Remote sensing methods for their detection. Agric. Water Manag. 210, 171–184. https://doi.org/10.1016/j.agwat.2018.08.004

Koch, A., Chappell, A., Eyres, M., Scott, E., 2015. Monitor Soil Degradation or Triage for Soil Security? An Australian Challenge. Sustainability 7, 4870–4892. https://doi.org/10.3390/su7054870

Kohzuma, K., Tamaki, M., Hikosaka, K., 2021. Corrected photochemical reflectance index (PRI) is an effective tool for detecting environmental stresses in agricultural crops under light conditions. J. Plant Res. 134, 683–694. https://doi.org/10.1007/s10265-021-01316-1

Kuhn, M., 2020. caret: Classification and Regression Training. R package.

Kuhn, M., Weston, S., Keefer, C., Coulter, N., Quinlan, R., 2023. Cubist: Rule- And Instance-Based Regression Modeling.

Kuhn, M., Wickham, H., 2020. Tidymodels: a collection of packages for modeling and machine learning using tidyverse principles.

Kuusk, A., 1991. The Hot Spot Effect in Plant Canopy Reflectance, in: Myneni, R.B., Ross, J. (Eds.), Photon-Vegetation Interactions: Applications in Optical Remote Sensing and Plant Ecology. Springer, Berlin, Heidelberg, pp. 139–159. https://doi.org/10.1007/978-3-642-75389-3_5

Labuschagne, M.T., Elago, O., Koen, E., 2009. The influence of temperature extremes on some quality and starch characteristics in bread, biscuit and durum wheat. J. Cereal Sci. 49, 184–189. https://doi.org/10.1016/j.jcs.2008.09.001

Ladha, J.K., Tirol-padre, A., Reddy, C.K., Cassman, K.G., Verma, S., Powlson, D.S., Van Kessel, C., De B Richter, D., Chakraborty, D., Pathak, H., 2016. Global nitrogen budgets in cereals: A 50-year assessment for maize, rice, and wheat production systems. Sci. Rep. Nat. Publ. Group Lond. 6, 19355. http://dx.doi.org.ezp.lib.unimelb.edu.au/10.1038/srep19355

Lai, Y.R., Pringle, M.J., Kopittke, P.M., Menzies, N.W., Orton, T.G., Dang, Y.P., 2018. An empirical model for prediction of wheat yield, using time-integrated Landsat NDVI. Int. J. Appl. Earth Obs. Geoinformation 72, 99–108. https://doi.org/10.1016/j.jag.2018.07.013

Lemaire, G., Jeuffroy, M.-H., Gastal, F., 2008. Diagnosis tool for plant and crop N status in vegetative stage: Theory and practices for crop N management. Eur. J. Agron. 28, 614–624. https://doi.org/10.1016/j.eja.2008.01.005

Li, F., Miao, Y., Hennig, S.D., Gnyp, M.L., Chen, X., Jia, L., Bareth, G., 2010. Evaluating hyperspectral vegetation indices for estimating nitrogen concentration of winter wheat at different growth stages. Precis. Agric. 11, 335–357. https://doi.org/10.1007/s11119-010-9165-6

Li, F., Mistele, B., Hu, Y., Chen, X., Schmidhalter, U., 2014. Reflectance estimation of canopy nitrogen content in winter wheat using optimised hyperspectral spectral indices and partial least squares regression. Eur. J. Agron. 52, 198–209. https://doi.org/10.1016/j.eja.2013.09.006

Li, G., Chen, T., Feng, B., Peng, S., Tao, L., Fu, G., 2021. Respiration, Rather Than Photosynthesis, Determines Rice Yield Loss Under Moderate High-Temperature Conditions. Front. Plant Sci. 12, 678653. https://doi.org/10.3389/fpls.2021.678653

Li, X., Lv, X., Wang, X., Wang, L., Zhang, M., Ren, M., 2018. Effects of abiotic stress on anthocyanin accumulation and grain weight in purple wheat. Crop Pasture Sci. 69, 1208–1214. https://doi.org/10.1071/CP18341

Li, X., Pu, H., Liu, F., Zhou, Q., Cai, J., Dai, T., Cao, W., Jiang, D., 2015. Winter Wheat Photosynthesis and Grain Yield Responses to Spring Freeze. Agron. J. 107, 1002–1010. https://doi.org/10.2134/agronj14.0460

Li, Z., Jin, X., Wang, J., Yang, G., Nie, C., Xu, X., Feng, H., 2015. Estimating winter wheat (Triticum aestivum) LAI and leaf chlorophyll content from canopy reflectance data by integrating agronomic prior knowledge with the PROSAIL model. Int. J. Remote Sens. 36, 2634–2653. https://doi.org/10.1080/01431161.2015.1041176

Li, Z., Jin, X., Yang, G., Drummond, J., Yang, H., Clark, B., Li, Zhenhong, Zhao, C., 2018. Remote Sensing of Leaf and Canopy Nitrogen Status in Winter Wheat (Triticum aestivum L.) Based on N-PROSAIL Model. Remote Sens. 10, 1463. https://doi.org/10.3390/rs10091463

Li, Zhenhai, Li, Zhenhong, Fairbairn, D., Li, N., Xu, B., Feng, H., Yang, G., 2019. Multi-LUTs method for canopy nitrogen density estimation in winter wheat by field and UAV hyperspectral. Comput. Electron. Agric. 162, 174–182. https://doi.org/10.1016/j.compag.2019.04.005

Li, Zhenhai, Taylor, J., Yang, H., Casa, R., Jin, X., Li, Zhenhong, Song, X., Yang, G., 2020. A hierarchical interannual wheat yield and grain protein prediction model using spectral vegetative indices and meteorological data. Field Crops Res. 248, 107711. https://doi.org/10.1016/j.fcr.2019.107711

Liang, L., Di, L., Huang, T., Wang, J., Lin, L., Wang, L., Yang, M., 2018. Estimation of Leaf Nitrogen Content in Wheat Using New Hyperspectral Indices and a Random Forest Regression Algorithm. Remote Sens. 10, 1940. https://doi.org/10.3390/rs10121940

Liang, L., Qin, Z., Zhao, S., Di, L., Zhang, C., Deng, M., Lin, H., Zhang, L., Wang, L., Liu, Z., 2016. Estimating crop chlorophyll content with hyperspectral vegetation indices and the hybrid inversion method. Int. J. Remote Sens. 37, 2923–2949. https://doi.org/10.1080/01431161.2016.1186850

Liang, X., Lam, S.K., Zhang, X., Oenema, O., Chen, D., 2021. Pursuing sustainable nitrogen management following the "5 Ps" principles: Production, People, Planet, Policy and Partnerships. Glob. Environ. Change 70, 102346. https://doi.org/10.1016/j.gloenvcha.2021.102346

Liang, X., Suter, H., Lam, S.K., Walker, C., Khalil, R., Chen, D., 2022. Sustainable nitrogen management in australian agroecosystems: challenges and opportunities. Front. Agric. Sci. Eng. 9, 366–372. https://doi.org/10.15302/J-FASE-2022447

Liao, C., Wang, J., Shan, B., Song, Y., He, Y., Dong, T., 2022. Near real-time yield forecasting of winter wheat using Sentinel-2 imagery at the early stages. Precis. Agric. https://doi.org/10.1007/s11119-022-09975-3

Lichtenthaler, H.K., 1996. Vegetation Stress: an Introduction to the Stress Concept in Plants. J. Plant Physiol. 148, 4–14. https://doi.org/10.1016/S0176-1617(96)80287-2

Liu, H.Q., Huete, A., 1995. A feedback based modification of the NDVI to minimize canopy background and atmospheric noise. IEEE Trans. Geosci. Remote Sens. 33, 457–465. https://doi.org/10.1109/TGRS.1995.8746027

Liu, L., Wang, J., Bao, Y., Huang, W., Ma, Z., Zhao, C., 2006. Predicting winter wheat condition, grain yield and protein content using multi-temporal EnviSat-ASAR and Landsat TM satellite images. Int. J. Remote Sens. 27, 737–753. https://doi.org/10.1080/01431160500296867

Llewellyn, R., Ouzman, J., 2014. Adoption of precision agriculture-related practices: status, opportunities and the role of farm advisers. CSIRO.

Llewellyn, R., Whitbread, A., Jones, B., Davoren, B., 2008. The role for EM mapping in precision agriculture in the Mallee. Glob. Issues Paddock ActionAustralian Soc. Agron. Adel.

Lobell, D.B., Gourdji, S.M., 2012. The Influence of Climate Change on Global Crop Productivity. Plant Physiol. 160, 1686–1697. https://doi.org/10.1104/pp.112.208298

Lobry de Bruyn, L., Andrews, S., 2016. Are Australian and United States Farmers Using Soil Information for Soil Health Management? Sustainability 8, 304. https://doi.org/10.3390/su8040304

Locherer, M., Hank, T., Danner, M., Mauser, W., 2015. Systematic analysis of the LUT-based inversion of PROSAIL using full range hyperspectral data for the retrieval of leaf area index in view of the future EnMAP mission, in: 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). Presented at the 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 4013–4016. https://doi.org/10.1109/IGARSS.2015.7326705

Lollato, R.P., Jaenisch, B.R., Silva, S.R., 2021. Genotype-specific nitrogen uptake dynamics and fertilizer management explain contrasting wheat protein concentration. Crop Sci. 61, 2048–2066. https://doi.org/10.1002/csc2.20442

Long, D.S., Engel, R.E., Carpenter, F.M., 2005. On-Combine Sensing and Mapping of Wheat Protein Concentration. Crop Manag. 4, 1–9. https://doi.org/10.1094/CM-2005-0527-01-RS

Longmire, A.R., Poblete, T., Hunt, J.R., Chen, D., Zarco-Tejada, P.J., 2022. Assessment of crop traits retrieved from airborne hyperspectral and thermal remote sensing imagery to predict wheat grain protein content. ISPRS J. Photogramm. Remote Sens. 193, 284–298. https://doi.org/10.1016/j.isprsjprs.2022.09.015

Lopez-Bellido, R.J., Shepherd, C.E., Barraclough, P.B., 2004. Predicting post-anthesis N requirements of bread wheat with a Minolta SPAD meter. Eur. J. Agron. 20, 313–320. https://doi.org/10.1016/S1161-0301(03)00025-X

Lowenberg-DeBoer, J., Erickson, B., 2019. Setting the Record Straight on Precision Agriculture Adoption. Agron. J. 111, 1552–1569. https://doi.org/10.2134/agronj2018.12.0779

Lunagaria, M.M., Patel, H.R., 2019. Evaluation of PROSAIL inversion for retrieval of chlorophyll, leaf dry matter, leaf angle, and leaf area index of wheat using spectrodirectional measurements. Int. J. Remote Sens. 40, 8125–8145. https://doi.org/10.1080/01431161.2018.1524608

Ma, J., Zheng, B., He, Y., 2022. Applications of a Hyperspectral Imaging System Used to Estimate Wheat Grain Protein: A Review. Front. Plant Sci. 13.

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., Johnson, B.A., 2019. Deep learning in remote sensing applications: A meta-analysis and review. ISPRS J. Photogramm. Remote Sens. 152, 166–177. https://doi.org/10.1016/j.isprsjprs.2019.04.015

Maccioni, A., Agati, G., Mazzinghi, P., 2001. New vegetation indices for remote measurement of chlorophylls based on leaf directional reflectance spectra. J. Photochem. Photobiol. B 61, 52–61. https://doi.org/10.1016/S1011-1344(01)00145-2

Magney, T.S., Eitel, J.U.H., Huggins, D.R., Vierling, L.A., 2016a. Proximal NDVI derived phenology improves in-season predictions of wheat quantity and quality. Agric. For. Meteorol. 217, 46–60. https://doi.org/10.1016/j.agrformet.2015.11.009

Magney, T.S., Vierling, L.A., Eitel, J., 2014. Remote detection of water stress conditions via a diurnal photochemical reflectance index (PRI) improves yield prediction in rainfed wheat 2014, B51F-0091.

Magney, T.S., Vierling, L.A., Eitel, J.U.H., Huggins, D.R., Garrity, S.R., 2016b. Response of high frequency Photochemical Reflectance Index (PRI) measurements to environmental conditions in wheat. Remote Sens. Environ. 173, 84–97. https://doi.org/10.1016/j.rse.2015.11.013

Mahlein, A.-K., Rumpf, T., Welke, P., Dehne, H.-W., Plümer, L., Steiner, U., Oerke, E.-C., 2013. Development of spectral indices for detecting and identifying plant diseases. Remote Sens. Environ. 128, 21–30. https://doi.org/10.1016/j.rse.2012.09.019

Masoni, A., Ercoli, L., Mariotti, M., Arduini, I., 2007. Post-anthesis accumulation and remobilization of dry matter, nitrogen and phosphorus in durum wheat as affected by soil type. Eur. J. Agron. 26, 179–186. https://doi.org/10.1016/j.eja.2006.09.006

Maxwell, A.E., Warner, T.A., Fang, F., 2018. Implementation of machine-learning classification in remote sensing: an applied review. Int. J. Remote Sens. 39, 2784–2817. https://doi.org/10.1080/01431161.2018.1433343

McCallum, M., Peirce, C., Porker, K., 2019. What drives the yield gap between durum and bread wheat?, in: Proceedings of the 19th Australian Agronomy Conference. Australian Society of Agronomy, Wagga Wagga, N.S.W., p. 4.

McLellan, E.L., Cassman, K.G., Eagle, A.J., Woodbury, P.B., Sela, S., Tonitto, C., Marjerison, R.D., van Es, H.M., 2018. The Nitrogen Balancing Act: Tracking the Environmental Performance of Food Production. BioScience 68, 194–203. https://doi.org/10.1093/biosci/bix164

McNeal, F.H., McGuire, C.F., Berg, M.A., 1978. Recurrent Selection for Grain Protein Content in Spring Wheat. Crop Sci. 18, 779–782. https://doi.org/10.2135/cropsci1978.0011183X001800050022x

Meroni, M., Rossini, M., Guanter, L., Alonso, L., Rascher, U., Colombo, R., Moreno, J., 2009. Remote sensing of solar-induced chlorophyll fluorescence: Review of methods and applications. Remote Sens. Environ. 113, 2037–2051. https://doi.org/10.1016/j.rse.2009.05.003

Miao, G., Guan, K., Yang, X., Bernacchi, C.J., Berry, J.A., DeLucia, E.H., Wu, J., Moore, C.E., Meacham, K., Cai, Y., Peng, B., Kimm, H., Masters, M.D., 2018. Sun-Induced Chlorophyll Fluorescence, Photosynthesis, and Light Use Efficiency of a Soybean Field from Seasonally Continuous Measurements. J. Geophys. Res. Biogeosciences 123, 610–623. https://doi.org/10.1002/2017JG004180

Milborrow, S., 2023. earth: Multivariate Adaptive Regression Splines.

Miller, R.O., Jacobsen, J.S., Skogley, E.O., 1993. Aerial accumulation and partitioning of nutrients by hard red spring wheat. Commun. Soil Sci. Plant Anal. 24, 2389–2407. https://doi.org/10.1080/00103629309368963

Miraglio, T., Adeline, K., Huesca, M., Ustin, S., Briottet, X., 2020. Monitoring LAI, Chlorophylls, and Carotenoids Content of a Woodland Savanna Using Hyperspectral Imagery and 3D Radiative Transfer Modeling. Remote Sens. 12, 28. https://doi.org/10.3390/rs12010028

Moffitt, E., 2020. Utilising new technologies to better manage within-paddock nitrogen variability and sustainably close the yield gap in southern NSW. FARMLINK Res. Rep. 2020.

Mohammed, G.H., Colombo, R., Middleton, E.M., Rascher, U., van der Tol, C., Nedbal, L., Goulas, Y., Pérez-Priego, O., Damm, A., Meroni, M., Joiner, J., Cogliati, S., Verhoef, W., Malenovský, Z., Gastellu-Etchegorry, J.-P., Miller, J.R., Guanter, L., Moreno, J., Moya, I., Berry, J.A., Frankenberg, C., Zarco-Tejada, P.J., 2019. Remote sensing of solar-induced chlorophyll fluorescence (SIF) in vegetation: 50 years of progress. Remote Sens. Environ. 231, 111177. https://doi.org/10.1016/j.rse.2019.04.030

Monjardino, M., McBeath, T., Ouzman, J., Llewellyn, R., Jones, B., 2015. Farmer risk-aversion limits closure of yield and profit gaps: A study of nitrogen management in the southern Australian wheatbelt. Agric. Syst. 137, 108–118. https://doi.org/10.1016/j.agsy.2015.04.006

Monjardino, M., McBeath, T.M., Brennan, L., Llewellyn, R.S., 2013. Are farmers in low-rainfall cropping regions under-fertilising with nitrogen? A risk analysis. Agric. Syst. 116, 37–51. https://doi.org/10.1016/j.agsy.2012.12.007

Moran, M.S., Clarke, T.R., Inoue, Y., Vidal, A., 1994. Estimating crop water deficit using the relation between surface-air temperature and spectral vegetation index. Remote Sens. Environ. 49, 246–263. https://doi.org/10.1016/0034-4257(94)90020-5

Murty, D., Kirschbaum, M.U.F., McMurtrie, R.E., McGilvray, H., 2002. Does conversion of forest to agricultural land change soil carbon and nitrogen? a review of the literature. Glob. Change Biol. 8, 105–123. https://doi.org/10.1046/j.1354-1013.2001.00459.x

Naing, A.H., Kim, C.K., 2021. Abiotic stress-induced anthocyanins in plants: Their role in tolerance to abiotic stresses. Physiol. Plant. 172, 1711–1723. https://doi.org/10.1111/ppl.13373

Nakazawa, M., 2022. Functions for Medical Statistics Book with some Demographic Data. R package "fsmb."

Nearing, G.S., Crow, W.T., Thorp, K.R., Moran, M.S., Reichle, R.H., Gupta, H.V., 2012. Assimilating remote sensing observations of leaf area index and soil moisture for wheat yield estimates: An observing system simulation experiment. Water Resour. Res. 48. https://doi.org/10.1029/2011WR011420

Nguy-Robertson, A., Gitelson, A., Peng, Y., Viña, A., Arkebauer, T., Rundquist, D., 2012. Green Leaf Area Index Estimation in Maize and Soybean: Combining Vegetation Indices to Achieve Maximal Sensitivity. Agron. J. 104, 1336–1347. https://doi.org/10.2134/agronj2012.0065

Nguy-Robertson, A.L., Peng, Y., Gitelson, A.A., Arkebauer, T.J., Pimstein, A., Herrmann, I., Karnieli, A., Rundquist, D.C., Bonfil, D.J., 2014. Estimating green LAI in four crops: Potential of determining optimal spectral bands for a universal algorithm. Agric. For. Meteorol. 192–193, 140–148. https://doi.org/10.1016/j.agrformet.2014.03.004

Novelli, L.E., Caviglia, O.P., Jobbágy, E.G., Sadras, V.O., 2023. Diversified crop sequences to reduce soil nitrogen mining in agroecosystems. Agric. Ecosyst. Environ. 341, 108208. https://doi.org/10.1016/j.agee.2022.108208

Ntinyari, W., Gweyi-Onyango, J., Giweta, M., Mutegi, J., Mochoge, B., Nziguheba, G., Masso, C., 2022. Nitrogen budgets and nitrogen use efficiency as agricultural performance indicators in Lake Victoria basin. Front. Sustain. Food Syst. 6.

Nuttall, J.G., Armstrong, R.D., 2010. Impact of subsoil physicochemical constraints on crops grown in the Wimmera and Mallee is reduced during dry seasonal conditions. Soil Res. 48, 125. https://doi.org/10.1071/SR09075

Nuttall, J.G., Armstrong, R.D., Connor, D.J., Matassa, V.J., 2003. Interrelationships between edaphic factors potentially limiting cereal growth on alkaline soils in north-western Victoria. Aust. J. Soil Res. 41, 277. https://doi.org/10.1071/SR02022

O'Brien, R.M., 2007. A Caution Regarding Rules of Thumb for Variance Inflation Factors. Qual. Quant. 41, 673–690. https://doi.org/10.1007/s11135-006-9018-6

Odilbekov, F., Armoniené, R., Henriksson, T., Chawade, A., 2018. Proximal Phenotyping and Machine Learning Methods to Identify Septoria Tritici Blotch Disease Symptoms in Wheat. Front. Plant Sci. 9.

Omara, P., Aula, L., Oyebiyi, F., Raun, W.R., 2019. World Cereal Nitrogen Use Efficiency Trends: Review and Current Knowledge. Agrosystems Geosci. Environ. 2, 180045. https://doi.org/10.2134/age2018.10.0045

Ottman, M.J., Doerge, T.A., Martin, E.C., 2000. Durum Grain Quality as Affected by Nitrogen Fertilization near Anthesis and Irrigation During Grain Fill. Agron. J. 92, 1035–1041. https://doi.org/10.2134/agronj2000.9251035x

Øvergaard, S.I., Isaksson, T., Korsaeth, A., 2013. Prediction of Wheat Yield and Protein Using Remote Sensors on Plots—Part I: Assessing near Infrared Model Robustness for Year and Site Variations. J. Infrared Spectrosc. 21, 117–131. https://doi.org/10.1255/jnirs.1042

Ozturk, A., Aydin, F., 2004. Effect of Water Stress at Various Growth Stages on Some Quality Characteristics of Winter Wheat. J. Agron. Crop Sci. 190, 93–99. https://doi.org/10.1046/j.1439-037X.2003.00080.x

Pan, H., Chen, Z., Ren, J., Li, H., Wu, S., 2019. Modeling Winter Wheat Leaf Area Index and Canopy Water Content With Three Different Approaches Using Sentinel-2 Multispectral Instrument Data. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 12, 482–492. https://doi.org/10.1109/JSTARS.2018.2855564

Pan, W.L., Kidwell, K.K., McCracken, V.A., Bolton, R.P., Allen, M., 2020. Economically Optimal Wheat Yield, Protein and Nitrogen Use Component Responses to Varying N Supply and Genotype. Front. Plant Sci. 10.

Pancorbo, J.L., Alonso-Ayuso, M., Camino, C., Raya-Sereno, M.D., Zarco-Tejada, P.J., Molina, I., Gabriel, J.L., Quemada, M., 2023. Airborne hyperspectral and Sentinel imagery to quantify winter wheat traits through ensemble modeling approaches. Precis. Agric. 24, 1288–1311. https://doi.org/10.1007/s11119-023-09990-y

Pancorbo, J.L., Camino, C., Alonso-Ayuso, M., Raya-Sereno, M.D., Gonzalez-Fernandez, I., Gabriel, J.L., Zarco-Tejada, P.J., Quemada, M., 2021. Simultaneous assessment of nitrogen and water status in winter wheat using hyperspectral and thermal sensors. Eur. J. Agron. 127, 126287. https://doi.org/10.1016/j.eja.2021.126287

Panozzo, J.F., Eagles, H.A., Wootton, M., 2001. Changes in protein composition during grain development in wheat. Aust. J. Agric. Res. 52, 485–493. https://doi.org/10.1071/ar00101

Panozzo, J.F., Walker, C.K., Partington, D.L., Neumann, N.C., Tausz, M., Seneweera, S., Fitzgerald, G.J., 2014. Elevated carbon dioxide changes grain protein concentration and composition and compromises baking quality. A FACE study. J. Cereal Sci. 60, 461–470. https://doi.org/10.1016/j.jcs.2014.08.011

Paul, M.J., Driscoll, S.P., 1997. Sugar repression of photosynthesis: the role of carbohydrates in signalling nitrogen deficiency through source:sink imbalance. Plant Cell Environ. 20, 110–116. https://doi.org/10.1046/j.1365-3040.1997.d01-17.x

Pearson, R.L., Miller, L.D., 1972. Remote Mapping of Standing Crop Biomass for Estimation of the Productivity of the Shortgrass Prairie, in: Proceedings of the Eighth International Symposium on Remote Sensing of Environment. Presented at the Eighth International Symposium on Remote Sensing of Environment, p. 1355.

Peel, M.C., Finlayson, B.L., McMahon, T.A., 2007. Updated world map of the Köppen-Geiger climate classification. Hydrol Earth Syst Sci 11, 1633–1644. https://doi.org/10.5194/hess-11-1633-2007

Peng, B., Guan, K., Zhou, W., Jiang, C., Frankenberg, C., Sun, Y., He, L., Köhler, P., 2020. Assessing the benefit of satellite-based Solar-Induced Chlorophyll Fluorescence in crop yield prediction. Int. J. Appl. Earth Obs. Geoinformation 90, 102126. https://doi.org/10.1016/j.jag.2020.102126

Peñuelas, J., Baret, F., Filella, I., 1995. Semi-empirical indices to assess carotenoids/chlorophyll a ratio from leaf spectral reflectance. Photosynthetica 31, 221–230.

Peñuelas, J., Filella, I., Biel, C., Savé, R., 1993. The reflectance at the 950–970 nm region as an indicator of plant water status. Int. J. Remote Sens. 14. https://doi.org/10.1080/01431169308954010

Peñuelas, J., Gamon, J.A., Fredeen, A.L., Merino, J., Field, C.B., 1994. Reflectance indices associated with physiological changes in nitrogen- and water-limited sunflower leaves. Remote Sens. Environ. 48, 135–146. https://doi.org/10.1016/0034-4257(94)90136-8

Ping, J.L., Dobermann, A., 2005. Processing of Yield Map Data. Precis. Agric. 6, 193–212. https://doi.org/10.1007/s11119-005-1035-2

Plascyk, J.A., 1975. The MK II Fraunhofer Line Discriminator (FLD-II) for Airborne and Orbital Remote Sensing of Solar-Stimulated Luminescence. Opt. Eng. 14, 339–346.

Plascyk, J.A., Gabriel, F.C., 1975. The Fraunhofer Line Discriminator MKII-An Airborne Instrument for Precise and Standardized Ecological Luminescence Measurement. IEEE Trans. Instrum. Meas. 24, 306–313. https://doi.org/10.1109/TIM.1975.4314448

Poblete, T., Camino, C., Beck, P.S.A., Hornero, A., Kattenborn, T., Saponari, M., Boscia, D., Navas-Cortes, J.A., Zarco-Tejada, P.J., 2020a. Detection of Xylella fastidiosa infection symptoms with airborne multispectral and thermal imagery: Assessing bandset reduction performance from hyperspectral analysis. ISPRS J. Photogramm. Remote Sens. 162, 27–40. https://doi.org/10.1016/j.isprsjprs.2020.02.010

Poblete, T., Camino, C., Beck, P.S.A., Hornero, A., Kattenborn, T., Saponari, M., Boscia, D., Navas-Cortes, J.A., Zarco-Tejada, P.J., 2020b. Detection of Xylella fastidiosa infection symptoms with airborne multispectral and thermal imagery: Assessing bandset reduction performance from hyperspectral analysis. ISPRS J. Photogramm. Remote Sens. 162, 27–40. https://doi.org/10.1016/j.isprsjprs.2020.02.010

Poblete, T., Navas-Cortes, J.A., Camino, C., Calderon, R., Hornero, A., Gonzalez-Dugo, V., Landa, B.B., Zarco-Tejada, P.J., 2021. Discriminating Xylella fastidiosa from Verticillium dahliae infections in olive trees using thermal- and hyperspectral-based plant traits. ISPRS J. Photogramm. Remote Sens. 179, 133–144. https://doi.org/10.1016/j.isprsjprs.2021.07.014

Porcar-Castell, A., Tyystjärvi, E., Atherton, J., van der Tol, C., Flexas, J., Pfündel, E.E., Moreno, J., Frankenberg, C., Berry, J.A., 2014. Linking chlorophyll a fluorescence to photosynthesis for remote sensing applications: mechanisms and challenges. J. Exp. Bot. 65, 4065–4095. https://doi.org/10.1093/jxb/eru191

Porter, J.R., Gawith, M., 1999. Temperatures and the growth and development of wheat: a review. Eur. J. Agron. 10, 23–36. https://doi.org/10.1016/S1161-0301(98)00047-1

Preece, C., Livarda, A., Christin, P., Wallace, M., Martin, G., Charles, M., Jones, G., Rees, M., Osborne, C.P., 2017. How did the domestication of Fertile Crescent grain crops increase their yields? Funct. Ecol. 31, 387–397. https://doi.org/10.1111/1365-2435.12760

Prey, L., Schmidhalter, U., 2019. Simulation of satellite reflectance data using high-frequency ground based hyperspectral canopy measurements for in-season estimation of grain yield and grain nitrogen status in winter wheat. ISPRS J. Photogramm. Remote Sens. 149, 176–187. https://doi.org/10.1016/j.isprsjprs.2019.01.023

QGIS Development Team, 2020. QGIS Geographic Information System.

Qi, J., Chehbouni, A., Huete, A.R., Kerr, Y.H., Sorooshian, S., 1994. A modified soil adjusted vegetation index. Remote Sens. Environ. 48, 119–126. https://doi.org/10.1016/0034-4257(94)90134-1

R Core Team, 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Rab, M.A., Fisher, P.D., Armstrong, R.D., Abuzar, M., Robinson, N.J., Chandra, S., 2009. Advances in precision agriculture in south-eastern Australia. IV. Spatial variability in plant-available water capacity of soil and its relationship with yield in site-specific management zones. Crop Pasture Sci. 60, 885–900. https://doi.org/10.1071/CP08350

Rahman, M.S., Wilson, J.H., 1978. Determination of spikelet number in wheat. III.* Effect of varying temperature on ear development. Aust. J. Agric. Res. 29, 459–467. https://doi.org/10.1071/ar9780459

Ranjan, R., Chopra, U.K., Sahoo, R.N., Singh, A.K., Pradhan, S., 2012. Assessment of plant nitrogen stress in wheat (Triticum aestivum L.) through hyperspectral indices. Int. J. Remote Sens. 33, 6342–6360. https://doi.org/10.1080/01431161.2012.687473

Rao, A.C.S., Smith, J.L., Jandhyala, V.K., Papendick, R.I., Parr, J.F., 1993. Cultivar and Climatic Effects on the Protein Content of Soft White Winter Wheat. Agron. J. 85, 1023–1028. https://doi.org/10.2134/agronj1993.00021962008500050013x

Raun, W.R., Solie, J.B., Johnson, G.V., Stone, M.L., Mullen, R.W., Freeman, K.W., Thomason, W.E., Lukina, E.V., 2002. Improving Nitrogen Use Efficiency in Cereal Grain Production with Optical Sensing and Variable Rate Application. Agron. J. 94, 815–820. https://doi.org/10.2134/agronj2002.8150

Raya-Sereno, M.D., Alonso-Ayuso, M., Pancorbo, J.L., Gabriel, J.L., Camino, C., Zarco-Tejada, P.J., Quemada, M., 2022. Residual Effect and N Fertilizer Rate Detection by High-Resolution VNIR-SWIR Hyperspectral Imagery and Solar-Induced Chlorophyll Fluorescence in Wheat. IEEE Trans. Geosci. Remote Sens. 60, 1–17. https://doi.org/10.1109/TGRS.2021.3099624

Raya-Sereno, M.D., Ortiz-Monasterio, J.I., Alonso-Ayuso, M., Rodrigues, F.A., Rodríguez, A.A., González-Perez, L., Quemada, M., 2021. High-Resolution Airborne Hyperspectral Imagery for Assessing Yield, Biomass, Grain N Concentration, and N Output in Spring Wheat. Remote Sens. 13, 1373. https://doi.org/10.3390/rs13071373

Rayment, G.E., Lyons, D.J. (Eds.), 2010. Soil chemical methods - Australasia. CSIRO Publishing, Melbourne, Australia.

Ren, S., Chen, X., An, S., 2017. Assessing plant senescence reflectance index-retrieved vegetation phenology and its spatiotemporal response to climate change in the Inner Mongolian Grassland. Int. J. Biometeorol. 61, 601–612. https://doi.org/10.1007/s00484-016-1236-6

Rharrabti, Y., Villegas, D., Royo, C., Martos-Núñez, V., García del Moral, L.F., 2003. Durum wheat quality in Mediterranean environments: II. Influence of climatic variables and relationships between quality parameters. Field Crops Res. 80, 133–140. https://doi.org/10.1016/S0378-4290(02)00177-6

Richter, K., Atzberger, C., Vuolo, F., D'Urso, G., 2011. Evaluation of Sentinel-2 Spectral Sampling for Radiative Transfer Model Based LAI Estimation of Wheat, Sugar Beet, and Maize. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 4, 458–464. https://doi.org/10.1109/JSTARS.2010.2091492

Ripley, B., 2023. nnet: Feed-Forward Neural Networks and Multinomial Log-Linear Models.

Rivera-Caicedo, J.P., Verrelst, J., Muñoz-Marí, J., Moreno, J., Camps-Valls, G., 2014. Toward a Semiautomatic Machine Learning Retrieval of Biophysical Parameters. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 7, 1249–1259. https://doi.org/10.1109/JSTARS.2014.2298752

Robertson, M.J., Llewellyn, R.S., Mandel, R., Lawes, R., Bramley, R.G.V., Swift, L., Metz, N., O'Callaghan, C., 2012. Adoption of variable rate fertiliser application in the Australian grains industry: status, issues and prospects. Precis. Agric. 13, 181–199. https://doi.org/10.1007/s11119-011-9236-3

Rodrigues, F., Blasch, G., Defourny, P., Ortiz-Monasterio, J., Schulthess, U., Zarco-Tejada, P., Taylor, J., Gérard, B., Rodrigues, F.A., Blasch, G., Defourny, P., Ortiz-Monasterio, J.I., Schulthess, U., Zarco-Tejada, P.J., Taylor, J.A., Gérard, B., 2018. Multi-Temporal and Spectral Analysis of High-Resolution Hyperspectral Airborne Imagery for Precision Agriculture: Assessment of Wheat Grain Yield and Grain Protein Content. Remote Sens. 10, 930. https://doi.org/10.3390/rs10060930

Rodriguez, D., Fitzgerald, G.J., Belford, R., Christensen, L.K., 2006. Detection of nitrogen deficiency in wheat from spectral reflectance indices and basic crop eco-physiological concepts. Aust. J. Agric. Res. 57, 781–789. https://doi.org/10.1071/AR05361

Rondeaux, G., Steven, M., Baret, F., 1996. Optimization of soil-adjusted vegetation indices. Remote Sens. Environ. 55, 95–107. https://doi.org/10.1016/0034-4257(95)00186-7

Roujean, J.-L., Breon, F.-M., 1995. Estimating PAR absorbed by vegetation from bidirectional reflectance measurements. Remote Sens. Environ. 51, 375–384. https://doi.org/10.1016/0034-4257(94)00114-3

Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.W., 1974. Monitoring vegetation systems in the Great Plains with ERTS. NASA Spec. Publ. 351, 309.

Ruan, G., Li, X., Yuan, F., Cammarano, D., Ata-UI-Karim, S.T., Liu, X., Tian, Y., Zhu, Y., Cao, W., Cao, Q., 2022. Improving wheat yield prediction integrating proximal sensing and weather data with machine learning. Comput. Electron. Agric. 195, 106852. https://doi.org/10.1016/j.compag.2022.106852

Russell, J.S., 1963. Nitrogen content of wheat grain as as indication of potential yield response to nitrogen fertilizer. Aust. J. Exp. Agric. 3, 319–325. https://doi.org/10.1071/ea9630319

Russell, K., Lee, C., Van Sanford, D., 2017. Interaction of Genetics, Environment, and Management in Determining Soft Red Winter Wheat Yields. Agron. J. 109, 2463–2473. https://doi.org/10.2134/agronj2017.02.0126

Sadras, V., Roget, D., O'Leary, G., 2002. On-farm assessment of environmental and management constraints to wheat yield and efficiency in the use of rainfall in the Mallee. Aust. J. Agric. Res. 53, 587–598. https://doi.org/10.1071/AR01150

Sadras, V.O., 2004. Yield and water-use efficiency of water- and nitrogen-stressed wheat crops increase with degree of co-limitation. Eur. J. Agron., Water Limited Agriculture 21, 455–464. https://doi.org/10.1016/j.eja.2004.07.007

Sadras, V.O., 2003. Influence of size of rainfall events on water-driven processes. I. Water budget of wheat crops in south-eastern Australia. Aust. J. Agric. Res. 54, 341–351. https://doi.org/10.1071/ar02112

Sadras, V.O., Angus, J.F., 2006. Benchmarking water-use efficiency of rainfed wheat in dry environments. Aust. J. Agric. Res. 57, 847–856. https://doi.org/10.1071/AR05359

Saint Pierre, C., Peterson, C.J., Ross, A.S., Ohm, J.B., Verhoeven, M.C., Larson, M., Hoefer, B., 2008. Winter wheat genotypes under different levels of nitrogen and water stress: Changes in grain protein composition. J. Cereal Sci. 47, 407–416. https://doi.org/10.1016/j.jcs.2007.05.007

Sanchez, P.A., 2002. Soil Fertility and Hunger in Africa. Science 295, 2019–2020.

Sanderman, J., Farquharson, R., Baldock, J., 2010. Soil Carbon Sequestration Potential: A review for Australian agriculture. Department of Climate Change and Energy Efficiency.

Schiefer, F., Schmidtlein, S., Kattenborn, T., 2021. The retrieval of plant functional traits from canopy spectra through RTM-inversions and statistical models are both critically affected by plant phenology. Ecol. Indic. 121, 107062. https://doi.org/10.1016/j.ecolind.2020.107062

Schliep, K., Hechenbichler, K., Lizee, A., 2016. kknn: Weighted k-Nearest Neighbors.

