

# ADVANCES IN THE STUDY OF BIOCHEMICAL, MORPHOLOGICAL AND PHYSIOLOGICAL TRAITS OF WHEAT AND SORGHUM CROPS IN AUSTRALIA USING HYPERSPECTRAL DATA AND MACHINE LEARNING

*Potgieter A. B.<sup>1</sup>, Camino C.<sup>2</sup>, Poblete T.<sup>4,3</sup>, Zhi X.<sup>5,6</sup>, Reynolds-Massey-Reed S.<sup>5</sup>, Zhao Y<sup>1</sup>, Belwalkar A.<sup>3</sup>, Ruizhu J.<sup>1</sup>, George-Jaeggli B.<sup>5,7</sup>, Chapman S.<sup>8</sup>, Jordan D.<sup>5</sup>, Wu A.<sup>1</sup>, Hammer G. L.<sup>1</sup> and Zarco-Tejada P.J.<sup>4,3,9</sup>*

<sup>1</sup>The University of Queensland, Queensland Alliance for Agriculture and Food Innovation, St Lucia, 4072, Queensland, Australia.

<sup>2</sup>European Commission, Joint Research Centre, Via E. Fermi 2749 – TP 262, I-21027, Ispra (VA)/Italy.

<sup>3</sup>Faculty of Engineering and Information Technology (FEIT), The University of Melbourne, Melbourne, Victoria, Australia

<sup>4</sup>School of Agriculture, Food and Ecosystem Sciences (SAFES), Faculty of Science, The University of Melbourne, Melbourne, Victoria, Australia

<sup>5</sup>The University of Queensland, Queensland Alliance for Agriculture and Food Innovation, Hermitage Research Facility, Warwick, 4370, Queensland, Australia.

<sup>6</sup>State Key Laboratory of Cotton Biology, Institute of Cotton Research of Chinese Academy of Agricultural Sciences, Henan Anyang 455000, China.

<sup>7</sup>Agri-Science Queensland, Department of Agriculture and Fisheries, Hermitage Research Facility, Warwick, 4370 Queensland, Australia.

<sup>8</sup>The University of Queensland, School of Agriculture and Food Sciences, St Lucia, 4072, Queensland, Australia.

<sup>9</sup>Instituto de Agricultura Sostenible (IAS), Consejo Superior de Investigaciones Científicas (CSIC), Córdoba, Spain

Correspondence: Email: [a.potgieter@uq.edu.au](mailto:a.potgieter@uq.edu.au)

## ABSTRACT

In this paper, we discuss the integration of systems such as multi-dimensional radiative transfer models (RTM) with deep learning (DL) algorithms to estimate plant biochemical, physiological, and morphological traits at canopy level using high-resolution hyperspectral imagery (361 bands in the 400-1000 nm spectral range). We applied the approaches to two case studies for dryland cropping in Australia (i.e., wheat and sorghum). Crop type averages for the early flight for leaf area index (LAI) varied between 2, for Canola, to as high as 4.3 for Lentils. Wheat and Barley had LAI of 4.1 and 3.8 (m<sup>2</sup>/m<sup>2</sup>), respectively. Chlorophyll *a+b* (C<sub>a+b</sub>) averages for emerged crops were 18, 41, 44, 51 and 59 µg/cm<sup>2</sup> for Faba beans, Wheat, Canola, Barley and Oats, respectively. The pigment Anthocyanin varied from 4.9 to 15.9 µg/cm<sup>2</sup> for Lentils and Canola, respectively. Similar patterns were observed in the Carotenoid (C<sub>x+c</sub>) levels (as high as 16.5 µg/cm<sup>2</sup> for Oats). For sorghum plots, the integrated DL approaches showed significant high correlation in predicting sorghum LAI (R<sup>2</sup> = 0.84, RMSE = 0.65 m<sup>2</sup>/m<sup>2</sup>) and C<sub>a+b</sub> (R<sup>2</sup> = 0.94, RMSE = 4.94 µgcm<sup>-2</sup>). The maximum velocity carboxylation rates (V<sub>cmax</sub>) varied between 45-75 µmol m<sup>-2</sup>s<sup>-1</sup>. For both studied periods, we yielded a R<sup>2</sup> > 0.78 and RMSE ≤ 5.35 µmol m<sup>-2</sup>s<sup>-1</sup>, being the RMSE lower when using the modelled fluorescence emission for retrieving the V<sub>cmax</sub>. In addition, we derived the solar induced fluorescence emission hyperspectral narrowband (5.8 nm) sensing and radiative transfer models (RTM).

