TECHNICAL INNOVATION

Integrating a crop growth model and radiative transfer model to improve estimation of crop traits based on deep learning

Qiaomin Chen1,2,*, Bangyou Zheng2,*, Tong Chen3 and Scott C. Chapman1,*

1 School of Agriculture and Food Sciences, The University of Queensland, St Lucia, 4067, QLD, Australia
2 CSIRO Agriculture and Food, Queensland Biosciences Precinct 306 Carmody Road, St Lucia, 4067, QLD, Australia
3 School of Information Technology and Electrical Engineering, The University of Queensland, St Lucia, 4067, QLD, Australia

* Correspondence: qiaomin.chen@uq.edu.au or scott.chapman@uq.edu.au

Received 27 January 2022; Editorial decision 27 June 2022; Accepted 29 June 2022

Editor: Greg Rebetzke, CSIRO Agriculture and Food, Australia

Abstract

A major challenge for the estimation of crop traits (biophysical variables) from canopy reflectance is the creation of a high-quality training dataset. To address this problem, this research investigated a conceptual framework by integrating a crop growth model with a radiative transfer model to introduce biological constraints in a synthetic training dataset. In addition to the comparison of two datasets without and with biological constraints, we also investigated the effects of observation geometry, retrieval method, and wavelength range on estimation accuracy of four crop traits (leaf area index, leaf chlorophyll content, leaf dry matter, and leaf water content) of wheat. The theoretical analysis demonstrated potential advantages of adding biological constraints in synthetic training datasets as well as the capability of deep learning. Additionally, the predictive models were validated on real unmanned aerial vehicle-based multispectral images collected from wheat plots contrasting in canopy structure. The predictive model trained over a synthetic dataset with biological constraints enabled the prediction of leaf water content from using wavelengths in the visible to near infrared range based on the correlations between crop traits. Our findings presented the potential of the proposed conceptual framework in simultaneously retrieving multiple crop traits from canopy reflectance for applications in precision agriculture and plant breeding.

Keywords: APSIM, biological constraints, canopy and leaf biophysical traits, canopy reflectance, PROSAIL, unmanned aerial vehicle.

Abbreviations: Cab, leaf Chl a and b content per leaf area; Cant, leaf anthocyanin content per leaf area; Car, leaf carotenoid content per leaf area; Cbrown, leaf brown pigment concentration; cdata, synthetic dataset with biological constraints of which PROSAIL input parameters are constrained to the crop growth pattern defined in APSIM; CGM, crop growth model; Cm, leaf dry matter content per leaf area; Cw, leaf water content per leaf area or leaf equivalent water thickness; FFNN, feedforward neural network; hsop, hot spot size parameter; LAI, leaf area index; LIDF, leaf inclination distribution function; LUT, look-up table; mVNIR, visible to near infrared (multispectral); Ms, leaf mesophyll structure parameter; r, Pearson correlation coefficient; RAA, relative azimuth angle; rdata, synthetic dataset without biological constraints of which PROSAIL input parameters can be combined randomly; RF, random forest; RMSE, root mean squared error; RRMSE, relative root mean squared error; rsol, reflectance of soil as a libertarian surface; RTM, radiative transfer model; SWIR, shortwave infrared; SZA, solar zenith angle; UAV, unmanned aerial vehicle; VIS, visible; VNIR, visible to near infrared; VNSWIR, visible to shortwave infrared; VZA, viewing zenith angle.

© The Author(s) 2022. Published by Oxford University Press on behalf of the Society for Experimental Biology.