Scholes, M.C., Scholes, R.J., 2013. Dust Unto Dust. Science 342, 565–566. https://doi.org/10.1126/science.1244579

Schulte-Uebbing, L.F., Beusen, A.H.W., Bouwman, A.F., de Vries, W., 2022. From planetary to regional boundaries for agricultural nitrogen pollution. Nature 610, 507–512. https://doi.org/10.1038/s41586-022-05158-2

Schwenke, G., Beange, L., Cameron, J., Bell, M., Harden, S., 2019. What soil information do crop advisors use to develop nitrogen fertilizer recommendations for grain growers in New South Wales, Australia? Soil Use Manag. 35, 85–93. https://doi.org/10.1111/sum.12469

Segarra, J., Buchaillot, M.L., Araus, J.L., Kefauver, S.C., 2020. Remote Sensing for Precision Agriculture: Sentinel-2 Improved Features and Applications. Agronomy 10, 641. https://doi.org/10.3390/agronomy10050641

Sehgal, V.K., Chakraborty, D., Sahoo, R.N., 2016. Inversion of radiative transfer model for retrieval of wheat biophysical parameters from broadband reflectance measurements. Inf. Process. Agric. 3, 107–118. https://doi.org/10.1016/j.inpa.2016.04.001

Shah, S.H., Angel, Y., Houborg, R., Ali, S., McCabe, M.F., 2019. A Random Forest Machine Learning Approach for the Retrieval of Leaf Chlorophyll Content in Wheat. Remote Sens. 11, 920. https://doi.org/10.3390/rs11080920

Shah, S.H., Houborg, R., McCabe, M.F., 2017. Response of Chlorophyll, Carotenoid and SPAD-502 Measurement to Salinity and Nutrient Stress in Wheat (Triticum aestivum L.). Agronomy 7, 61. https://doi.org/10.3390/agronomy7030061

Shanahan, J.F., Kitchen, N.R., Raun, W.R., Schepers, J.S., 2008. Responsive in-season nitrogen management for cereals. Comput. Electron. Agric., Emerging Technologies For Real-time and Integrated Agriculture Decisions 61, 51–62. https://doi.org/10.1016/j.compag.2007.06.006

Shcherbak, I., Millar, N., Robertson, G.P., 2014. Global metaanalysis of the nonlinear response of soil nitrous oxide (N2O) emissions to fertilizer nitrogen. Proc. Natl. Acad. Sci. 111, 9199–9204. https://doi.org/10.1073/pnas.1322434111

Shewry, P.R., 2009. Wheat. J. Exp. Bot. 60, 1537–1553. https://doi.org/10.1093/jxb/erp058

Shiferaw, B., Smale, M., Braun, H.-J., Duveiller, E., Reynolds, M., Muricho, G., 2013. Crops that feed the world 10. Past successes and future challenges to the role played by wheat in global food security. Food Secur. 5, 291–317. https://doi.org/10.1007/s12571-013-0263-y

Shoeva, O.Y.U., Gordeeva, E.I., Arbuzova, V.S., Khlestkina, E.K., 2017. Anthocyanins Participate in Protection of Wheat Seedlings from Osmotic Stress. Cereal Res. Commun. 45, 47–56. https://doi.org/10.1556/0806.44.2016.044

Sims, D.A., Gamon, J.A., 2002. Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. Remote Sens. Environ. 81, 337–354. https://doi.org/10.1016/S0034-4257(02)00010-X

Sims, D.A., Luo, H., Hastings, S., Oechel, W.C., Rahman, A.F., Gamon, J.A., 2006. Parallel adjustments in vegetation greenness and ecosystem CO2 exchange in response to drought in a Southern California chaparral ecosystem. Remote Sens. Environ., Spectral Network 103, 289–303. https://doi.org/10.1016/j.rse.2005.01.020

Sishodia, R.P., Ray, R.L., Singh, S.K., 2020. Applications of Remote Sensing in Precision Agriculture: A Review. Remote Sens. 12, 3136. https://doi.org/10.3390/rs12193136

Sloat, L.L., Lin, M., Butler, E.E., Johnson, D., Holbrook, N.M., Huybers, P.J., Lee, J.-E., Mueller, N.D., 2021. Evaluating the benefits of chlorophyll fluorescence for in-season crop productivity forecasting. Remote Sens. Environ. 260, 112478. https://doi.org/10.1016/j.rse.2021.112478

Smil, V., 2002. Nitrogen and Food Production: Proteins for Human Diets. AMBIO J. Hum. Environ. 31, 126–131. https://doi.org/10.1579/0044-7447-31.2.126

Song, X., Yang, G., Yang, C., Wang, J., Cui, B., 2017. Spatial Variability Analysis of Within-Field Winter Wheat Nitrogen and Grain Quality Using Canopy Fluorescence Sensor Measurements. Remote Sens. 9, 237. https://doi.org/10.3390/rs9030237

Song, Y., Wang, J., Wang, L., 2020. Satellite Solar-Induced Chlorophyll Fluorescence Reveals Heat Stress Impacts on Wheat Yield in India. Remote Sens. 12, 3277. https://doi.org/10.3390/rs12203277

Sonobe, R., Miura, Y., Sano, T., Horie, H., 2018. Estimating leaf carotenoid contents of shade-grown tea using hyperspectral indices and PROSPECT–D inversion. Int. J. Remote Sens. 39, 1306–1320. https://doi.org/10.1080/01431161.2017.1407050

Sonobe, R., Wang, Q., 2018. Nondestructive assessments of carotenoids content of broadleaved plant species using hyperspectral indices. Comput. Electron. Agric. 145, 18–26. https://doi.org/10.1016/j.compag.2017.12.022

Sowers, K.E., Miller, B.C., Pan, W.L., 1994. Optimizing Yield and Grain Protein in Soft White Winter Wheat with Split Nitrogen Applications. Agron. J. 86, 1020–1025. https://doi.org/10.2134/agronj1994.00021962008600060017x

Steffen, W., Richardson, K., Rockström, J., Cornell, S.E., Fetzer, I., Bennett, E.M., Biggs, R., Carpenter, S.R., de Vries, W., de Wit, C.A., Folke, C., Gerten, D., Heinke, J., Mace, G.M., Persson, L.M., Ramanathan, V., Reyers, B., Sörlin, S., 2015. Planetary boundaries: Guiding human development on a changing planet. Science 347, 1259855. https://doi.org/10.1126/science.1259855

Stein, L.Y., Klotz, M.G., 2016. The nitrogen cycle. Curr. Biol. 26, R94–R98. https://doi.org/10.1016/j.cub.2015.12.021

Stoy, P.C., Khan, A.M., Wipf, A., Silverman, N., Powell, S.L., 2022. The spatial variability of NDVI within a wheat field: Information content and implications for yield and grain protein monitoring. PLOS ONE 17, e0265243. https://doi.org/10.1371/journal.pone.0265243

Strzałka, K., Kostecka-Gugała, A., Latowski, D., 2003. Carotenoids and Environmental Stress in Plants: Significance of Carotenoid-Mediated Modulation of Membrane Physical Properties. Russ. J. Plant Physiol. 50, 168–173. https://doi.org/10.1023/A:1022960828050

Suárez, L., Zarco-Tejada, P.J., Sepulcre-Cantó, G., Pérez-Priego, O., Miller, J.R., Jiménez-Muñoz, J.C., Sobrino, J., 2008. Assessing canopy PRI for water stress detection with diurnal airborne imagery. Remote Sens. Environ., Soil Moisture Experiments 2004 (SMEX04) Special Issue 112, 560–575. https://doi.org/10.1016/j.rse.2007.05.009

Subedi, K.D., Gregory, P.J., Summerfield, R.J., Gooding, M.J., 1998. Cold temperatures and boron deficiency caused grain set failure in spring wheat (Triticum aestivum L.). Field Crops Res. 57, 277–288. https://doi.org/10.1016/S0378-4290(97)00148-2

Sun, Z., Li, Q., Jin, S., Song, Y., Xu, S., Wang, X., Cai, J., Zhou, Q., Ge, Y., Zhang, R., Zang, J., Jiang, D., 2022. Simultaneous Prediction of Wheat Yield and Grain Protein Content Using Multitask Deep Learning from Time-Series Proximal Sensing. Plant Phenomics 2022, 2022/9757948. https://doi.org/10.34133/2022/9757948

Sutton, M.A., Oenema, O., Erisman, J.W., Leip, A., van Grinsven, H., Winiwarter, W., 2011. Too much of a good thing. Nature 472, 159–161. https://doi.org/10.1038/472159a

Sutton, M.A., van Dijk, N., Levy, P.E., Jones, M.R., Leith, I.D., Sheppard, L.J., Leeson, S., Sim Tang, Y., Stephens, A., Braban, C.F., Dragosits, U., Howard, C.M., Vieno, M., Fowler, D., Corbett, P., Naikoo, M.I., Munzi, S., Ellis, C.J., Chatterjee, S., Steadman, C.E., Móring, A., Wolseley, P.A., 2020. Alkaline air: changing perspectives on nitrogen and air pollution in an ammonia-rich world. Philos. Trans. R. Soc. Math. Phys. Eng. Sci. 378, 20190315. https://doi.org/10.1098/rsta.2019.0315

Tan, C., Zhou, X., Zhang, P., Wang, Z., Wang, D., Guo, W., Yun, F., 2020. Predicting grain protein content of field-grown winter wheat with satellite images and partial least square algorithm. PLOS ONE 15, e0228500. https://doi.org/10.1371/journal.pone.0228500

Tan, C.-W., Zhang, P.-P., Zhou, X.-X., Wang, Z.-X., Xu, Z.-Q., Mao, W., Li, W.-X., Huo, Z.-Y., Guo, W.-S., Yun, F., 2020. Quantitative monitoring of leaf area index in wheat of different plant types by integrating NDVI and Beer-Lambert law. Sci. Rep. 10, 929. https://doi.org/10.1038/s41598-020-57750-z

Taulemesse, F., Gouis, J.L., Gouache, D., Gibon, Y., Allard, V., 2016. Bread Wheat (Triticum aestivum L.) Grain Protein Concentration Is Related to Early Post-Flowering Nitrate Uptake under Putative Control of Plant Satiety Level. PLOS ONE 11, e0149668. https://doi.org/10.1371/journal.pone.0149668

Terman, G.L., 1979. Yields and Protein Content of Wheat Grain as Affected by Cultivar, N, and Environmental Growth Factors1. Agron. J. 71, 437–440. https://doi.org/10.2134/agronj1979.00021962007100030014x

Thakur, P., Kumar, S., Malik, J.A., Berger, J.D., Nayyar, H., 2010. Cold stress effects on reproductive development in grain crops: An overview. Environ. Exp. Bot. 67, 429–443. https://doi.org/10.1016/j.envexpbot.2009.09.004

Thenot, F., Méthy, M., Winkel, T., 2002. The Photochemical Reflectance Index (PRI) as a water-stress index. Int. J. Remote Sens. 23, 5135–5139. https://doi.org/10.1080/01431160210163100

Therneau, T., Atkinson, B., 2023. rpart: Recursive Partitioning and Regression Trees.

Thorne, G.N., Wood, D.W., 1987. Effects of Radiation and Temperature on Tiller Survival, Grain Number and Grain Yield in Winter Wheat. Ann. Bot. 59, 413–426. https://doi.org/10.1093/oxfordjournals.aob.a087330

Toscano, P., Castrignanò, A., Di Gennaro, S.F., Vonella, A.V., Ventrella, D., Matese, A., 2019. A Precision Agriculture Approach for Durum Wheat Yield Assessment Using Remote Sensing Data and Yield Mapping. Agronomy 9, 437. https://doi.org/10.3390/agronomy9080437

Triboï, E., Martre, P., Triboï-Blondel, A., 2003. Environmentally-induced changes in protein composition in developing grains of wheat are related to changes in total protein content. J. Exp. Bot. 54, 1731–1742. https://doi.org/10.1093/jxb/erg183

Ugarte, C., Calderini, D.F., Slafer, G.A., 2007. Grain weight and grain number responsiveness to pre-anthesis temperature in wheat, barley and triticale. Field Crops Res. 100, 240–248. https://doi.org/10.1016/j.fcr.2006.07.010

Upreti, D., Huang, W., Kong, W., Pascucci, S., Pignatti, S., Zhou, X., Ye, H., Casa, R., 2019. A Comparison of Hybrid Machine Learning Algorithms for the Retrieval of Wheat Biophysical Variables from Sentinel-2. Remote Sens. 11, 481. https://doi.org/10.3390/rs11050481

Ustin, S.L., Gitelson, A.A., Jacquemoud, S., Schaepman, M., Asner, G.P., Gamon, J.A., Zarco-Tejada, P., 2009. Retrieval of foliar information about plant pigment systems from high resolution spectroscopy. Remote Sens. Environ., Imaging Spectroscopy Special Issue 113, S67–S77. https://doi.org/10.1016/j.rse.2008.10.019

Vallentin, C., Dobers, E.S., Itzerott, S., Kleinschmit, B., Spengler, D., 2020. Delineation of management zones with spatial data fusion and belief theory. Precis. Agric. 21, 802–830. https://doi.org/10.1007/s11119-019-09696-0

Van Grinsven, H.J.M., Holland, M., Jacobsen, B.H., Klimont, Z., Sutton, M. a., Jaap Willems, W., 2013. Costs and Benefits of Nitrogen for Europe and Implications for Mitigation. Environ. Sci. Technol. 47, 3571–3579. https://doi.org/10.1021/es303804g

van Herwaarden, A.F., Angus, J.F., Richards, R.A., Farquhar, G.D., 1998. “Haying-off”, the negative grain yield response of dryland wheat to nitrogen fertiliser II. Carbohydrate and protein dynamics. Aust. J. Agric. Res. 49, 1083–1094. https://doi.org/10.1071/a97040

van Ittersum, M.K., Howden, S.M., Asseng, S., 2003. Sensitivity of productivity and deep drainage of wheat cropping systems in a Mediterranean environment to changes in CO2, temperature and precipitation. Agric. Ecosyst. Environ. 97, 255–273. https://doi.org/10.1016/S0167-8809(03)00114-2

van Klompenburg, T., Kassahun, A., Catal, C., 2020. Crop yield prediction using machine learning: A systematic literature review. Comput. Electron. Agric. 177, 105709. https://doi.org/10.1016/j.compag.2020.105709

Verhoef, W., 1984. Light scattering by leaf layers with application to canopy reflectance modeling: The SAIL model. Remote Sens. Environ. 16, 125–141. https://doi.org/10.1016/0034-4257(84)90057-9

Verhoef, W., Jia, L., Xiao, Q., Su, Z., 2007. Unified Optical-Thermal Four-Stream Radiative Transfer Theory for Homogeneous Vegetation Canopies. IEEE Trans. Geosci. Remote Sens. 45, 1808–1822. https://doi.org/10.1109/TGRS.2007.895844

Verrelst, J., Camps-Valls, G., Muñoz-Marí, J., Rivera, J.P., Veroustraete, F., Clevers, J.G.P.W., Moreno, J., 2015a. Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties – A review. ISPRS J. Photogramm. Remote Sens. 108, 273–290. https://doi.org/10.1016/j.isprsjprs.2015.05.005

Verrelst, J., Malenovský, Z., Van der Tol, C., Camps-Valls, G., Gastellu-Etchegorry, J.-P., Lewis, P., North, P., Moreno, J., 2019. Quantifying Vegetation Biophysical Variables from Imaging Spectroscopy Data: A Review on Retrieval Methods. Surv. Geophys. 40, 589–629. https://doi.org/10.1007/s10712-018-9478-y

Verrelst, J., Rivera, J.P., van der Tol, C., Magnani, F., Mohammed, G., Moreno, J., 2015b. Global sensitivity analysis of the SCOPE model: What drives simulated canopy-leaving sun-induced fluorescence? Remote Sens. Environ. 166, 8–21. https://doi.org/10.1016/j.rse.2015.06.002

Vitousek, P.M., Aber, J.D., Howarth, R.W., Likens, G.E., Matson, P.A., Schindler, D.W., Schlesinger, W.H., Tilman, D.G., 1997. Human Alteration of the Global Nitrogen Cycle: Sources and Consequences. Ecol. Appl. 7, 737–750. https://doi.org/10.1890/1051-0761(1997)007[0737:HAOTGN]2.0.CO;2

Vogelmann, J.E., Rock, B.N., Moss, D.M., 1993. Red edge spectral measurements from sugar maple leaves. Int. J. Remote Sens. 14, 1563–1575. https://doi.org/10.1080/01431169308953986

von Caemmerer, S., Lawson, T., Oxborough, K., Baker, N.R., Andrews, T.J., Raines, C.A., 2004. Stomatal conductance does not correlate with photosynthetic capacity in transgenic tobacco with reduced amounts of Rubisco. J. Exp. Bot. 55, 1157–1166. https://doi.org/10.1093/jxb/erh128

Waldner, F., Horan, H., Chen, Y., Hochman, Z., 2019. High temporal resolution of leaf area data improves empirical estimation of grain yield. Sci. Rep. 9, 15714. https://doi.org/10.1038/s41598-019-51715-7

Walker, A.P., Beckerman, A.P., Gu, L., Kattge, J., Cernusak, L.A., Domingues, T.F., Scales, J.C., Wohlfahrt, G., Wullschleger, S.D., Woodward, F.I., 2014. The relationship of leaf photosynthetic traits – Vcmax and Jmax – to leaf nitrogen, leaf phosphorus, and specific leaf area: a meta-analysis and modeling study. Ecol. Evol. 4, 3218–3235. https://doi.org/10.1002/ece3.1173

Walsh, O.S., Marshall, J.M., Nambi, E., Jackson, C.A., Ansah, E.O., Lamichhane, R., McClintick-Chess, J., Bautista, F., 2023. Wheat Yield and Protein Estimation with Handheld and Unmanned Aerial Vehicle-Mounted Sensors. Agronomy 13, 207. https://doi.org/10.3390/agronomy13010207

Wang, L., Tian, Y., Yao, X., Zhu, Y., Cao, W., 2014. Predicting grain yield and protein content in wheat by fusing multi-sensor and multi-temporal remote-sensing images. Field Crops Res. 164, 178–188. https://doi.org/10.1016/j.fcr.2014.05.001

Wang, W., Yao, Xia, Yao, XinFeng, Tian, Y., Liu, X., Ni, J., Cao, W., Zhu, Y., 2012. Estimating leaf nitrogen concentration with three-band vegetation indices in rice and wheat. Field Crops Res. 129, 90–98. https://doi.org/10.1016/j.fcr.2012.01.014

Wang, Y., Suarez, L., Poblete, T., Gonzalez-Dugo, V., Ryu, D., Zarco-Tejada, P.J., 2022. Evaluating the role of solar-induced fluorescence (SIF) and plant physiological traits for leaf nitrogen assessment in almond using airborne hyperspectral imagery. Remote Sens. Environ. 279, 113141. https://doi.org/10.1016/j.rse.2022.113141

Wang, Z., Skidmore, A.K., Darvishzadeh, R., Heiden, U., Heurich, M., Wang, T., 2015. Leaf Nitrogen Content Indirectly Estimated by Leaf Traits Derived From the PROSPECT Model. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 8, 3172–3182. https://doi.org/10.1109/JSTARS.2015.2422734

Wang, Z.J., Wang, J.H., Liu, L.Y., Huang, W.J., Zhao, C.J., Wang, C.Z., 2004. Prediction of grain protein content in winter wheat (Triticum aestivum L.) using plant pigment ratio (PPR). Field Crops Res. 90, 311–321. https://doi.org/10.1016/j.fcr.2004.04.004

Wardlaw, I.F., 1967. The effect of water stress on translocation in relation to photosynthesis and growth. I. Effect during grain development in wheat. Aust. J. Biol. Sci. 20, 25–39. https://doi.org/10.1071/bi9670025

Weiss, M., Baret, F., 1999. Evaluation of Canopy Biophysical Variable Retrieval Performances from the Accumulation of Large Swath Satellite Data. Remote Sens. Environ. 70, 293–306. https://doi.org/10.1016/S0034-4257(99)00045-0

Weiss, M., Condon, J., Tavakkoli, E., Whatmuff, M., Armstrong, R., 2022. Amending multiple soil constraints throughout the profile on a layer basis, in: Proceedings of the 20th Agronomy Australia Conference. Toowoomba, Australia.

Whelan, B.M., 2019. On-the-go protein sensors, in: GRDC Grains Research Update. Grains Research and Development Corporation, Goondiwindi.

Whelan, B.M., McBratney, A., 2003. Definition and interpretation of potential management zones in Australia.

Whelan, B.M., Taylor, J.A., Hassall, J.A., 2009. Site-specific variation in wheat grain protein concentration and wheat grain yield measured on an Australian farm using harvester-mounted on-the-go sensors. Crop Pasture Sci. 60, 808–817. https://doi.org/10.1071/CP08343

Wiegand, C.L., Richardson, A.J., Escobar, D.E., Gerbermann, A.H., 1991. Vegetation indices in crop assessments. Remote Sens. Environ. 35, 105–119. https://doi.org/10.1016/0034-4257(91)90004-P

Wohlfahrt, G., Gerdel, K., Migliavacca, M., Rotenberg, E., Tatarinov, F., Müller, J., Hammerle, A., Julitta, T., Spielmann, F.M., Yakir, D., 2018. Sun-induced fluorescence and gross primary productivity during a heat wave. Sci. Rep. 8, 14169. https://doi.org/10.1038/s41598-018-32602-z

Wolanin, A., Camps-Valls, G., Gómez-Chova, L., Mateo-García, G., van der Tol, C., Zhang, Y., Guanter, L., 2019. Estimating crop primary productivity with Sentinel-2 and Landsat 8 using machine learning methods trained with radiative transfer simulations. Remote Sens. Environ. 225, 441–457. https://doi.org/10.1016/j.rse.2019.03.002

Wolanin, A., Mateo-García, G., Camps-Valls, G., Gómez-Chova, L., Meroni, M., Duveiller, G., Liangzhi, Y., Guanter, L., 2020. Estimating and understanding crop yields with explainable deep learning in the Indian Wheat Belt. Environ. Res. Lett. 15, 024019. https://doi.org/10.1088/1748-9326/ab68ac

Wong, C.Y.S., Mercado, L.M., Arain, M.A., Ensminger, I., 2022. Remotely sensed carotenoid dynamics improve modelling photosynthetic phenology in conifer and deciduous forests. Agric. For. Meteorol. 321, 108977. https://doi.org/10.1016/j.agrformet.2022.108977

Wright, D.L., Rasmussen, V.P., Ramsey, R.D., Baker, D.J., Ellsworth, J.W., 2004. Canopy Reflectance Estimation of Wheat Nitrogen Content for Grain Protein Management. GIScience Remote Sens. 41, 287–300. https://doi.org/10.2747/1548-1603.41.4.287

Wright, M.N., Ziegler, A., 2017. ranger: A fast implementation of random forests for high dimensional data in C++ and R. J. Stat. Softw. 77, 1–17. https://doi.org/10.18637/jss.v077.i01

Wu, C., Niu, Z., Tang, Q., Huang, W., 2008. Estimating chlorophyll content from hyperspectral vegetation indices: Modeling and validation. Agric. For. Meteorol. 148, 1230–1241. https://doi.org/10.1016/j.agrformet.2008.03.005

Wu, D., Hur, K., Xiao, Z., 2021. A GAN-Enhanced Ensemble Model for Energy Consumption Forecasting in Large Commercial Buildings. IEEE Access 9, 158820–158830. https://doi.org/10.1109/ACCESS.2021.3131185

Xia, L., Lam, S.K., Chen, D., Wang, J., Tang, Q., Yan, X., 2017. Can knowledge-based N management produce more staple grain with lower greenhouse gas emission and reactive nitrogen pollution? A meta-analysis. Glob. Change Biol. 23, 1917–1925. https://doi.org/10.1111/gcb.13455

Xia, L., Yan, X., 2023. How to feed the world while reducing nitrogen pollution. Nature 613, 34–35. https://doi.org/10.1038/d41586-022-04490-x

Xu, G., Fan, X., Miller, A.J., 2012. Plant Nitrogen Assimilation and Use Efficiency. Annu. Rev. Plant Biol. 63, 153–182. https://doi.org/10.1146/annurev-arplant-042811-105532

Xu, X.Q., Lu, J.S., Zhang, N., Yang, T.C., He, J.Y., Yao, X., Cheng, T., Zhu, Y., Cao, W.X., Tian, Y.C., 2019. Inversion of rice canopy chlorophyll content and leaf area index based on coupling of radiative transfer and Bayesian network models. ISPRS J. Photogramm. Remote Sens. 150, 185–196. https://doi.org/10.1016/j.isprsjprs.2019.02.013

Xue, L.-H., Cao, W.-X., Yang, L.-Z., 2007. Predicting Grain Yield and Protein Content in Winter Wheat at Different N Supply Levels Using Canopy Reflectance Spectra. Pedosphere 17, 646–653. https://doi.org/10.1016/S1002-0160(07)60077-0

Yamuangmorn, S., Dell, B., Rerkasem, B., Prom-u-thai, C., 2018. Applying nitrogen fertilizer increased anthocyanin in vegetative shoots but not in grain of purple rice genotypes. J. Sci. Food Agric. 98, 4527–4532. https://doi.org/10.1002/jsfa.8978

Yang, B., Ma, J., Yao, X., Cao, W., Zhu, Y., 2021. Estimation of Leaf Nitrogen Content in Wheat Based on Fusion of Spectral Features and Deep Features from Near Infrared Hyperspectral Imagery. Sensors 21, 613. https://doi.org/10.3390/s21020613

Yao, X., Huang, Y., Shang, G., Zhou, C., Cheng, T., Tian, Y., Cao, W., Zhu, Y., 2015. Evaluation of Six Algorithms to Monitor Wheat Leaf Nitrogen Concentration. Remote Sens. 7, 14939–14966. https://doi.org/10.3390/rs71114939

Yendrek, C.R., Tomaz, T., Montes, C.M., Cao, Y., Morse, A.M., Brown, P.J., McIntyre, L.M., Leakey, A.D.B., Ainsworth, E.A., 2017. High-Throughput Phenotyping of Maize Leaf Physiological and Biochemical Traits Using Hyperspectral Reflectance. Plant Physiol. 173, 614–626. https://doi.org/10.1104/pp.16.01447

Yoder, B.J., Pettigrew-Crosby, R.E., 1995. Predicting nitrogen and chlorophyll content and concentrations from reflectance spectra (400–2500 nm) at leaf and canopy scales. Remote Sens. Environ. 53, 199–211. https://doi.org/10.1016/0034-4257(95)00135-N

Zadoks, J.C., Chang, T.T., Konzak, C.F., 1974. A decimal code for the growth stages of cereals. Weed Res. 14, 415–421. https://doi.org/10.1111/j.1365-3180.1974.tb01084.x

Zarco-Tejada, P.J., Berjón, A., Miller, J.R., 2004. Stress Detection in Crops with Hyperspectral Remote Sensing and Physical Simulation Models. Presented at the Airborne Imaging Spectroscopy Workshop, Bruges, Belgium, p. 5.

Zarco-Tejada, P.J., Camino, C., Beck, P.S.A., Calderon, R., Hornero, A., Hernández-Clemente, R., Kattenborn, T., Montes-Borrego, M., Susca, L., Morelli, M., Gonzalez-Dugo, V., North, P.R.J., Landa, B.B., Boscia, D., Saponari, M., Navas-Cortes, J.A., 2018. Previsual symptoms of Xylella fastidiosa infection revealed in spectral plant-trait alterations. Nat. Plants 4, 432. https://doi.org/10.1038/s41477-018-0189-7

Zarco-Tejada, P.J., Catalina, A., González, M.R., Martín, P., 2013a. Relationships between net photosynthesis and steady-state chlorophyll fluorescence retrieved from airborne hyperspectral imagery. Remote Sens. Environ. 136, 247–258. https://doi.org/10.1016/j.rse.2013.05.011

Zarco-Tejada, P.J., González-Dugo, M.V., Fereres, E., 2016. Seasonal stability of chlorophyll fluorescence quantified from airborne hyperspectral imagery as an indicator of net photosynthesis in the context of precision agriculture. Remote Sens. Environ. 179, 89–103. https://doi.org/10.1016/j.rse.2016.03.024

Zarco-Tejada, P.J., González-Dugo, V., Berni, J.A.J., 2012. Fluorescence, temperature and narrow-band indices acquired from a UAV platform for water stress detection using a micro-hyperspectral imager and a thermal camera. Remote Sens. Environ., Remote Sensing of Urban Environments 117, 322–337. https://doi.org/10.1016/j.rse.2011.10.007

Zarco-Tejada, P.J., Guillén-Climent, M.L., Hernández-Clemente, R., Catalina, A., González, M.R., Martín, P., 2013b. Estimating leaf carotenoid content in vineyards using high resolution hyperspectral imagery acquired from an unmanned aerial vehicle (UAV). Agric. For. Meteorol. 171–172, 281–294. https://doi.org/10.1016/j.agrformet.2012.12.013

Zarco-Tejada, Pablo J, Miller, J.R., Mohammed, G.H., Noland, T.L., Sampson, P.H., 2001. Estimation of chlorophyll fluorescence under natural illumination from hyperspectral data. Int. J. Appl. Earth Obs. Geoinformation 3, 321–327. https://doi.org/10.1016/S0303-2434(01)85039-X

Zarco-Tejada, P.J., Miller, J.R., Noland, T.L., Mohammed, G.H., Sampson, P.H., 2001. Scaling-up and model inversion methods with narrowband optical indices for chlorophyll content estimation in closed forest canopies with hyperspectral data. IEEE Trans. Geosci. Remote Sens. 39, 1491–1507. https://doi.org/10.1109/36.934080

Zarco-Tejada, P.J., Morales, A., Testi, L., Villalobos, F.J., 2013c. Spatio-temporal patterns of chlorophyll fluorescence and physiological and structural indices acquired from hyperspectral imagery as compared with carbon fluxes measured with eddy covariance. Remote Sens. Environ. 133, 102–115. https://doi.org/10.1016/j.rse.2013.02.003

Zarco-Tejada, P.J., Poblete, T., Camino, C., Gonzalez-Dugo, V., Calderon, R., Hornero, A., Hernandez-Clemente, R., Román-Écija, M., Velasco-Amo, M.P., Landa, B.B., Beck, P.S.A., Saponari, M., Boscia, D., Navas-Cortes, J.A., 2021. Divergent abiotic spectral pathways unravel pathogen stress signals across species. Nat. Commun. 12, 6088. https://doi.org/10.1038/s41467-021-26335-3

Zarco-Tejada, P.J., Ustin, S.L., Whiting, M.L., 2005. Temporal and Spatial Relationships between Within-Field Yield Variability in Cotton and High-Spatial Hyperspectral Remote Sensing Imagery. Agron. J. Madison 97, 641–653.

Zebarth, B.J., Botha, E.J., Rees, H., 2007. Rate and time of fertilizer nitrogen application on yield, protein and apparent efficiency of fertilizer nitrogen use of spring wheat. Can. J. Plant Sci. 87, 709–718. https://doi.org/10.4141/CJPS06001

Zhang, C., Kovacs, J.M., Wachowiak, M.P., Flores-Verdugo, F., 2013. Relationship between Hyperspectral Measurements and Mangrove Leaf Nitrogen Concentrations. Remote Sens. 5, 891–908. https://doi.org/10.3390/rs5020891

Zhang, C., Pattey, E., Liu, J., Cai, H., Shang, J., Dong, T., 2018. Retrieving Leaf and Canopy Water Content of Winter Wheat Using Vegetation Water Indices. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 11, 112–126. https://doi.org/10.1109/JSTARS.2017.2773625

Zhang, G., Liu, S., Dong, Y., Liao, Y., Han, J., 2022. A nitrogen fertilizer strategy for simultaneously increasing wheat grain yield and protein content: Mixed application of controlled-release urea and normal urea. Field Crops Res. 277, 108405. https://doi.org/10.1016/j.fcr.2021.108405

Zhang, H., Turner, N.C., Poole, M.L., Asseng, S., 2007. High ear number is key to achieving high wheat yields in the high-rainfall zone of south-western Australia. Aust. J. Agric. Res. 58, 21–27. https://doi.org/10.1071/AR05170

Zhang, Y., Guanter, L., Joiner, J., Song, L., Guan, K., 2018. Spatially-explicit monitoring of crop photosynthetic capacity through the use of space-based chlorophyll fluorescence data. Remote Sens. Environ. 210, 362–374. https://doi.org/10.1016/j.rse.2018.03.031

Zhang, Y., Ma, J., Liang, S., Li, X., Li, M., 2020. An Evaluation of Eight Machine Learning Regression Algorithms for Forest Aboveground Biomass Estimation from Multiple Satellite Data Products. Remote Sens. 12, 4015. https://doi.org/10.3390/rs12244015

Zhao, C., Liu, L., Wang, J., Huang, W., Song, X., Li, C., 2005. Predicting grain protein content of winter wheat using remote sensing data based on nitrogen status and water stress. Int. J. Appl. Earth Obs. Geoinformation 7, 1–9. https://doi.org/10.1016/j.jag.2004.10.002

Zhao, H., Song, X., Yang, G., Li, Z., Zhang, D., Feng, H., 2019. Monitoring of Nitrogen and Grain Protein Content in Winter Wheat Based on Sentinel-2A Data. Remote Sens. 11, 1724. https://doi.org/10.3390/rs11141724

Zhao, K., Tao, Y., Liu, M., Yang, D., Zhu, M., Ding, J., Zhu, X., Guo, W., Zhou, G., Li, C., 2022. Does temporary heat stress or low temperature stress similarly affect yield, starch, and protein of winter wheat grain during grain filling? J. Cereal Sci. 103, 103408. https://doi.org/10.1016/j.jcs.2021.103408

Zhao, Y., Potgieter, A.B., Zhang, M., Wu, B., Hammer, G.L., 2020. Predicting Wheat Yield at the Field Scale by Combining High-Resolution Sentinel-2 Satellite Imagery and Crop Modelling. Remote Sens. 12, 1024. https://doi.org/10.3390/rs12061024

Zhou, X., Huang, W., Kong, W., Ye, H., Dong, Y., Casa, R., 2017. Assessment of leaf carotenoids content with a new carotenoid index: Development and validation on experimental and model data. Int. J. Appl. Earth Obs. Geoinformation 57, 24–35. https://doi.org/10.1016/j.jag.2016.12.005

Zhou, X., Huang, W., Zhang, J., Kong, W., Casa, R., Huang, Y., 2019. A novel combined spectral index for estimating the ratio of carotenoid to chlorophyll content to monitor crop physiological and phenological status. Int. J. Appl. Earth Obs. Geoinformation 76, 128–142. https://doi.org/10.1016/j.jag.2018.10.012

Zhou, X., Kono, Y., Win, A., Matsui, T., Tanaka, T.S.T., 2021. Predicting within-field variability in grain yield and protein content of winter wheat using UAV-based multispectral imagery and machine learning approaches. Plant Prod. Sci. 24, 137–151. https://doi.org/10.1080/1343943X.2020.1819165

Zörb, C., Ludewig, U., Hawkesford, M.J., 2018. Perspective on Wheat Yield and Quality with Reduced Nitrogen Supply. Trends Plant Sci. 23, 1029–1037. https://doi.org/10.1016/j.tplants.2018.08.012

Appendix A

Table A1. Vegetation indices (VI) calculated from hyperspectral reflectance captured over commercial wheat fields and assessed for multicollinearity with inverted traits and contribution to GPC estimation.

