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

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\textsubscript{a} \textsubscript{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\textsubscript{a} \textsubscript{b}, leaf dry matter, leaf water content (C\textsubscript{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\textsuperscript{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\textsubscript{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\textsubscript{a} \textsubscript{b}) content strongly influence assimilation (Wolanin et al., 2019), hence protein dilution, while above-ground N correlates with both C\textsubscript{a} \textsubscript{b} (Evans, 1989) and GPC (Feng et al., 2014; Xue et al., 2007; Zhao et al., 2005). C\textsubscript{a} \textsubscript{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\textsubscript{a} \textsubscript{b}/N may be only moderately correlated with GPC (Longmire et al., 2022; Zhao et al.,...
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 and yield, as seen in lower estimation accuracies (e.g. Rodriguez 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; Feret 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; Houles 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 ($C_x+C_c$), 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., 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 et al., 2002) 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 RMSE $= 0.141–0.144$. 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 $> 265$ 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 facilitative 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), very greatly across tens to hundreds of metres in clay fraction, subsoil constraints including B and Al, calcere, 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 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), a scientific-grade instrument radiometrically calibrated by the manufacturer. A further indirect correction was carried out during flights using ground observations from a handheld infrared thermometer (LaserSight from Optiris GmbH, Berlin, Germany), after Calderon 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 (Calderon 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, Vill, 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 $54$HWE (cz1), $54$HXG and $54$HYG (cz2) were downloaded from the Copernicus Open Access Hub and atmospherically corrected to surface reflectance with Sen2Cor v. 2.3.1. Bands of GSD $= 510$.

Table 1

Location, rainfall, and 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).

<table>
<thead>
<tr>
<th>Zone</th>
<th>Lat.</th>
<th>Lon.</th>
<th>Alt. (m)</th>
<th>AAR</th>
<th>Clim. zone</th>
<th>Year</th>
<th>Rain</th>
<th>GSR</th>
<th>Area (Ha)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>−36.371</td>
<td>141.241</td>
<td>142</td>
<td>449</td>
<td>Cb</td>
<td>2019</td>
<td>288</td>
<td>238</td>
<td>662</td>
</tr>
<tr>
<td>2</td>
<td>−35.051</td>
<td>142.881</td>
<td>55</td>
<td>316</td>
<td>Bsk</td>
<td>2019</td>
<td>194</td>
<td>135</td>
<td>2341</td>
</tr>
</tbody>
</table>

A. Longmire et al.
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.

### 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.

<table>
<thead>
<tr>
<th>Zone</th>
<th>Image</th>
<th>Year</th>
<th>Date</th>
<th>GDDAS</th>
<th>Z stage</th>
<th>Z name</th>
<th>Durum</th>
<th>GDDAS</th>
<th>Z stage</th>
<th>Z name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Bread</td>
<td>2019</td>
<td>18-Jul</td>
<td>481</td>
<td>15</td>
<td>seedling</td>
<td>684</td>
<td>16</td>
<td>seedling</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>23-Jul</td>
<td>537</td>
<td>15</td>
<td></td>
<td></td>
<td>739</td>
<td>17</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>17-Aug</td>
<td>770</td>
<td>17</td>
<td></td>
<td></td>
<td>972</td>
<td>31</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>11-Sep</td>
<td>1008</td>
<td>32</td>
<td>stem elong.</td>
<td>1211</td>
<td>37</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1-Oct</td>
<td>1237</td>
<td>43</td>
<td>booting</td>
<td>1439</td>
<td>52</td>
<td>ear emerg.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>21-Oct</td>
<td>1508</td>
<td>69</td>
<td>anthesis</td>
<td>1710</td>
<td>74</td>
<td>grain fill</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1-Oct</td>
<td>452</td>
<td>14</td>
<td>seedling</td>
<td>560</td>
<td>15</td>
<td>seedling</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1-Aug</td>
<td>581</td>
<td>15</td>
<td></td>
<td>689</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>26-Aug</td>
<td>803</td>
<td>31</td>
<td>stem elong.</td>
<td>910</td>
<td>31</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>10-Sep</td>
<td>979</td>
<td>32</td>
<td></td>
<td>1087</td>
<td>33</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>15-Sep</td>
<td>1375</td>
<td>51</td>
<td>ear emerg.</td>
<td>1483</td>
<td>67</td>
<td>anthesis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Durum</td>
<td>2019</td>
<td>17-Jul</td>
<td>904</td>
<td>17</td>
<td>seedling</td>
<td>1776</td>
<td>54</td>
<td>ear emerg.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>28-Jul</td>
<td>1028</td>
<td>31</td>
<td>stem elong.</td>
<td>803</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>12-Aug</td>
<td>1169</td>
<td>31</td>
<td></td>
<td>1109</td>
<td>32</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>17-Aug</td>
<td>1220</td>
<td>32</td>
<td></td>
<td>1325</td>
<td>32</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>27-Aug</td>
<td>1776</td>
<td>54</td>
<td></td>
<td>1882</td>
<td>33</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1-Oct</td>
<td>719</td>
<td>17</td>
<td></td>
<td>1237</td>
<td>32</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>26-Aug</td>
<td>1099</td>
<td>32</td>
<td>stem elong.</td>
<td>1307</td>
<td>34</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>31-Aug</td>
<td>1161</td>
<td>34</td>
<td></td>
<td>1385</td>
<td>42</td>
<td>booting</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>10-Sep</td>
<td>1307</td>
<td>34</td>
<td></td>
<td>1385</td>
<td>42</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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...
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 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 (Verhoeof 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 TCARI1610 (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 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).
Fig. 5. Plant traits $C_{w}$, $C_m$, $C_w$ 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.)