This paper is available online free of all access charges (see https://academic.oup.com/jxb/pages/openaccess for further details).
Introduction

Since the 1960s, satellite imagery to measure reflectance of the Earth’s surface at a scale of tens of metres has been used to monitor vegetation health and estimate and forecast changes in vegetation cover (see related reviews, e.g. Dorigo et al., 2007; Baret and Buis, 2008; Thenkabail et al., 2019). More recently, these imagery methods have been deployed in proximal sensors (e.g. planes, drones, and vehicles) to analyse vegetation at higher resolutions (i.e. sub-centimetre scales) in a research domain sometimes referred as ‘high-throughput phenotyping’ (e.g. Fahlgren et al., 2015; Chapman et al., 2018). Nanosatellites with higher temporal (revisit time down to 1 d), spatial (0.3–1.0 m), and spectral (narrower hyperspectral bandwidth, 1–10 nm) resolutions are also being deployed (e.g. McCabe et al., 2016; Sozzi et al., 2018). Since sensors and imaging technologies can rapidly measure multiple crop traits across time and space in a cost- and labour-efficient way, they present potential to benefit applications in precision agriculture and in plot experiments conducted in plant breeding and agronomy.

Depending on the electromagnetic spectrum used in sensors or cameras, different crop traits can be retrieved from sensing data. The visible range at ~400–750 nm (from RGB cameras) can be used to measure the morphological, geometric, and colour properties, such as ground cover (e.g. Duan et al., 2017; Wan et al., 2021), ear/head number, or density (e.g. Madec et al., 2019; David et al., 2021). From spectral cameras, narrow bands in the visible range from 400 nm to 700 nm can also be used to measure some plant biochemical properties such as pigment content (e.g. chlorophyll, carotenoid, and anthocyanins) (e.g. Ustin et al., 2009) and nitrogen content (e.g. Jay et al., 2017). The visible to near infrared (NIR) range at ~400–1100 nm is generally used to estimate the fractional ground cover of vegetation and the leaf area index (LAI) (e.g. Hu et al., 2021). The shortwave infrared range at ~1200–2500 nm is suitable to measure plant water content (e.g. Jin et al., 2013). The longwave infrared range (sometimes called thermal infrared) at ~7.5–13 µm can be used to measure plant temperature and is usually proposed for detection of water-related or heat-related stresses (e.g. Zhang et al., 2019). Based on the estimation of variables associated with stress status, sensing data have been used for abiotic/biotic stress detection and disease diagnoses in crop protection (e.g. Mahlein et al., 2012).

The retrieval of crop traits from spectral signals (canopy reflectance and its derived variables such as vegetation index) is based on the structural and biochemical relationships between traits and spectra. These can directly come from the cause–effect relationship expressed in radiative transfer models (RTMs) and be learned from training data collected from field experiments (experimental data) or generated by RTMs (synthetic data). Compared with experimental data, the main advantage of synthetic data is to allow construction of a training dataset to represent the diverse range of situations generated by interactions of genotype, environment, and management, as well as observation conditions (Dorigo et al., 2007; Baret and Buis, 2008), and provides a capability to generalize the prediction of crop traits in field conditions. There is an increasing interest in the application of RTMs in crop trait retrieval studies, in which PROSAIL is a popular model and has been widely used for variable retrieval, as reviewed by Berger et al. (2018).

For those retrieval studies based on the inversion of the cause–effect relationship between crop traits and reflectance expressed in RTMs, numerical optimization and look-up table (LUT) are the two major retrieval methods. Numerical optimization is realized by comparing the observed and simulated spectral signal to find the optimal solution (e.g. Danner et al., 2019; Lunagaria and Patel, 2019), while LUT is achieved by searching the entire LUT for optimal solutions stored in the LUT (e.g. Quan et al., 2017; X.Q. Xu et al., 2019; Zhu et al., 2019). Machine learning methods including neural networks are also increasingly used to rebuild (learn) the inverse relationship between reflectance and crop traits (e.g. Bacour et al., 2006; Upreti et al., 2019). As summarized in the literature, different methods have advantages and limitations, with no obvious global solution, but machine learning methods are assumed to be superior in representing relationships between model input and output from training data, and are more computationally efficient once the model is fully trained (e.g. Dorigo et al., 2007; Verrelst et al., 2015).