Name	Formula	Reference
Anthocyanin Reflectance Index 1	$ARI1 = 1/R_{550} - 1/R_{700}$	Gitelson *et al*. (2001)
Blue Index	$B = R_{450}/R_{490}$	-
BF1	$BF1 = R_{400}/R_{410}$	-
BF2	$BF2 = R_{400}/R_{420}$	-
BF3	$BF3 = R_{400}/R_{430}$	-
BF4	$BF4 = R_{400}/R_{440}$	-
BF5	$BF5 = R_{400}/R_{450}$	-
Blue/green indices1	$BGI1 = R_{400}/R_{550}$	Zarco-Tejada *et al*. (2012, 2005)
Blue/green indices2	$BGI2 = R_{450}/R_{550}$	Zarco-Tejada *et al*. (2012, 2005)
Blue/red indices1	$BRI1 = R_{400}/R_{690}$	-
Blue/red indices2	$BRI2 = R_{450}/R_{690}$	-
Carotenoid Index	$CAR = R_{515}/R_{570}$	Hernández-Clemente *et al*. (2012)
Canopy Chlorophyll Index	$CCI = R_{720}/R_{700}$	Sims *et al*. (2006)
Chlorophyll Index	$CI = R_{750}/R_{710}$	Zarco-Tejada *et al*. (2001)
Chlorophyll Index - green	$CI_g = R_{840}/R_{560} - 1$	Gitelson *et al*. (2003)
Chlorophyll Index - red edge	$CI_{re} = R_{750}/R_{700} - 1$	Gitelson *et al*. (2003)
Carotenoid Reflectance Indices 500	$CRI_{550} = 1/R_{510} - 1/R_{550}$	Gitelson *et al*. (2006, 2003)
Carotenoid Reflectance Indices 550, 515	$CRI_{550,\,515} = 1/R_{515} - 1/R_{550}$	Gitelson *et al*. (2006, 2003)
Carotenoid Reflectance Indices 700	$CRI_{700} = 1/R_{510} - 1/R_{700}$	Gitelson *et al*. (2006, 2003)
Carotenoid Reflectance Indices 700, 515	$CRI_{700,\,515} = 1/R_{515} - 1/R_{700}$	Gitelson *et al*. (2006, 2003)
Carter indices1	$Ctr1 = R_{695}/R_{420}$	Carter (1994)
Reflectance Curvature Index	$CUR = (R_{675} \cdot R_{690}) / R_{683}^2$	Zarco-Tejada *et al*. (2000)
Chlorophyll Vegetation Index	$CVI = (R_{840} \cdot R_{670}) / R_{550}^2$	Vincini *et al*. (2008)

Reflectance band ratio indices	DCabCxc = R672 / (R_{550} · 3 · R_{708})	Datt (1998)
Double-peak Canopy Nitrogen Index	DCNI = [(R_{720} − R_{700}) / (R_{700} − R_{670})] / (R_{720} − R_{670} + 0.03)	Chen *et al*. (2010)
Enhanced Vegetation Index	EVI = 2.5 · (R_{800} − R_{670}) / (R_{800} + 6 · R_{670} − 7.5 · R_{450}) + 1	Liu & Huete (1995)
Enhanced VI (NIR)	EVI_{NIR} = 2.5 · (R_{800} − R_{670}) / (R_{800} + 6 · R_{670} − 7.5 · R_{800}) + 1	Longmire *et al*. (2022)
Greenness Index	G = R_{570}/R_{670}	Zarco-Tejada *et al*. (2001)
Green Leaf Index	GLI = ((2 · (R_{550})) − R_{670} − R_{450}) / ((2 · (R_{550})) + R_{670} + R_{450}	Louhaichi *et al*. (2001)
Gitelson & Merzlyak indices1	GM1 = R_{750}/R_{550}	Gitelson and Merzlyak (1996)
Gitelson & Merzlyak indices2	GM2 = R_{750}/R_{700}	Gitelson and Merzlyak (1996)
Green Normalized Difference Vegetation Index	gNDVI = (R_{840} − R_{550}) / (R_{840} + R_{550})	Gitelson *et al*. (1994a)
Healthy-index	HI = ((R_{534} − R_{698}) / (R_{534} + R_{698})) − 0.5 · R_{704}	Mahlein *et al*. (2013)
Lichtenthaler Index 2	LIC2 = R_{440}/R_{690}	Lichtenthaler (1996)
Lichtenthaler Index 3	LIC3 = R_{440}/R_{740}	Lichtenthaler (1996)
Modified Chlorophyll Absorbance Index	MCARI = (R_{700} − R_{670}) − 0.2 · (R_{700} − R_{550}) · (R_{700}/R_{670})	Haboudane *et al*. (2002)
MCARI/MTVI2 ratio	MCARI / MTVI2 = ((R_{700} − R_{670}) − 0.2 · (R_{700} − R_{550}) · (R_{700}/R_{670})) / (1.5 · (1.2 · (R_{800} − R_{550}) − 2.5 · (R_{670} − R_{550})) / √(((2 · R800 + 1)^2) − (6 · R_{800} − 5 · √(R_{670})) − 0.5))	Eitel *et al*. (2008)
Modified Chlorophyll Absorbance Index 1	MCARI1 = 1.2(2.5(R_{800} − R_{670}) − 1.3 · (R_{800} − R_{550}))	Haboudane *et al*. (2004)
Modified Chlorophyll Absorbance Index 2	MCARI2 = 1.5 · (2.5 · (R_{800} − R_{670}) − 1.3 · (R_{800} − R_{550})) / √(((2 · R_{800} + 1) · (2 · R_{800} + 1)) − (6 · R_{800} − 5 · √(R_{670})) − 0.5)	Haboudane *et al*. (2004)
Modified Soil Adjusted Vegetation Index	MSAVI = (1 + *L*) · ((R_{800} − R_{670}) / (R_{800} + R_{670} + *L*)	Qi *et al*. (1994)
MERIS total chlorophyll index	MTCI = (R_{754} – R_{709}) / (R_{709} – R_{681})	Dash and Curran (2004)
Modified Triangular Vegetation Index 1	MTVI1 = 1.2 · (1.2 · (R_{800} − R_{550}) − 2.5 · (R_{670} − R_{550}))	Haboudane *et al*. (2004)
Modified Triangular Vegetation Index 2	MTVI2 = 1.5 · (1.2 · (R_{800} − R_{550}) − 2.5 · (R_{670} − R_{550})) / √(((2 · R_{800} + 1)^2) − (6 · R_{800} − 5 · √(R_{670})) − 0.5)	Haboudane *et al*. (2004)
Normalized Difference Index (red edge)	NDREI = (R_{674} – R_{712}) / (R_{674} + R_{712})	Delegido *et al*. (2013)
Normalised Difference Vegetation Index	NDVI = (R_{800} − R_{670}) / (R_{800} + R_{670})	Rouse *et al*. (1974)
Normalized difference vegetation index green-blue	$NDVI_{gb}$ = (R_{573} − R_{440}) / (R_{573} + R_{440})	Hansen and Schjoerring (2003)
Normalized green-red difference index	NGRDI = (R_{550} − R_{670}) / (R_{550} + R_{670})	Tucker (1979)
Normalized Pigments Index	NPCI = (R_{680} − R_{430}) / (R_{680} + R_{430})	Peñuelas *et al*. (1994)

Normalized Phaeophytinization Index	$NPQI = (R_{415} - R_{435}) / (R_{415} + R_{435})$	Barnes *et al*. (1992)
Optimized Soil Adjusted Vegetation Index	$OSAVI = ((1 + 0.16) \cdot (R_{840} - R_{670})) / (R_{840} + R_{670} + 0.16)$	Rondeaux *et al*. (1996)
Photochemical Reflectance Index	$PRI = (R_{550} - R_{531}) / (R_{550} + R_{531})$	Gamon *et al*. (1992)
Photochemical Refl. Index (515)	$PRI_{515} = (R_{515} - R_{531}) / (R_{515} + R_{531})$	Sukhova *et al*. (2022)
Photochemical Refl. Index (570)	$PRI_{570} = (R_{570} - R_{531}) / (R_{570} + R_{531})$	Sukhova *et al*. (2022)
Carotenoid/Chlorophyll Ratio Index	$PRI \cdot CI = (R_{570} - R_{530}) / (R_{570} + R_{530}) \cdot ((R_{760}/R_{700}) - 1)$	Garrity *et al*. (2011)
Photochemical Refl. Index (512)	$PRI_{m1} = (R_{512} - R_{531}) / (R_{512} + R_{531})$	Hernández-Clemente *et al*. (2011)
Photochemical Refl. Index (600)	$PRI_{m2} = (R_{600} - R_{531}) / (R_{600} + R_{531})$	Hernández-Clemente *et al*. (2011)
Photochemical Refl. Index (670)	$PRI_{m3} = (R_{670} - R_{531}) / (R_{670} + R_{531})$	Hernández-Clemente *et al*. (2011)
Photochemical Refl. Index (670 and 570)	$PRI_{m4} = (R_{570} - R_{531} - R_{670}) / (R_{570} + R_{531} + R_{670})$	Hernández-Clemente *et al*. (2011)
Normalized Photoch. Refl. Index	$PRIn = PRI / (RDVI\ R_{700}/R_{670})$	Zarco-Tejada *et al*. (2013c)
Pigment Specific Normalized Difference	$PSNDc = (R_{800} - R_{470}) / (R_{800} + R_{470})$	Blackburn (1998)
Plant Senescence Reflectance Index	$PSRI = (R_{648} - R_{858}) / R_{555}$	Ren *et al*. (2017)
Pigment Specific Simple Ratio a	$PSSRa = R_{800}/R_{675}$	Blackburn (1998)
Pigment Specific Simple Ratio b	$PSSRb = R_{800}/R_{650}$	Blackburn (1998)
Pigment Specific Simple Ratio c	$PSSRc = R_{800}/R_{500}$	Blackburn (1998)
Plant Water Index	$PWI = R_{970}/R_{900}$	Peñuelas *et al*. (1993)
R_{920}/R_{729}	R_{920}/R_{729}	-
R_{950}/R_{770}	R_{950}/R_{770}	-
Ratio Analysis of Reflectance Spectra	$RARS = R_{746}/R_{513}$	Chappelle *et al*. (1992)
Renormalised Difference Vegetation Index	$RDVI = (R_{840} - R_{670}) / \sqrt{(R_{840} + R_{670})}$	Roujean & Breon (1995)
Relative greenness index	$RGI = R_{690}/R_{550}$	Ceccato *et al*. (2001)
RNIR Carotenoid Reflectance Indices 510, 550	$RNIR/CRI_{510,\ 550} = (1/R_{510}) - (1/R_{700}) \cdot R_{770}$	Gitelson *et al*. (2006, 2003)
RNIR Carotenoid Reflectance Indices 510, 700	$RNIR/CRI_{510,\ 700} = (1/R_{510}) - (1/R_{550}) \cdot R_{770}$	Gitelson *et al*. (2006, 2003)
Ratio vegetation index (= simple ratio)	$RVI = R_{840}/R_{670}$	Jordan (1969)
Ratio Vegetation Index I	$RVI1 = R_{810}/R_{660}$	Zhu *et al*. (2008)
Ratio Vegetation Index II	$RVI2 = R_{810}/R_{560}$	Xue *et al*. (2004)

Soil Adjusted Vegetation Index	$SAVI = ((1.5) \cdot (R_{840} - R_{670})) / (R_{840} + R_{670} + 0.5)$	Huete (1988)
Structure-Intensive Pigment Index	$SIPI = (R_{800} - R_{445}) / (R_{800} + R_{680})$	Peñuelas *et al*. (1995)
Simple Ratio Pigment Index	$SRPI = R_{430}/R_{680}$	Peñuelas *et al*. (1995)
Transformed Chlorophyll Absorption in Reflectance Index	$TCARI = 3 \cdot [(R_{700} - R_{670}) - 0.2 \cdot (R_{700} - R_{550}) \cdot (R_{700} / R_{670})]$	Haboudane *et al*. (2002)
TCARI/OSAVI	TCARI/OSAVI	Haboudane *et al*. (2002)
Triangular Chlorophyll Index	$TCI = (1.2 \cdot (R_{700} - R_{550})) - (1.5 - (R_{670} - R_{550})) \cdot (\sqrt{(R_{700} / R_{670})})$	Broge and Leblanc (2001)
Triangular Greenness Index	$TGI = -0.5 \cdot (190 \cdot (R_{670} - R_{550})) - 120 \cdot (R_{670} - R_{480})$	Hunt *et al*. (2011)
Triangular Vegetation Index	$TVI = 0.5 \cdot (120 \cdot (R_{750} - R_{550}) - 200 \cdot (R_{670} - R_{550}))$	Broge & Leblanc (2001)
Visible atmospherically resistant index	$VARI = (R_{550} - R_{670}) / (R_{550} + R_{670} - R_{450})$	Gitelson *et al*. (2002)
Vogelmann1	$VOG1 = R_{740}/R_{720}$	Vogelmann *et al*. (1993)
Vogelmann2	$VOG2 = (R_{734} - R_{747}) / (R_{715} + R_{726})$	Vogelmann *et al*. (1993)
Vogelmann3	$VOG3 = (R_{734} - R_{747}) / (R_{715} + R_{720})$	Vogelmann *et al*. (1993)
Zarco-Tejada & Miller Index	$ZMI = R_{750}/R_{710}$	Zarco-Tejada *et al*. (2001)

Table A2. Vegetation indices (VI) calculated from Sentinel-2 reflectance captured over commercial wheat fields and assessed for multicollinearity with inverted traits and contribution to GPC estimation. Solar-induced fluorescence (SIF) and crop water stress index (CWSI) are from airborne hyperspectral and thermal imagery.

Name	Formula	Reference
Adjusted Transformed Soil-Adjusted VI	$ATSAVI = a \cdot (R_{800} - a \cdot R_{670} - b / a \cdot R_{800} + R_{670} - a \cdot b + X(1 + a^2))$	Baret and Guyot (1991)
Aerosol Free VI_{1600}	$AFRI_{1510} = R_{800} - 0.66 \cdot (R_{1600}/R_{800} + 0.66 \cdot R_{1600})$	Karnieli *et al*. (2001)
Aerosol Free VI_{2100}	$AFRI_{2100} = R_{800} - 0.66 \cdot (R_{2100}/R_{800} + 0.66 \cdot R_{2100})$	Karnieli *et al*. (2001)
Atmospherically Resistant VI	$ARVI = R_{800} - R_{670} - y \cdot (R_{670} - R_{450}) / R_{800} + R_{670} - y \cdot (R_{670} - R_{450})$	Bannari *et al*. (1995)
Chlorophyll Index	$CI = R_{750}/R_{710}$	Zarco-Tejada *et al*. (2001)
Chlorophyll Index (green)	$CI_g = R_{783}/R_{560} - 1$	Gitelson *et al*. (2003)
Chlorophyll Index (red edge)	$CI_{re} = R_{783}/R_{705} - 1$	Gitelson *et al*. (2003)
Difference VI	$DVI = g \cdot R_{800} - R_{670}$	Richardson and Wiegand (1991)
Global Environment Monitoring Index	$GEMI = n \cdot (1 - 0.25n) \cdot (R_{670} - 0.125 / 1 - R_{670})$	Bannari *et al*. (1995)
Double-peak Canopy Nitrogen Index	$DCNI = [(R_{720} - R_{700}) / (R_{700} - R_{670})] / (R_{720} - R_{670} + 0.03)$	Chen *et al*. (2010)
Enhanced VI	$EVI = 2.5 \cdot (R_{800} - R_{670}) / (R_{800} + 6 \cdot R_{670} - 7.5 \cdot R_{450}) + 1$	Liu & Huete (1995)
Enhanced VI (NIR)	$EVI_{NIR} = 2.5 \cdot (R_{800} - R_{670}) / (R_{800} + 6 \cdot R_{670} - 7.5 \cdot R_{800}) + 1$	Longmire *et al*. (2022)
Green Normalized Difference VI	$gNDVI = (R_{800} - R_{550}) / (R_{800} + R_{550})$	Gitelson *et al*. (1996)
Inverted Red-Edge Chlorophyll Index	$IRECI = (R_{783} - R_{665}) / (R_{705} + R_{740})$	Frampton *et al*. (2013)
Maccioni Index	$Macc = (R_{780} - R_{710}) / (R_{780} - R_{680})$	Maccioni *et al*. (2001)
Modified Chlorophyll Abs. Ratio Index	$MCARI = ((R_{700} - R_{670}) - 0.2 \cdot (R_{700} - R_{550})) \cdot (R_{700}/R_{670})$	Haboudane *et al*. (2004)
Modified Chlorophyll Abs. Ratio $Index_{1510}$	$MCARI_{1510} = ((R_{700} - R_{1510}) - 0.2 \cdot (R_{700} - R_{550})) \cdot (R_{700}/R_{1510})$	Herrmann *et al*. (2010)
MCARI/OSAVI $_{705, 750}$	R_{705}, R_{750} for red, NIR bands of MCARI/OSAVI	Wu *et al*. (2008)
Modified Simple Ratio	$MSR = R_{800}/R_{670} - 1 / (R_{800}/R_{670})^2 + 1$	Chen *et al*. (1996)
Modified Soil Adjusted VI	$MSAVI = 2 \cdot R_{800} + 1 - \sqrt{(2 \cdot R_{800}+1)^2 - 8 \cdot (R_{800} - R_{670})} / 2$	Qi *et al*. (1994)
MERIS Terrestrial Chlorophyll Index	$MTCI = (R_{754}/R_{709}) / (R_{709} - R_{681})$	Dash and Curran (2004)
Normalised Difference Index	$NDI = (R_{706} - R_{664}) / (R_{706} + R_{664})$	Delegido *et al*. (2011)
Normalised Difference VI	$NDVI = (R_{800} - R_{670}) / (R_{800} + R_{670})$	Rouse *et al*. (1974)
Normalised Difference RE 1	$NDRE1 = (R_{740} - R_{705}) / (R_{740} + R_{705})$	Barnes *et al*. (2000)

Normalised Difference RE 2	$NDRE2 = (R_{783} - R_{705}) / (R_{783} + R_{705})$	Sims and Gamon (2002)
Normalised Difference Water Index	$NDWI_{1510} = (R_{800} - R_{1510}) / (R_{800} + R_{1510})$	Gao (1996)
Optimized Soil Adjusted VI	$OSAVI = (1 + 0.16) \cdot (R_{800} - R_{670} / R_{800} + R_{670} + L)$	Rondeaux *et al*. (1996)
Optimized Soil Adjusted VI_{1510}	$OSAVI_{1510} = (1 + 0.16) \cdot (R_{800} - R_{1510} / R_{800} + R_{1510} + L)$	Herrmann *et al*. (2010)
Pigment Specific Simple Ratio Chl_a	$PSSR_a = R_{800}/R_{680}$	Blackburn (1998)
Ratio VI - Simple Ratio 800/670	$RVI = R_{800}/R_{670}$	Pearson and Miller (1972)
Red, Green Ratio Index	$IRG = R_{670} - R_{550}$	Gamon and Surfus (1999)
Renormalised Difference VI	$RDVI = R_{800} - R_{670} / \sqrt{R_{800} + R_{670}}$	Roujean and Breon (1995)
Sentinel-2 Red-Edge Position	$S2REP = 705 + 35 \cdot (R_{783} + R_{665}) / 2 - R_{705} / (R_{740} - R_{705})$	Frampton *et al*. (2013)
Transformed Chlorophyll Abs. in Reflectance Index	$TCARI = 3 \cdot [(R_{700} - R_{670}) - 0.2 \cdot (R_{700} - R_{550}) \cdot (R_{700}/R_{670})]$	Haboudane *et al*. (2002)
Transformed Chlorophyll Abs. in Reflectance $Index_{1610}$	$TCARI_{1510} = 3 \cdot [(R_{700} - R_{1510}) - 0.2 \cdot (R_{700} - R_{550}) \cdot (R_{700}/R_{1510})]$	Herrmann *et al*. (2010)
TCARI/OSAVI	TCARI/OSAVI	Haboudane *et al*. (2002)
$TCARI/OSAVI_{705, 750}$	R_{705}, R_{750} for red, NIR bands of TCARI/OSAVI	Wu *et al*. (2008)
Plant Water Index	$PWI = R_{970}/R_{900}$	Peñuelas *et al*. (1993)
Crop Water Stress Index	$CWSI = \frac{(Tc-Ta)-(Tc-Ta)_{LL}}{(Tc-Ta)_{UL}-(Tc-Ta)_{LL}}$ where: $(Tc - Ta)_{LL} = -3.25 \cdot VPD + 3.38$, T_c = canopy temperature, T_a = air temperature, VPD = vapour pressure deficit	Idso (1982)
Solar-induced fluorescence ($mW/m^2/nm/sr$)	$SIF_{FLD2} = d\text{-}R \cdot b$, where: $R = (c\text{-}d)/(a\text{-}b)$, $a = E_{750}$, $b = E_{762}$, $c = L_{750}$ and $d = L_{762}$	Plascyk and Gabriel (1975)

Appendix B

Longmire, A.R., Poblete, T., Hunt, J.R., Chen, D., Zarco-Tejada, P.J., 2022. Assessment of crop traits retrieved from airborne hyperspectral and thermal remote sensing imagery to predict wheat grain protein content. *ISPRS Journal of Photogrammetry and Remote Sensing 193, 284–298*. https://doi.org/10.1016/j.isprsjprs.2022.09.015

—

See overleaf.

ISPRS Journal of Photogrammetry and Remote Sensing 193 (2022) 284–298

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Assessment of crop traits retrieved from airborne hyperspectral and thermal remote sensing imagery to predict wheat grain protein content

A.R. Longmire [a], T. Poblete [b], J.R. Hunt [a,c], D. Chen [a], P.J. Zarco-Tejada [a,b,d,*]

[a] School of Agriculture and Food, Faculty of Veterinary and Agricultural Sciences (FVAS), University of Melbourne, Melbourne 3010, Victoria, Australia
[b] Department of Infrastructure Engineering, Faculty of Engineering and Information Technology (FEIT), University of Melbourne, Melbourne 3010, Victoria, Australia
[c] Department of Animal, Plant and Soil Sciences, AgriBio Centre for AgriBiosciences, La Trobe University, Bundoora 3086, Victoria, Australia
[d] Instituto de Agricultura Sostenible (IAS), Consejo Superior de Investigaciones Científicas (CSIC), Avenida Menéndez Pidal s/n, 14004 Córdoba, Spain

ARTICLE INFO

Keywords:
Hyperspectral
Radiative transfer
Stress detection
Grain protein content
Machine learning
Thermal

ABSTRACT

Wheat (*Triticum* spp.) is among the world's most widely grown crops and receives large quantities of nitrogen (N) fertiliser. Grain protein content (GPC) is influenced by genetic, agronomic and weather conditions affecting crops' physiological status and stress levels; accurate GPC prediction has potential to reduce N losses and improve profit. Success in GPC estimation from remotely sensed plant traits has been limited. For progress to be made, it is necessary to robustly identify imaging spectroscopy-based physiological traits most closely associated with GPC in both experimental and commercial contexts. We present results from piloted hyperspectral flights and ground campaigns at two dryland field experiments with divergent water supply and wide-ranging N treatments, and from two years' flights over 17 commercial fields planted to either bread (*T. aestivum*) or durum (*T. durum*) wheat, in the southern Australian wheat belt. Imagery was acquired with airborne hyperspectral and thermal sensors, with spatial resolutions of approx. 0.3 m and 0.5 m for experimental plots and 1 m/1.7 m in commercial fields. Leaf clip measurements, leaf and grain samples were collected and, in commercial fields, ~40,000 records obtained from harvester-mounted protein monitors. Crop Water Stress Index (CWSI), solar-induced fluorescence (SIF), reflectance indices and PRO4SAIL radiative transfer model inverted parameters were retrieved for each plot and GPC record location. The photochemical reflectance index (PRI) related to xanthophyll pigments was consistently associated with GPC at both leaf and canopy scale in the plots and transect. In the commercial crops, a gradient boosted machine learning algorithm (GBM) ranked CWSI as the strongest indicator of GPC under severe water stress, while SIF, PRI and inverted biochemical constituents anthocyanins and carotenoids were consistently important under more benign conditions. Structural parameters inverted from the hyperspectral reflectance imagery were not prominent except under severe drought. We attained statistically significant results estimating GPC in unseen samples, with best relationships between predicted and observed GPC of $r^2 = 0.80$ in a model built with thermal and physiological traits obtained from the hyperspectral and thermal imagery.

1. Introduction

Combined, bread (*Triticum aestivum* L.) and durum wheat (*T. durum* Desf.) provide >20% of humans' carbohydrate (CHO) and protein needs (Shiferaw et al., 2013). Bread wheat receives ~17% of global nitrogen (N) fertiliser (Heffer and Prud'homme, 2020), but in Australia, a major wheat producer, only ~40% of fertiliser N is assimilated in the year of application (Angus and Grace, 2017). This is a major expense for farmers (Monjardino et al., 2015), and has severe environmental costs (Sutton et al., 2011). Grain protein content (GPC) determines the economic and nutritional value of grain and the rheological properties of flour. While most of the N and protein ultimately translocated to the grain is already in the plant before anthesis (Giuliani et al., 2011; Lopez-Bellido et al., 2004; Masoni et al., 2007), GPC is also influenced by the amount of soil N plants can extract, especially during grain filling (Gooding et al., 2007; Jamieson and Semenov, 2000; Ottman et al., 2000). Post-anthesis, CHO are accumulated while photosynthesis proceeds, but drought or heat may depress assimilation, reducing protein dilution (Gooding et al.,

* Corresponding author at: School of Agriculture and Food, Faculty of Veterinary and Agricultural Sciences (FVAS), University of Melbourne, Melbourne 3010, Victoria, Australia.
E-mail address: pablo.zarco@unimelb.edu.au (P.J. Zarco-Tejada).

https://doi.org/10.1016/j.isprsjprs.2022.09.015
Received 3 February 2022; Received in revised form 3 August 2022; Accepted 23 September 2022
Available online 5 October 2022

2007). Further, early season N oversupply can exacerbate drought stress if vigorous early growth exhausts soil moisture (Angus and Fischer, 1991; van Herwaarden et al., 1998). Pre-harvest knowledge of GPC could allow forecasting for grain segregation or blending and hence access to price premiums (Apan et al., 2006), while GPC monitoring at harvest opens opportunities for analysis (Whelan et al., 2009) so far little explored. Piloted light aircraft can carry multiple sensors to capture thermal and hyperspectral images concurrently. This allows efficient operation over individual fields of up to 300 Ha and up to 6500 Ha in equivalent light conditions on a single day; UAV payloads are insufficient for such large-scale operations particularly when advanced hyperspectral imagers and several instruments are used concurrently.

Vegetation indices (VI) from remote sensing (RS) have often been used to estimate wheat yield. The normalised difference VI (NDVI; (Rouse et al., 1974)), the most-used index in agriculture (Herrmann et al., 2010) correlates with yield, especially at coarse spatial scales (Becker-Reshef et al., 2010; Dempewolf et al., 2014; Lai et al., 2018; Toscano et al., 2019), through its sensitivity to crop cover and leaf area index (LAI). However, NDVI saturates at high biomass, causing models to fail (Baret and Guyot, 1991; Eitel et al., 2008; Zarco-Tejada et al., 2005), is unreliable where other factors intervene between biomass and yield, and lacks transferability. Other VIs have been used in yield estimation, especially those based in the red edge which can account for chlorophyll concentration as a proxy for crop physiological status.

Grain quality assessment, in particular GPC and grain N content (GNC), is less advanced than yield assessment, but hyperspectral RS is promising as it quantifies variables with physiological links to grain quality. GPC estimations have been based on satellite (Feng et al., 2014; Wang et al., 2014; Wright et al., 2004; Zhao et al., 2005) and airborne imagery (Jensen et al., 2007; Prey and Schmidhalter, 2019; Raya-Sereno et al., 2021; Rodrigues et al., 2018), mostly in experimental plots. This typically induces GPC variability but limits the scope of conclusions by omission of commercial settings. Where tested over multiple seasons, estimates of wheat GPC, and the specific predictors most suited to doing so, have lacked consistency, but for both yield and GPC, observations around early grain filling ((Zadoks et al., 1974); development stage Z65—73) appear optimal (Apan et al., 2006; Jensen et al., 2007; Lopez-Bellido et al., 2004; Prey and Schmidhalter, 2019; Raya-Sereno et al., 2021). To the best of our knowledge, the objective of assessing wheat GPC from hyperspectral RS data at commercial scale has been addressed only in a single-year study with ~200 sampling points in one 86 ha field (Rodrigues et al., 2018). We build on previous work (Rodrigues et al. (2018) and make progress focusing on the understanding of specific spectrally derived plant traits such as biochemical constituents and structural traits to evaluate their importance across experimental and commercial fields, over two years and including advanced traits such as thermal indicators of water stress and solar-induced fluorescence.

Empirical VI-based models predominate for estimating biophysical parameters including pigment concentrations, leaf N and structural aspects related to crop condition and GPC from hyperspectral data (Ferwerda et al., 2005; Herrmann et al., 2010). Narrow band hyperspectral reflectance indices (NBHI) have been used successfully in chlorophyll (C_{a+b}) and N estimation in wheat (Li et al., 2014; Wang et al., 2012); some are designed to disentangle leaf biochemistry from structural variables and water status. Many successful VIs incorporate the red edge for its strong correlations with C_{a+b} (Clevers and Gitelson, 2013; Haboudane et al., 2008; Prey and Schmidhalter, 2019). These include the normalised difference red edge index (NDRE; (Gitelson and Merzlyak, 1994a)), used to retrieve C_{a+b} in wheat (Li et al., 2015). Devised to overcome canopy effects, the Transformed Chlorophyll Absorption in Reflectance Index (TCARI; (Haboudane et al., 2002)) normalised by the Optimised Soil-Adjusted Vegetation Index (OSAVI; (Rondeaux et al., 1996)) estimated C_{a+b} (Gonzalez-Dugo et al., 2015) and total N (Klem et al., 2018) in wheat. (Zhao et al., 2020) predicted yield in commercial wheat using two indices focused in the red edge: the chlorophyll index (CI; (Gitelson et al., 2003)), and OSAVI. Raya-Sereno et al. (2021) found that RE indices consistently correlated better than other indices with GPC, and that the NDRE was a relatively stable indicator of GNC over four years. Similarly, Li et al. (2020) obtained the best of moderate results estimating GPC with the double-peak canopy N index (DCNI; (Chen et al., 2010)) and CI. In contrast, in a study estimating GPC from airborne multispectral data, greater importance was attributed to the red and NIR domains, than to the RE (Zhou et al., 2021). Rodrigues et al. (2018) found that combinations of green- and RE bands performed best.

Indices also target anthocyanins (Anth) and carotenoids (C_{x+c}), whose stress responses suggest them as indicators of GPC. Anth are upregulated under water stress (Chalker-Scott, 2002; Naing and Kim, 2021), can be high in senescing leaves (Sims and Gamon, 2002), and post-anthesis drought increases their accumulation in grain (Li et al., 2018). C_{x+c} are produced in response to photooxidative, water and heat stress (Groth et al., 2020; Janeczko et al., 2018). The photochemical reflectance index (PRI; (Gamon et al., 1992)) is sensitive to C_{x+c} and reacts to instantaneous changes in water stress and photosynthetic rate (Feng et al., 2017; Magney et al., 2016). PRI is a pre-visual indicator of water stress and recovery and is proposed as an alternative to thermal RS for water stress detection (Kohzuma et al., 2021; Suárez et al., 2008). PRI has been used to improve yield estimates and discriminate water and disease stress in wheat (Feng et al., 2017; Magney et al., 2014). Alongside PRI, the carotenoid index CAR (Zarco-Tejada et al., 2013b) has shown moderately strong relationships with yield across rainfed and irrigated bread and durum wheat (Gonzalez-Dugo et al., 2015). Despite the links between stress and GPC and use of CAR and PRI in stress detection, neither index seems to have been tested as a predictor of GPC in wheat.

As an alternative to VIs, whose empirical relationships with crop performance measures may vary across seasons and locations, parameters retrieved by radiative transfer models (RTM) are more robust across location, phenological stage and crop type (Clevers and Kooistra, 2012; Dorigo et al., 2007; Jacquemoud et al., 1995). By linking leaf and canopy models, biochemical and structural parameters at each level can be estimated concurrently. Quantities directly associated with yield and/or GPC, including C_{a+b} and LAI, and others that change more dynamically with stress such as Anth and C_{x+c} have been quantified by RTM inversion in wheat under water/N stress (Botha et al., 2010; Camino et al., 2018). Moreover, retrieval accuracies for PROSPECT and SAIL have been shown since early in RTM development (Bacour et al., 2002; Féret et al., 2008; Jacquemoud et al., 2009, 1995; Li et al., 2015; Ustin et al., 2009).

Solar-induced fluorescence (SIF) is a proxy for the functional status of vegetation, instantaneous photosynthetic rate and assimilation (Genty et al., 1989; Meroni et al., 2009; Mohammed et al., 2019). SIF shows diurnal and seasonal variations in photosynthetic rate and is key to stress diagnosis (Poblete et al., 2020; Zarco-Tejada et al., 2018, 2016, 2013a). In bread and durum wheat, SIF has been combined with C_{a+b} and leaf structural traits to diagnose N deficiency and water stress (Camino et al., 2018). SIF variability should therefore parallel relative CHO availability during grain filling in wheat. While C_{x+c}, SIF, and PRI are sensitive to short-term changes induced by stress (Gamon et al., 1992; Peñuelas et al., 1994), C_{a+b} and structural measures related to LAI reflect more temporally stable canopy traits with influence on GPC.

Reduced stomatal conductance lowers leaf cooling and assimilation, making thermal data useful for tracking the effects of water stress (Gonzalez-Dugo et al., 2015; Grant et al., 2007; Idso, 1982). High temperatures and drought during grain filling correlate with higher GPC because plants' capacity to dilute protein with newly assimilated CHO is lower (Daniel and Triboï, 2002; Gooding et al., 2003); these also shorten grain filling by accelerating phenology and starch insolubilisation (Daniel and Triboï, 2002). Water stress indicators have consistently improved estimates of wheat yield (e.g. Gonzalez-Dugo et al., 2015; Zhao et al., 2020).

Gradient boosting/boosted machines (GBM) are a machine learning algorithm based on work by Friedman (2001). With good predictive skill, robustness to multicollinearity and the ability to deal with large

numbers of both input features and observations in a computationally efficient way (Ruan et al., 2022), GBMs assess input features' importance in terms of the gain they bring to a model (Grinberg et al., 2020). GBM use with RS data has included estimation of leaf N (Yang et al., 2021) and yield (Cao et al., 2020; Ruan et al., 2022) in wheat, but GPC estimation to date has been based solely on genotype data (Grinberg et al., 2020). To advance GPC estimation from RS, the variables most closely associated with it should be identified and their stability across cropping situations assessed; tree-based algorithms such as GBM and RF have the advantages both of offering feature importance assessment, the primary focus of our work, and of being among the most accurate ML algorithms (Abdi, 2020; Ruan et al., 2022; Zhang et al., 2020).

2. Materials and methods

2.1. Study sites

We consider experimental plots and commercial crops, grown under rainfed conditions at three locations in the southern Australian wheat belt with a variety of climatic regimes and soils. The plot trials were located at Birchip (site 1; 35.969°S, 142.822°E; altitude 102 m; average annual rainfall (AAR) = 353 mm) and Yarrawonga (site 2; 36.050°S, 145.983°E; altitude 129 m; AAR = 470 mm; Fig. 1) and were planted in randomised complete block designs with hard white bread wheat (cv. Scepter) for N fertiliser treatment trials in 2019 (Fig. 1). Site 1 takes the Köppen-Geiger climate classification Bsk and site 2, Cfa (Peel et al., 2007). Soils are classified as calcarosol and sodosols, respectively, based on the Australian Soil Classification (ASC; Isbell, 2002). Site 1 was sown on 2019-05-16 and N rates were adjusted according to in-season rainfall to target yield modelled with the Yield Prophet® decision support tool (Hochman et al., 2009; Hunt et al., 2006) and applied on 2019-08-10 at Z31. This produced small increments in N rates, some of which were grouped for analysis (Table 1). Site 2 was sown on 2019-05-09 and fertiliser was applied in equal doses at Z23 and Z31. All plots were approx. 12 × 2 m, treatments were replicated four times and agronomic procedures were equivalent.

The commercial crops included hard white bread and durum wheat cultivars in 17 fields across 30 km of latitude around Kaniva (36.37°S, 141.24°E; altitude 142 m; AAR = 451 mm; Fig. 2a). There, the dominant soils are vertosols, but both sodosols and chromosols are also common; the Köppen-Geiger climate classification for Kaniva is Cfb. Seven fields totalling 815 ha of bread and durum wheat were sown between 2019-05-15 and 2019-06-03; similar crops were sown from 2019-05-12–2019–06–05 across 10 fields (1039 ha). Fertiliser was applied 1–3 times each season, usually as urea. Dualex and SpectraPen leaf clip observations and soil samples were taken in a transect across one field, M01 (36.30°S 141.35°E), concurrent with the 2020 flight. Absent differential fertiliser treatments in the crops, K-means clustering (Hartigan and Wong, 1979) by GPC was used to divide the transect data into three levels for analysis.

Table 1
Nitrogen fertiliser applied to experimental wheat plots at Birchip (site 1) and Yarrawonga (site 2). Pre-sowing soil N to 120 cm (site 1) and 100 cm (site 2). Fertiliser applied at sowing, on 2019-08-10 (Z31; site 1) and 2019-06-26 (Z23) and 2019-08-06 (Z31; site 2).

Location	Soil N (mg kg^{-1})	Total fertiliser (kg N/ha)	Treatment (aggregated)	Plots (n)
Site 1 Birchip	46.8	0	B0	4
		30	B1	8
		37		
		98	B2	8
		104		
		162	B3	12
		167		
		171		
Site 2 Yarrawonga	46.1	0	Y1	4
		46	Y2	
		92	Y3	
		138	Y4	
		184	Y5	

2.2. Ground data collection and laboratory processes

Dualex (FORCE-A, Orsay, France) and SpectraPen (PSI, Drasov, Czech Republic) measurements were made on the adaxial surface of

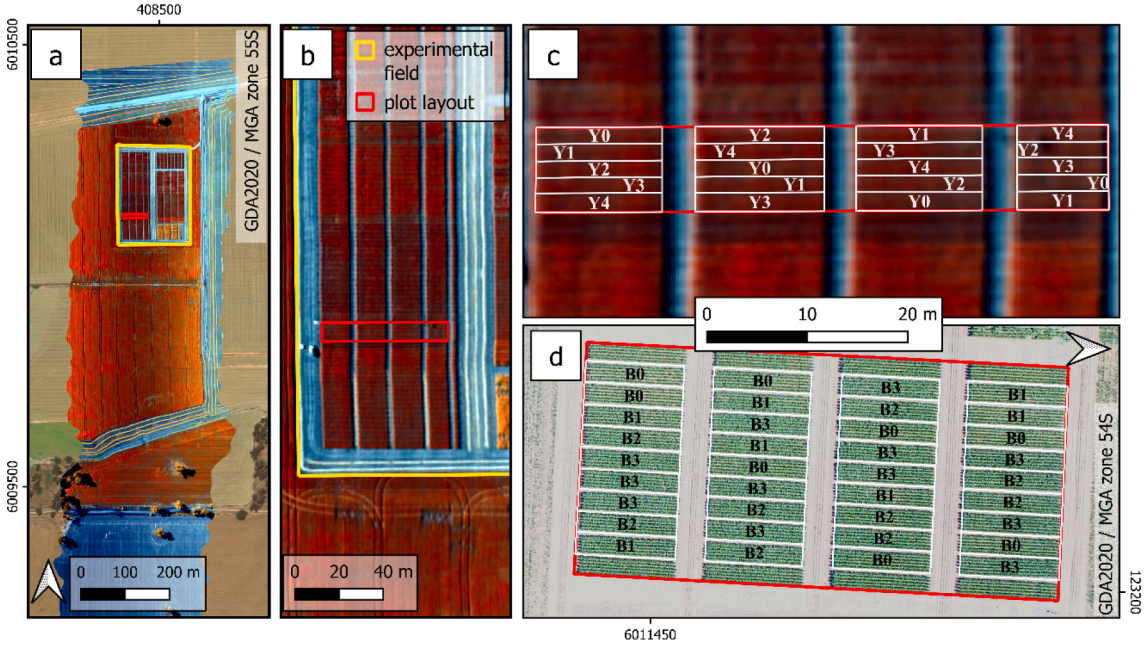

Fig. 1. Yarrawonga (site 2) experimental area, Victoria, Australia (a); site 2 plot layout on false colour hyperspectral image (b); site 2 (c) and site 1 (d) study plots with N fertiliser treatment levels (c). N application rates (kg N/ha): Y0 = 0, Y1 = 46, Y2 = 92, Y3 = 138, Y4 = 184; B0 = 0, B1 = 30 – 37, B2 = 98 – 104, B3 = 162 – 171.