**Index Terms** — Machine learning, Hyperspectral data, AI, SCOPE, Pro4SAIL, VcMax, SIF

Food security and sustainable agriculture are vital concerns to society and key topics in assessments of climate variability and change on crop production. However, the drivers of crop yield are complex and differ across spatial scales as a result of Genotype x Environment x Management (GxExM) combinations [1]. One such driver determining crop yield in interaction with environmental and management factors, is leaf photosynthetic capacity. During the last two decades, advances in global cereal yields have significantly slowed down [2]. Photosynthesis is the process by which chlorophyll in living plants converts sunlight, water, and CO<sub>2</sub> into energy and is a recent target in improving crop yield [3]. This is phenometrically manifested through growth in roots, leaves, inflorescence, and ensuing biomass. The advent of hyperspectral sensing will enable us to determine the effect of GxExM more accurately on photosynthetic function at plant, canopy, and field scales [4-6]. Here we exemplified the utility of hyperspectral sensing data for two dry land cropping examples for Australia. Firstly, for 108 plots with 93 diverse sorghum genotypes (*Sorghum bicolor* (L.) Moench) grown in a breeding experiment during the 2019/20 summer season at the Gatton Research Facility in Queensland and secondly, 16 fields of six winter crops (including Wheat *Triticum aestivum* L.) in central Victoria. Hyperspectral imagers (full width at half-maximum resolutions of 0.1–0.2 nm and 5.8 nm) were flown in tandem around flowering and senescence for the winter crop fields. A narrowband imager (5.8 nm) was used to capture hyperspectral data at weekly intervals for the sorghum breeding plots. Specifically, we explore the ability of *machine learning* (ML) algorithms to retrieve maximum velocity of carboxylation rate (V<sub>cmax</sub>), structural parameters (e.g., Leaf area index and Leaf inclination distribution

## 3. INTRODUCTION

function), and leaf photosynthetic pigments (e.g., chlorophyll content) derived from RTM [7] and hyperspectral imagery.

In addition, we explored their relationships with chlorophyll fluorescence emission calculated at the 760 nm oxygen absorption region and spectral indices centered on the red edge spectral region (SIF). Although top-of-canopy (TOC) SIF is associated with plant photosynthetic capacity, it is also influenced by incident radiation and leaf and canopy characteristics, such as leaf chlorophyll concentration and Leaf Area Index (LAI). Consequently, SIF differs among plant species, with croplands exhibiting the highest SIF, followed by broadleaf and needleleaf species. In terms of the seasonal dynamics of SIF, the early stage of canopy development is characterized by low SIF values owing to the presence of young leaves. As the canopy develops and the leaf area index and proportion of mature leaves increase, SIF levels rise. In the later phases, when senescence and flowering occur, SIF eventually decreases [8]. Under stressed conditions, the three pathways of absorbed energy regulation (i.e., photochemistry, heat dissipation, and fluorescence emissions) perform differently, with fluorescence emission tending to decrease as a result of an increase in plant protection mechanisms [9]. Finally, to avoid potential ill-posed solutions in the simulations, we constrained all inputs based on field measurements, information from literature, and preliminary simulations to ensure that the resulting look-up tables (LUT) were within the range of the observations made by the hyperspectral sensor.

## 2. METHODS

### 2.1. Modelling methods

Leaf biochemical constituents and canopy structural properties were retrieved using two inversion approaches based on the (i) PRO4SAIL radiative transfer model [10] PRO4SAIL couples the PROSPECT-D [11] leaf radiative transfer model and the 4SAIL [12] canopy radiative transfer model and (ii) SCOPE model [13, 14] to derive  $V_{cmax}$  and modelled SIF at 760 nm in the sorghum breeding plots. Leaf pigments such as  $C_{a+b}$ , Anth, and  $C_{x+c}$  were quantified. Structural traits such as LAI were estimated for each plot. This procedure was carried out for each specific study site considering the intrinsic variations between each dataset and time of collection (Table 1).