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 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, 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).

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 $C_m$ became predominant; neither $C_{w}$ 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 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 $C_m$ 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 ($\approx -4^\circ$ 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 (Z23–51). Frost around anthesis severely reduces grain count, and durum is more susceptible to both frost and heat damage than bread wheat (Berera et al., 2020; McCallum et al., 2019). Immediately after these frosts, four days had maxima of 35–38 $^\circ$ 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 $T > 34^\circ$ C, senescence is accelerated by a factor of three, and sixfold at
T $\geq$ 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 $C_{w}$, $C_{m}$, and SIF, diminishing these as GPC estimators, while likely enhancing the importance of moisture sufficiency measures, as we observed in both CWSI and SIF. 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 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 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.

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: $C_{w}$, $C_{m}$, $C_{a-b}$ and LAI, ± SIF. Blue background: EVI, PRI, Anth, $C_{a-c}$, $C_{x-c}$, LAI and LIDF, 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.)
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 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 $C_m$, 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 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 \( (\text{Longmire et al., 2022}) \), except in very dry conditions \( (\text{Fig. 7}) \). The lower S2 spectral resolution explains this: indicators linked to GPC through their detection of mild stress, inverted Anth and \( C_{a+b} \), and the PRI, cannot be calculated from S2 data so are absent from these analyses, but made substantial contributions to HS predictions. Here the advantages of S2 TS become clear, whereby TS prediction metrics were as good, and sometimes better than those from HS \( (\text{Fig. 7}) \). The improvement over single-date S2 estimates was also large, especially in cz2, where they came off a low base \( (\text{Table 4}) \). 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.

\( \text{Cai et al. (2019)} \) 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 \( (\text{Tan et al., 2020; Zhao et al., 2019}) \) and that the addition of extra S2 TS data after June gave little improvement \( (\text{yield; England; Hunt et al., 2019}) \). 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 \( (\text{Zhao et al., 2019}) \). Nevertheless our accuracies \( (\text{Table 4}) \) 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 (\(R^2 = 0.86, \text{RMSE} = 0.56\%\)). The order and relative importance of S2 plant traits were similar to airborne HS: S2 importance was dominated by CA 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 competing financial interests or personal relationships that could have influenced 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


Adam, E., Deng, H., Odindi, J., Abdel-Rahman, E.M., Mutanga, O., 2017. Detecting the importance of S2 plant traits were similar to airborne HS: S2 importance was dominated by CA 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.


Adam, E., Deng, H., Odindi, J., Abdel-Rahman, E.M., Mutanga, O., 2017. Detecting the importance of S2 plant traits were similar to airborne HS: S2 importance was dominated by CA 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.


Adam, E., Deng, H., Odindi, J., Abdel-Rahman, E.M., Mutanga, O., 2017. Detecting the importance of S2 plant traits were similar to airborne HS: S2 importance was dominated by CA 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.


doi.org/10.3390/agronomy13010027.