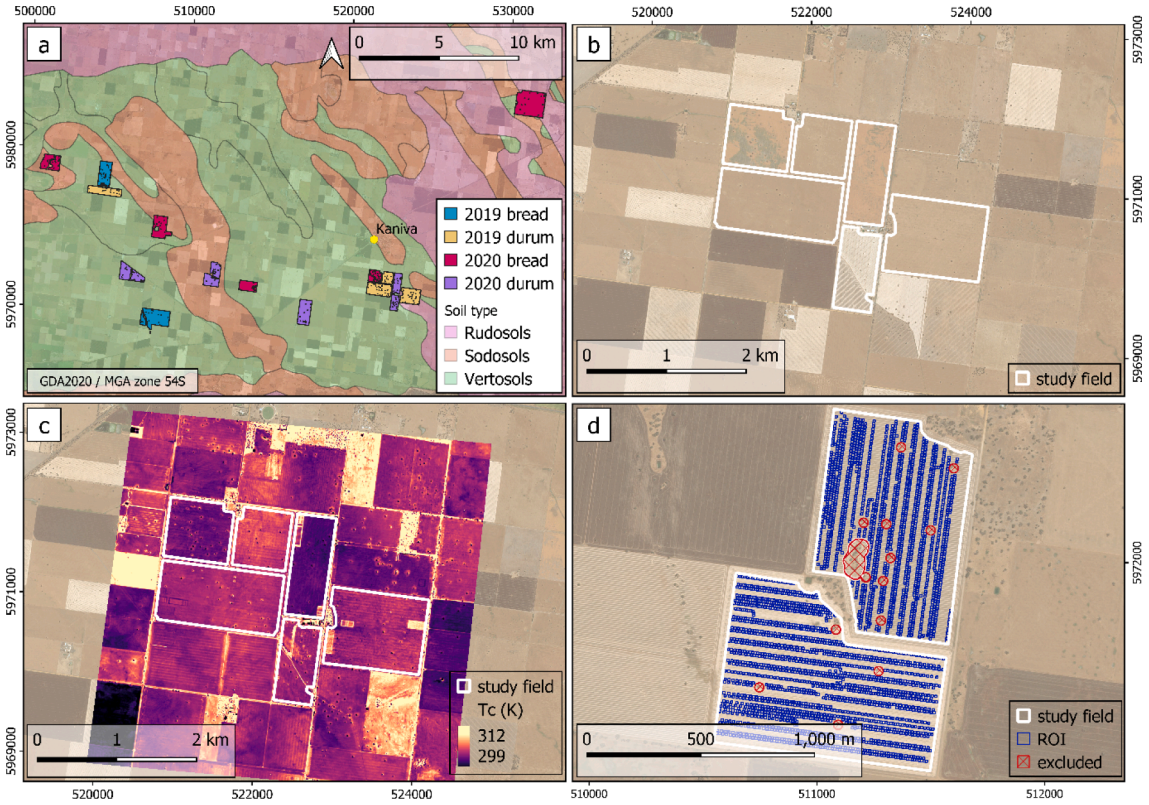

Fig. 2. Wheat fields studied at Kaniva in 2019/2020, with soil types a), false-colour hyperspectral image (R = 775 nm, G = 710 nm, B = 678 nm; b), canopy temperature in Kelvin (c) and fields showing regions of interest (ROI) and areas excluded (d).

upper sun-adapted leaves central in plots. Ten measurements per plot were taken with each instrument, quasi-concurrently with flights and in equivalent light and meteorological conditions. Approximately 80 g of entire flag or other upper, sun-exposed leaves were cut from the central area of plots, sealed in plastic, refrigerated in transit then kept at −20 °C until processing. Ten subsamples were taken from the central area of these cut leaves with leaf tissue punches of known diameter so that areal N content could be calculated; these discs were weighed before and after drying. Leaf tissue was dried at 65 °C for 48 h, then ground to a uniform powder in a ball mill and analysed by Dumas combustion for total N (mass %). Plots were machine harvested and the grain assessed for protein content by near infrared (NIR) spectroscopy (CropScan 3000B Grain Analyser, Next Instruments, Sydney Australia). Ambient temperature, barometric pressure and incoming shortwave and longwave radiation were recorded on the ground using a portable weather station (model WXT510, Vaisala, Helsinki, Finland). During commercial harvesting, NIR spectrometers (CropScan 3000/3300H, Next Instruments), mounted on combine harvesters, collected GPC with location (±0.01 m) from real-time kinetic GPS. Dualex and SpectraPen measurements were also made during a transect across one field at Miram (M01) in 2020. Three soil samples from the top 15 cm were taken with a hand auger and mixed to represent each location, which was marked with a hand-held GPS (Garmin, Olathe, Kansas, USA). Samples were oven-dried at 65 °C for 48 h and mineral N extracted as per Rayment and Lyons (2010) on a Skalar San++ SFA (FlowAccess V 3.2).

2.3. Airborne data collection

Airborne hyperspectral and thermal images were collected by sensors flown on a light aircraft over plots at site 1 on 2019–10-03 (1409° days after sowing (DDAS)), site 2 on 2019-10-09 (1559 DDAS), and commercial fields near Kaniva on 2019-10-22 (bread wheat 1514 DDAS, durum 1736 DDAS) and 2020-10-28 (bread 1592 DDAS, durum 1742 DDAS). Thermal time, and ambient temperature for CWSI calculation, were based on data from the Australian Bureau of Meteorology's (BoM) recording station #78015 (Bureau of Meteorology, 2021). The hyperspectral data were collected in the visible and near infrared (VNIR) domains with a hyperspectral VNIR sensor (VNIR E-Series model; Headwall Photonics, Fitchburg, MA, USA), capturing 371 bands from 400 to 1001 nm at 8 nm per pixel, yielding 7 nm FWHM with a 25 μm slit. At 12-bit radiometric resolution, storage rate was 50 frames per second with exposure time of 18 ms and an 8 mm focal length. Radiometric and spectral calibration was completed in the laboratory prior to flights. Atmospheric correction of radiance was applied with the Simple Model of Atmospheric Radiative Transfer of Sunshine (SMARTS) model (Gueymard, 1995), using aerosol optical depth (AOD) observed at time of flight with a Micro-Tops II sunphotometer (Solar LIGHT Co., Philadelphia, PA, USA). This method has previously been implemented for hyperspectral data (Calderón et al., 2015; Poblete et al., 2020; Zarco-Tejada et al., 2018). Orthorectification was performed using Parametric Geocoding & Orthorectification for Airborne Optical Scanner Data (PARGE; ReSe applications GmbH, Wil, Switzerland) using the IMU and GPS flight data obtained from a VN-300 (VectorNav Technologies LLC, Dallas, TX, USA). Thermal images were collected in the 7.5–14 μm region with an A655c camera (FLIR systems, Wilsonville, Oregon, USA). Over commercial fields, images were acquired at ≈2000 m above ground level (AGL), yielding pixels of 1.0 m (hyperspectral) and 1.7 m (thermal) ground sampling distance (GSD). Images acquired at 350 m and 400 m AGL gave GSD = 0.15 m for hyperspectral at experimental site 1 and GSD = 0.2 m at site 2, respectively. Radiance (L), reflectance (R; Fig. 2b) and thermal (canopy temperature; T_c; Fig. 2c) values were aggregated to pixel mean per plot. Mean L and R spectra by fertiliser treatment level are shown for both plot sites in Fig. 3.

The crop water stress index (CWSI) was calculated according to Idso et al. (1981), normalising canopy temperature (T_c) with air temperature (T_a) and vapour pressure deficit (VPD).

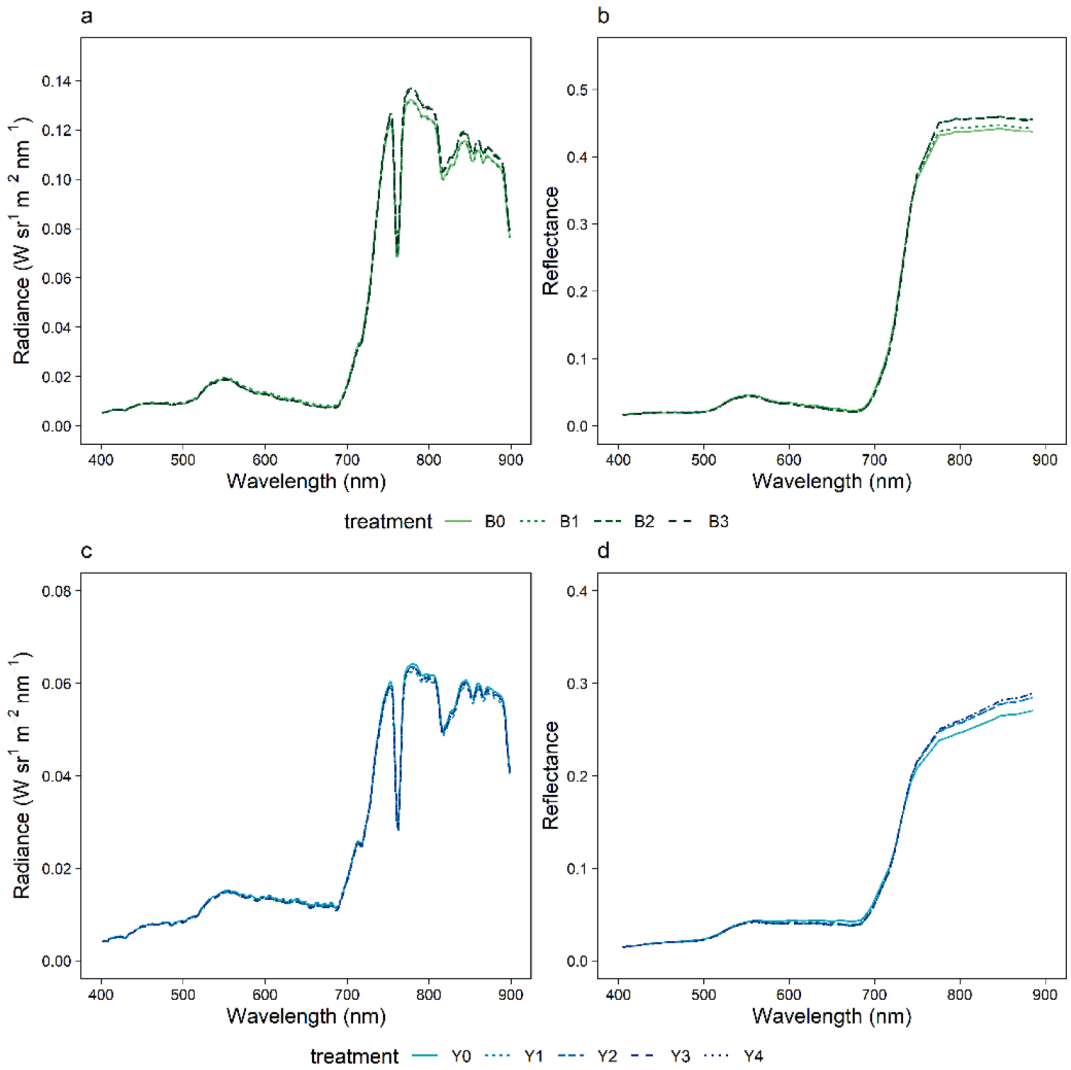

Fig. 3. Mean radiance (W/sr m^{-2} nm^{-1}) and reflectance spectra captured by airborne hyperspectral sensors at Birchip (site 1; a, b) and Yarrawonga (site 2; c, d) by treatment.

$$CWSI = \frac{(Tc - Ta) - (Tc - Ta)_{LL}}{(Tc - Ta)_{UL} - (Tc - Ta)_{LL}} \quad (1)$$

The lower limit $(Tc - Ta)_{LL}$ represents the canopy/air temperature differential in a canopy transpiring at its maximum potential rate for a given VPD while the upper limit (Tc – Ta)UL represents the same for a canopy in which the transpiration flux is zero. The lower limit used here was defined by Idso (1982) and adopted by Gonzalez-Dugo et al. (2015).

$$(Tc - Ta)_{LL} = -3.25 \cdot \mathrm{VPD} + 3.38 \quad (2)$$

2.4. Data processing

After their collection, the datasets were combined at each level's primary study unit: plots, transect waypoints and regions of interest (ROI). Dualex data from the plots and transect were screened for outliers ≥2 standard deviations from the plot or waypoint mean, and for erroneous SpectraPen spectra, then aggregated to mean values and spectra per study unit. Data analysis for commercial fields was based on the geolocated GPC records collected during harvest. A 100 m2 ROI was established around each GPC point by buffering to a 5 m radius, then drawing bounding geometry for each point, and the GPC value adopted for the ROI (Fig. 2d). The 5 m radius was chosen so that ROI width was less than the harvester swath width (12 m) and to increase spatial independence between ROIs. Areas within 20 m of perimeter fences, dams, trees and cloud shadow, and all ROIs intersecting these, were excluded (Fig. 2b–d). A Wilcoxon test (Bauer, 1972) was applied to assess the significance and effect size of differences between GPC observations across years and wheat types. Mean L and R spectra and Tc values were calculated from image pixels contained within each ROI; for L and R, n ≈ 100 pixels, Tc n ≈ 36 pixels. NBHI, inverted leaf and canopy parameters, solar-induced fluorescence (SIF) and CWSI were retrieved from airborne R, L, and Tc, respectively for all plots and ROIs. NBHI were also calculated from SpectraPen R at plot and transect scales. Where transect waypoints intersected with ROIs, airborne GPC, NBHI, inverted parameters, SIF and CWSI values retrieved for the ROI were assigned to waypoints. The NBHI relevant to later procedures are detailed in Table 2. Spatial analysis was done in QGIS (QGIS Development Team, 2020) and R (R Core Team, 2020).

Analogously to Poblete et al. (2021), inverted leaf and canopy parameters were retrieved with the PROSPECT-D (Féret et al., 2017) and 4SAILH (Verhoef et al., 2007) radiative transfer models (RTM), linked as PRO4SAIL. PROSPECT-D was used to retrieve the leaf pigments C_{x+c}, C_{a+b} and Anth, while 4SAILH was used to estimate canopy structural traits LAI and leaf inclination distribution function (LIDFa) from mean plot and ROI spectra. RTM parameters were randomly sampled from uniform distributions in the ranges given in Table 3 to construct a look-up table (LUT) of 200,000 reflectance spectra simulations with their

Table 2
Indices calculated from spectra observed with the SpectraPen leaf clip instrument in plots at Birchip (site 1) and Yarrawonga (site 2) and in a transect of field M01 near Kaniva, and from airborne hyperspectral and thermal imagery captured in flights over the same plots and over commercial wheat fields near Kaniva. The table shows only those indices/calculations relevant to subsequent analyses.

Type	Name	Formula	Reference
Chlorophyll $a+b$	Red edge chlorophyll index	$CI = (R_{750}/R_{700}) - 1$	Gitelson et al. (2003)
	Normalised Pigments Index	$NPCI = (R_{680} - R_{430})/(R_{680} + R_{430})$	Peñuelas et al. (1994)
	Zarco-Tejada and Miller index	$ZMI = R_{750}/R_{710}$	Zarco-Tejada et al. (2001)
	Vogelmann index 1	$VOG1 = R_{740}/R_{720}$	Vogelmann et al. (1993)
Other pigments	Photochemical Refectance Index	$PRI = (R_{531} - R_{570})/(R_{531} + R_{570})$	Gamon et al. (1992)
	Photochemical Refl. Index (670)	$PRI_{m3} = (R_{670} - R_{531})/(R_{670} + R_{531})$	(Hernández-Clemente et al., 2011)
	Photochemical Refl. Index (670 and 570)	$PRI_{m4} = (R_{570} - R_{531} - R_{670})/(R_{570} + R_{531} + R_{670})$	Hernández-Clemente et al. (2011)
Structure	Enhanced VI (NIR)	$EVI_{NIR} = 2.5*(R_{800}\text{-}R_{670})/(R_{800} + 6*R_{670}\text{-}7.5*R_{800})$	This study
	Normalised Difference VI	$NDVI = (R_{840} - R_{670})/(R_{840} + R_{670})$	Rouse et al. (1974)
Structure/ water	R_{920}/R_{729}	R_{920}/R_{729}	L. Suárez (personal communication)
Fluorescence	Solar-induced fluorescence (mW/ m^2/nm/sr)	FLD2 = d-R*b, where R = (c-d)/(a-b), a = E_{750}, b = E_{762}, c = L_{750} and d = L_{762}	Plascyk and Gabriel (1975)
Water status	Crop Water Stress Index	Refer to section 2.3	Idso (1982)

Table 3
Values and ranges of leaf and canopy traits used for PRO4SAIL (PROSPECT-D + 4SAIL) radiative transfer model inversion and look-up table generation.

Parameter	Abbreviation	Value/range
Chlorophyll $a+b$ content [μg/cm^2]	C_{a+b}	3–70
Carotenoid content [μg/cm^2]	C_{x+c}	1–20
Anthocyanin content [μg/cm^2]	Anth	1–10
Dry matter content [g/cm^2]	C_m	0.001–0.035
Water content [g/cm^2]	C_w	0.001–0.035
Mesophyll struct. Coef.	N	0.5–3.0
Leaf area index [m^2/m^2]	LAI	1–5
Leaf Inclination Dist. Func. [°]	$LIDF_a$	0–90
Hot spot parameter	h	0.01
Observer angle [deg.]	tt_o	0
Sun zenith angle [deg.]	tt_s	47.7
Relative azimuth angle [deg.]	ψ	0

associated leaf and canopy values. To compare synthetic and airborne spectra, the LUT was interrogated using support vector machine (SVM) algorithms run in MATLAB (MATLAB; Statistics and Machine Learning, Deep Learning and Parallel Computing toolboxes; Mathworks Inc., Natick, MA, USA), applying a radial basis function. Hyperparameters were optimised for each target variable during training, and inversions took simulated reflectance spectra as SVM inputs and plant/canopy characters as outputs.

Fluorescence was retrieved using bands inside and outside the O_2–A Fraunhofer line (FLD2; Plascyk and Gabriel, 1975) and based on irradiance simulated with the SMARTS software (Gueymard, 1995), then convolved to the FWHM and spectral sampling interval (SSI) of the hyperspectral sensor. The inside band (762 nm) is the local minimum incoming irradiance at the relevant SSI while the outside band (750 nm) is on the shoulder of the O_2–A line.

To evaluate collinearity between inputs and reduce dimensionality, each year's data were assessed by variance inflation factor analysis (VIF; R package fmsb; (Nakazawa, 2022)) with threshold = 5 (Akinwande et al., 2015). Multicollinear features were first excluded among NBHI; those surviving (VIF < 5) then underwent a second VIF analysis with the inverted parameters, CWSI and SIF. Variables thus chosen were kept only if they improved the ML model; these final input features were categorised into three layers: physiological (Anth, C_{a+b}, C_{x+c}, SIF, PRI); structural (EVI_{NIR}; hereafter 'EVI', LAI, LIDFa); and thermal (CWSI).

2.5. Algorithms for GPC assessment

A gradient boosted machine (GBM) machine learning (ML) algorithm was used to estimate, through supervised learning, relationships between leaf and canopy traits and the target variable GPC. Our primary objective was to assess input features' relative importance to GPC estimation (Chen and Guestrin, 2016). Feature importance, and the performance of models containing them, was assessed stepwise: physiological + structural + thermal (Model 1); physiological + structural (Model 2); physiological (Model 2), as proposed by Friedman and Meulman (2003).

The GBM used decision trees as its base learners, sub-models with weak predictive skill which learn from their predecessors' error in estimating the target variable to iteratively improve the estimate (Chen and Guestrin, 2016). Datasets from each site were randomly split 70:30 into training and test sets at each model run and passed to a linear function. Four model hyperparameters were varied across either three or four levels: learning rate, tree depth, minimum node size and stochastic gradient descent (SGD), and this space was searched by full factorial sampling. In SGD, each iteration runs on a randomly sampled subset of rows, introducing noise and improving robustness to overfitting. Randomised K-fold (K = 5) cross-validation (CV) was implemented, also to reduce overfitting. The combination of hyperparameters that minimised root mean square error (RMSE) was adopted as the top model for each training set, then applied to predict GPC in the relevant test set. Data analysis and ML were done in R (R Core Team, 2020) using the packages xgboost (Chen et al., 2021) for gradient boosting and caret (Kuhn, 2020) for model tuning.

3. Results

3.1. Plot experiments

The 2019 growing season rainfall (GSR) at experimental site 1 was 65% of long-term AAR, but with good soil moisture from 2018, while at site 2 GSR was 54% of AAR with low starting soil moisture. A strong gradient in GPC was seen at both sites, parallel with N dosing; GPC saturated under high N and was higher overall at the droughted site 2 (Figs. 4 and 5b).

Several leaf-level indicators were associated with higher GPC along the N treatment gradient at both sites 1 and 2. The NPCI, VOG1, ZMI, R_{920}/R_{729} and PRI_{m3} (Hernández-Clemente et al., 2011) increased with fertiliser N rate (Fig. 5a, b) although trends with GPC were less distinct at site 2. At image level, structural indices NDVI and EVI increased in

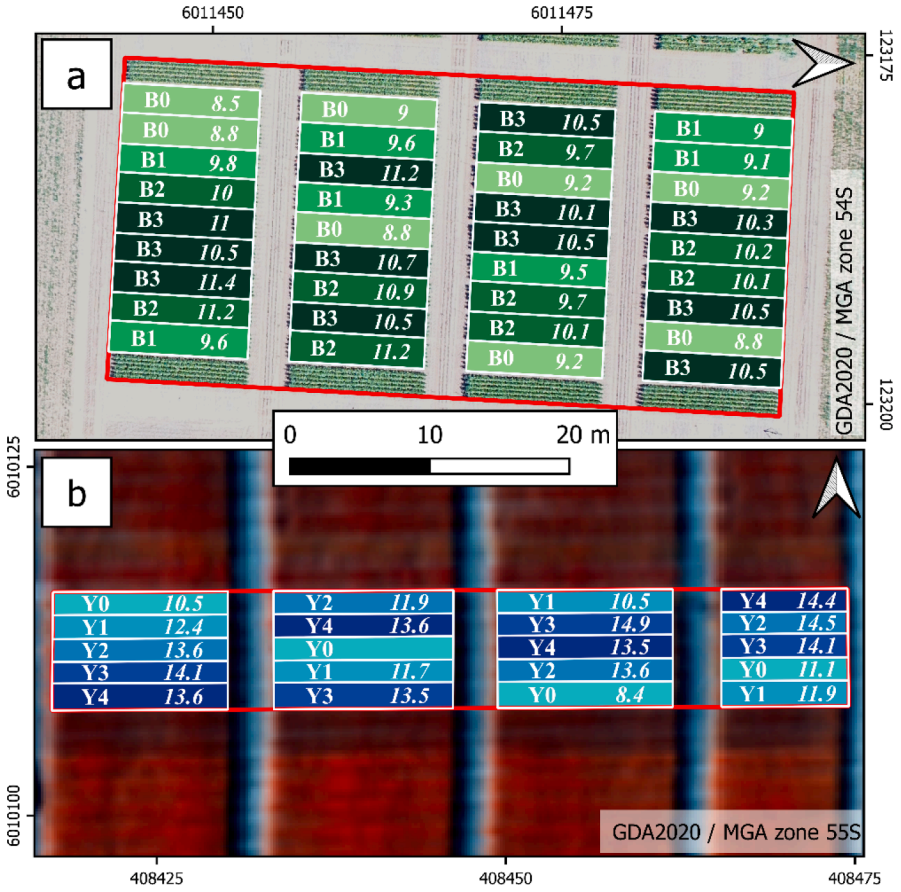

Fig. 4. Plots at site 1 and site 2 with treatment (left of plots) and GPC (%, in italics; right) over false-colour hyperspectral image. Colours represent N rates (kg N/ha): B0 = 0, B1 = 30 – 37, B2 = 98 – 104, B3 = 162 – 171; Y0 = 0, Y1 = 46, Y2 = 92, Y3 = 138, Y4 = 184.

line with N treatment and GPC but saturated after the first treatment level at site 2 and after the second at site 1. The CI, PRI_{m4} and, as at leaf level, R_{920}/R_{729} trended with GPC. Agreement between PRI_{m3} (leaf) and PRI_{m4} (canopy) was close, and leaf-level readings were lower at both sites (Fig. 5a–d). Each PRI_{m*} tracked as well with GPC as with leaf N from optical and, where conducted, destructive sampling (Fig. 5e, f). Associations between image-level indicators and GPC were weaker at site 2 (Fig. 5c, d). Dualex observations were similar across sites: C_{a+b} and the Nitrogen Balance Index (NBI) had positive associations with GPC, with higher absolute readings at site 1, while Anth were inverted with respect to GPC. Leaf N from destructive sampling generally increased with N treatment at both sites, but like GPC declined in treatment Y4.

Among parameters retrieved by RTM inversion, C_{a+b} trended with GPC at both sites, though with saturation at higher values, particularly at site 2 where C_{a+b} concentration was lower (Fig. 6a, b). Inverted C_{a+b} was significantly correlated with Dualex C_{a+b} ($r^2 = 0.61$, $p < 0.0001$). C_{x+c} showed a minor trend parallel to GPC at site 2 but saturated after one N treatment and was lower range than at site 1, where no trend was seen. Anth trended higher with N treatment, and were in higher concentration overall at site 1, but this relationship was inverted at site 2. LAI trended higher with N treatment at site 1, while at site 2 the opposite was seen, with saturation, and LAI values were very high. No alignment between SIF observations and treatment was seen at site 1, while SIF declined with higher N treatment, against GPC at site 2. CWSI was in a substantially higher range at site 2. The following relationships at site 1 were statistically significant: Anth ($R = 0.61$, $p < 0.01$), C_{a+b} ($R = 0.63$, $p < 0.01$) and CWSI ($R = -0.52$, $p < 0.05$). At site 2 these were: Anth ($R = -0.62$, $p < 0.01$) and LAI ($R = -0.55$, $p < 0.05$).

3.2. Field transect

The range of GPC at M01 was intermediate between sites 1 and 2, the crop had not senesced as much as at site 2 and retained soil moisture during our campaigns. At leaf level, VOG1, ZMI, R_{920}/R_{729} (not shown) and PRI_{m3} trended with GPC, showing some saturation, while NPCI showed little association (Fig. 7a). Airborne NDVI, EVI, CI, R_{920}/R_{729} (not shown) and PRI_{m4} trended with GPC, all with saturation (Fig. 7b). The two PRI versions agreed closely. Inverted C_{a+b} was negatively associated with GPC, while C_{x+c} and Anth were positively associated with GPC (Fig. 7c). Dualex NBI and soil mineral N followed GPC (Fig. 7d).

3.3. Commercial fields

A large difference in growing conditions, especially low rainfall in 2019 (280 mm) compared to 2020 (443 mm), also affected the commercial fields (Bureau of Meteorology 2021). Rainfall from December 2018 until sowing in early May 2019 was also very low (87 mm) compared to the subsequent equivalent period (164 mm). Weather conditions changed suddenly around anthesis 2019, whereby frost was recorded on 9 October, and daily maxima of >35 °C were recorded in mid-late October. Flights over the 2019 bread wheat crop took place 1514 DDAS in a year in the lowest decile of long-term AAR: the crop was under severe water stress. Remote sensing of bread wheat was done 1592 DDAS in 2020, a year of rainfall at the long-term AAR. Data capture for durum wheat was 1736 and 1742 DDAS in the respective years. Such conditions, especially the moisture contrast between soil types and years, can have large effects on grain protein.

Mean commercial bread wheat GPC was higher in 2019 (mean =

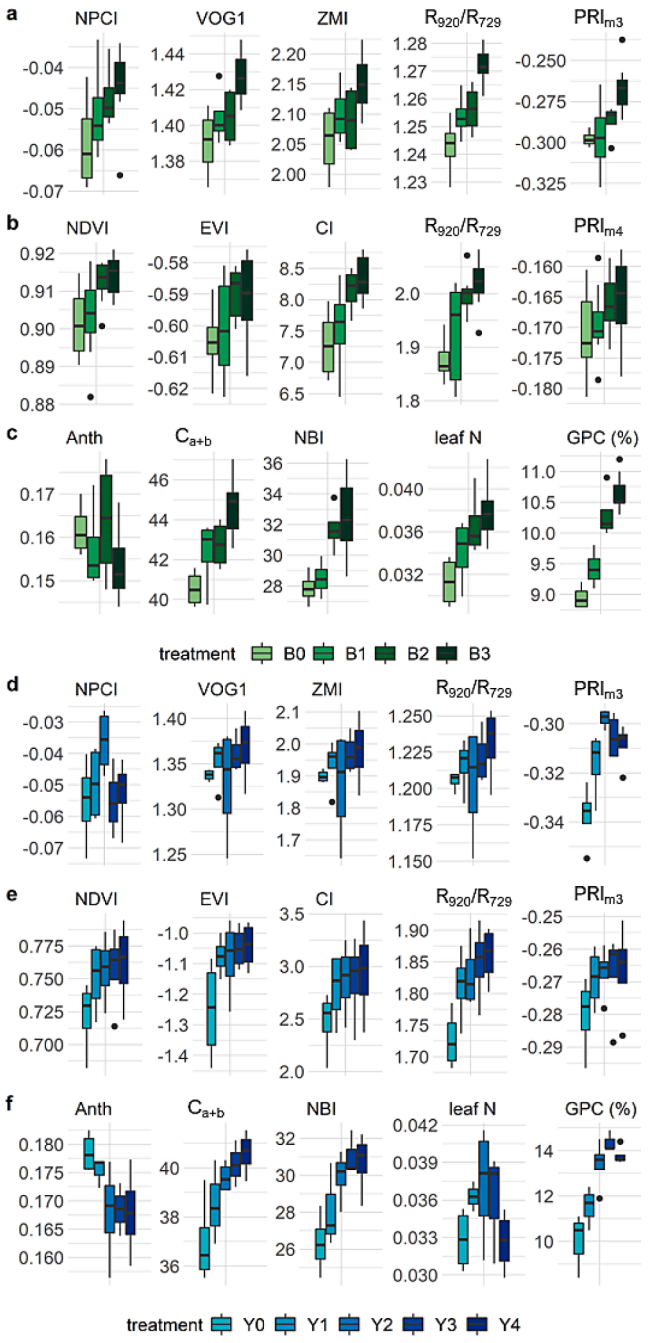

Fig. 5. Reflectance indices at leaf (a, d) and canopy level (b, e) and ground-truth indicators (c, f) by N fertiliser treatment at sites 1 (upper) and 2 (lower). At site 1, n = 20 for ground observations (a) and n = 36 for airborne indices (c); at site 2, n = 19. Anth, C_{a+b}. and NBI in Dualex proprietary units; leaf N in mg N cm^{-2}.

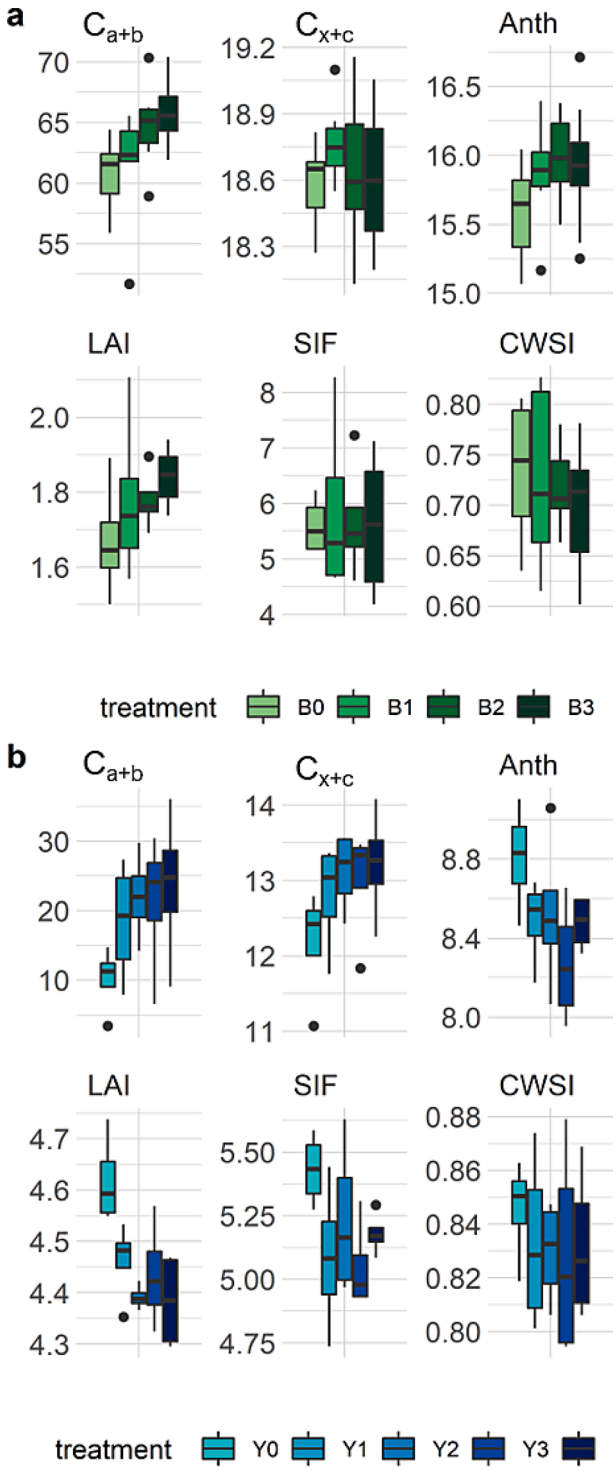

Fig. 6. Chlorophyll a + b, Carotenes, Anthocyanins and LAI (C_{a+b}, C_{x+c}, Anth; μg cm^{-2}, LAI; $m2/m^{-2}(-|-)$), retrieved by physical model inversion, plus solar-induced fluorescence (SIF; mW m^{-2} nm sr) and crop water stress index (CWSI) from hyperspectral and thermal data at sites 1 (a; n = 36) and 2 (b; n = 19).

11.6, SD = 1.52) than 2020 (mean = 11.3, SD = 1.05; Wilcoxon's $p < 0.0001$, effect size r = 0.489). Mean durum wheat GPC was higher in 2020 (mean = 12.7, SD = 0.94) than in 2019 (mean = 11.9, SD = 1.18; Wilcoxon's $p < 0.0001$, r = 0.360). GPC was significantly higher in durum than bread wheat in each year also (2019: $p < 0.0001$, r = 0.112; 2020: $p < 0.0001$, r = 0.564). Wide GPC variability was also seen between and within fields of each product type (Fig. 8). In a subset of our data, PRO4SAIL C_{a+b} and CI were strongly correlated ($r^2 = 0.86$, $p < 0.0001$).

Two reflectance indices, EVI and PRI, were both a) selected after VIF and b) robust across the two years of the study; these were retained as ML input features, supporting our finding of their correlations with GPC (Section 3.1). The relative importance of features to GPC estimation was quantified for input layers comprising: (i) physiological indicators Anth, C_{a+b} and C_{x+c} from model inversion, SIF and the photochemical reflectance index (PRI); (ii) structural indicators EVI, LAI and LIDFa; and (iii) the thermal-based CWSI. These inputs were supplied to the GBM algorithm as separate models: physiological + structural + CWSI (Model

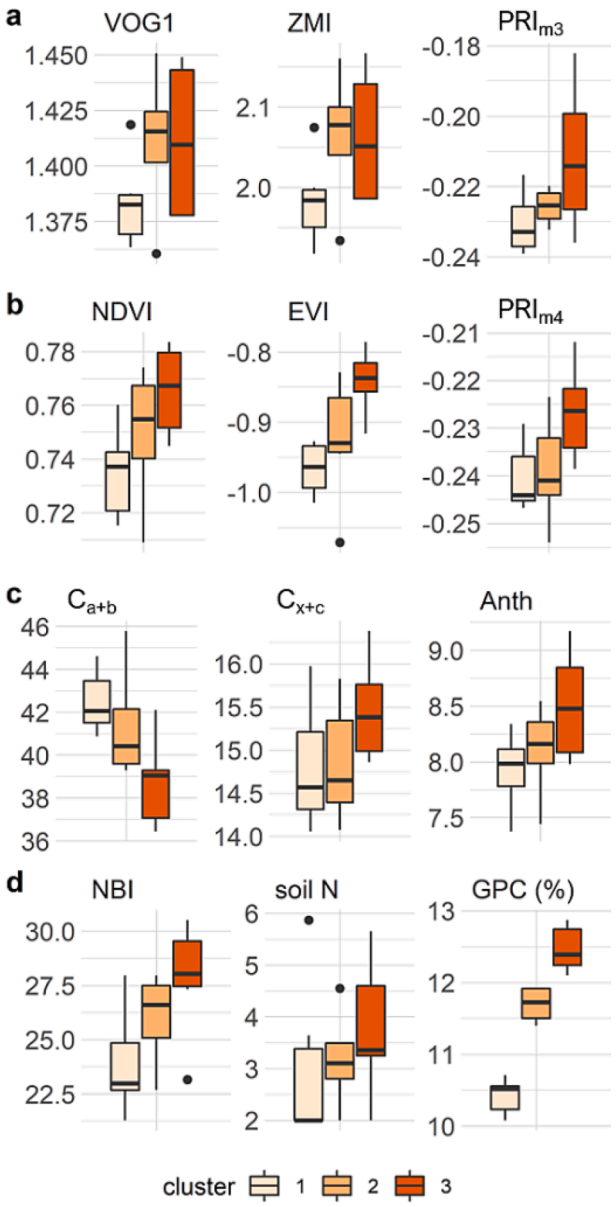

Fig. 7. Leaf-level (a) and airborne (b) reflectance indices; Chlorophyll, Carotenes, and Anthocyanins (C_{a+b}, C_{x+c}, Anth; μg cm^{-2}), retrieved by physical model inversion (c) and ground-truth indicators Dualex Nitrogen Balance Index (NBI), soil mineral N (mg kg^{-1}; d), and grain protein content (GPC; %; d) by GPC-based k-means clusters in a transect of commercial wheat field M01 (n = 20).

1); physiological + structural (Model 2); physiological (Model 3). Regarding Model 1 for bread wheat under drought in 2019, CWSI contributed 69% of total information, EVI 13% and the remainder was split (Fig. 9a). For model 1 in 2020, CWSI was second to C_{x+c} and EVI was also prominent in a more even distribution; C_{a+b}, PRI and SIF had low importance. Under Model 2, LAI and C_{x+c} were most important: C_{x+c} was most important feature in 2020 under all models (Fig. 9b).

In durum wheat under Model 1, CWSI was important for GPC estimation consistently across years (Fig. 9d), while SIF ranked highly in both years and in 2020 was quasi-equivalent to PRI and C_{x+c}. Pigments C_{a+b} and C_{x+c} had high importance in 2019 but C_{a+b} ranked lower in 2020. Structural features LAI, LIDFa and EVI, with Anth, occupied the lowest rankings in both 2019 and 2020. Under Model 2, importance was again evenly distributed and rankings were largely retained across inputs and years (Fig. 9e): SIF was consistently ranked highest and the structural indicators lowest. Anth and C_{a+b} ranked highly in 2019 but declined in 2020, while PRI and LIDFa had higher importance in 2020. In Model 3 (Fig. 9f), SIF retained top rank across years while C_{a+b} was lower in 2020 than 2019 and PRI higher.

3.4. Model predictive skill

The skill of each model in predicting GPC in an unseen 30% of observations was assessed for each combination of year/product/input layer by coefficient of determination (r^2) and relative root mean square error (rRMSE; GPC (%); Table 4). In every case, more information gave better predictions: Model 1, with all three layers of input data, outperformed Models 2 and 3. The best GPC prediction was seen in the 2019 bread wheat crop ($r^2 = 0.80$, rRMSE = 0.62), when CWSI contributed a mean 69% of total predictive power. However, this year and crop also had the highest r^2 when models 2 or 3 were used. In less water-stressed scenarios, the contribution of CWSI relative to total model skill was lower, confirming the tested physiological quantities as important indicators of GPC. Added to the physiological layer, the structural layer, comprising LAI, LIDFa and EVI, also increased model skill by between 11 and 21% despite their low importance rankings. This was higher than the respective CWSI contribution in any year/product/model combination.