**Table 1. Ranges of parameters used for the simulations with the PRO4SAIL radiative transfer model for winter crops (Left) and sorghum using SCOPE (Right).**

Parameter	Abbreviation	Units	Value / Range	Range (sorghum)
Chlorophyll content	$C_{a+b}$	[ $\mu\text{g}/\text{cm}^2$ ]	[3,70]	[0.5-95]
Carotenoid content	$C_{x+c}$	[ $\mu\text{g}/\text{cm}^2$ ]	[1,20]	[0.1-40]
Anthocyanin content	Anth.	[ $\mu\text{g}/\text{cm}^2$ ]	[1,10]	[0-7]
Dry matter content	$C_m$	[ $\text{g}/\text{cm}^2$ ]	[0.001,0.035]	[0.02]
Water content	$C_w$	[ $\text{g}/\text{cm}^2$ ]	[0.001,0.035]	[0-0.02]
Mesophyll struct. coeff.	N	-	[0.5,3]	[1.5-2.5]
Leaf area index	LAI	[ $\text{m}^2/\text{m}^2$ ]	[0.5,5]	[0-7]
Average leaf angle	LIDF	[deg.]	[1,90]	[-0.5 -0.5] a,b
Hot spot parameter	hot	-	0.01	0.0
Observer angle	tto	[deg.]	0	0.15
Sun zenith angle	tts	[deg.]	37.7	[15-35]
Relative azimuth angle	psi	[deg.]	0	0

For PRO4SAIL, a look-up table of 200,000 simulations varying uniformly distributing the parameters shown in Table 1 left and right, and the rest of the parameters were used in their default ranges. The simulated reflectance was resampled to the spectral resolution of the hyperspectral sensor by a convolution using a Gaussian spectral response function. The plant traits were inverted using SVM algorithms [15]. For regression, SVM is called SVR, which finds the function with maximum deviation between observed responses. SVR models were trained in parallel using MATLAB ([www.mathworks.com](http://www.mathworks.com)). The resampled reflectance was used as input, and the biochemical and structural traits were used as output. The SVR algorithms were trained 10-fold using a radial basis function as kernel and optimizing the hyperparameters during training for each variable.

For the SCOPE model, we designed a Multi-tier ML approach to retrieve plant traits coupling hyperspectral imagery with the SCOPE model (version v2.1) in R software (R Core Team, 2022, [www.R-project.org](http://www.R-project.org)). We made 20,000 simulations with the main plant trait ranges described in Table 1 (right). In addition, we applied a uniform distribution transformation for  $V_{cmax}$  and varied each plant trait within the specified range of 0-175  $\mu\text{mol m}^{-2}\text{s}^{-1}$ , while incoming radiation was 200-1000 and 100-500 ( $\text{W.m}^{-2}$ ) for shortwave and longwave radiation, respectively. Additionally, we assumed correlation between main plant traits to avoid combinations of plant traits, which are not present in the studied forest canopies. Similar to PRO4SAIL approached, we distributed the data with a uniform distribution function for the inputs of Table 1 (right). To avoid noise in the data affecting the model inversion, we applied smoothing of the simulated spectra at 1 nm resolution using a Savitzky-Golay derived calculation method [16]. We then resampled each simulated spectrum to adjust its resolution to the bandwidth of the sensors using Gaussian spectral response functions defined by the FWHM values of the hyperspectral sensors.

To make the  $V_{cmax}$  estimates, the resampled spectra simulated via SCOPE model were first randomly partitioned into two groups, the training sample with 80% of the simulations and the testing sample with the 20%. We then trained a one-dimensional convolutional neural network model (1D CNN) to retrieve  $V_{cmax}$  and fluorescence emission at 760 nm using the LUT configuration described in Tables 1 (Sorghum, right).

### 2.2. Generation of SIF outputs

The total incoming irradiance was continuously measured throughout the duration of the airborne campaign using the radiance reflected from a white reference panel ([www.labsphere.com/](http://www.labsphere.com/)) using a 0.065 nm FWHM Ocean Optics HR-2000 spectrometer ([www.oceaninsight.com](http://www.oceaninsight.com)) installed at the field site. The acquired irradiance was then convolved assuming a gaussian spectral response function matching the narrow-band imager's spectral characteristics. Using the radiance imagery from the narrow-band imager and the convolved irradiance from the HR-2000 spectrometer, a

SIF map was generated utilizing the O<sub>2</sub>-A band *in-filling* method through the Fraunhofer Line Depth (FLD) principle and a total of three spectral bands (3FLD) [17]. Further scaling of SIF maps between 0 and 5 arbitrary units (a.u.) was performed to investigate the relative pattern across the study region.