4. Discussion

4.1. Plot experiments

Few studies have focused on GPC estimation using airborne hyperspectral remote sensing. Our objective was to identify traits related to harvest GPC in bread and durum wheat in experimental and commercial settings. High GPC is common in conditions unfavourable for CHO production and translocation, observations verified through our focus on stress indicators. At site 2, N supply at the upper treatment levels manifestly exceeded plant requirements given the low starting soil moisture and rainfall. This N excess is shown by the high LAI, which was extreme for Australian wheat (Waldner et al., 2019) and far exceeded LAI at site 1 (Fig. 6). Similarly, high GPC and declining LAI and leaf N, especially at high fertiliser rates, are signs of excessive N at site 2 but not site 1 (Figs. 5 and 6). While a greater GPC response to fertiliser is expected when grain filling is water limited (Angus and Fischer, 1991; Holford et al., 1992), and is seen in the comparison of our two sites, the site 2 GPC response saturated at heavy N applications, and at the highest level was below the next highest N treatment. Site 2 was also more advanced at the time of flights, and little extractable soil moisture remained: at the highest N treatment levels, high biomass worsened water stress, likely further restricting N uptake. Evidence for this is seen in declining leaf N and GPC in treatments Y3 and Y4, not seen in the more moderate moisture and calibrated N dosing of site 1. The close alignment of CI, VOG1 and ZMI with C_{a+b} and NBI at leaf level, inverted canopy C_{a+b}, and GPC supports findings of these indices' links with leaf N (Ustin et al., 2009; Vogelmann et al., 1993; Zarco-Tejada et al., 2001), and GPC (Li et al., 2020). These associations were more pronounced at site 1 than at site 2. The VOG1 index correlated strongly with leaf N at site 1 and with C_{a+b}, NBI and GPC at both sites. This, and close associations between NPCI and leaf N, support other findings in wheat (Ranjan et al., 2012), although NPCI was less closely associated with GPC. Further, higher ranges of CI, VOG1, ZMI and leaf- and canopy-level C_{a+b} at site 1, but higher GPC at site 2, suggest more advanced translocation at site 2, driven by senescence. This latter hypothesis is also supported by the reduction in LAI along the N gradient at site 2.

While most traits and indices discussed above are stable over days to weeks, the PRI family of indices can change on much shorter timescales; however, given the stable weather, here PRI likely shows stable stresses. Especially at site 2, evidence for the physiological link between PRI and photosynthesis was seen in the similarity between PRI and GPC, via a

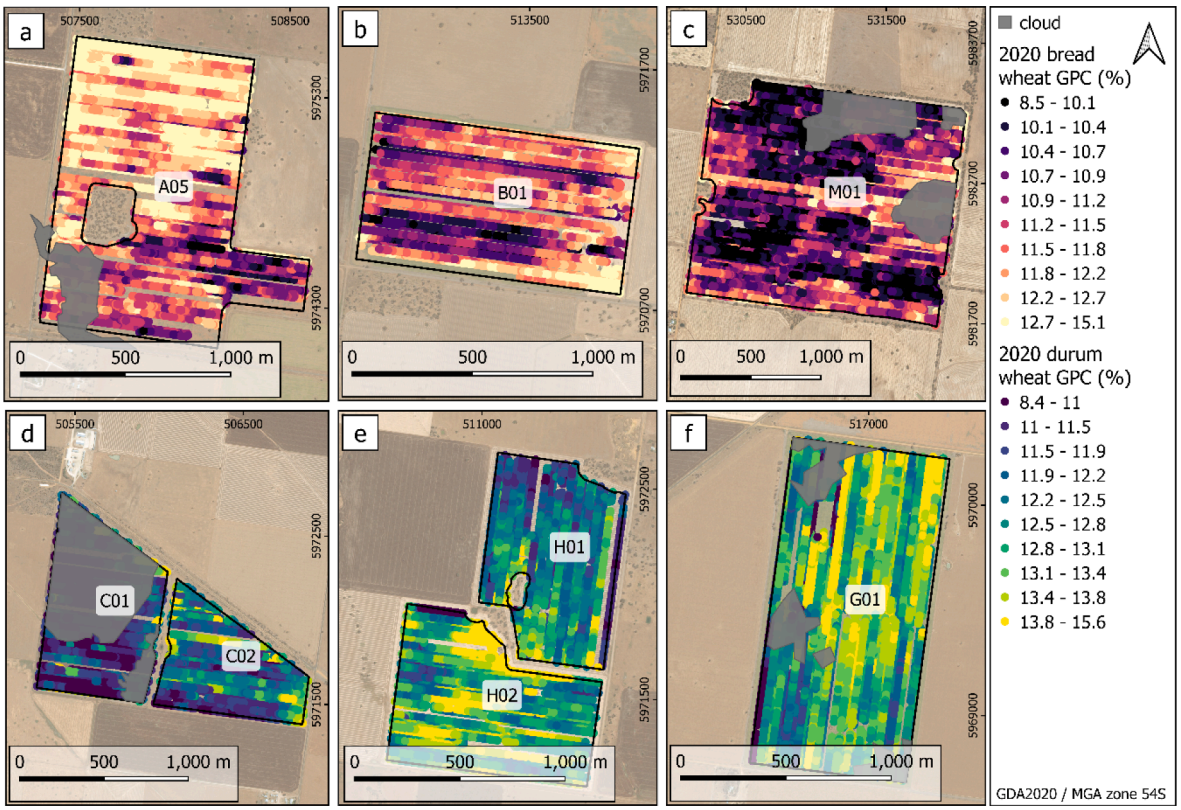

Fig. 8. Spatial heterogeneity in grain protein content (GPC; %) in bread (a–c) and durum wheat (d–f) fields in the Kaniva region, 2020.

clear inverse relationship between airborne SIF and GPC, and in lower overall SIF. Our findings accord with the inverse PRI ~ SIF relationship seen by Magney et al. (2016), Peñuelas et al. (1994) and Suárez et al. (2008).

Although no conclusive CWSI trend was seen at either site, its higher range at site 2 aligned with that site's lower SIF: SIF also declined at higher N treatments, likely due to greater drought stress. Seen together, high GPC, distinct alignments in the PRI_{m3} ~ GPC and CWSI ~ SIF responses show strongly constrained assimilation at site 2 but not at site 1. This was also seen in the C_{x+c} response, which despite saturation was also associated with GPC. However C_{x+c} were also parallel with leaf N and C_{a+b}, so their response may simply reflect N nutrition, as seen by Shah et al. (2017) in wheat. At image capture, C_{x+c} were lower overall at site 2, likely due to prior remobilisation to the grain. Low C_{a+b} values at site 2, compared with site 1 and other work (Hamblin et al., 2014), also suggest that C_{a+b} declined before C_{x+c} (Gitelson and Merzlyak, 1994a, 1994b). Given Anth upregulation under stress, one may expect high Anth to correspond to high GPC, but this was seen only in canopy Anth at site 1, which aligned with both leaf N and GPC. Alternatively, these observations may show a positive Anth response to N fertiliser (Yamuangmorn et al., 2018). At site 2, similar Anth responses at leaf and canopy levels, and their inversion with respect to N treatment and GPC, suggest that like C_{a+b} and C_{x+c}, Anth translocation had begun before data collection, especially in the high N treatments. Assessed on C_{x+c} and Anth, and if these compounds' concentrations are taken as a function of stress, site 1 appears more stressed. But C_{x+c} and Anth also respond positively to high N supply, and this suggests more accessible soil N at site 1 due to higher soil moisture. At site 2, strong declines in Anth at high N appear to confirm lower soil N accessibility at high N treatments and/or earlier senescence. NDVI and EVI are inferior to the indices discussed above as stress indicators, as seen previously (Gamon et al., 1992; Peñuelas et al., 1994) and showed little association with GPC. The consistency of the R_{920}/R_{729}. index with C_{a+b}, NBI, and GPC suggests it could be applied in GPC estimation, due to its component bands' sensitivity to water, structure and C_{a+b}.

4.2. Field transect

As at site 1, inverted canopy Anth increased with GPC in the field transect. Also as at both plot sites, C_{x+c} was parallel with GPC, and with soil N in M01, supporting an effect of N nutrition, stress or both on GPC. The association of PRI with lower assimilation is evident at M01 through its alignment with GPC across the transect. Given that wheat in field M01 was at least as advanced in phenology during RS campaigns as the plot sites were, its substantially lower CWSI likely shows a better match of N supply to soil moisture than at either plot site, reflecting the contrasting experimental and commercial objectives.

4.3. Commercial fields

The influence of heterogeneous soil moisture and N availability is strong and operates through mechanisms discernible in the leaf and canopy traits we retrieved. Our division of input features into thermal, structural and physiological layers, and their sequential removal from the model, allowed us to assess each input's contribution to GPC estimation, and each layer's influence on predictive skill. In each situation, especially high water stress, CWSI was an important indicator of GPC via its relationship with assimilation, but physiological features also showed commonalities across moisture conditions. In severe stress, in 2019 bread wheat, physiological components contributed little. With stomatal conductance and photosynthesis universally depressed, the physiological links of C_{a+b}, SIF, PRI with GPC lacked power and were deemphasised as predictive features for ML. In these conditions, C_{x+c} was highest among low importances for physiological indicators and was joined by Anth and C_{a+b} on removal of thermal and structural layers. When CWSI was excluded as a model input for severely stressed crops, LAI importance was high, confirming the greater relative influence of structure in drought than in benign conditions. The displacement of EVI

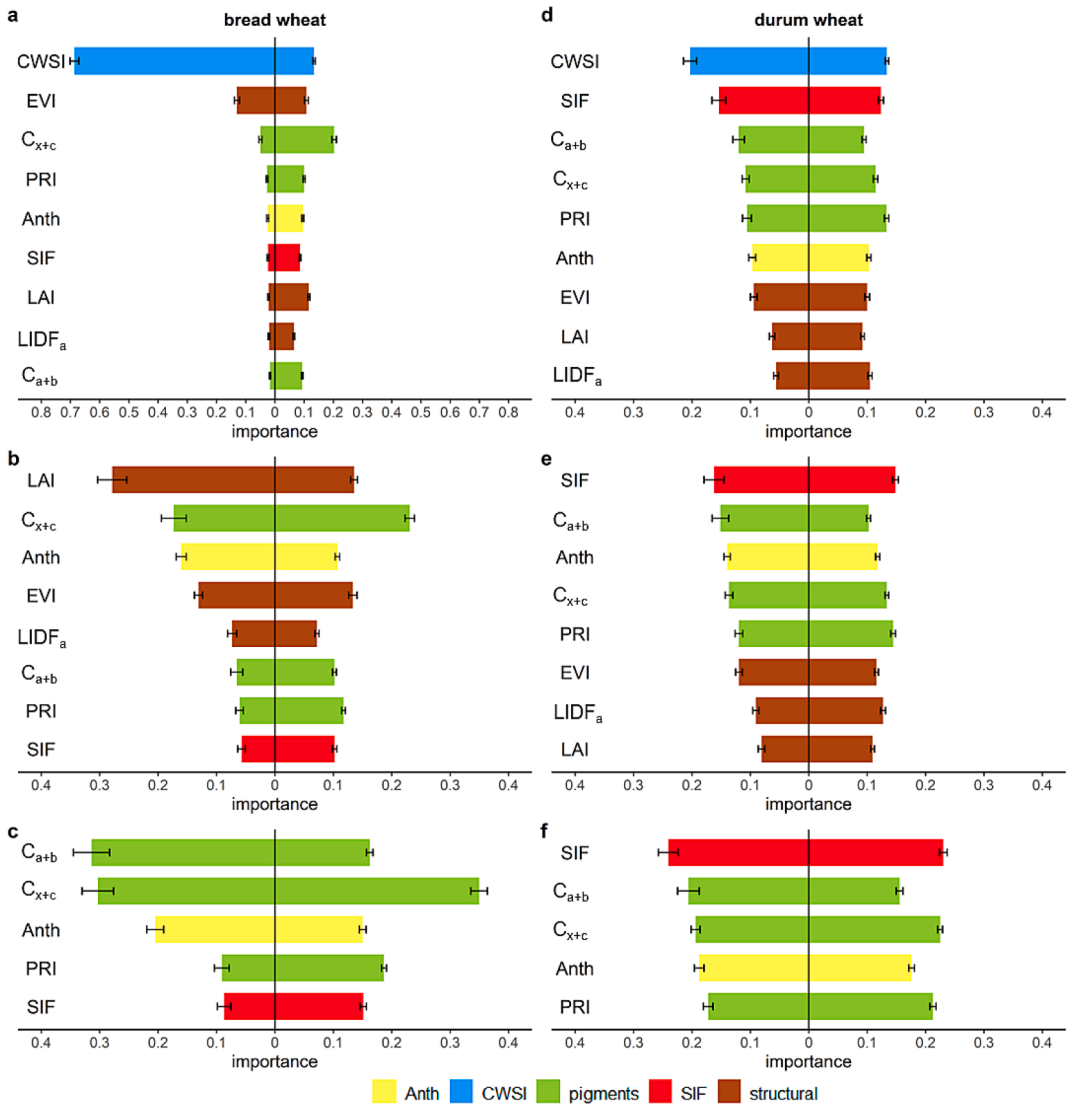

Fig. 9. Relative importance (proportion) of input features to a gradient boosting machine estimating grain protein content (GPC; %) in bread (left) and durum wheat (right) in commercial fields near Kaniva, Australia. Three models are shown: physiological + structural + CWSI (model 1; a, d); physiological + structural (model 2; b, e); physiological (model 3; c, f). Each sub-figure represents the 2019 (left; bread n = 7213, durum n = 5030) and 2020 seasons (right; bread n = 11060, durum n = 17310). Error bars show standard deviation of the mean proportional importance over 80 runs.

Table 4
Predictive skill (r^2, rRMSE; %) for Model 1, built with physiological + structural + CWSI layers, Model 2 (physiological + structural) and Model 3 (physiological only) across bread and durum wheat. Each model/product/year combination was run 80 times.

Model	Bread wheat				Durum wheat			
	2019		2020		2019		2020	
	r^2	rRMSE	r^2	rRMSE	r^2	rRMSE	r^2	rRMSE
1	0.8	0.62	0.54	0.66	0.54	0.81	0.49	0.67
2	0.7	0.76	0.5	0.69	0.49	0.84	0.46	0.69
3	0.57	0.91	0.39	0.76	0.43	0.9	0.37	0.75

by LAI in drought suggests redundancy between them; each may indicate canopy variability established prior to the onset of water stress and both are less affected by drought than are physiological traits.

In more moderate conditions, soil moisture heterogeneity drives variability in photosynthetic rate and hence physiological indicators across and between fields. This greater variability allows physiological features to convey more information, giving higher importance in GPC prediction. For example, at moderate stress in the 2019 durum crop, SIF approached CWSI in importance and all remaining physiological features were close together below SIF, none individually prominent but each more informative than structural traits. The order of physiological features changed little on removal of thermal and structural layers, showing their robustness and utility as GPC predictors in such conditions. Meanwhile in 2020, when water stress was lowest, importance was more evenly spread between physiological, thermal and structural components and physiological indicators C_{x+c} and PRI reached higher importance than CWSI. In these conditions, C_{x+c} was the best indicator overall in bread wheat, and in durum was below only CWSI, SIF and PRI,

likely showing a water stress differential between the two wheat types.

Results at moderate and low stress confirm that C_{x+c}, Anth, PRI and SIF are sensitive to water stress that is yet insufficient to strongly diminish photosynthesis. Moreover, each of these consistently had equal or higher importance than C_{a+b} in low stress, showing that short-term stress indicators also indicate GPC. Overall, physiological features were stable across conditions, and consistently supplied more information than structural features. This concurs with our own findings at plot scale and those of others that Anth, C_{x+c}, PRI and SIF, or combinations thereof, were crucial in stress diagnosis (Poblete et al., 2021; Suárez et al., 2008; Zarco-Tejada et al., 2018). The PRI ~ SIF alignment also mirrors our plot and transect studies and commends PRI as a proxy for photosynthesis in wheat (Magney et al., 2016). Anth importance was moderate in all conditions and stable between wheat types in low stress; this shows a differential Anth response under variable water stress and was supported at plot scale and in the literature (Chalker-Scott, 2002; Li et al., 2018). Unstable importance in EVI, LAI and LIDFa limits their value for GPC estimation, except in extreme stress, but together they contribute to model performance and generalisability. In benign conditions, LAI and LIDFa generally show low importance.

4.4. Model predictive skill

Though a minor focus of our work, model skill improved on some previous results for wheat GPC. Using leave-one-out (LOO) validation, Øvergaard et al. (2013) obtained pred ~ obs r^2 = 0.16–0.68, while Hansen et al. (2002) achieved $r^2 = 0.56$ and Apan et al. (2006) $r^2 = 0.92$. Li et al. (2020) obtained r^2 = 0.13–0.85, testing on 33% of their observations, and Zhou et al. (2021) also realised r^2 = 0.55–0.63 in an unseen 31% of their full (n = 327) dataset. Rodrigues et al. (2018) obtained $r^2 = 0.21$. Like Øvergaard et al. (2013), Li et al. (2020) and Zhou et al. (2021), we tested on a substantial unseen hold-out of observations across all fields, a more robust model proof than the LOO methods often seen. We also tested prediction with a field-wise LOO method, such that each field's data were successively used as the unseen test set for a model calibrated on the rest. Successful demonstration of this is important for many potential applications of our methods. When zero data from the LOO field were included in training, predictive skill was very poor (not shown). We then reduced stepwise from 70% to 10% the availability of training data from the LOO field; predictive skill declined, but not dramatically, at each step down to 10%.

That our physiological layer predicted GPC with acceptable accuracy without thermal or structural features attests to its coverage of GPC-relevant traits, from instantaneous SIF to the relatively stable C_{a+b}. These results agree with findings that SIF and C_{a+b} were far better than structural measures in estimating wheat leaf N (Camino et al., 2018). Because the indicators we use to estimate GPC are proxies of the water and nutrient stresses present in the region of study, our methods will probably work in regions with similar water and nutrient stress levels. In regions which are not water- or nutrient-limited, plant traits other than those described in this paper will be sensitive to GPC. Each of these proposals should be tested; our methods should also be tested with RS data of lower resolution, from spaceborne sensors, and in diverse seasonal, soil, cultivar, and agronomic conditions. It would also be valuable to test timeseries data captured within and across seasons. Further inputs such as year-to-date rainfall, soil or agronomic data may improve model predictive skill, as may training on multi-year databases. Further investigation of field-wise LOO is also needed to assess model transferability to unseen paddocks.

5. Conclusions

This study identified the most important hyperspectral-based plant traits related to grain protein content in rainfed wheat under variable stress levels. In experimental plots, two variants of the PRI index related to the xanthophyll pigment cycle showed consistent trends very similar to GPC along the induced N nutrition gradient, and in this respect performed better than any other spectral trait. In commercial crops, we implemented a gradient boosted machine to investigate relationships between input features and GPC. The thermal CWSI indicator of canopy transpiration contributed strongly to the model under water stress conditions, while Anth, C_{x+c}, PRI and SIF consistently showed high importance in GPC estimation under more benign conditions. Structural indicators such as LAI or its proxy NDVI contributed significantly less. We obtained promising results using gradient boosted machine learning to estimate GPC from hyperspectral and thermal images. Results yielded $r^2 = 0.80$ with rRMSE = 0.62% between predicted and observed GPC using a model built with thermal and physiological traits quantified by radiative transfer modelling methods.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors gratefully acknowledge the Birchip Cropping Group, Riverine Plains Incorporated and Foundation for Arable Research Australia for their provision and management of the plot trials, collection of GPC data and support of field operations. We also thank Jonathan Dyer, grain grower at Kaniva for his enthusiastic participation, advice and for access to GPC data. A. Gracia-Romero and A. Belwalkar are acknowledged for their technical support during the field and airborne campaigns. Andrew Longmire received PhD funding from the Grains Research and Development Corporation Australia. We also thank three anonymous reviewers for their thoughtful contributions, which improved this manuscript.

Author contributions

AL conceived the project, did the experimental design, collected field data at ground level, did the machine learning and data analyses, and wrote the manuscript. TP performed model inversions and supervised the machine learning algorithms. JH led the plot experiment at Birchip. AL, PZ and TP contributed to the data interpretation, and to manuscript drafting. PZ supervised the remote sensing work and led the airborne hyperspectral campaigns. DC supervised the work. JH, PZ and TP reviewed and edited the manuscript.

References

Abdi, A.M., 2020. Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data. GIScience Remote Sens. 57, 1–20. https://doi.org/10.1080/15481603.2019.1650447.

Akinwande, M.O., Dikko, H.G., Samson, A., 2015. Variance Inflation Factor: As a Condition for the Inclusion of Suppressor Variable(s) in Regression Analysis. Open J. Stat. 05, 754. https://doi.org/10.4236/ojs.2015.57075.

Angus, J.F., Fischer, R.A., 1991. Grain protein responses to nitrogen applied to wheat growing on a red earth. Aust. J. Agric. Res. 42, 735–746. https://doi.org/10.1071/ar9910735.

Angus, J.F., Grace, P.R., 2017. Nitrogen balance in Australia and nitrogen use efficiency on Australian farms. Soil Res. 55, 435–450. https://doi.org/10.1071/SR16325.

Apan, A., Kelly, R., Phinn, S., Strong, W., Lester, D., Butler, D., Robson, A., 2006. Predicting grain protein content in wheat using hyperspectral sensing of in-season crop canopies and partial least squares regression. Int. J. Geoinformatics 2, 93–108.

Bacour, C., Jacquemoud, S., Leroy, M., Hautecœur, O., Weiss, M., Prévot, L., Bruguier, N., Chauki, H., 2002. Reliability of the estimation of vegetation characteristics by inversion of three canopy reflectance models on airborne POLDER data. Agronomie 22, 555–565. https://doi.org/10.1051/agro:2002039.

Baret, F., Guyot, G., 1991. Potentials and limits of vegetation indices for LAI and APAR assessment. Remote Sens. Environ. 35, 161–173. https://doi.org/10.1016/0034-4257(91)90009-U.

Bauer, D.F., 1972. Constructing Confidence Sets Using Rank Statistics. J. Am. Stat. Assoc. 67, 687–690. https://doi.org/10.1080/01621459.1972.10481279.

Becker-Reshef, I., Vermote, E., Lindeman, M., Justice, C., 2010. A generalized regression-based model for forecasting winter wheat yields in Kansas and Ukraine using MODIS

data. Remote Sens. Environ. 114, 1312–1323. https://doi.org/10.1016/j.rse.2010.01.010.
Botha, E.J., Leblon, B., Zebarth, B.J., Watmough, J., 2010. Non-destructive estimation of wheat leaf chlorophyll content from hyperspectral measurements through analytical model inversion. Int. J. Remote Sens. 31, 1679–1697. https://doi.org/10.1080/01431160902926574.
Bureau of Meteorology, 2021. Climate Data Online [WWW Document]. URL http://www.bom.gov.au/climate/data/ (accessed 9.24.21).
Calderón, R., Navas-Cortés, J.A., Zarco-Tejada, P.J., 2015. Early Detection and Quantification of Verticillium Wilt in Olive Using Hyperspectral and Thermal Imagery over Large Areas. Remote Sens. 7, 5584–5610. https://doi.org/10.3390/rs70505584.
Camino, C., González-Dugo, V., Hernández, P., Sillero, J.C., Zarco-Tejada, P.J., 2018. Improved nitrogen retrievals with airborne-derived fluorescence and plant traits quantified from VNIR-SWIR hyperspectral imagery in the context of precision agriculture. Int. J. Appl. Earth Obs. Geoinformation 70, 105–117. https://doi.org/10.1016/j.jag.2018.04.013.
Cao, J., Zhang, Z., Tao, F., Zhang, L., Luo, Y., Han, J., Li, Z., 2020. Identifying the contributions of multi-source data for winter wheat yield prediction in China. Remote Sens. 12, 750. https://doi.org/10.3390/rs12050750.
Chalker-Scott, L., 2002. Do anthocyanins function as osmoregulators in leaf tissues? Adv. Bot. Res. 37, 103–127. https://doi.org/10.1016/S0065-2296(02)37046-0.
Chen, T., Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16. Association for Computing Machinery, New York, NY, USA, pp. 785–794. https://doi.org/10.1145/2939672.2939785.
Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., Zhou, T., Li, M., Xie, J., Lin, M., Geng, Y., Li, Y., 2021. xgboost: Extreme Gradient Boosting.
Chen, P., Haboudane, D., Tremblay, N., Wang, J., Vigneault, P., Li, B., 2010. New spectral indicator assessing the efficiency of crop nitrogen treatment in corn and wheat. Remote Sens. Environ. 114, 1987–1997. https://doi.org/10.1016/j.rse.2010.04.006.
Clevers, J.G.P.W., Gitelson, A.A., 2013. Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on Sentinel-2 and -3. Int. J. Appl. Earth Obs. Geoinformation 23, 344–351. https://doi.org/10.1016/j.jag.2012.10.008.
Clevers, J.G.P.W., Kooistra, L., 2012. Using Hyperspectral Remote Sensing Data for Retrieving Canopy Chlorophyll and Nitrogen Content. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 5, 574–583. https://doi.org/10.1109/JSTARS.2011.2176468.
Daniel, C., Triboï, E., 2002. Changes in wheat protein aggregation during grain development: effects of temperatures and water stress. Eur. J. Agron. 16, 1–12. https://doi.org/10.1016/S1161-0301(01)00114-9.
Dempewolf, J., Adusei, B., Becker-Reshef, I., Hansen, M., Potapov, P., Khan, A., Barker, B., 2014. Wheat Yield Forecasting for Punjab Province from Vegetation Index Time Series and Historic Crop Statistics. Remote Sens. 6, 9653–9675. https://doi.org/10.3390/rs6109653.
Dorigo, W.A., Zurita-Milla, R., de Wit, A.J.W., Brazile, J., Singh, R., Schaepman, M.E., 2007. A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. Int. J. Appl. Earth Obs. Geoinformation 9, 165–193. https://doi.org/10.1016/j.jag.2006.05.003.
Eitel, J.U.H., Long, D.S., Gessler, P.E., Hunt, E.R., 2008. Combined Spectral Index to Improve Ground-Based Estimates of Nitrogen Status in Dryland Wheat. Agron. J. 100, 1694–1702. https://doi.org/10.2134/agronj2007.0362.
Feng, W., Qi, S., Heng, Y., Zhou, Y., Wu, Y., Liu, W., He, L., Li, X., 2017. Canopy Vegetation Indices from In situ Hyperspectral Data to Assess Plant Water Status of Winter Wheat under Powdery Mildew Stress. Front. Plant Sci. https://doi.org/10.3389/fpls.2017.01219.
Feng, M., Xiao, L., Zhang, M., Yang, W., Ding, G., 2014. Integrating Remote Sensing and GIS for Prediction of Winter Wheat (Triticum aestivum) Protein Contents in Linfen (Shanxi). China. PLOS ONE 9, e80989. https://doi.org/10.1371/journal.pone.0080989.
Féret, J.-B., François, C., Asner, G.P., Gitelson, A.A., Martin, R.E., Bidel, L.P.R., Ustin, S. L., le Maire, G., Jacquemoud, S., 2008. PROSPECT-4 and 5: Advances in the leaf optical properties model separating photosynthetic pigments. Remote Sens. Environ. 112, 3030–3043. https://doi.org/10.1016/j.rse.2008.02.012.
Féret, J.-B., Gitelson, A.A., Noble, S.D., Jacquemoud, S., 2017. PROSPECT-D: Towards modeling leaf optical properties through a complete lifecycle. Remote Sens. Environ. 193, 204–215. https://doi.org/10.1016/j.rse.2017.03.004.
Ferwerda, J.G., Skidmore, A.K., Mutanga, O., 2005. Nitrogen detection with hyperspectral normalized ratio indices across multiple plant species. Int. J. Remote Sens. 26, 4083–4095. https://doi.org/10.1080/01431160500181044.
Friedman, J.H., 2001. Greedy Function Approximation: A Gradient Boosting Machine. Ann. Stat. 29, 1189–1232.
Friedman, J.H., Meulman, J.J., 2003. Multiple additive regression trees with application in epidemiology. Stat. Med. 22, 1365–1381. https://doi.org/10.1002/sim.1501.
Gamon, J.A., Peñuelas, J., Field, C.B., 1992. A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. Remote Sens. Environ. 41, 35–44. https://doi.org/10.1016/0034-4257(92)90059-S.
Genty, B., Briantais, J.-M., Baker, N.R., 1989. The relationship between the quantum yield of photosynthetic electron transport and quenching of chlorophyll fluorescence. Biochim. Biophys. Acta BBA - Gen. Subj. 990, 87–92. https://doi.org/10.1016/S0304-4165(89)80016-9.
Gitelson, A.A., Gritz, Y., Merzlyak, M.N., 2003. Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. J. Plant Physiol. 160, 271–282. https://doi.org/10.1078/0176-1617-00887.
Gitelson, A., Merzlyak, M.N., 1994a. Quantitative estimation of chlorophyll-a using reflectance spectra: Experiments with autumn chestnut and maple leaves. J. Photochem. Photobiol. B 22, 247–252. https://doi.org/10.1016/1011-1344(93)06963-4.
Gitelson, A., Merzlyak, M.N., 1994b. Spectral Reflectance Changes Associated with Autumn Senescence of Aesculus hippocastanum L. and Acer platanoides L. Leaves. Spectral Features and Relation to Chlorophyll Estimation. J. Plant Physiol. 143, 286–292. https://doi.org/10.1016/S0176-1617(11)81633-0.
Giuliani, M.M., Giuzio, L., Caro, A.D., Flagella, Z., 2011. Relationships between Nitrogen Utilization and Grain Technological Quality in Durum Wheat: I. Nitrogen Translocation and Nitrogen Use Efficiency for Protein. Agron. J. 103, 1487–1494. https://doi.org/10.2134/agronj2011.0153.
Gonzalez-Dugo, V., Hernandez, P., Solis, I., Zarco-Tejada, P.J., 2015. Using High-Resolution Hyperspectral and Thermal Airborne Imagery to Assess Physiological Condition in the Context of Wheat Phenotyping. Remote Sens. 7, 13586–13605. https://doi.org/10.3390/rs71013586.
Gooding, M.J., Ellis, R.H., Shewry, P.R., Schofield, J.D., 2003. Effects of Restricted Water Availability and Increased Temperature on the Grain Filling, Drying and Quality of Winter Wheat. J. Cereal Sci. 37, 295–309. https://doi.org/10.1006/jcrs.2002.0501.
Gooding, M.J., Gregory, P.J., Ford, K.E., Ruske, R.E., 2007. Recovery of nitrogen from different sources following applications to winter wheat at and after anthesis. Field Crops Res. 100, 143–154. https://doi.org/10.1016/j.fcr.2006.06.002.
Grant, O.M., Tronina, Ł., Jones, H.G., Chaves, M.M., 2007. Exploring thermal imaging variables for the detection of stress responses in grapevine under different irrigation regimes. J. Exp. Bot. 58, 815–825. https://doi.org/10.1093/jxb/erl153.
Grinberg, N.F., Orhobor, O.I., King, R.D., 2020. An evaluation of machine-learning for predicting phenotype: studies in yeast, rice, and wheat. Mach. Learn. 109, 251–277. https://doi.org/10.1007/s10994-019-05848-5.
Groth, S., Wittmann, R., Longin, C.F.H., Böhm, V., 2020. Influence of variety and growing location on carotenoid and vitamin E contents of 184 different durum wheat varieties (Triticum turgidum ssp. durum) in Germany. Eur. Food Res. Technol. 246, 2079–2092. https://doi.org/10.1007/s00217-020-03557-1.
Gueymard, C., 1995. Simple Model for the Atmospheric Radiative Transfer of Sunshine (SMARTS2) Algorithms and performance assessment. Florida Solar Energy Center/University of Central Florida.
Haboudane, D., Miller, J.R., Tremblay, N., Zarco-Tejada, P.J., Dextraze, L., 2002. Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. Remote Sens. Environ. 81, 416–426. https://doi.org/10.1016/S0034-4257(02)00018-4.
Haboudane, D., Tremblay, N., Miller, J.R., Vigneault, P., 2008. Remote Estimation of Crop Chlorophyll Content Using Spectral Indices Derived From Hyperspectral Data. IEEE Trans. Geosci. Remote Sens. 46, 423–437. https://doi.org/10.1109/TGRS.2007.904836.
Hamblin, J., Stefanova, K., Angessa, T.T., 2014. Variation in Chlorophyll Content per Unit Leaf Area in Spring Wheat and Implications for Selection in Segregating Material. PLOS ONE 9, e92529. https://doi.org/10.1371/journal.pone.0092529.
Hansen, P.M., Jørgensen, J.R., Thomsen, A., 2002. Predicting grain yield and protein content in winter wheat and spring barley using repeated canopy reflectance measurements and partial least squares regression. J. Agric. Sci. 139, 307–318. https://doi.org/10.1017/S0021859602002320.
Hartigan, J.A., Wong, M.A., 1979. Algorithm AS 136: A K-Means Clustering Algorithm. Appl. Stat. 28, 100. https://doi.org/10.2307/2346830.
Heffer, P., Prud'homme, M., 2020. Global nitrogen fertiliser demand and supply: trend, current level and outlook. Presented at the 7th International Nitrogen Initiative Conference (INI2016), Melbourne, Australia.
Hernández-Clemente, R., Navarro-Cerrillo, R.M., Suárez, L., Morales, F., Zarco-Tejada, P. J., 2011. Assessing structural effects on PRI for stress detection in conifer forests. Remote Sens. Environ. 115, 2360–2375. https://doi.org/10.1016/j.rse.2011.04.036.
Herrmann, I., Karnieli, A., Bonfil, D.J., Cohen, Y., Alchanatis, V., 2010. SWIR-based spectral indices for assessing nitrogen content in potato fields. Int. J. Remote Sens. 31, 5127–5143. https://doi.org/10.1080/01431160903283892.
Hochman, Z., van Rees, H., Carberry, P.S., Hunt, J.R., McCown, R.L., Gartmann, A., Holzworth, D., van Rees, S., Dalgliesh, N.P., Long, W., Peake, A.S., Poulton, P.L., McClelland, T., 2009. Re-inventing model-based decision support with Australian dryland farmers. 4. Yield Prophet® helps farmers monitor and manage crops in a variable climate. Crop Pasture Sci. 60, 1057–1070. https://doi.org/10.1071/CP09020.
Holford, I.C.R., Doyle, A.D., Leckie, C.C., 1992. Nitrogen response characteristics of wheat protein in relation to yield responses and their interactions with phosphorus. Aust. J. Agric. Res. 43, 969–986. https://doi.org/10.1071/ar9920969.
Hunt, J., van Rees, H., Hochman, Z., Carberry, P.S., Holzworth, D., Dalgliesh, N., Brennan, L., Poulton, P., van Rees, S., Huth, N.I., 2006. Yield Prophet®: An online crop simulation service. In: Proceedings of the 13th Australian Agronomy Conference. pp. 10–14.
Idso, S.B., 1982. Non-water-stressed baselines: A key to measuring and interpreting plant water stress. Agric. Meteorol. 27, 59–70. https://doi.org/10.1016/0002-1571(82)90020-6.
Idso, S.B., Jackson, R.D., Pinter, P.J., Reginato, R.J., Hatfield, J.L., 1981. Normalizing the stress-degree-day parameter for environmental variability. Agric. Meteorol. 24, 45–55. https://doi.org/10.1016/0002-1571(81)90032-7.
Isbell, R., 2002. The Australian Soil Classification. CSIRO Publishing.
Jacquemoud, S., Baret, F., Andrieu, B., Danson, F.M., Jaggard, K., 1995. Extraction of vegetation biophysical parameters by inversion of the PROSPECT + SAIL models on sugar beet canopy reflectance data. Application to TM and AVIRIS sensors. Remote Sens. Environ. 52, 163–172. https://doi.org/10.1016/0034-4257(95)00018-V.

Jacquemoud, S., Verhoef, W., Baret, F., Bacour, C., Zarco-Tejada, P.J., Asner, G.P., François, C., Ustin, S.L., 2009. PROSPECT+SAIL models: A review of use for vegetation characterization. Remote Sens. Environ. Imaging Spectrosc. Special Issue 113, S56–S66. https://doi.org/10.1016/j.rse.2008.01.026.

Jamieson, P.D., Semenov, M.A., 2000. Modelling nitrogen uptake and redistribution in wheat. Field Crops Res. 68, 21–29. https://doi.org/10.1016/S0378-4290(00)00103-9.

Janeczko, A., Dziurka, M., Pociecha, E., 2018. Increased leaf tocopherol and β-carotene content is associated with the tolerance of winter wheat cultivars to frost. J. Agron. Crop Sci. 204, 594–602. https://doi.org/10.1111/jac.12287.

Jensen, T., Apan, A., Young, F., Zeller, L., 2007. Detecting the attributes of a wheat crop using digital imagery acquired from a low-altitude platform. Comput. Electron. Agric. 59, 66–77. https://doi.org/10.1016/j.compag.2007.05.004.

Klem, K., Záhora, J., Zemek, F., Trunda, P., Tůma, I., Novotná, K., Hodaňová, P., Rapantová, B., Hanuš, J., Vavříková, J., Holub, P., 2018. Interactive effects of water deficit and nitrogen nutrition on winter wheat. Remote sensing methods for their detection. Agric. Water Manag. 210, 171–184. https://doi.org/10.1016/j.agwat.2018.08.004.

Kohzuma, K., Tamaki, M., Hikosaka, K., 2021. Corrected photochemical reflectance index (PRI) is an effective tool for detecting environmental stresses in agricultural crops under light conditions. J. Plant Res. 134, 683–694. https://doi.org/10.1007/s10265-021-01316-1.

Kuhn, M., 2020. caret: Classification and Regression Training. R package.

Lai, Y.R., Pringle, M.J., Kopittke, P.M., Menzies, N.W., Orton, T.G., Dang, Y.P., 2018. An empirical model for prediction of wheat yield, using time-integrated Landsat NDVI. Int. J. Appl. Earth Obs. Geoinformation 72, 99–108. https://doi.org/10.1016/j.jag.2018.07.013.

Li, Z., Jin, X., Wang, J., Yang, G., Nie, C., Xu, X., Feng, H., 2015. Estimating winter wheat (Triticum aestivum) LAI and leaf chlorophyll content from canopy reflectance data by integrating agronomic prior knowledge with the PROSAIL model. Int. J. Remote Sens. 36, 2634–2653. https://doi.org/10.1080/01431161.2015.1041176.

Li, X., Lv, X., Wang, X., Wang, L., Zhang, M., Ren, M., 2018. Effects of abiotic stress on anthocyanin accumulation and grain weight in purple wheat. Crop Pasture Sci. 69, 1208–1214. https://doi.org/10.1071/CP18341.

Li, F., Mistele, B., Hu, Y., Chen, X., Schmidhalter, U., 2014. Reflectance estimation of canopy nitrogen content in winter wheat using optimised hyperspectral spectral indices and partial least squares regression. Eur. J. Agron. 52, 198–209. https://doi.org/10.1016/j.eja.2013.09.006.

Li, Z., Taylor, J., Yang, H., Casa, R., Jin, X., Li, Z., Song, X., Yang, G., 2020. A hierarchical interannual wheat yield and grain protein prediction model using spectral vegetative indices and meteorological data. Field Crops Res. 248, 107711. https://doi.org/10.1016/j.fcr.2019.107711.

Lopez-Bellido, R.J., Shepherd, C.E., Barraclough, P.B., 2004. Predicting post-anthesis N requirements of bread wheat with a Minolta SPAD meter. Eur. J. Agron. 20, 313–320. https://doi.org/10.1016/S1161-0301(03)00025-X.

Magney, T.S., Vierling, L.A., Eitel, J., 2014. Remote detection of water stress conditions via a diurnal photochemical reflectance index (PRI) improves yield prediction in rainfed wheat 2014, B51F-0091.

Magney, T.S., Vierling, L.A., Eitel, J.U.H., Huggins, D.R., Garrity, S.R., 2016. Response of high frequency Photochemical Reflectance Index (PRI) measurements to environmental conditions in wheat. Remote Sens. Environ. 173, 84–97. https://doi.org/10.1016/j.rse.2015.11.013.

Masoni, A., Ercoli, L., Mariotti, M., Arduini, I., 2007. Post-anthesis accumulation and remobilization of dry matter, nitrogen and phosphorus in durum wheat as affected by soil type. Eur. J. Agron. 26, 179–186. https://doi.org/10.1016/j.eja.2006.09.006.

Meroni, M., Rossini, M., Guanter, L., Alonso, L., Rascher, U., Colombo, R., Moreno, J., 2009. Remote sensing of solar-induced chlorophyll fluorescence: Review of methods and applications. Remote Sens. Environ. 113, 2037–2051. https://doi.org/10.1016/j.rse.2009.05.003.

Mohammed, G.H., Colombo, R., Middleton, E.M., Rascher, U., van der Tol, C., Nedbal, L., Goulas, Y., Pérez-Priego, O., Damm, A., Meroni, M., Joiner, J., Cogliati, S., Verhoef, W., Malenovský, Z., Gastellu-Etchegorry, J.-P., Miller, J.R., Guanter, L., Moreno, J., Moya, I., Berry, J.A., Frankenberg, C., Zarco-Tejada, P.J., 2019. Remote sensing of solar-induced chlorophyll fluorescence (SIF) in vegetation: 50 years of progress. Remote Sens. Environ. 231, 111177. https://doi.org/10.1016/j.rse.2019.04.030.

Monjardino, M., McBeath, T., Ouzman, J., Llewellyn, R., Jones, B., 2015. Farmer risk-aversion limits closure of yield and profit gaps: A study of nitrogen management in the southern Australian wheatbelt. Agric. Syst. 137, 108–118. https://doi.org/10.1016/j.agsy.2015.04.006.

Naing, A.H., Kim, C.K., 2021. Abiotic stress-induced anthocyanins in plants: Their role in tolerance to abiotic stresses. Physiol. Plant. 172, 1711–1723. https://doi.org/10.1111/ppl.13373.

Nakazawa, M., 2022. Functions for Medical Statistics Book with some Demographic Data.

Ottman, M.J., Doerge, T.A., Martin, E.C., 2000. Durum Grain Quality as Affected by Nitrogen Fertilization near Anthesis and Irrigation During Grain Fill. Agron. J. 92, 1035–1041. https://doi.org/10.2134/agronj2000.9251035x.

Øvergaard, S.I., Isaksson, T., Korsaeth, A., 2013. Prediction of Wheat Yield and Protein Using Remote Sensors on Plots—Part I: Assessing near Infrared Model Robustness for Year and Site Variations. J. Infrared Spectrosc. 21, 117–131. https://doi.org/10.1255/jnirs.1042.

Peel, M.C., Finlayson, B.L., McMahon, T.A., 2007. Updated world map of the Köppen-Geiger climate classification. Hydrol. Earth Syst. Sci. 11, 1633–1644. https://doi.org/10.5194/hess-11-1633-2007.

Peñuelas, J., Gamon, J.A., Fredeen, A.L., Merino, J., Field, C.B., 1994. Reflectance indices associated with physiological changes in nitrogen- and water-limited sunflower leaves. Remote Sens. Environ. 48, 135–146. https://doi.org/10.1016/0034-4257(94)90136-8.

Plascyk, J.A., 1975. The MK II Fraunhofer Line Discriminator (FLD-II) for Airborne and Orbital Remote Sensing of Solar-Stimulated Luminescence. Opt. Eng. 14, 339–346.

Poblete, T., Camino, C., Beck, P.S.A., Hornero, A., Kattenborn, T., Saponari, M., Boscia, D., Navas-Cortes, J.A., Zarco-Tejada, P.J., 2020. Detection of Xylella fastidiosa infection symptoms with airborne multispectral and thermal imagery: Assessing bandset reduction performance from hyperspectral analysis. ISPRS J. Photogramm. Remote Sens. 162, 27–40. https://doi.org/10.1016/j.isprsjprs.2020.02.010.

Poblete, T., Navas-Cortes, J.A., Camino, C., Calderon, R., Hornero, A., Gonzalez-Dugo, V., Landa, B.B., Zarco-Tejada, P.J., 2021. Discriminating Xylella fastidiosa from Verticillium dahliae infections in olive trees using thermal- and hyperspectral-based plant traits. ISPRS J. Photogramm. Remote Sens. 179, 133–144. https://doi.org/10.1016/j.isprsjprs.2021.07.014.

Prey, L., Schmidhalter, U., 2019. Simulation of satellite reflectance data using high-frequency ground based hyperspectral canopy measurements for in-season estimation of grain yield and grain nitrogen status in winter wheat. ISPRS J. Photogramm. Remote Sens. 149, 176–187. https://doi.org/10.1016/j.isprsjprs.2019.01.023.

QGIS Development Team, 2020. QGIS Geographic Information System. Open Source Geospatial Foundation.

R Core Team, 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Ranjan, R., Chopra, U.K., Sahoo, R.N., Singh, A.K., Pradhan, S., 2012. Assessment of plant nitrogen stress in wheat (*Triticum aestivum* L.) through hyperspectral indices. Int. J. Remote Sens. 33, 6342–6360. https://doi.org/10.1080/01431161.2012.687473.

Raya-Sereno, M.D., Ortiz-Monasterio, J.I., Alonso-Ayuso, M., Rodrigues, F.A., Rodríguez, A.A., González-Perez, L., Quemada, M., 2021. High-Resolution Airborne Hyperspectral Imagery for Assessing Yield, Biomass, Grain N Concentration, and N Output in Spring Wheat. Remote Sens. 13, 1373. https://doi.org/10.3390/rs13071373.

Rayment, G.E., Lyons, D.J. (Eds.), 2010. Soil Chemical Methods - Australasia. CSIRO Publishing, Melbourne, Australia.

Rodrigues, F., Blasch, G., Defourny, P., Ortiz-Monasterio, J., Schulthess, U., Zarco-Tejada, P., Taylor, J., Gérard, B., Rodrigues, F.A., Blasch, G., Defourny, P., Ortiz-Monasterio, J.I., Schulthess, U., Zarco-Tejada, P.J., Taylor, J.A., Gérard, B., 2018. Multi-Temporal and Spectral Analysis of High-Resolution Hyperspectral Airborne Imagery for Precision Agriculture: Assessment of Wheat Grain Yield and Grain Protein Content. Remote Sens. 10, 930. https://doi.org/10.3390/rs10060930.

Rondeaux, G., Steven, M., Baret, F., 1996. Optimization of soil-adjusted vegetation indices. Remote Sens. Environ. 55, 95–107. https://doi.org/10.1016/0034-4257(95)00186-7.

Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.W., 1974. Monitoring vegetation systems in the Great Plains with ERTS. NASA Spec. Publ. 351, 309.

Ruan, G., Li, X., Yuan, F., Cammarano, D., Ata-UI-Karim, S.T., Liu, X., Tian, Y., Zhu, Y., Cao, W., Cao, Q., 2022. Improving wheat yield prediction integrating proximal sensing and weather data with machine learning. Comput. Electron. Agric. 195, 106852. https://doi.org/10.1016/j.compag.2022.106852.

Shah, S.H., Houborg, R., McCabe, M.F., 2017. Response of Chlorophyll, Carotenoid and SPAD-502 Measurement to Salinity and Nutrient Stress in Wheat (Triticum aestivum L.). Agronomy 7, 61. https://doi.org/10.3390/agronomy7030061.

Shiferaw, B., Smale, M., Braun, H.-J., Duveiller, E., Reynolds, M., Muricho, G., 2013. Crops that feed the world 10. Past successes and future challenges to the role played by wheat in global food security. Food Secur. 5, 291–317. https://doi.org/10.1007/s12571-013-0263-y.

Sims, D.A., Gamon, J.A., 2002. Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. Remote Sens. Environ. 81, 337–354. https://doi.org/10.1016/S0034-4257(02)00010-X.

Suárez, L., Zarco-Tejada, P.J., Sepulcre-Cantó, G., Pérez-Priego, O., Miller, J.R., Jiménez-Muñoz, J.C., Sobrino, J., 2008. Assessing canopy PRI for water stress detection with diurnal airborne imagery. Remote Sens. Environ. 112, 560–575. https://doi.org/10.1016/j.rse.2007.05.009.

Sutton, M.A., Oenema, O., Erisman, J.W., Leip, A., van Grinsven, H., Winiwarter, W., 2011. Too much of a good thing. Nature 472, 159–161. https://doi.org/10.1038/472159a.

Toscano, P., Castrignanò, A., Di Gennaro, S.F., Vonella, A.V., Ventrella, D., Matese, A., 2019. A Precision Agriculture Approach for Durum Wheat Yield Assessment Using Remote Sensing Data and Yield Mapping. Agronomy 9, 437. https://doi.org/10.3390/agronomy9080437.

Ustin, S.L., Gitelson, A.A., Jacquemoud, S., Schaepman, M., Asner, G.P., Gamon, J.A., Zarco-Tejada, P., 2009. Retrieval of foliar information about plant pigment systems from high resolution spectroscopy. Remote Sens. Environ. Imaging Spectrosc. Special Issue 113, S67–S77. https://doi.org/10.1016/j.rse.2008.10.019.

van Herwaarden, A.F., Angus, J.F., Richards, R.A., Farquhar, G.D., 1998. "Haying-off", the negative grain yield response of dryland wheat to nitrogen fertiliser II. Carbohydrate and protein dynamics. Aust. J. Agric. Res. 49, 1083–1094. https://doi.org/10.1071/a97040.

Verhoef, W., Jia, L., Xiao, Q., Su, Z., 2007. Unified Optical-Thermal Four-Stream Radiative Transfer Theory for Homogeneous Vegetation Canopies. IEEE Trans. Geosci. Remote Sens. 45, 1808–1822. https://doi.org/10.1109/TGRS.2007.895844.

Vogelmann, J.E., Rock, B.N., Moss, D.M., 1993. Red edge spectral measurements from sugar maple leaves. Int. J. Remote Sens. 14, 1563–1575. https://doi.org/10.1080/01431169308953986.

Waldner, F., Horan, H., Chen, Y., Hochman, Z., 2019. High temporal resolution of leaf area data improves empirical estimation of grain yield. Sci. Rep. 9, 15714. https://doi.org/10.1038/s41598-019-51715-7.

Wang, L., Tian, Y., Yao, X., Zhu, Y., Cao, W., 2014. Predicting grain yield and protein content in wheat by fusing multi-sensor and multi-temporal remote-sensing images. Field Crops Res. 164, 178–188. https://doi.org/10.1016/j.fcr.2014.05.001.

Wang, W., Yao, X., Yao, XinFeng, Tian, Y., Liu, X., Ni, J., Cao, W., Zhu, Y., 2012. Estimating leaf nitrogen concentration with three-band vegetation indices in rice and wheat. Field Crops Res. 129, 90–98. https://doi.org/10.1016/j.fcr.2012.01.014.

Whelan, B.M., Taylor, J.A., Hassall, J.A., 2009. Site-specific variation in wheat grain protein concentration and wheat grain yield measured on an Australian farm using harvester-mounted on-the-go sensors. Crop Pasture Sci. 60, 808–817. https://doi.org/10.1071/CP08343.

Wright, D.L., Rasmussen, V.P., Ramsey, R.D., Baker, D.J., Ellsworth, J.W., 2004. Canopy Reflectance Estimation of Wheat Nitrogen Content for Grain Protein Management. GIScience Remote Sens. 41, 287–300. https://doi.org/10.2747/1548-1603.41.4.287.

Yamuangmorn, S., Dell, B., Rerkasem, B., Prom-U-Thai, C., 2018. Applying nitrogen fertilizer increased anthocyanin in vegetative shoots but not in grain of purple rice genotypes. J. Sci. Food Agric. 98, 4527–4532. https://doi.org/10.1002/jsfa.8978.

Yang, B., Ma, J., Yao, X., Cao, W., Zhu, Y., 2021. Estimation of leaf nitrogen content in wheat based on fusion of spectral features and deep features from near infrared hyperspectral imagery. Sensors 21, 613. https://doi.org/10.3390/s21020613.

Zadoks, J.C., Chang, T.T., Konzak, C.F., 1974. A decimal code for the growth stages of cereals. Weed Res. 14, 415–421. https://doi.org/10.1111/j.1365-3180.1974.tb01084.x.

Zarco-Tejada, P.J., Miller, J.R., Noland, T.L., Mohammed, G.H., Sampson, P.H., 2001. Scaling-up and model inversion methods with narrowband optical indices for chlorophyll content estimation in closed forest canopies with hyperspectral data. IEEE Trans. Geosci. Remote Sens. 39, 1491–1507. https://doi.org/10.1109/36.934080.

Zarco-Tejada, P.J., Ustin, S.L., Whiting, M.L., 2005. Temporal and Spatial Relationships between Within-Field Yield Variability in Cotton and High-Spatial Hyperspectral Remote Sensing Imagery. Agron. J. Madison 97, 641–653.

Zarco-Tejada, P.J., Catalina, A., González, M.R., Martín, P., 2013a. Relationships between net photosynthesis and steady-state chlorophyll fluorescence retrieved from airborne hyperspectral imagery. Remote Sens. Environ. 136, 247–258. https://doi.org/10.1016/j.rse.2013.05.011.

Zarco-Tejada, P.J., Guillén-Climent, M.L., Hernández-Clemente, R., Catalina, A., González, M.R., Martín, P., 2013b. Estimating leaf carotenoid content in vineyards using high resolution hyperspectral imagery acquired from an unmanned aerial vehicle (UAV). Agric. For. Meteorol. 171–172, 281–294. https://doi.org/10.1016/j.agrformet.2012.12.013.

Zarco-Tejada, P.J., González-Dugo, M.V., Fereres, E., 2016. Seasonal stability of chlorophyll fluorescence quantified from airborne hyperspectral imagery as an indicator of net photosynthesis in the context of precision agriculture. Remote Sens. Environ. 179, 89–103. https://doi.org/10.1016/j.rse.2016.03.024.

Zarco-Tejada, P.J., Camino, C., Beck, P.S.A., Calderon, R., Hornero, A., Hernández-Clemente, R., Kattenborn, T., Montes-Borrego, M., Susca, L., Morelli, M., Gonzalez-Dugo, V., North, P.R.J., Landa, B.B., Boscia, D., Saponari, M., Navas-Cortes, J.A., 2018. Previsual symptoms of Xylella fastidiosa infection revealed in spectral plant-trait alterations. Nat. Plants 4, 432. https://doi.org/10.1038/s41477-018-0189-7.

Zhang, Y., Ma, J., Liang, S., Li, X., Li, M., 2020. An Evaluation of Eight Machine Learning Regression Algorithms for Forest Aboveground Biomass Estimation from Multiple Satellite Data Products. Remote Sens. 12, 4015. https://doi.org/10.3390/rs12244015.

Zhao, C., Liu, L., Wang, J., Huang, W., Song, X., Li, C., 2005. Predicting grain protein content of winter wheat using remote sensing data based on nitrogen status and water stress. Int. J. Appl. Earth Obs. Geoinformation 7, 1–9. https://doi.org/10.1016/j.jag.2004.10.002.

Zhao, Y., Potgieter, A.B., Zhang, M., Wu, B., Hammer, G.L., 2020. Predicting Wheat Yield at the Field Scale by Combining High-Resolution Sentinel-2 Satellite Imagery and Crop Modelling. Remote Sens. 12, 1024. https://doi.org/10.3390/rs12061024.

Zhou, X., Kono, Y., Win, A., Matsui, T., Tanaka, T.S.T., 2021. Predicting within-field variability in grain yield and protein content of winter wheat using UAV-based multispectral imagery and machine learning approaches. Plant Prod. Sci. 24, 137–151. https://doi.org/10.1080/1343943X.2020.1819165.

Appendix C

Longmire, A.R., Poblete, T., Hornero, A., Chen, D., Zarco-Tejada, P.J., (accepted 2023-10-27). Estimation of grain protein content in commercial bread and durum wheat fields via traits inverted by radiative transfer modelling from Sentinel-2 timeseries. *ISPRS Journal of Photogrammetry and Remote Sensing*.

—

See overleaf.

ISPRS Journal of Photogrammetry and Remote Sensing xxx (xxxx) 1–14

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Estimation of grain protein content in commercial bread and durum wheat fields via traits inverted by radiative transfer modelling from Sentinel-2 timeseries

A Longmire [a], T Poblete [a, b], A Hornero [b, c], D Chen [a], P.J Zarco-Tejada [a, b, c]

[a] School of Agriculture, Food, and Ecosystem Sciences, Faculty of Science, University of Melbourne, Parkville 3010, Victoria, Australia
[b] Department of Infrastructure Engineering, Faculty of Engineering and Information Technology (FEIT), University of Melbourne, Parkville 3010, Victoria, Australia
[c] Instituto de Agricultura Sostenible (IAS), Consejo Superior de Investigaciones Científicas (CSIC), Avenida Menéndez Pidal s/n, 14004 Córdoba, Spain

ARTICLE INFO

Keywords:
Wheat
Grain protein content
Stress detection
Sentinel-2 time series
Radiative transfer model inversion
Machine learning

ABSTRACT

Wheat (*Triticum* spp.) is crucial to food security. Grain protein content (GPC) is key to its nutritional and economic value and is controlled by genetic and agronomic factors, soil properties and weather. GPC prediction from remote sensing could reduce nitrogen (N) losses, help management decisions, and improve profit. However, GPC prediction is complex because multiple plant traits influence GPC and their effects change through the growing season. Traits with known physiological links to GPC, which can be retrieved from imaging spectroscopy, include leaf area index (LAI), chlorophyll (C_{a+b}), and stress indicators. Further inspection of these and other traits retrieved from satellite data can advance research relevant to precision agriculture. Sentinel-2 (S2) timeseries (TS) were acquired for 6,355 ha of commercial dryland bread (*T. aestivum*) and durum (*T. durum*) wheat fields in south-east Australia through two consecutive years with dissimilar rainfall. Wheat growers provided ~ 92,000 GPC data points from harvester-mounted protein monitors. For each, C_{a+b}, leaf dry matter, leaf water content (C_w) and LAI were retrieved from the S2 images by radiative transfer model inversion. A gradient boosted machine learning algorithm was applied to analyse these traits' importance to GPC and to predict GPC in 30% of samples unseen by the algorithm in training. The strongest relationships between predicted and observed GPC (R^2 = 0.86, RMSE = 0.56 %), in a model built from five S2 images across a season, were better than those from single-date hyperspectral (HS). In severe water stress, LAI was the main predictor of GPC early in the season, but this switched to C_w later. Trait importance was more evenly distributed in milder conditions. S2 TS had a clear accuracy advantage over single-date S2 and HS, especially in benign conditions, emphasising the potential of S2 TS for large-scale GPC monitoring.

1. Introduction

Wheat (*Triticum* spp.) supplies around 20% of humans' carbohydrate (CHO) and protein intake (FAO, 2022). Grain protein content (GPC; %) dictates the price paid to growers and is influenced by the amount of nitrogen (N) in the canopy at anthesis (Zadoks stage Z65; Giuliani et al., 2011; Masoni et al., 2007; Zadoks et al., 1974), and the N accessible for uptake during grain filling (Gooding et al., 2007). However, total photosynthesis after Z65 controls protein dilution by new assimilates, a long-recognised inverse yield ~ GPC relationship (McNeal et al., 1978). Fertiliser N is a major cost and risky investment for grain growers (Monjardino et al., 2015), and entails a heavy environmental footprint (Galloway et al., 2017). The global urea price increased by > 400 % in two years to April 2022, (Baffes and Koh, 2022); relief is unlikely as fossil fuels are deeply embedded in fertiliser supply. Timely GPC estimates could improve farmer access to price premiums (Apan et al., 2006; Skerritt et al., 2002), guide N applications aimed at increasing protein and improve farm environmental and economic sustainability.

Leaf area index (LAI) and chlorophyll (C_{a+b}) content strongly influence assimilation (Wolanin et al., 2019), hence protein dilution, while above-ground N correlates with both C_{a+b} (Evans, 1989) and GPC (Feng et al., 2014; Xue et al., 2007; Zhao et al., 2005). C_{a+b}/N concentration and LAI affect both the total protein available for translocation (Masclaux-Daubresse et al., 2010) and, via photosynthetic capacity, the CHO source size. Despite these relationships, C_{a+b}/N may be only moderately correlated with GPC (Longmire et al., 2022; Zhao et al., 2005) because other factors, especially water, nutrient and temperature stress, also affect N/CHO sources and sinks, the rate and duration of post-anthesis photosynthesis and the fate of assimilates during grain filling (Asseng et al., 2002; Masclaux-Daubresse et al., 2010). These dynamics complicate GPC estimation relative to that of its contributory variables

https://doi.org/10.1016/j.isprsjprs.2023.10.018
Received 3 May 2023; Received in revised form 15 October 2023; Accepted 27 October 2023
0924-2716/

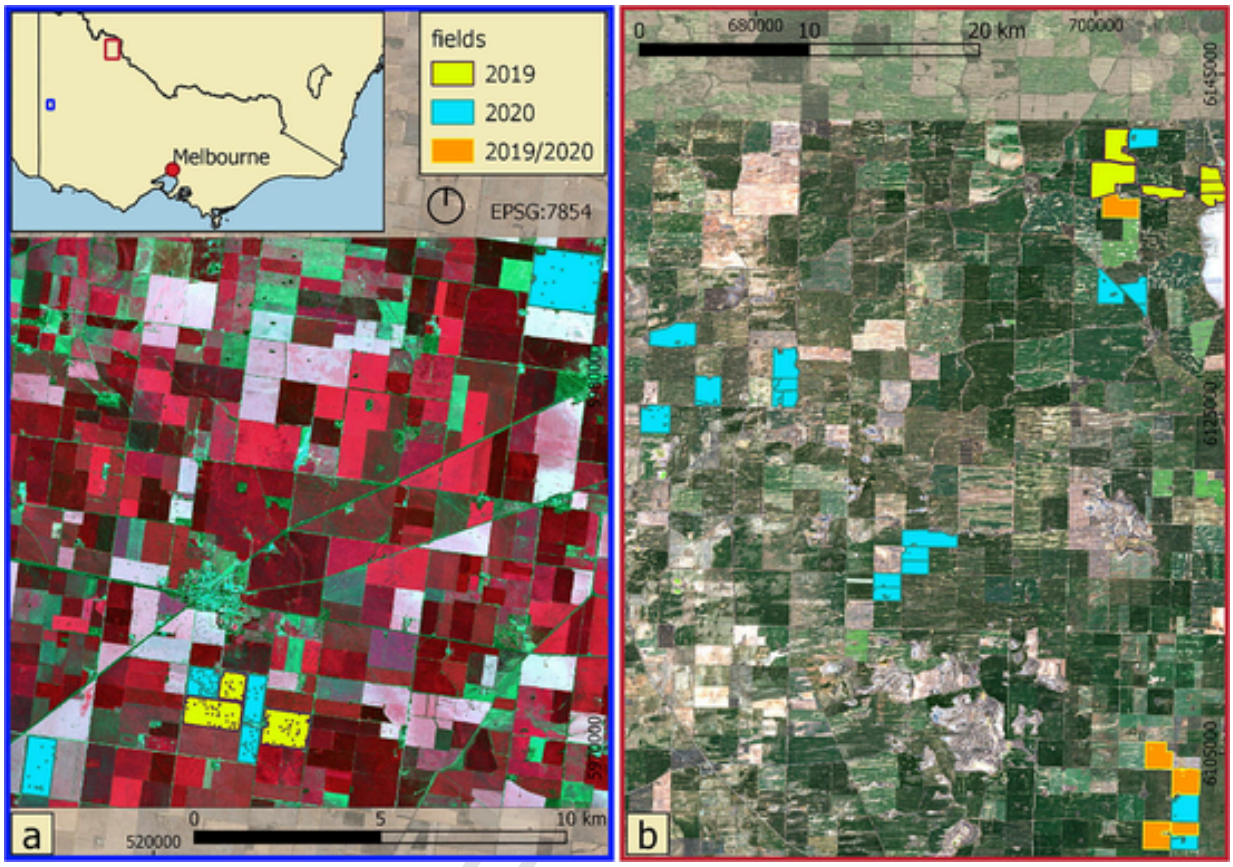

Fig. 1. Zone 1 (Kaniva; a) and zone 2 (Manangatang; b) commercial wheat fields considered in this study. Subfigure a: Sentinel-2, 2020–09-10 (B08, B04, B03); b; 2019–09-11 (B04, B03, B02).

Table 1
Location, rainfall, climate and crop areas for Kaniva (Zone 1) and Manangatang (Zone 2) in 2019–20. AAR/GSR = long term average annual / growing season (Apr–Oct) rainfall; rain = annual total (mm).

Zone	Lat.	Lon.	Alt. (m)	AAR	Clim. zone	Year	Rain	GSR	Area (Ha)
1	−36.37°	141.24°	142	449	Cfb	2019	288	238	662
						2020	444	291	858
2	−35.05°	142.88°	55	316	Bsk	2019	194	135	2341
						2020	342	277	2494

and yield, as seen in lower estimation accuracies (e.g. Rodrigues et al., 2018; Wright et al., 2004; Zhao et al., 2005).

Radiative transfer models (RTM) elucidate mechanistic relationships between leaf and canopy traits and light interactions (Jacquemoud et al., 1995; Jacquemoud and Baret, 1990). Because they are robust to non-linearity in these relationships and account for factors including leaf optical properties, soil reflectance and sun position, RTMs are transferable across crop types, phenology and agronomic situations (Clevers and Kooistra, 2012; Dorigo et al., 2007; Jacquemoud et al., 1995). Through inversion, they allow accurate estimation of plant traits relevant to GPC (Bacour et al., 2002; Féret et al., 2008; Jacquemoud et al., 2009, 1995; Li et al., 2015), where structural and physiological traits can be retrieved concurrently (Poblete et al., 2020; Zarco-Tejada et al., 2018). Recent studies have retrieved leaf C_{a+b}, canopy C_{a+b} content (CCC; Clevers et al., 2017; Houlès et al., 2007) and LAI from Sentinel-2 (S2) data by RTM inversion (Delloye et al., 2018; Pan et al., 2019; Upreti et al., 2019; Wolanin et al., 2020), but the work of Longmire et al. (2022) appears to remain unique in using RTM inversions to estimate GPC.

The multispectral instruments (MSI) aboard the S2 satellites offer cost-free images optimised for observing vegetation, with 13 bands at a ground resolution of 10–20 m in the visible, red edge (RE) and near-infrared domains (Drusch et al., 2012). The strong RE focus of the MSI, where it records three bands, is optimised for C_{a+b}, N content and LAI estimation (Frampton et al., 2013; Herrmann et al., 2011). A five-day revisit time facilitates timeseries (TS) observations. Hunt et al. (2019) obtained a substantial improvement in wheat yield estimation on adding a second S2 TS image, diminishing returns for further images but better performance closer to harvest, as Wolanin and colleagues (2020) saw. Similarly, Wang et al. (2014) reported poor correlation between early-season broadband vegetation indices (VI) and wheat GPC; relationships improved as the season progressed and stacked TS images were best. Cumulative VIs have also improved yield but not GPC estimates (Xue et al., 2007).

Longmire et al. (2022) demonstrated that stress-sensitive hyperspectral (HS) plant traits anthocyanins (Anth), C_{a+b}, and carotenoids (C_{x+c}), the Photochemical Reflectance Indices PRI (Gamon *et al.* 1992), PRI_{m3} and PRI_{m4} (Hernández-Clemente *et al.*, 2011) and solar-induced fluorescence (SIF) are associated with GPC. In water-stressed crops, the thermal crop water stress index (CWSI; Idso, 1982) also showed a strong, positive association with GPC. While the use of S2 data precludes retrieval of stress-related HS and thermal traits, it offers potentially major advantages for GPC estimation at yet larger scales and through TS.

The use of machine- and deep-learning algorithms with RS data is widespread in agriculture, including for disease detection and monitoring (e.g. Adam et al., 2017; Poblete et al., 2020; Zarco-Tejada et al., 2018), weed recognition (Gao et al., 2018), crop- and land use classification (Abdi, 2020; Ji et al., 2018), primary productivity and yield (e.g. Cheng et al., 2022; Gómez et al., 2021; Hunt et al., 2019; Wolanin et al.,

Table 2
Cloud-free Sentinel-2 images available in zones 1 and 2 in 2019 and 2020 with associated growing degree days after sowing (GDDAS; °C day) and Zadoks (Z) stage/name. Entries in bold are those compared against airborne hyperspectral analyses.

	Image		Bread			Durum		
Zone	Year	Date	GDDAS	Z stage	Z name	GDDAS	Z stage	Z name
1	2019	18-Jul	481	15	seedling	684	16	seedling
		23-Jul	537	15		739	17	
		17-Aug	770	17		972	31	stem elong.
		11-Sep	1008	32	stem elong.	1211	37	
		1-Oct	1237	43	booting	1439	52	ear emerg.
		21-Oct	**1508**	**69**	**anthesis**	**1710**	**74**	**grain fill**
	2020	17-Jul	452	14	seedling	560	15	seedling
		1-Aug	581	15		689	17	
		26-Aug	803	31	stem elong.	910	31	stem elong.
		10-Sep	979	32		1087	33	
		10-Oct	**1375**	**51**	**ear emerg.**	**1483**	**67**	**anthesis**
2	2019	17-Jul	904	17	seedling			
		28-Jul	1028	31	stem elong.			
		12-Aug	1169	31				
		17-Aug	1220	32				
		27-Aug	1325	32				
		1-Oct	**1778**	**54**	**ear emerg.**			
	2020	17-Jul	719	17	seedling			
		27-Jul	803	17				
		26-Aug	1099	32	stem elong.			
		31-Aug	1161	32				
		10-Sep	1307	34				
		15-Sep	**1385**	**42**	**booting**			

2020, 2019), and GPC estimation (Longmire et al., 2022; Tan et al., 2020; Zhou et al., 2021). Zhou et al. (2021) obtained their best GPC predictions with a random forest (RF) and found machine learning (ML) superior to traditional statistical methods. Gradient boosting machines (GBM) are a supervised ML algorithm based on the work of Friedman (2002, 2001) and developed by Chen and Guestrin (2016). The GBM has seen relatively little use in agriculture (van Klompenburg et al., 2020), but performed as well as other algorithms in estimating LAI and CCC in wheat (Upreti et al., 2019). The tree-based GBM algorithm is able to assess the relative contribution, termed importance or gain, of input features to target variable estimation (Abdi, 2020; Hunt et al., 2019). This greatly improves the algorithms' interpretability, which is crucial to current research into the dynamic and interacting plant traits influencing GPC.

Some studies estimating wheat GPC from satellites have had moderate success. Wright et al. (2004) found the green NDVI (Gitelson and Merzlyak, 1998) best for GPC (airborne R^2 = 0.53, satellite R^2 = 0.48). Liu *et al.* (2005) combined synthetic aperture radar with the structure insensitive pigment index (SIPI; Peñuelas et al., 1995), from Landsat data, with R^2 = 0.56. Also from Landsat, Zhao et al. (2005) reported relationships between VI_{green} (Gitelson et al., 2002) and GPC (R^2 = 0.46), while Feng et al. (2014) estimated GPC at field level in commercial wheat, combining MODIS satellite NDVI from two growth stages and obtaining R^2 = 0.567–0.632 and rRMSE = 0.141–0.144. From S2 data, Zhao et al. (2019) retrieved several plant N indicators with good fidelity, but these predicted GPC with lower skill (R^2 = 0.428–0.467); like others (Wang et al., 2014; Zhao et al., 2005, 2019), this study reported that growth stages ≥ Z65 offered the best estimations.

A large majority of studies estimating GPC from spectroscopy have used plot experiments (e.g. Raya-Sereno et al., 2021; Walsh et al., 2023; Wang et al., 2004; Zhao et al., 2005, 2019), and to date these appear largely to have used VIs, with no reference to RTM inversions. This research gap demands attention, particularly given the lack of consistency among VIs (Raya-Sereno et al., 2021). Moreover, few studies look at GPC variability within commercial fields; while Rodrigues et al. (2018) and Stoy et al. (2022) do so, considering natural soil variability, only Longmire et al. (2022) sample multiple fields. Finally, the additive combination of information from TS images is rare in the canon, has had limited success and, due to reliance on VIs, is likely poorly transferable between agronomic situations (Feng et al., 2014; Rodrigues et al., 2018; Xue et al., 2007).

To advance precision agriculture, there is a need to improve GPC prediction within fields but across large extents. The current research addresses the gaps identified above by combining satellite TS and RTM inversion to predict GPC within many commercial wheat fields, across regions with very different climatic and soil characteristics, diverse seasons and both bread- and durum cultivars. Leveraging the GBM's flexibility, interpretability and predictive skill, plant trait importances to GPC estimation and model performance are compared between S2 and airborne HS/thermal RS. The effects of bandset reduction and the facultative inclusion of airborne CWSI and/or SIF to S2 traits are tested. In the temporal dimension, trait importance dynamics and skill are comprehensively assessed through seasons, first with TS elements as separate models, then additively stacking them within site-years to form single predictive models.

2. Materials and methods

2.1. Study sites

This study considers 6355 Ha of rainfed commercial hard white bread (cv. Scepter, Vixen, Catapult) and durum (cv. Aurora, Bitalli) wheat crops grown in two areas of the southern Australian wheat belt. Both varieties were sown in late May and mid-June in 2019 and 2020 around Kaniva (zone 1; Fig. 1a), while bread wheat only was sown in similar periods around Manangatang (zone 2; Fig. 1b). Hereafter, the cropping zones are designated cz1 and cz2, with the year appended, e.g., cz1-19. Commercial cropping soils and fields in cz1 are described in Longmire et al. (2022); cz2 soils, Calcarosols in the Australian Soil Classification (Isbell, 2002), vary greatly across tens to hundreds of metres in clay fraction, subsoil constraints including B and Al, calcrete, and hardpans. These influence plant-available water (PAW) and N availability, root growth and harvest outcomes (Nuttall et al., 2003; Sadras et al., 2002). Climate, including Köppen-Geiger classification (Peel et al., 2007), and location details are provided in Table 1. Fertiliser was applied 1–3 times each season in each field, according to grower assessment of conditions, usually as urea, but data relating to these applications were not consistently available.

2.2. Data collection

Harvester-mounted NIR spectrometers (CropScan 3000/3300H, Next Instruments, Sydney, Australia) collected GPC during harvesting, with geolocation (±0.01 m) by real-time kinetic GPS. These devices use near infrared (NIR) spectroscopy (720–1100 nm) to estimate GPC for ~ 400 ml grain samples temporarily removed from the combine's clean grain elevator. These devices assess GPC with accuracy sufficient to meet legal requirements for weights and measures in the United States and Australia (Clancy and Heiken, n.d.). Readings are taken every 8–10 s, representing GSD ≈ 10–25 m parallel to harvester travel and, perpendicularly, equivalent to the swath width (12 m). Over cz1

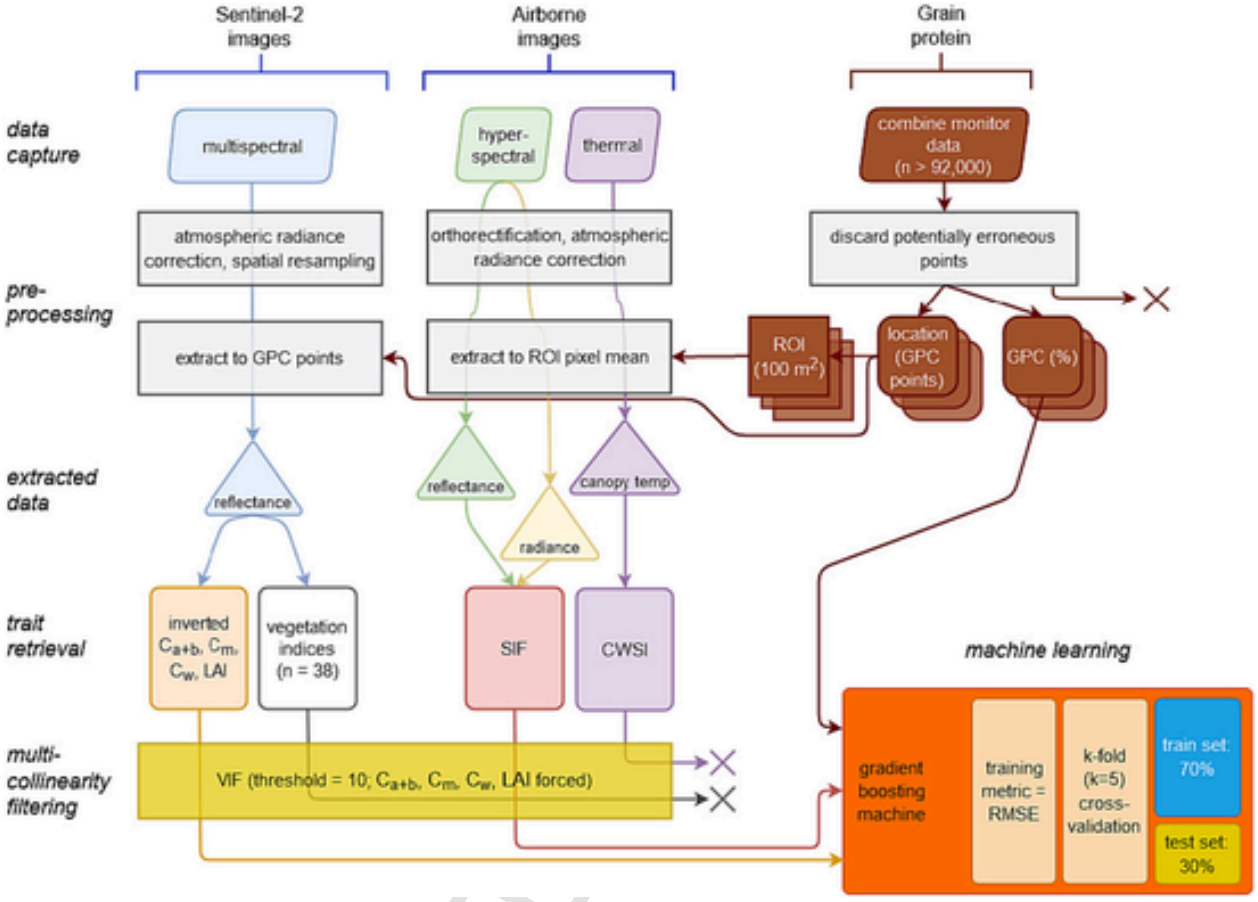

Fig. 2. Schematic summary of data handling and machine learning processes.