### 3. RESULTS

#### 3.1. Determining of canopy traits for winter crop fields.

The results of the C<sub>a+b</sub> estimations (Fig. 1) showed that Faba beans had the highest increase in C<sub>a+b</sub>, from 15.54 μg/cm<sup>2</sup> to 70.61 μg/cm<sup>2</sup>. Barley had the largest decrease in C<sub>a+b</sub>, from 54.11, 57.82, and 42.47 μg/cm<sup>2</sup> to 7.4, 17.05, and 35.23 μg/cm<sup>2</sup>. Lentils had similar concentrations for both fields in both years, with an average of 67.3 μg/cm<sup>2</sup> in the second campaign. Wheat had the highest concentration of C<sub>a+b</sub> in the first campaign, with 52.3 μg/cm<sup>2</sup>, and the lowest concentration in the second campaign, with 29.17 μg/cm<sup>2</sup>. Canola had similar values for both campaigns, with 50.21 and 43.54 μg/cm<sup>2</sup>. Oats hay had a sharp decrease in C<sub>a+b</sub> from 59.41 μg/cm<sup>2</sup> to 4.4 μg/cm<sup>2</sup> from the first to the second campaign. Similar patterns in trends were observed for C<sub>x+c</sub> and Anth. content for all crops and dates. However, the exception was the trends in LAIs, which had values around 4 m<sup>2</sup>/m<sup>2</sup> for all crops except for Canola and Oats, which had LAI ~2 m<sup>2</sup>/m<sup>2</sup>. for first date (e.g., flowering), for barley and oats, respectively. LAI during the second flight ranged reduced to around 2 m<sup>2</sup>/m<sup>2</sup>.

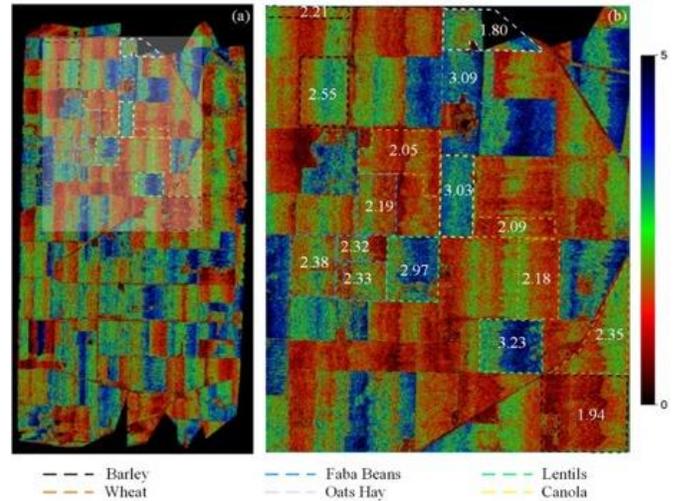
□ Flight 09-08-2021 ■ Flight 01-11-2021

■ Oats hay ■ Barley ■ Canola ■ Wheat ■ Faba Beans ■ Lentils

**Figure 1: Prediction of Chlorophyll (C<sub>a+b</sub>) content for Oats hay, Barley, Canola, Wheat, Faba beans and lentils crops using hyperspectral imagery captured on 09<sup>th</sup> of August 2021 and 1<sup>st</sup> of November 2021.**

#### 3.2. SIF for winter crops from airborne campaigns.

Figure 2 shows the values of modelled SIF in a.u. The lowest SIF was observed for Oats with an average of 1.8 a.u.. For wheat the mean was 2.17 a.u., while in Barley it was 2.16 a.u.. Canola had the highest average of 3.03 a.u. Faba beans had an average of 2.49 a.u.. Variation in SIF were observed between crop types and within fields.



**Figure 2: (a) Pixel-scale (0.5 m x 0.5m) SIF map showing the spatial pattern across the fields for the different crops (15,000 hectares). (b) The zoom-in map showing the average SIF values across fields. For cross scale comparison units were scaled to arbitrary units (a.u.).**