Table 3
Values and ranges of leaf and canopy traits used for PRO4SAIL (PROSPECT-D + 4SAIL) radiative transfer model inversion and look-up table generation.

Parameter	Abbreviation	Unit	Value/range
Anthocyanins	Anth	μg/cm²	1–10
Carotenoids	C_{x+c}		1–20
Chlorophyll *a* + *b*	C_{a+b}		3–70
Dry matter	C_m	g/cm²	0.001–0.035
Hot spot parameter	h	–	0.01
Leaf area index	LAI	m²/ m²	1–5
Leaf Inclination Dist. Func.	$LIDF_a$	°	0–90
Mesophyll struct. Coef.	N	–	0.5–3.0
Observer angle	tt_o	°	0
Relative azimuth angle	ψ		0
Solar zenith angle	tt_s		varied with date
Water content	C_w	g/cm²	0.001–0.035

only, HS and thermal data were collected by sensors on a light aircraft at ~ 2000 m above ground level (AGL) on 2019–10-22 and 2020–10-28. This gave pixels of 1.0 m (HS) and 1.7 m (thermal) ground sampling distance (GSD). HS data were collected with a VNIR E-Series model (Headwall Photonics, Fitchburg, MA, USA), capturing 371 bands from 400 to 1000 nm at 8 nm per pixel, yielding 5.8 nm FWHM with a 25 μm slit. At 12-bit radiometric resolution, the storage rate was 50 frames per second with an exposure time of 18 ms and an 8 mm focal length. The hyperspectral imager was calibrated using an integrating sphere (Labsphere XTH2000C, Labsphere Inc., North Sutton, NH, USA), deriving coefficients at four illumination levels. Thermal images were collected from 7.5 to 14 μm with an A655c camera (Teledyne FLIR LLC, Wilsonville, OR, USA), a scientific-grade instrument radiometrically calibrated by the manufacturer. A further indirect calibration was carried out during flights using ground observations from a handheld infrared thermometer (LaserSight from Optris GmbH, Berlin, Germany), after Calderón et al. (2015). Atmospheric correction of radiance was applied with the Simple Model of Atmospheric Radiative Transfer of Sunshine (SMARTS) model (Gueymard, 1995) using aerosol optical depth (AOD) observed at the time of flight (Micro-Tops II sunphotometer, Solar LIGHT Co., Philadelphia, PA, USA), as done before (Calderón et al., 2015; Poblete et al., 2020; Zarco-Tejada et al., 2018). Orthorectification was performed using Parametric Geocoding and Orthorectification for Airborne Optical Scanner Data (PARGE; ReSe applications GmbH, Wil, Switzerland) using an inertial measurement unit and GPS data from a VN-300 (VectorNav Technologies LLC, Dallas, TX, USA). As for S2 data, Level 1C orthorectified top-of-atmosphere reflectance (Richter *et al.*, 2011) rasters for tiles 54HWE (cz1), 54HXG and 54HYG (cz2) were downloaded from the Copernicus Open Access Hub () and atmospherically corrected to surface reflectance with Sen2Cor v. 2.3.1. Bands of GSD = 20 m were resampled to 10 m prior to stacking as a multiband raster. All images between 1 July and 31 October in each season were assessed; 23, in which all subject fields were cloud free (Table 2), were retained.

2.3. Data extraction and processing

Potentially erroneous GPC points were discarded: those with unrealistic GPC values, within 20 m of trees, dams, fences and headlands, or in harvester turn/slow travel zones. A Wilcoxon test (Bauer, 1972) was applied across a) sites within year/wheat type combinations and b) wheat types within site/year combinations to test for differences in median GPC. S2 spectra were extracted to the GPC points, then compatible

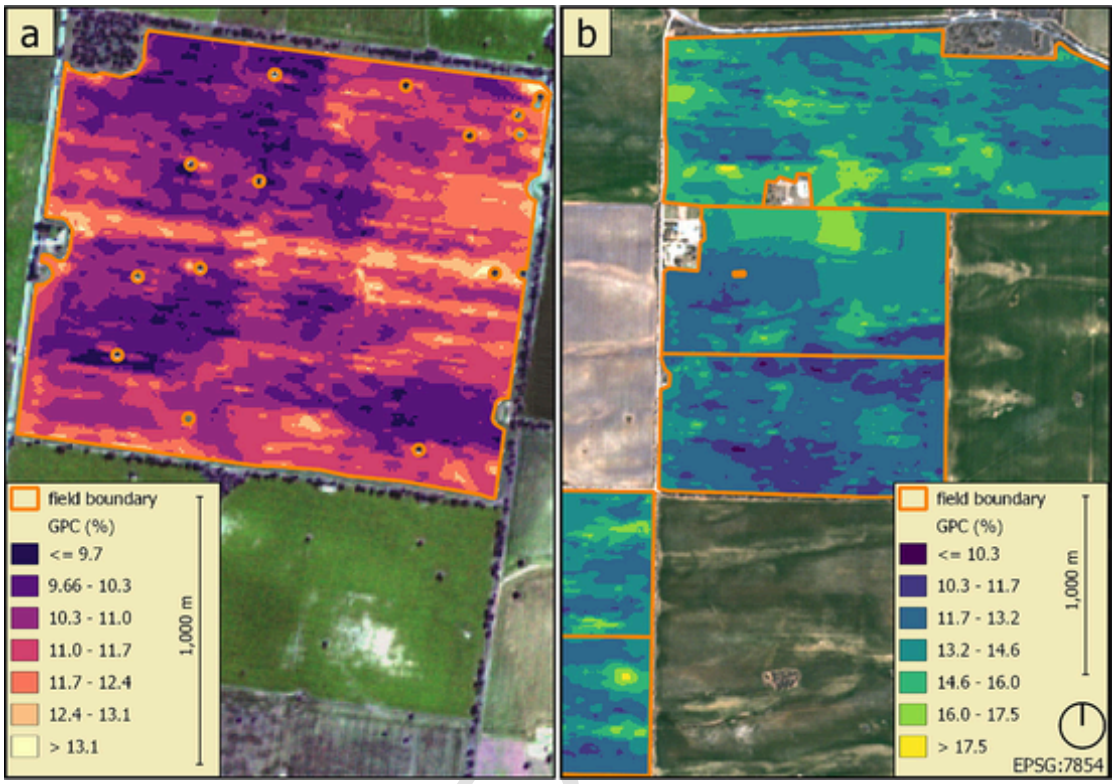

Fig. 3. Spatial variability in grain protein content (GPC; %) in wheat fields in zones 1 (a) and 2 (b) on Sentinel-2 images (R = band 2, G = band 3, B = band 4).

vegetation indices (VIs; Supplementary Table 1) were calculated – and inverted parameters retrieved (see Section 2.4) – for each point. Airborne narrow-band HS indices and inverted parameters, against which multispectral satellite results are compared, are detailed in Longmire et al. (2022); for these, CWSI and SIF (Supplementary Table 1), mean pixel values were calculated per 100 m² regions of interest (ROI). Each ROI had as its centroid a GPC point record, giving geolocation and GSD equivalent to the GPC data and quasi-equivalent to S2 data. SIF was calculated by the Fraunhofer line depth (FLD2) method (Plascyk and Gabriel, 1975). Data-handling processes are summarised in Fig. 2.

2.4. Radiative transfer model inversion

Leaf properties C_{a+b}, C_m and C_w were retrieved with PROSPECT-D (Féret et al., 2017), while LAI was modelled with 4SAILH (Verhoef et al., 2007), coupled as PRO4SAIL. Simulated spectra were generated by randomly sampling the full range of model parameters, from uniform distributions in appropriate ranges for wheat (Table 3; Camino et al., 2018; Li et al., 2015). The hybrid PRO4SAIL inversions, after Xu et al. (2019), used look-up tables (LUT) of 200,000 simulations, shown sufficient previously (Longmire et al., 2022; Poblete et al., 2021; Xu et al., 2019; Zarco-Tejada et al., 2018). The LUTs were interrogated with support vector machine (SVM) ML algorithms to retrieve each trait independently, using as input the simulated reflectance spectra, convolved to the S2 spectral specifications and the target traits as outputs; such hybrid methods effectively address the ill-posed problem (Verrelst et al., 2015). The SVM models were built in MATLAB (MATLAB; Statistics and Machine Learning toolbox and Deep Learning toolbox; Mathworks Inc., Natick, MA, USA) and trained using a radial basis function and SVM hyperparameters optimised during training for each variable. With these trained SVM models, plant traits were inverted from observed reflectance at each GPC point or ROI in each site/image combination, varying solar zenith angle for the changing dates.

2.5. Variance inflation factor analysis

Multicollinearity between potential model input features crop traits was inspected by variance inflation factor analysis (VIF; R package *fsmb*; (Nakazawa, 2022)). Like other recent work, this study used VIF thresholds (t) of 5–10 (Akinwande et al., 2015; Magney et al., 2016; Poblete et al., 2021; Zarco-Tejada et al., 2018). VIF was repeated, with forced inclusion of inverted parameters, for each permutation of site, year and product. None of the 38 VIs calculated was kept at t = 5; those kept at t = 10 were added stepwise to ML models, to assess their contributions to skill. VIs that survived the VIF analysis were: Global Environment Monitoring Index (GEMI; Pinty and Verstraete, 1992), Maccioni Index (Macc; Maccioni et al., 2001), MERIS Terrestrial Chlorophyll Index (MTCI; Dash and Curran, 2004), Transformed Chlorophyll Absorption in Reflectance Index / Optimized Soil Adjusted VI (TCARI/OSAVI; Haboudane et al., 2002) and $TCARI_{1610}$ (Herrmann et al., 2010). Each of these improved GPC estimation by $R^2 \leq 0.03$ over inverted traits only, so all were discarded. Airborne SIF was linearly independent so it was included where available. Features were also VIF tested along TS and between features within dates. Minor collinearity between inverted traits from close image dates (e.g. < 14 days apart) was disregarded in order to assess feature importance evolution.

2.6. Application of machine learning algorithm to estimate GPC

The study used a GBM to estimate relationships of inverted leaf and canopy traits – input features – with the target variable GPC, focusing on the features' relative importance. Feature importance is a unitless quantity in the range 0–1, expressing the relative gain contributed by each model input feature to estimation of the target variable. In each GBM run, data were randomly split 70%:30% into training and test sets (Hunt et al., 2019; Wu et al., 2021). Stochastic gradient descent (SGD) reduces the likelihood of overfitting by training models on random subsets of observations, introducing randomness (Friedman, 2002); here, models were trained on either 65% or 85% of rows, but all columns

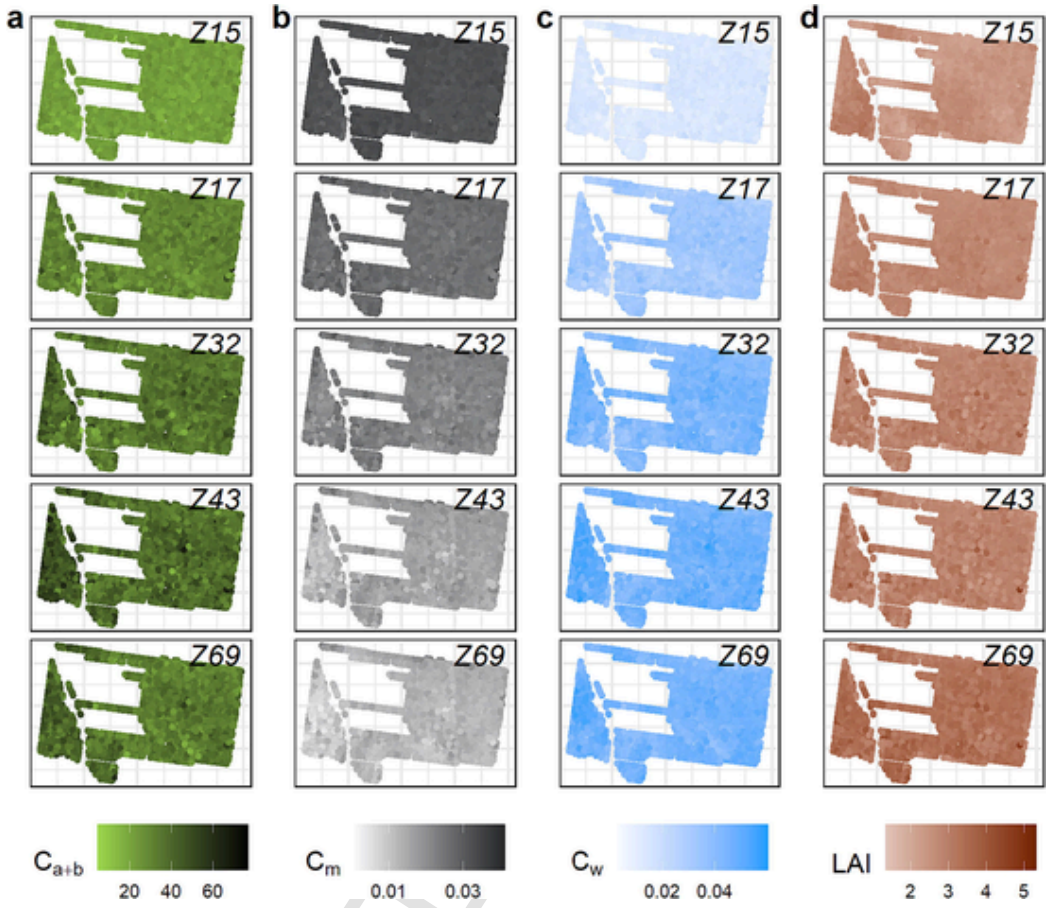

Fig. 4. Spatial variability in C_{a+b} (μg/cm²; a), C_m (g/cm²; b), C_w (g/cm²; c) and LAI (m²/m²; d) retrieved by radiative transfer model inversion from S2 images of a bread wheat field, 2019. Plots arranged by Zadoks growth stage; n = 4897.

were included in every run. Randomised K-fold cross-validation (K = 5) provided further protection against overfitting. In addition to SGD, learning rate, tree depth and minimum node size were varied, with full factorial hyperparameter searching. The GBM was tuned at each run by finding the hyperparameter combination that minimised root mean square error (RMSE) of prediction (GPC %). These procedures mirror previous work (Longmire et al., 2022).

The ML algorithm was first run for each site/year/product combination, using as inputs the plant traits from the S2 image closest to the HS flight date, the last in the TS. To these features, airborne SIF was facultatively added. Each date was analysed in turn, whereby the input feature set contained only the four traits retrieved from the relevant image. Finally, TS were combined such that each permutation of inverted parameter and image date was taken as an individual input feature, disregarding minor collinearity between inverted parameters in the temporal dimension. To facilitate comparison across years, sites and crop types, growing degree days after sowing (GDDAS) were calculated after Asseng *et al.* (2010), based on daily temperatures and precipitation specific to each location, drawn from the SILO dataset (Jeffrey et al., 2001). Given the large potential for differences in phenological advance between locations and cultivars, and the need to compare these directly, APSIM Next Generation (Holzworth et al., 2018) was used to model phenology in Zadoks stages. For APSIM, met data were from SILO, sowing date was the mean of fields in each year/location, and cultivars were those most planted (cz1 bread, cv. Scepter; cz1 durum, cv. Aurora; cz2-19, cv. Scepter; cz2-20, cv. Kord).

3. Results

3.1. Fields, GPC, retrieved parameters

Growing conditions differed strongly between years at both locations: both total and growing season rainfall were extremely low in 2019 while 2020 was above average (Table 1). A Wilcoxon test (Bauer, 1972) showed significant differences in GPC between all zone/year/product combinations; effect sizes were small to moderate (not shown). Large GPC differences were seen within and between paddocks, often over short distances (Fig. 3).

Plant traits for GPC points showed spatial heterogeneity and phenological progression for all site/year/product combinations. Example fields (cz1-19) are plotted to map traits over time for bread (Fig. 4) and to show progression with phenological advance, relationships between traits and density distributions within trait and stage combinations in durum (Fig. 5).

3.2. Feature importance and model performance

3.2.1. End-of-season Sentinel-2 images against hyperspectral images

The following considers first ML models built with traits inverted from S2 images captured as temporally close as possible to airborne HS missions detailed in Longmire et al. (2022). In cz1 bread, C_w was dominant in the very dry conditions of 2019, while in 2020 importance was spread evenly in a tight range (0.26–0.27) between C_{a+b}, C_w and C_m. In each year, LAI was the least important (Fig. 6a). cz1-19 durum had even importance across feature types, with C_w most important and C_{a+b} marginally higher than C_m, and in 2020 durum, feature importance was

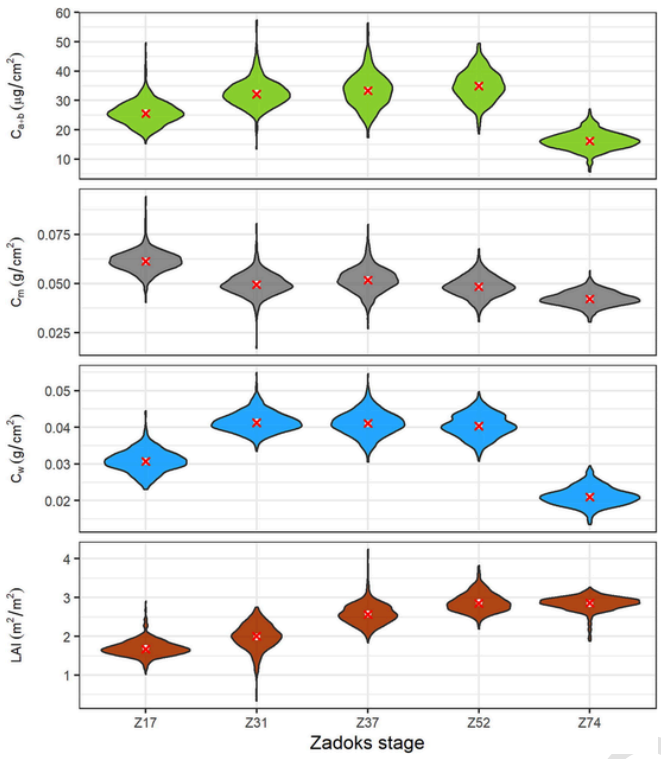

Fig. 5. Plant traits C_{a+b}, C_m, C_w and LAI inverted from S2 images of a durum wheat field, 2019. Violin plots show distribution within Zadoks (Z) growth stage; white circle = mean, red cross = median; n = 782. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

shared very evenly (Fig. 6b). Airborne SIF was added to S2 inverted features in each wheat type and year. SIF had low importance in cz1-19 bread and did not disturb the feature order relative to the base model, while in cz1-20 importance remained evenly distributed among C_m, C_w and C_{a+b} despite the inclusion of SIF (Fig. 6c). For cz1-19 durum, SIF was marginally more important than C_w but again feature order was otherwise unchanged compared to without SIF, while in cz1-20 durum SIF was most important and other features were approximately equal (Fig. 6d). In cz2–19 (bread wheat only), C_{a+b} was most important, with C_m and C_w intermediate and LAI lowest (not shown). In cz2-20, C_w took > 40% of importance over C_{a+b}, while C_m and C_w were the lowest.

Relative feature importance was similar between S2 and HS. Across conditions and wheat types, the relative importance of S2 traits was parallel with that of HS features grouped by their type: C_{a+b} (S2) with HS pigment indicators C_{a+b}, C_{x+c} and PRI; C_m with the HS structural reflectance index EVI; C_w with CWSI and LAI (S2) with HS LAI and $LIDF_a$ (Fig. 6).

Predictive skill for S2 ± SIF was highest in cz1-19 bread, as it was for HS ± CWSI previously (Longmire et al., 2022). Overall, S2 models were better predictors than HS models only in cz1-19 bread (Fig. 7a). Adding SIF to S2 traits made only a minor difference in that context but improved GPC estimation relatively more from a lower base in each other site/year/product combination (Fig. 7).

3.2.2. Assessment with individual images in timeseries

Analysis of individual S2 TS components showed that for cz1-19 bread, feature importance was concentrated in LAI until around anthesis (Z60—Z69), when C_w became predominant; neither C_m nor C_{a+b} took importance at any time (Fig. 8). In the more benign 2020, in zone 1 durum wheat in both years and in zone 2 (not shown), feature importance was spread evenly at each stage, changing relatively little and without a discernible pattern through the season.

Model predictive skill in cz1-19 bread wheat was high during Z15 but diminished from Z17 until after anthesis when the final image was captured (Fig. 9). All site/year/product combinations showed better predictive performance in early development than in the mid-season, usually with an increase late in the season.

3.2.3. Stacked timeseries images

Use of all retrieved parameters from all crop stages together revealed patterns consistent with analysis of individual images: high LAI importance early in the cz1-19 bread wheat season, switching to C_w around anthesis, and a relatively even distribution of importance across traits and stages in other crops (Fig. 10).

Predictive skill was compared between single image models and those incorporating all available trait/date combinations; the latter brought large improvements in all site/year/product combinations except cz1-19 bread (Table 4).

4. Discussion

4.1. Growing conditions and protein variability

GPC is a complex variable under genetic, environmental and management control (Zhao et al., 2019). While rainfall can be presumed invariant across fields within a region, and genetics within a field, soil properties vary widely within fields and have large effects on GPC independent of GSR. PAW differences influence GPC via lowered CHO assimilation, hence dilution, especially during grain filling. However, grain count is influenced by earlier conditions, especially around anthesis, is a primary determinant of sink size for both proteins and photosynthate and hence is a strong driver of GPC. Further, excessive early vigour in rainfed crops can dry soil so much that later photosynthesis is restricted, a phenomenon known as 'haying off' in which grain ends with high protein because it fails to fill with CHO (van Herwaarden et al., 1998).

Besides extreme dryness, cz1 saw other weather extremes during critical periods of 2019. From mid-September into early October, frosts (≈-4 °C) occurred at crucial stages (Z42—69; booting, ear emergence and anthesis) for the durum crop, while bread wheat fields were less affected due to their location higher in the landscape and less susceptible growth stages (Z33—51). Frost around anthesis severely reduces grain count, and durum is more susceptible to both frost and heat damage than bread wheat (Beres et al., 2020; McCallum et al., 2019). Immediately after these frosts, four days had maxima of 35–38 °C, imposing high respiratory loads (Heskel et al., 2016) and with potential to cause permanent damage, reduce total N uptake (van Ittersum et al., 2003) and/or simply exceed the optimum for photosynthesis (Asseng et al., 2011; Lobell and Gourdji, 2012). Heat stress can also severely alter phenology: At T > 34 °C, senescence is accelerated by a factor of three, and sixfold at T ≥ 36 °C, foreshortening grain filling (Asseng et al., 2011; Porter and Gawith, 1999). Moreover, droughted plants accumulate more heat. These effects reduce yield, especially if cumulative, and should therefore increase GPC (Asseng et al., 2011). Heat stress may therefore increase mean GPC but reduce its variability through a generalised reduction of assimilation. Lower assimilation during grain fill (cz1-19), associated with senescence, would also reduce variability in C_{a+b}, C_m, and SIF, diminishing these as GPC estimators, while likely enhancing the importance of moisture sufficiency measures, as we observed in both CWSI and C_w. These conditions likely affected both GPC and the difficulty of its prediction, differentially between bread and durum wheat.

In cz2, where rainfall is lower and crops are grown on former dunefields, phenomena influencing PAW and N, including both osmotic and physical impediments to root growth, arise from the dune-swale mor-

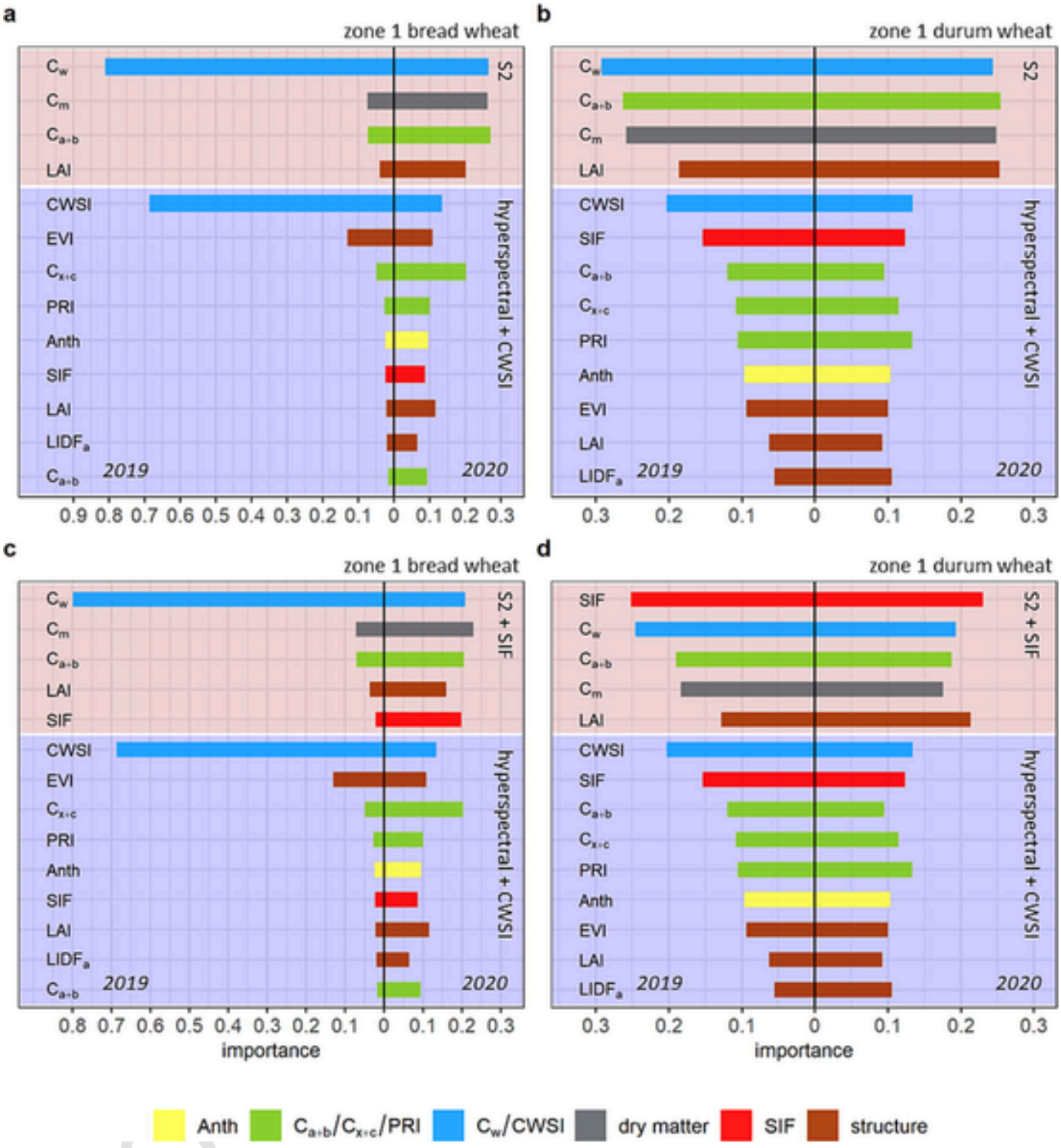

Fig. 6. Importance (proportion) of inverted traits to GPC (%) estimation in bread (a, c) and durum wheat (b, d) grown in zone 1, 2019–20. Brown background: S2 inverted traits: C_{a+b}, C_m, C_w and LAI, ± SIF. Blue background: EVI, PRI, Anth, C_{a+b}, C_{x+c}, LAI and $LIDF_a$ inverted from airborne hyperspectral images. CWSI and SIF were also from airborne data. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

phology and can be severe (Nuttall et al., 2003; Sadras et al., 2002). Though high-resolution soil data were unavailable, the influence of the cz2 dune-swale systems is seen in GPC (Fig. 3b); the dune effect on biomass can also be seen in the S2 RGB areas of this figure. In these areas, height within the dunes can be strongly discernible at harvest, including complete reversal of yield response between dune and swale between wet and dry years; mid-slope areas, which are relatively unaffected by soil variability, can occupy a high proportion of fields (Armstrong et al., 2009; Hoffmann et al., 2016). The severe soil variability, known yield response differences, and the proportion of fields affected, appear also to be reflected both in less definitive feature importance dynamics, and lower GPC estimation skill, as seen in cz2. This is supported by Rab et al. (2009): a large majority of the ~ 80 Ha they studied across four years in the same region showed great yield variability.

4.2. Plant trait contributions to protein estimation

To be important for GPC determination, a plant trait needs to vary across the crop in some physiological relationship with GPC. This biophysical reality is seen in traits' relative importance to GPC prediction across and between years. For single-image analyses of cz1 bread and

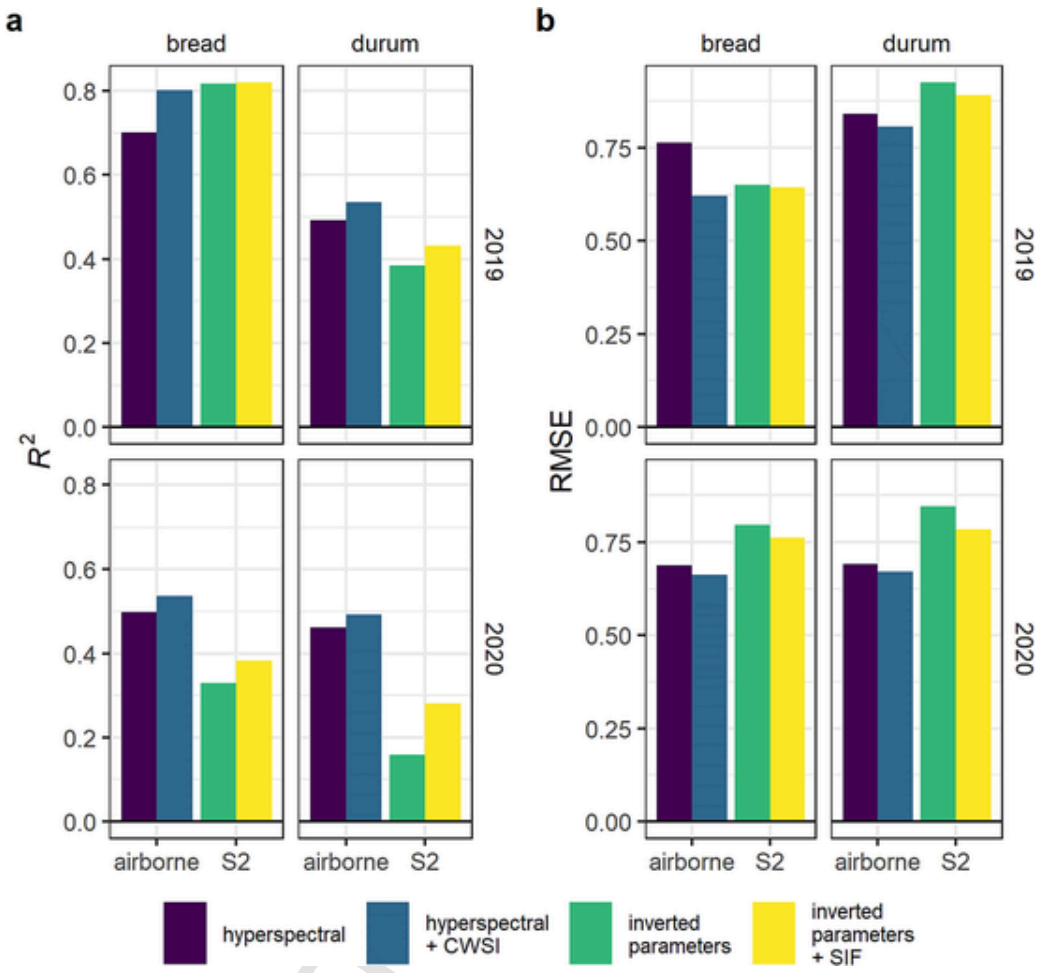

Fig. 7. Mean skill in GPC prediction of models comprising S2 traits C_{a+b}, C_m, C_w and LAI, ± airborne SIF and models comprising airborne EVI, PRI, Anth, C_{a+b}, C_{x+c}, LAI, $LIDF_a$ and SIF, ± CWSI from commercial bread and durum wheat fields in zone 1, 2019–20.

durum wheat, and regarding both relative trait importance and predictive skill, our results mirrored those from airborne HS data (Fig. 6, Fig. 7). As no VI was both linearly independent and a strong contributor to model skill, it is concluded that S2 data, with retrieval as shown, provide information adequate to estimate GPC. This accords with the findings of Wolanin and colleagues (2019) in estimating complex traits via RTM inversion and ML. While it contrasts with work based on HS imaging (Longmire et al. 2022), where several narrow-band indices complemented inverted parameters, it is as expected for data of lower spectral and spatial resolution. In lieu of CWSI, excluded as collinear, C_w assumed high importance in the very dry conditions of cz1-19 (Fig. 5a, c), suggesting that PROSAIL C_w retrieval can replace thermal observations. Under dry conditions, C_w apparently retains more variability and predictive power than other traits, especially C_{a+b} and C_m, whose low importance can be understood as complementarity but also highlights their low variability in such conditions. Physiologically, this is reasonable because they were established earlier in the season when water and nutrients were not limiting. C_{a+b}, C_m, and LAI contribute to the pools of both N and CHO available for translocation, a process impeded by water stress. Hence low water stress should increase these factors' influence on GPC, while conversely under high stress it is less than that of the ongoing photosynthetic rate. C_{a+b} and C_m importances are low throughout cz1-19 and vary least through all crops and seasons.

In contrast to VIs and CWSI, and despite its close links to C_{a+b} and C_w, SIF was linearly independent of all inverted parameters and where available was included in models. This independence, and the substantial extra skill it conferred to our predictions, are important findings in themselves. SIF proxies instantaneous photosynthesis, improving estimates of other complex physiological quantities (Camino et al., 2019); it follows that, as a measure of assimilation and hence protein dilution, it improves GPC estimation. Substantial improvements seen with SIF inclusion, especially where base model accuracy was low (Fig. 7), suggest that TS SIF could substantially improve GPC prediction, though perhaps with a strong role limited to grain filling and to benign moisture conditions. Indeed, like those of C_{a+b} and C_m, SIF contribution was minor in very dry conditions, as found by others (Cai et al., 2019; Sloat et al., 2021). Further, the relatively minor skill improvement on adding SIF in 2019 durum may relate to weather damage not picked up in the SIF signal but crucial to GPC. In mild conditions, the combined relationships of C_{a+b}, C_m and SIF with GPC were insufficient to offset the strong C_w ~ stomatal conductance ~ photosynthesis ~ GPC dynamic in drought. LAI had low importance in single-image S2 analyses, as it and other structural components did in our HS models.

Structure played a far greater role in TS analyses because it encompasses the early part of seasons. For example, when TS images were considered separately for cz1-19 bread, LAI took 60–80% of total importance at each stage up to Z65, whereafter the emphasis switched to C_w (Fig. 8a). When all dates were pooled, Z15 LAI took > 40% of importance, and C_w at anthesis took > 30% (Fig. 10a). This sudden change was seen only under drought, but a gradual switch from LAI to C_w was seen in cz1-19 durum also, despite a likely reduced crop water demand after frosting and noise from weather damage. In other situations, the distribution of importance between dates and traits was quite seen whether image dates were separate or pooled, but always included

Fig. 8. Relative importance (proportion; sum = 1 within each year/stage) of C_{a+b}, C_m, C_w and LAI inverted from timeseries S2 images and used as separate feature sets for GPC estimation in commercial bread (a) and durum wheat (b) in zone 1 in 2019 and 2020. Image capture dates by Zadoks stage from top down.

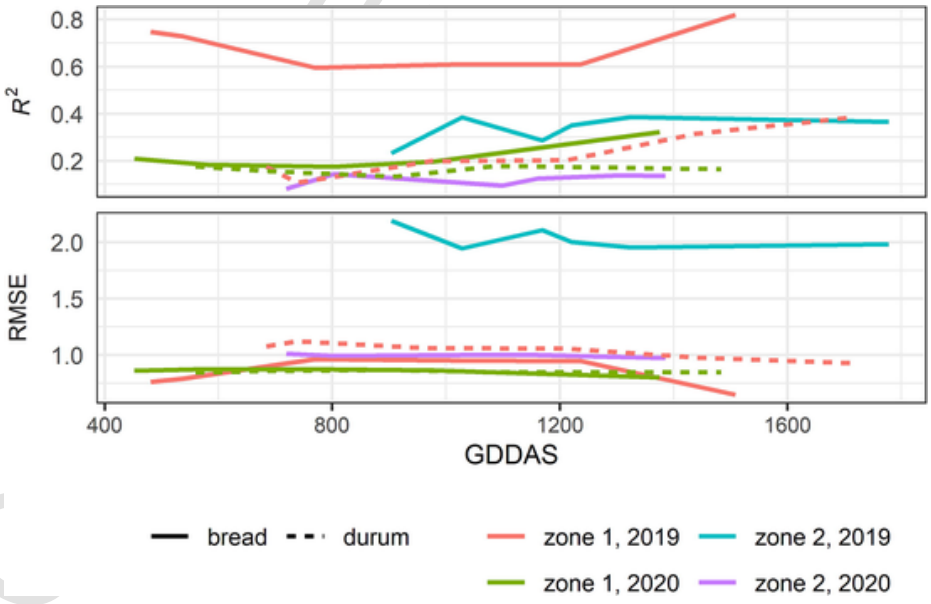

Fig. 9. Performance metrics (R^2 and RMSE) for GPC estimation from plant traits inverted from single S2 images of bread and durum wheat fields in cz1, 2019–2020. Metrics are shown as a function of growing degree days after sowing (GDDAS) at image capture.

substantial contributions from LAI. This, and the decline of LAI from a mid-season peak when all date/feature combinations were pooled, shows the high sensitivity of GPC to above-ground biomass, as a source of both proteins and CHO. However, PAW variability is lower early in the season partly because crops have had less time – and biomass – to withdraw water. The growth of C_w importance through seasons likely reflects increasingly variable soil moisture as plants differentially accumulate biomass, hence capacity to dry the soil, and soil properties exert more influence due to drying.