#### 3.3. Determining of V<sub>cmax</sub> for sorghum breeding plots.

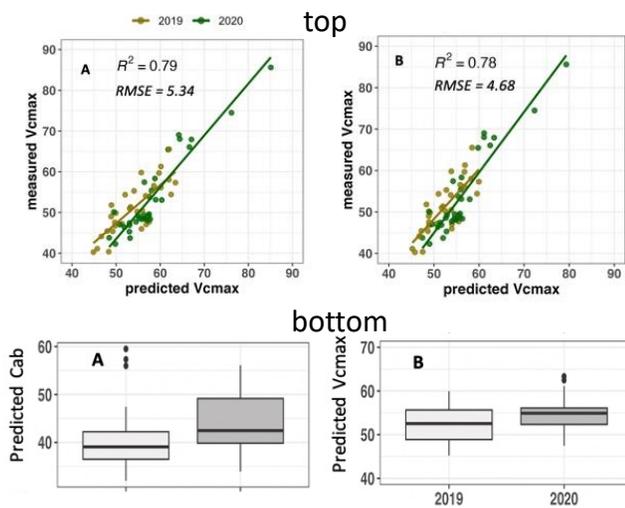
The results of the V<sub>cmax</sub> estimates (Fig. 3 top) showed that with the coupling between SCOPE and hyperspectral imagery we obtained a significant performance ( $R^2 > 0.78$ ) with the V<sub>cmax</sub> measured in the sorghum crop plots. However, by adding the modelled SIF (Fig. 3 top B), we reduced the RMSE and the outliers, compared to the predicted V<sub>cmax</sub> retrievals based exclusively on the applied CNN approach (Fig. 3 top A).

In the sorghum crop plots for measured V<sub>cmax</sub> and predicted plant traits, we noted a similar pattern for both inputs. In general, we observed higher plant trait values in the sorghum plots for the 2020 period than for sorghum plots placed in 2019. In this sense, the mean predicted C<sub>a+b</sub> values were 40.55 μg/cm<sup>2</sup> in 2020 lower than 44.68 μg/cm<sup>2</sup> for C<sub>a+b</sub> in the 2019 period (Fig. 3. bottom A). The predicted V<sub>cmax</sub> yielded similar trend as measured V<sub>cmax</sub> (Fig. 3 top) for both years. The predicted V<sub>cmax</sub> when adding modelled SIF ranged between 45 to 80 μmol m<sup>-2</sup>s<sup>-1</sup>, (Fig. 3 bottom B) close to the observed V<sub>cmax</sub> 40 and 85 μmol m<sup>-2</sup>s<sup>-1</sup>.

## 6. HARNESSING HIGH RESOLUTION HYPERSPECTRAL DATA AND AI IN CROPPING SYSTEMS

During the last decade, the fusion of biophysical models, climate forecast and hyperspectral remote sensing technologies have progressively become more achievable. This was mainly due to the acceleration in state-of-the-art data storage capacity and cloud computing and the evolution in artificial intelligence to solve convoluted “*real-world*” problems. In addition, high-resolution hyperspectral remote sensing platforms have rapidly advanced in temporal, spatial and spectral resolutions with global coverage. enable for more accurate estimation of biochemical, morphological, and

physiological traits in crop canopies across field scales. Furthermore, outputs from such integrated systems will result into objective, accurate and timely estimation of concomitant functional traits. These morphological, biochemical, and physiological traits is a direct outcome of plant responses for various GxExM combinations. Thus, illuminating our understanding of the impact of abiotic, biotic stresses and photosynthesis on ensuing crop growth and production. Finally, it is anticipated that outputs generated from high-resolution earth observation sensors combined with targeted AI algorithms and linked to dynamic crop models [18] will advance the development of a functional and integrated systems approaches to elite high yielding varieties, mitigate risk and enhancing resilience.



**Figure 3: (Top):** Relationship between measured  $V_{cmax}$  in sorghum breeding plots and  $V_{cmax}$  predicted by coupling the SCOPE model and hyperspectral imagery using two modelling strategies: A) a CNN model for estimating  $V_{cmax}$ ; and B) a CNN model for estimating  $V_{cmax}$  + adding modeled fluorescence emission. **(Bottom):** boxplots of chlorophyll content (A; in  $\mu\text{g}/\text{cm}^2$ ) and  $V_{cmax}$  (B; in  $\mu\text{mol m}^{-2}\text{s}^{-1}$ ) predicted coupling the SCOPE, modelled SIF, and hyperspectral imagery in sorghum breeding plots.

## 7. ACKNOWLEDGEMENTS

The authors want to acknowledge the funding support from The University of Queensland, the Centre of Excellence for Translational Photosynthesis, Australian Research Council (grant CE140100015) and the ‘CropPhen: Remote mapping of grain crop type and phenology’ project, funded by The Australian Grains Research and Development Corporation.