4.3. Model predictive skill

GPC predictions from single S2 images late in seasons were substantially less accurate than those from HS data (Longmire et al., 2022), except in very dry conditions (Fig. 7). The lower S2 spectral resolution explains this: indicators linked to GPC through their detection of mild stress, inverted Anth and C_{x+c}, and the PRI, cannot be calculated from S2 data so are absent from those analyses, but made substantial contributions to HS predictions. Here the advantages of S2 TS become clear,

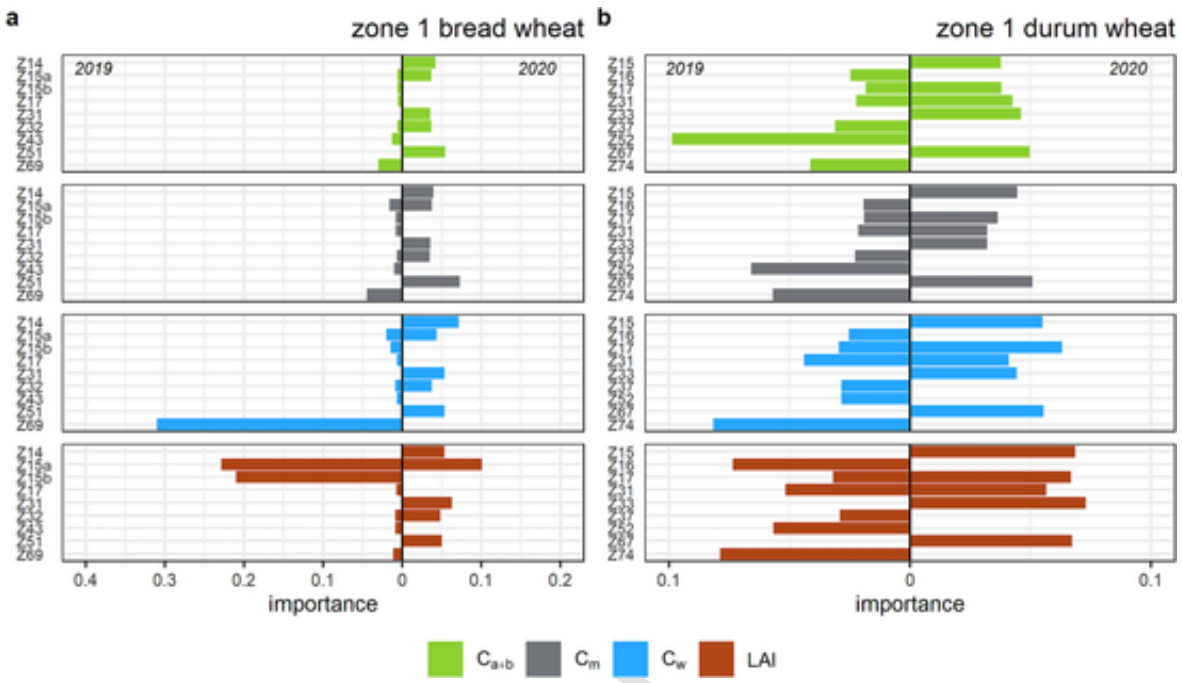

Fig. 10. Relative importance (proportion; sum = 1 within each year) of C_{a+b}, C_m, C_w and LAI inverted from timeseries S2 images and used together as a single training feature set for GPC estimation in commercial bread (a) and durum wheat (b) in zone 1, 2019–2020. Features are arranged by Zadoks growth stage at image capture from top down, within plant trait category.

Table 4
Model performance (R^2 and RMSE) for GPC estimation from single, late-season S2 images (base) and TS models from 5 to 6 S2 images across the same season, captured over commercial bread and durum wheat fields in cropping zones 1 and 2.

			base model		timeseries model	
zone	year	wheat type	R^2	RMSE	R^2	RMSE
CZ1	2019	bread	0.82	0.65	0.86	0.56
		durum	0.38	0.93	0.55	0.81
	2020	bread	0.33	0.80	0.52	0.67
		durum	0.16	0.85	0.43	0.70
CZ2	2019	bread	0.37	1.97	0.55	1.68
	2020	bread	0.14	0.98	0.31	0.90

whereby TS prediction metrics were as good, and sometimes better than those from HS (Fig. 7). The improvement over single-date S2 estimates was also large, especially in cz2, where they came off a low base (Table 4). These gains also come from the incorporation of early-season structural information and demonstrate that S2 TS can be used to predict GPC even in very difficult conditions. In all situations, predictive performance was better during early development than in the mid-season, confirming both that emergence and early vigour are important to GPC outcomes and that our ML approach is sensitive to the same biological reality.

Cai et al. (2019) assert that the optimal timing for Australian wheat yield prediction accuracy is before October; this may hold for GPC prediction also, given that the two quantities co-vary in opposition. Others contend that Z65 is effective (Tan et al., 2020; Zhao et al., 2019) and that the addition of extra S2 TS data after June gave little improvement (yield; England; Hunt et al., 2019). This study shows that despite the many factors that can intervene between potential and reality, observations of specific inverted parameters as early as Z15 can contribute to GPC prediction. It also confirms that later images bring higher accuracy, and that TS stacking further improves performance. A lack of late-season images leaves an unfortunate gap, notably in the milder, wetter 2020 when no image was available after Z50/1400 GDDAS in either zone but also in cz2-19. Performance may improve further with later cloud-free images, but the likelihood of finding these would not be substantially higher in any similarly mild season. Estimation of GPC is considerably more complex than estimation of intermediate quantities (Zhao et al., 2019). Nevertheless our accuracies (Table 4) are comparable to and often improve on other recent results from satellite RS: Zhao et al. (2019) recorded $0.428 \leq R^2 \leq 0.467$ for wheat GPC from S2, while Tan et al. (2020) achieved best metrics of $R^2 = 0.81$ and RMSE = 0.54 % from Landsat.

5. Conclusions

Plant and canopy traits, retrieved by radiative transfer modelling from Sentinel-2 image timeseries, were assessed for their contribution to, and ability to predict, wheat grain protein content in commercial fields and under diverse soil and weather. Using equivalent modelling methods from single images, GPC estimation accuracy was generally lower when based on S2 than on airborne HS traits. However, in very dry conditions, our best model using a single S2 image and made with PRO4SAIL-inverted C_{a+b}, C_m, C_w and LAI, outperformed that built with HS inputs and CWSI. Adding timeseries S2 inverted traits substantially improved all models over single-image versions; improvement was very strong in benign conditions and compensated for the accuracy reduction caused by switching from HS to S2. The best predictive performance was achieved by stacking retrieved parameters from all dates as inputs to a single model ($R^2 = 0.86$, RMSE = 0.56 %). The order and relative importance of S2 plant traits were similar to airborne HS: S2 importance was dominated by C_w in drought but evenly spread between structural and physiological features in benign conditions. The results obtained suggest potential applications in precision agriculture. Nevertheless, many refinements are possible: GPC and RS data from a wider range of seasonal and agronomic conditions, spaceborne SIF and ground-based sources such as soil moisture assessments should be tested for their impact on model skill.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors gratefully acknowledge Jonathan Dyer, Ryan O'Sullivan and Hayden Walters, grain growers and Simon Craig, grower/agronomist for access to GPC data. Andrew Longmire received PhD funding from the Grains Research and Development Corporation (GRDC) Australia. The GRDC had no role in designing or conducting the study.

Author Contributions

AL conceived the project, did the experimental design, negotiated access to GPC and other data at ground level, did the machine learning and data analyses, and wrote the manuscript. TP processed the image collections and performed the model inversions using radiative transfer models. AH obtained and pre-processed the Sentinel-2 data. AL, PZ and TP contributed to data interpretation, manuscript drafting and designed objectives and discussed results and conclusions. PZ supervised the work and led the airborne hyperspectral campaigns. DC supervised the work.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.isprsjprs.2023.10.018.

References

Abdi, A.M., 2020. Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data. GIScience & Remote Sensing 57, 1–20. https://doi.org/10.1080/15481603.2019.1650447.

Adam, E., Deng, H., Odindi, J., Abdel-Rahman, E.M., Mutanga, O., 2017. Detecting the Early Stage of Phaeosphaeria Leaf Spot Infestations in Maize Crop Using *In Situ* Hyperspectral Data and Guided Regularized Random Forest Algorithm. J. Spectrosc. 2017, 1. https://doi.org/10.1155/2017/6961387.

Akinwande, M.O., Dikko, H.G., Samson, A., 2015. Variance Inflation Factor: As a Condition for the Inclusion of Suppressor Variable(s) in Regression Analysis. Open J. Stat. 5, 754. https://doi.org/10.4236/ojs.2015.57075.

Apan, A., Kelly, R., Phinn, S., Strong, W., Lester, D., Butler, D., Robson, A., 2006. Predicting grain protein content in wheat using hyperspectral sensing of in-season crop canopies and partial least squares regression. Int. J. Geoinform. 2, 93–108.

Armstrong, R.D., Fitzpatrick, J., Rab, M.A., Abuzar, M., Fisher, P.D., O'Leary, G.J., 2009. Advances in precision agriculture in south-eastern Australia. III. Interactions between soil properties and water use help explain spatial variability of crop production in the Victorian Mallee. Crop Pasture Sci. 60, 870–884. https://doi.org/10.1071/CP08349.

Asseng, S., Bar-Tal, A., Bowden, J.W., Keating, B.A., Van Herwaarden, A., Palta, J.A., Huth, N.I., Probert, M.E., 2002. Simulation of grain protein content with APSIM-Nwheat. Eur. J. Agron. 16, 25–42. https://doi.org/10.1016/S1161-0301(01)00116-2.

Asseng, S., Foster, I., Turner, N.C., 2011. The impact of temperature variability on wheat yields. Glob. Chang. Biol. 17, 997–1012. https://doi.org/10.1111/j.1365-2486.2010.02262.x.

Bacour, C., Jacquemoud, S., Leroy, M., Hautecœur, O., Weiss, M., Prévot, L., Bruguier, N., Chauki, H., 2002. Reliability of the estimation of vegetation characteristics by inversion of three canopy reflectance models on airborne POLDER data. Agronomie 22, 555–565. https://doi.org/10.1051/agro:2002039.

Baffes, J., Koh, W.C., 2022. Fertilizer prices expected to remain higher for longer. World Bank Blogs. accessed 2022-05-26.

Bauer, D.F., 1972. Constructing Confidence Sets Using Rank Statistics. J. Am. Stat. Assoc. 67, 687–690. https://doi.org/10.1080/01621459.1972.10481279.

Beres, B.L., Rahmani, E., Clarke, J.M., Grassini, P., Pozniak, C.J., Geddes, C.M., Porker, K.D., May, W.E., Ransom, J.K., 2020. A Systematic Review of Durum Wheat: Enhancing Production Systems by Exploring Genotype, Environment, and Management (G × E × M) Synergies. Front. Plant Sci. 11. https://doi.org/10.3389/fpls.2020.568657.

Cai, Y., Guan, K., Lobell, D., Potgieter, A.B., Wang, S., Peng, J., Xu, T., Asseng, S., Zhang, Y., You, L., Peng, B., 2019. Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches. Agric. For. Meteorol. 274, 144–159. https://doi.org/10.1016/j.agrformet.2019.03.010.

Calderón, R., Navas-Cortés, J.A., Zarco-Tejada, P.J., 2015. Early Detection and Quantification of Verticillium Wilt in Olive Using Hyperspectral and Thermal Imagery over Large Areas. Remote Sens. (Basel) 7, 5584–5610. https://doi.org/10.3390/rs70505584.

Camino, C., González-Dugo, V., Hernández, P., Sillero, J.C., Zarco-Tejada, P.J., 2018. Improved nitrogen retrievals with airborne-derived fluorescence and plant traits quantified from VNIR-SWIR hyperspectral imagery in the context of precision agriculture. Int. J. Appl. Earth Obs. Geoinf. 70, 105–117. https://doi.org/10.1016/j.jag.2018.04.013.

Camino, C., Gonzalez-Dugo, V., Hernandez, P., Zarco-Tejada, P.J., 2019. Radiative transfer Vcmax estimation from hyperspectral imagery and SIF retrievals to assess photosynthetic performance in rainfed and irrigated plant phenotyping trials. Remote Sens. Environ. 231, 111186. https://doi.org/10.1016/j.rse.2019.05.005.

Chen, T., Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16. Association for Computing Machinery, New York, NY, USA, pp. 785–794. Doi: 10.1145/2939672.2939785.

Cheng, E., Zhang, B., Peng, D., Zhong, L., Yu, L., Liu, Y., Xiao, C., Li, C., Li, X., Chen, Y., Ye, H., Wang, H., Yu, R., Hu, J., Yang, S., 2022. Wheat yield estimation using remote sensing data based on machine learning approaches. Front. Plant Sci. 13. https://doi.org/10.3389/fpls.2022.1090970.

Clancy, P., Heiken, D., n.d. Montana Field Trials for On Combine NIR Analyzer. URL https://cropscanag.com/wp-content/uploads/2019/02/2015_Montana_Field_Trials_for_On_Combine_NIR_Analyzer.pdf.

Clevers, J.G.P.W., Kooistra, L., 2012. Using Hyperspectral Remote Sensing Data for Retrieving Canopy Chlorophyll and Nitrogen Content. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 5, 574–583. https://doi.org/10.1109/JSTARS.2011.2176468.

Clevers, J.G.P.W., Kooistra, L., Van den Brande, M.M.M., 2017. Using Sentinel-2 Data for Retrieving LAI and Leaf and Canopy Chlorophyll Content of a Potato Crop. Remote Sens. (Basel) 9, 405. https://doi.org/10.3390/rs9050405.

Dash, J., Curran, P.J., 2004. The MERIS terrestrial chlorophyll index. Int. J. Remote Sens. 25, 5403–5413. https://doi.org/10.1080/0143116042000274015.

Delloye, C., Weiss, M., Defourny, P., 2018. Retrieval of the canopy chlorophyll content from Sentinel-2 spectral bands to estimate nitrogen uptake in intensive winter wheat cropping systems. Remote Sens. Environ. 216, 245–261. https://doi.org/10.1016/j.rse.2018.06.037.

Dorigo, W.A., Zurita-Milla, R., de Wit, A.J.W., Brazile, J., Singh, R., Schaepman, M.E., 2007. A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. International Journal of Applied Earth Observation and Geoinformation, Advances in airborne electromagnetics and remote sensing of agro-ecosystems 9, 165–193. https://doi.org/10.1016/j.jag.2006.05.003.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., Bargellini, P., 2012. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. Remote Sensing of Environment, The Sentinel Missions - New Opportunities for Science 120, 25–36. https://doi.org/10.1016/j.rse.2011.11.026.

Evans, J.R., 1989. Photosynthesis and Nitrogen Relationships in Leaves of C Plants. Oecologia 78, 9–19.

FAO, 2022. FAOSTAT (production statistics), Crops and livestock products. Food and Agriculture Organization of the United Nations Statistics Division, Rome, Italy.

Feng, M., Xiao, L., Zhang, M., Yang, W., Ding, G., 2014. Integrating Remote Sensing and GIS for Prediction of Winter Wheat (*Triticum aestivum*) Protein Contents in Linfen (Shanxi). China. PLOS ONE 9, e80989.

Féret, J.-B., François, C., Asner, G.P., Gitelson, A.A., Martin, R.E., Bidel, L.P.R., Ustin, S.L., le Maire, G., Jacquemoud, S., 2008. PROSPECT-4 and 5: Advances in the leaf optical properties model separating photosynthetic pigments. Remote Sens. Environ. 112, 3030–3043. https://doi.org/10.1016/j.rse.2008.02.012.

Féret, J.-B., Gitelson, A.A., Noble, S.D., Jacquemoud, S., 2017. PROSPECT-D: Towards modeling leaf optical properties through a complete lifecycle. Remote Sens. Environ. 193, 204–215. https://doi.org/10.1016/j.rse.2017.03.004.

Frampton, W.J., Dash, J., Watmough, G., Milton, E.J., 2013. Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. ISPRS J. Photogramm. Remote Sens. 82, 83–92. https://doi.org/10.1016/j.isprsjprs.2013.04.007.

Friedman, J.H., 2001. Greedy Function Approximation: A Gradient Boosting Machine. The Annals of Statistics 29, 1189–1232.

Friedman, J.H., 2002. Stochastic gradient boosting. Computational Statistics & Data Analysis, Nonlinear Methods and Data Mining 38, 367–378. https://doi.org/10.1016/S0167-9473(01)00065-2.

Galloway, J.N., Leach, A.M., Erisman, J.W., Bleeker, A., 2017. Nitrogen: the historical progression from ignorance to knowledge, with a view to future solutions. Soil Res. 55, 417–424. https://doi.org/10.1071/SR16334.

Gao, J., Nuyttens, D., Lootens, P., He, Y., Pieters, J.G., 2018. Recognising weeds in a maize crop using a random forest machine-learning algorithm and near-infrared snapshot mosaic hyperspectral imagery. Biosyst. Eng. 170, 39–50. https://doi.org/10.1016/j.biosystemseng.2018.03.006.

Gitelson, A.A., Kaufman, Y.J., Stark, R., Rundquist, D., 2002. Novel algorithms for remote estimation of vegetation fraction. Remote Sens. Environ. 80, 76–87. https://doi.org/10.1016/S0034-4257(01)00289-9.

Gitelson, A.A., Merzlyak, M.N., 1998. Remote sensing of chlorophyll concentration in higher plant leaves. Advances in Space Research, Synergistic Use of Multisensor Data for Land Processes 22, 689–692. https://doi.org/10.1016/S0273-1177(97)01133-2.

Giuliani, M.M., Giuzio, L., Caro, A.D., Flagella, Z., 2011. Relationships between Nitrogen Utilization and Grain Technological Quality in Durum Wheat: I. Nitrogen Translocation and Nitrogen Use Efficiency for Protein. Agron. J. 103, 1487–1494. https://doi.org/10.2134/agronj2011.0153.

Gómez, D., Salvador, P., Sanz, J., Casanova, J.L., 2021. Modelling wheat yield with

antecedent information, satellite and climate data using machine learning methods in Mexico. Agric. For. Meteorol. 300, 108317. https://doi.org/10.1016/j.agrformet.2020.108317.
Gooding, M.J., Gregory, P.J., Ford, K.E., Ruske, R.E., 2007. Recovery of nitrogen from different sources following applications to winter wheat at and after anthesis. Field Crop Res 100, 143–154. https://doi.org/10.1016/j.fcr.2006.06.002.
Gueymard, C., 1995. Simple Model for the Atmospheric Radiative Transfer of Sunshine (SMARTS2) Algorithms and performance assessment. University of Central Florida, Florida Solar Energy Center.
Haboudane, D., Miller, J.R., Tremblay, N., Zarco-Tejada, P.J., Dextraze, L., 2002. Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. Remote Sens. Environ. 81, 416–426. https://doi.org/10.1016/S0034-4257(02)00018-4.
Herrmann, I., Karnieli, A., Bonfil, D.J., Cohen, Y., Alchanatis, V., 2010. SWIR-based spectral indices for assessing nitrogen content in potato fields. Int. J. Remote Sens. 31, 5127–5143. https://doi.org/10.1080/01431160903283892.
Herrmann, I., Pimstein, A., Karnieli, A., Cohen, Y., Alchanatis, V., Bonfil, D.J., 2011. LAI assessment of wheat and potato crops by VENμS and Sentinel-2 bands. Remote Sens. Environ. 115, 2141–2151. https://doi.org/10.1016/j.rse.2011.04.018.
Heskel, M.A., O'Sullivan, O.S., Reich, P.B., Tjoelker, M.G., Weerasinghe, L.K., Penillard, A., Egerton, J.J.G., Creek, D., Bloomfield, K.J., Xiang, J., Sinca, F., Stangl, Z.R., Martinez-de la Torre, A., Griffin, K.L., Huntingford, C., Hurry, V., Meir, P., Turnbull, M.H., Atkin, O.K., 2016. Convergence in the temperature response of leaf respiration across biomes and plant functional types. Proceedings of the National Academy of Sciences 113, 3832–3837. Doi: 10.1073/pnas.1520282113.
Hoffmann, M.P., Llewellyn, R.S., Davoren, C.W., Whitbread, A.M., 2016. Assessing the Potential for Zone-Specific Management of Cereals in Low-Rainfall South-Eastern Australia: Combining On-Farm Results and Simulation Analysis. J Agro Crop Sci n/a-n/a. https://doi.org/10.1111/jac.12159.
Holzworth, D., Huth, N.I., Fainges, J., Brown, H., Zurcher, E., Cichota, R., Verrall, S., Herrmann, N.I., Zheng, B., Snow, V., 2018. APSIM Next Generation: Overcoming challenges in modernising a farming systems model. Environ. Model. Softw. 103, 43–51. https://doi.org/10.1016/j.envsoft.2018.02.002.
Houlès, V., Guérif, M., Mary, B., 2007. Elaboration of a nitrogen nutrition indicator for winter wheat based on leaf area index and chlorophyll content for making nitrogen recommendations. Eur. J. Agron. 27, 1–11. https://doi.org/10.1016/j.eja.2006.10.001.
Hunt, M.L., Blackburn, G.A., Carrasco, L., Redhead, J.W., Rowland, C.S., 2019. High resolution wheat yield mapping using Sentinel-2. Remote Sens. Environ. 233, 111410. https://doi.org/10.1016/j.rse.2019.111410.
Isbell, R., 2002. The Australian soil classification. CSIRO publishing, Melbourne, Australia.
Jacquemoud, S., Baret, F., 1990. PROSPECT: A model of leaf optical properties spectra. Remote Sens. Environ. 34, 75–91. https://doi.org/10.1016/0034-4257(90)90100-Z.
Jacquemoud, S., Baret, F., Andrieu, B., Danson, F.M., Jaggard, K., 1995. Extraction of vegetation biophysical parameters by inversion of the PROSPECT + SAIL models on sugar beet canopy reflectance data. Application to TM and AVIRIS sensors. Remote Sens. Environ. 52, 163–172. https://doi.org/10.1016/0034-4257(95)00018-V.
Jacquemoud, S., Verhoef, W., Baret, F., Bacour, C., Zarco-Tejada, P.J., Asner, G.P., François, C., Ustin, S.L., 2009. PROSPECT+SAIL models: A review of use for vegetation characterization. Remote Sensing of Environment, Imaging Spectroscopy Special Issue 113, S56–S66. https://doi.org/10.1016/j.rse.2008.01.026.
Jeffrey, S.J., Carter, J.O., Moodie, K.B., Beswick, A.R., 2001. Using spatial interpolation to construct a comprehensive archive of Australian climate data. Environ. Model. Softw. 16, 309–330. https://doi.org/10.1016/S1364-8152(01)00008-1.
Ji, S., Zhang, C., Xu, A., Shi, Y., Duan, Y., 2018. 3D Convolutional Neural Networks for Crop Classification with Multi-Temporal Remote Sensing Images. Remote Sens. (Basel) 10, 75. https://doi.org/10.3390/rs10010075.
Li, Z., Jin, X., Wang, J., Yang, G., Nie, C., Xu, X., Feng, H., 2015. Estimating winter wheat (*Triticum aestivum*) LAI and leaf chlorophyll content from canopy reflectance data by integrating agronomic prior knowledge with the PROSAIL model. Int. J. Remote Sens. 36, 2634–2653. https://doi.org/10.1080/01431161.2015.1041176.
Lobell, D.B., Gourdji, S.M., 2012. The Influence of Climate Change on Global Crop Productivity. Plant Physiol. 160, 1686–1697. https://doi.org/10.1104/pp.112.208298.
Longmire, A.R., Poblete, T., Hunt, J.R., Chen, D., Zarco-Tejada, P.J., 2022. Assessment of crop traits retrieved from airborne hyperspectral and thermal remote sensing imagery to predict wheat grain protein content. ISPRS J. Photogramm. Remote Sens. 193, 284–298. https://doi.org/10.1016/j.isprsjprs.2022.09.015.
Maccioni, A., Agati, G., Mazzinghi, P., 2001. New vegetation indices for remote measurement of chlorophylls based on leaf directional reflectance spectra. J. Photochem. Photobiol. B Biol. 61, 52–61. https://doi.org/10.1016/S1011-1344(01)00145-2.
Magney, T.S., Eitel, J.U.H., Huggins, D.R., Vierling, L.A., 2016. Proximal NDVI derived phenology improves in-season predictions of wheat quantity and quality. Agric. For. Meteorol. 217, 46–60. https://doi.org/10.1016/j.agrformet.2015.11.009.
Masclaux-Daubresse, C., Daniel-Vedele, F., Dechorgnat, J., Chardon, F., Gaufichon, L., Suzuki, A., 2010. Nitrogen uptake, assimilation and remobilization in plants: challenges for sustainable and productive agriculture. Ann. Bot. 105, 1141–1157. https://doi.org/10.1093/aob/mcq028.
Masoni, A., Ercoli, L., Mariotti, M., Arduini, I., 2007. Post-anthesis accumulation and remobilization of dry matter, nitrogen and phosphorus in durum wheat as affected by soil type. Eur. J. Agron. 26, 179–186. https://doi.org/10.1016/j.eja.2006.09.006.
McCallum, M., Peirce, C., Porker, K., 2019. What drives the yield gap between durum and bread wheat?, in: Proceedings of the 19th Australian Agronomy Conference. Australian Society of Agronomy, Wagga Wagga, N.S.W., p. 4.
McNeal, F.H., McGuire, C.F., Berg, M.A., 1978. Recurrent Selection for Grain Protein Content in Spring Wheat. Crop Sci. 18, 779–782. https://doi.org/10.2135/cropsci1978.0011183X001800050022x.
Monjardino, M., McBeath, T., Ouzman, J., Llewellyn, R., Jones, B., 2015. Farmer risk-aversion limits closure of yield and profit gaps: A study of nitrogen management in the southern Australian wheatbelt. Agr. Syst. 137, 108–118. https://doi.org/10.1016/j.agsy.2015.04.006.
Nakazawa, M., 2022. Functions for Medical Statistics Book with some Demographic Data. R package "fsmb.".
Nuttall, J.G., Armstrong, R.D., Connor, D.J., Matassa, V.J., 2003. Interrelationships between edaphic factors potentially limiting cereal growth on alkaline soils in north-western Victoria. Aust. J. Soil Res. 41, 277. https://doi.org/10.1071/SR02022.
Pan, H., Chen, Z., Ren, J., Li, H., Wu, S., 2019. Modeling Winter Wheat Leaf Area Index and Canopy Water Content With Three Different Approaches Using Sentinel-2 Multispectral Instrument Data. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 12, 482–492. https://doi.org/10.1109/JSTARS.2018.2855564.
Peel, M.C., Finlayson, B.L., McMahon, T.A., 2007. Updated world map of the Köppen-Geiger climate classification. Hydrol. Earth Syst. Sci. 11, 1633–1644. https://doi.org/10.5194/hess-11-1633-2007.
Peñuelas, J., Filella, I., Gamon, J.A., 1995. Assessment of photosynthetic radiation-use efficiency with spectral reflectance. New Phytol. 131, 291–296. https://doi.org/10.1111/j.1469-8137.1995.tb03064.x.
Pinty, B., Verstraete, M.M., 1992. GEMI: a non-linear index to monitor global vegetation from satellites. Vegetatio 101, 15–20. https://doi.org/10.1007/BF00031911.
Plascyk, J.A., Gabriel, F.C., 1975. The Fraunhofer Line Discriminator MKII-An Airborne Instrument for Precise and Standardized Ecological Luminescence Measurement. IEEE Trans. Instrum. Meas. 24, 306–313. https://doi.org/10.1109/TIM.1975.4314448.
Poblete, T., Camino, C., Beck, P.S.A., Hornero, A., Kattenborn, T., Saponari, M., Boscia, D., Navas-Cortes, J.A., Zarco-Tejada, P.J., 2020. Detection of *Xylella fastidiosa* infection symptoms with airborne multispectral and thermal imagery: Assessing bandset reduction performance from hyperspectral analysis. ISPRS J. Photogramm. Remote Sens. 162, 27–40. https://doi.org/10.1016/j.isprsjprs.2020.02.010.
Poblete, T., Navas-Cortes, J.A., Camino, C., Calderon, R., Hornero, A., Gonzalez-Dugo, V., Landa, B.B., Zarco-Tejada, P.J., 2021. Discriminating *Xylella fastidiosa* from Verticillium dahliae infections in olive trees using thermal- and hyperspectral-based plant traits. ISPRS J. Photogramm. Remote Sens. 179, 133–144. https://doi.org/10.1016/j.isprsjprs.2021.07.014.
Porter, J.R., Gawith, M., 1999. Temperatures and the growth and development of wheat: a review. Eur. J. Agron. 10, 23–36. https://doi.org/10.1016/S1161-0301(98)00047-1.
Rab, M.A., Fisher, P.D., Armstrong, R.D., Abuzar, M., Robinson, N.J., Chandra, S., 2009. Advances in precision agriculture in south-eastern Australia. IV. Spatial variability in plant-available water capacity of soil and its relationship with yield in site-specific management zones. Crop Pasture Sci. 60, 885–900. https://doi.org/10.1071/CP08350.
Raya-Sereno, M.D., Ortiz-Monasterio, J.I., Alonso-Ayuso, M., Rodrigues, F.A., Rodríguez, A.A., González-Perez, L., Quemada, M., 2021. High-Resolution Airborne Hyperspectral Imagery for Assessing Yield, Biomass, Grain N Concentration, and N Output in Spring Wheat. Remote Sens. (Basel) 13, 1373. https://doi.org/10.3390/rs13071373.
Rodrigues, F., Blasch, G., Defourny, P., Ortiz-Monasterio, J., Schulthess, U., Zarco-Tejada, P., Taylor, J., Gérard, B., Rodrigues, F.A., Blasch, G., Defourny, P., Ortiz-Monasterio, J.I., Schulthess, U., Zarco-Tejada, P.J., Taylor, J.A., Gérard, B., 2018. Multi-Temporal and Spectral Analysis of High-Resolution Hyperspectral Airborne Imagery for Precision Agriculture: Assessment of Wheat Grain Yield and Grain Protein Content. Remote Sens. (Basel) 10, 930. https://doi.org/10.3390/rs10060930.
Sadras, V., Roget, D., O'Leary, G., 2002. On-farm assessment of environmental and management constraints to wheat yield and efficiency in the use of rainfall in the Mallee. Aust. J. Agr. Res. 53, 587–598. https://doi.org/10.1071/AR01150.
Skerritt, J.H., Adams, M.L., Cook, S.E., Naglis, G., 2002. Within-field variation in wheat quality: implications for precision agricultural management. Aust. J. Agr. Res. 53, 1229–1242. https://doi.org/10.1071/ar01204.
Sloat, L.L., Lin, M., Butler, E.E., Johnson, D., Holbrook, N.M., Huybers, P.J., Lee, J.-E., Mueller, N.D., 2021. Evaluating the benefits of chlorophyll fluorescence for in-season crop productivity forecasting. Remote Sens. Environ. 260, 112478. https://doi.org/10.1016/j.rse.2021.112478.
Stoy, P.C., Khan, A.M., Wipf, A., Silverman, N., Powell, S.L., 2022. The spatial variability of NDVI within a wheat field: Information content and implications for yield and grain protein monitoring. PLoS One 17, e0265243.
Tan, C., Zhou, X., Zhang, P., Wang, Z., Wang, D., Guo, W., Yun, F., 2020. Predicting grain protein content of field-grown winter wheat with satellite images and partial least square algorithm. PLoS One 15, e0228500.
Upreti, D., Huang, W., Kong, W., Pascucci, S., Pignatti, S., Zhou, X., Ye, H., Casa, R., 2019. A Comparison of Hybrid Machine Learning Algorithms for the Retrieval of Wheat Biophysical Variables from Sentinel-2. Remote Sens. (Basel) 11, 481. https://doi.org/10.3390/rs11050481.
van Herwaarden, A.F., Angus, J.F., Richards, R.A., Farquhar, G.D., 1998. "Haying-off", the negative grain yield response of dryland wheat to nitrogen fertiliser II. Carbohydrate and protein dynamics. Aust. J. Agr. Res. 49, 1083–1094. https://doi.org/10.1071/a97040.
van Ittersum, M.K., Howden, S.M., Asseng, S., 2003. Sensitivity of productivity and deep drainage of wheat cropping systems in a Mediterranean environment to changes in CO2, temperature and precipitation. Agr Ecosyst Environ 97, 255–273. https://doi.org/10.1016/S0167-8809(03)00114-2.
van Klompenburg, T., Kassahun, A., Catal, C., 2020. Crop yield prediction using machine learning: A systematic literature review. Comput. Electron. Agric. 177, 105709. https://doi.org/10.1016/j.compag.2020.105709.

Verhoef, W., Jia, L., Xiao, Q., Su, Z., 2007. Unified Optical-Thermal Four-Stream Radiative Transfer Theory for Homogeneous Vegetation Canopies. IEEE Trans. Geosci. Remote Sens. 45, 1808–1822. https://doi.org/10.1109/TGRS.2007.895844.

Verrelst, J., Camps-Valls, G., Muñoz-Marí, J., Rivera, J.P., Veroustraete, F., Clevers, J.G.P.W., Moreno, J., 2015. Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties – A review. ISPRS J. Photogramm. Remote Sens. 108, 273–290. https://doi.org/10.1016/j.isprsjprs.2015.05.005.

Walsh, O.S., Marshall, J.M., Nambi, E., Jackson, C.A., Ansah, E.O., Lamichhane, R., McClintick-Chess, J., Bautista, F., 2023. Wheat Yield and Protein Estimation with Handheld and Unmanned Aerial Vehicle-Mounted Sensors. Agronomy 13, 207. https://doi.org/10.3390/agronomy13010207.

Wang, L., Tian, Y., Yao, X., Zhu, Y., Cao, W., 2014. Predicting grain yield and protein content in wheat by fusing multi-sensor and multi-temporal remote-sensing images. Field Crop Res 164, 178–188. https://doi.org/10.1016/j.fcr.2014.05.001.

Wang, Z.J., Wang, J.H., Liu, L.Y., Huang, W.J., Zhao, C.J., Wang, C.Z., 2004. Prediction of grain protein content in winter wheat (*Triticum aestivum* L.) using plant pigment ratio (PPR). Field Crop Res 90, 311–321. https://doi.org/10.1016/j.fcr.2004.04.004.

Wolanin, A., Camps-Valls, G., Gómez-Chova, L., Mateo-García, G., van der Tol, C., Zhang, Y., Guanter, L., 2019. Estimating crop primary productivity with Sentinel-2 and Landsat 8 using machine learning methods trained with radiative transfer simulations. Remote Sens. Environ. 225, 441–457. https://doi.org/10.1016/j.rse.2019.03.002.

Wolanin, A., Mateo-García, G., Camps-Valls, G., Gómez-Chova, L., Meroni, M., Duveiller, G., Liangzhi, Y., Guanter, L., 2020. Estimating and understanding crop yields with explainable deep learning in the Indian Wheat Belt. Environ. Res. Lett. 15, 024019. https://doi.org/10.1088/1748-9326/ab68ac.

Wright, D.L., Rasmussen, V.P., Ramsey, R.D., Baker, D.J., Ellsworth, J.W., 2004. Canopy Reflectance Estimation of Wheat Nitrogen Content for Grain Protein Management. GISci. Remote Sens. 41, 287–300. https://doi.org/10.2747/1548-1603.41.4.287.

Wu, D., Hur, K., Xiao, Z., 2021. A GAN-Enhanced Ensemble Model for Energy Consumption Forecasting in Large Commercial Buildings. IEEE Access 9, 158820–158830. https://doi.org/10.1109/ACCESS.2021.3131185.

Xu, X.Q., Lu, J.S., Zhang, N., Yang, T.C., He, J.Y., Yao, X., Cheng, T., Zhu, Y., Cao, W.X., Tian, Y.C., 2019. Inversion of rice canopy chlorophyll content and leaf area index based on coupling of radiative transfer and Bayesian network models. ISPRS J. Photogramm. Remote Sens. 150, 185–196. https://doi.org/10.1016/j.isprsjprs.2019.02.013.

Xue, L.-H., Cao, W.-X., Yang, L.-Z., 2007. Predicting Grain Yield and Protein Content in Winter Wheat at Different N Supply Levels Using Canopy Reflectance Spectra. Pedosphere 17, 646–653. https://doi.org/10.1016/S1002-0160(07)60077-0.

Zadoks, J.C., Chang, T.T., Konzak, C.F., 1974. A decimal code for the growth stages of cereals. Weed Res. 14, 415–421. https://doi.org/10.1111/j.1365-3180.1974.tb01084.x.

Zarco-Tejada, P.J., Camino, C., Beck, P.S.A., Calderon, R., Hornero, A., Hernández-Clemente, R., Kattenborn, T., Montes-Borrego, M., Susca, L., Morelli, M., Gonzalez-Dugo, V., North, P.R.J., Landa, B.B., Boscia, D., Saponari, M., Navas-Cortes, J.A., 2018. Previsual symptoms of *Xylella fastidiosa* infection revealed in spectral plant-trait alterations. Nat. Plants 4, 432. https://doi.org/10.1038/s41477-018-0189-7.

Zhao, C., Liu, L., Wang, J., Huang, W., Song, X., Li, C., 2005. Predicting grain protein content of winter wheat using remote sensing data based on nitrogen status and water stress. Int. J. Appl. Earth Obs. Geoinf. 7, 1–9. https://doi.org/10.1016/j.jag.2004.10.002.

Zhao, H., Song, X., Yang, G., Li, Z., Zhang, D., Feng, H., 2019. Monitoring of Nitrogen and Grain Protein Content in Winter Wheat Based on Sentinel-2A Data. Remote Sens. (Basel) 11, 1724. https://doi.org/10.3390/rs11141724.

Zhou, X., Kono, Y., Win, A., Matsui, T., Tanaka, T.S.T., 2021. Predicting within-field variability in grain yield and protein content of winter wheat using UAV-based multispectral imagery and machine learning approaches. Plant Prod. Sci. 24, 137–151. https://doi.org/10.1080/1343943X.2020.1819165.

End of thesis.

***