## 8. REFERENCES

- [1] D. T. Smith, A. B. Potgieter, and S. C. Chapman, "Scaling up high-throughput phenotyping for abiotic stress selection in the field," *Theoretical and Applied Genetics*, vol. 134, no. 6, pp. 1845-1866, 2021/06/01 2021, doi: 10.1007/s00122-021-03864-5.
- [2] R. A. Fischer, D. Byerlee, and G. O. Edmeades, *Crop yields and global food security: will yield increase continue to feed the world?*

- Canberra: Australian Centre for International Agricultural Research, 2014, pp. 1 - 8.
- [3] J. Evans, "Improving Photosynthesis," *Plant Physiology*, vol. 162, pp. 1780-1793, 2013.
- [4] A. B. Potgieter *et al.*, "Evolution and application of digital technologies to predict crop type and crop phenology in agriculture," *in silico Plants*, vol. 3, no. 1, 2021, doi: 10.1093/insilicoplants/diab017.
- [5] X. Zhi *et al.*, "Estimating Photosynthetic Attributes from High-Throughput Canopy Hyperspectral Sensing in Sorghum," *Plant Phenomics*, vol. 2022, 2022, doi: 10.34133/2022/9768502.
- [6] R. T. Furbank, J. A. Jimenez-Berni, B. George-Jaeggli, A. B. Potgieter, and D. M. Deery, "Field crop phenomics: enabling breeding for radiation use efficiency and biomass in cereal crops," *New Phytologist*, vol. 0, no. 0, 2019, doi: 10.1111/nph.15817.
- [7] C. Camino *et al.*, "Plant trait retrievals using PROSAIL and SIF emission with hyperspectral sensors," *Remote Sensing and Environment*, To be submitted.
- [8] M. Rossini *et al.*, "Analysis of Red and Far-Red Sun-Induced Chlorophyll Fluorescence and Their Ratio in Different Canopies Based on Observed and Modeled Data," *Remote Sensing*, vol. 8, no. 5, p. 412, 2016. [Online]. Available: <https://www.mdpi.com/2072-4292/8/5/412>.
- [9] F. Pinto *et al.*, "Sun-induced chlorophyll fluorescence from high-resolution imaging spectroscopy data to quantify spatio-temporal patterns of photosynthetic function in crop canopies," *Plant, Cell & Environment*, vol. 39, no. 7, pp. 1500-1512, 2016, doi: <https://doi.org/10.1111/pce.12710>.
- [10] P. J. Zarco-Tejada *et al.*, "Divergent abiotic spectral pathways unravel pathogen stress signals across species," *Nature Communications*, vol. 12, no. 1, p. 6088, 2021/10/19 2021, doi: 10.1038/s41467-021-26335-3.
- [11] J. B. Féret, A. A. Gitelson, S. D. Noble, and S. Jacquemoud, "PROSPECT-D: Towards modeling leaf optical properties through a complete lifecycle," *Remote Sensing of Environment*, vol. 193, pp. 204-215, 2017/05/01/ 2017, doi: <https://doi.org/10.1016/j.rse.2017.03.004>.
- [12] W. Verhoef, L. Jia, Q. Xiao, and Z. Su, "Unified Optical-Thermal Four-Stream Radiative Transfer Theory for Homogeneous Vegetation Canopies," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 6, pp. 1808-1822, 2007, doi: 10.1109/TGRS.2007.895844.
- [13] C. van der Tol, W. Verhoef, J. Timmermans, A. Verhoef, and Z. Su, "An integrated model of soil-canopy spectral radiances, photosynthesis, fluorescence, temperature and energy balance," *Biogeosciences*, vol. 6, no. 12, pp. 3109-3129, 2009, doi: 10.5194/bg-6-3109-2009.
- [14] P. Yang, E. Prikaziuk, W. Verhoef, and C. van der Tol, "SCOPE 2.0: a model to simulate vegetated land surface fluxes and satellite signals," *Geosci. Model Dev.*, vol. 14, no. 7, pp. 4697-4712, 2021, doi: 10.5194/gmd-14-4697-2021.
- [15] V. N. Vapnik, "Setting of the Learning Problem," in *The Nature of Statistical Learning Theory*. New York, NY: Springer New York, 1995, pp. 15-32.
- [16] A. Savitzky and M. J. E. Golay, "Smoothing and Differentiation of Data by Simplified Least Squares Procedures," *Analytical Chemistry*, vol. 36, no. 8, pp. 1627-1639, 1964/07/01 1964, doi: 10.1021/ac60214a047.
- [17] S. Maier and K. Günther, *Sun-induced fluorescence - a new tool for precision farming*. 2001, pp. 90-93.
- [18] A. Wu, G. L. Hammer, A. Doherty, S. von Caemmerer, and G. D. Farquhar, "Quantifying impacts of enhancing photosynthesis on crop yield," *Nature Plants*, vol. 5, no. 4, pp. 380-388, 2019/04/01 2019, doi: 10.1038/s41477-019-0398-8.