A major challenge for synthetic data approaches is the ‘ill-posed’ problem to retrieve crop traits from sensing data (e.g. Baret and Buis, 2008; Jacquemoud et al., 2009). This problem is caused by the model uncertainty which results from its simplification of the canopy structure and biochemistry, so that more than one state situation of a canopy could result in exactly the same reflectance (e.g. Jacquemoud et al., 1995; Combal et al., 2003) However, in practical applications, this problem is caused by the local optimization of a criterion (i.e. objective function, cost function, or loss function) that can be obtained from different solutions (parameter combinations). In synthetic data, this problem is aggravated by the poor selection of combinations of input parameter values; that is, not accounting for biophysical limitations among combinations of parameter values that actually exist in the real-world. Without consideration of these biophysical constraints among parameters, the ill-posed nature of the inverse problem results in multiple solutions that may be spread across the whole parameter space rather than centred around the true solution (e.g. Atzberger, 2004).

The ill-posed problem can be partly alleviated by using prior knowledge to strengthen constraints on individual input parameters or among parameters. The simplest and most common method is to define the appropriate range of parameters with experimental data, which have been proved to improve estimation accuracy (e.g. Li et al., 2019). Inversion studies with LUT, using the medium or mean of the best $k$ ($k>1$) solutions is another common way to alleviate the ill-posed problem (e.g. Combal et al., 2003; Quan et al., 2017). Studies also provided evidence for using additional biological or spatial constraints to
improve estimation accuracy: for example, M. Xu et al. (2019) introduced a matrix-based LUT consisting of two vegetation indexes (one sensitive to leaf chlorophyll and the other sensitive to LAI) to consider the effect of LAI on leaf chlorophyll content estimation; X.Q. Xu et al. (2019) developed a LUT of maximum conditional probabilities of LAI and canopy chlorophyll content based on a Bayesian network for trait estimation; and Atzberger (2004) proposed an object-based method to introduce the spatial constraints of the target trait (LAI, leaf chlorophyll content, and leaf water content) to improve its estimation accuracy. In addition, multangular observations (e.g. Jay et al., 2017; Roosjen et al., 2018) were also investigated to address the ill-posed problem.

An alternative approach is to link a crop growth model (CGM) to RTM, which provides a more straightforward solution to address the ill-posed problem by directly constraining the combinations of RTM input parameters with crop physiology algorithms that are built into the CGM. Theoretically, the integration can be implemented in two ways by (i) calibrating/parameterizing the CGM using canopy reflectance and then predicting target crop traits using the calibrated CGM (Method 1); or (ii) converting CGM output variables into RTM input parameters and then applying retrieval methods (e.g. LUT) on these constrained crop traits and corresponding canopy reflectance (Method 2). Method 1 can directly retrieve those traits not included in RTM inputs such as crop yield and also provide estimation of traits across the whole growth season by running calibrated CGMs (e.g. Thorp et al., 2012; Machwitz et al., 2014; Guo et al., 2019). Method 2 provides a more direct way to predict target traits included in RTM inputs from reflectance data without additional needs of weather and management data used in CGMs, which is more applicable in practice but has been rarely discussed. In this study, for the first time, we implemented Method 2 to link a CGM and RTM for retrieving crop traits from canopy reflectance.

Previous studies have indicated that radiometric information is not sufficient to accurately estimate crop traits, and some prior information is needed to improve the estimation accuracy (e.g. Combal et al., 2003). The aim of this study is to compare synthetic datasets with and without biological constraints to retrieve four crop traits (i.e. LAI, leaf chlorophyll content, leaf dry matter, and leaf water content) of wheat. The biological constraints were introduced through integrating a CGM (e.g. Agricultural Production System Simulator, APSIM) and an RTM (e.g. PROSAIL) in the way described in Method 2, so that PROSAIL input parameters are constrained to APSIM biology. We further investigated the effects of observation geometries, wavelength ranges, and retrieval methods on estimation accuracy of the four crop traits.

Materials and methods

Overview

This research consists of several key steps (Fig. 1), summarized as follows. APSIM and PROSAIL models were integrated by passing variables from the former to the latter based on variable transformation relationships. A set of defined conditions for wheat growing (characterized by genotype, environment, and management) and observation conditions (determined by local latitude, day of year, and time of day) was set up to run APSIM and PROSAIL for simulation of crop traits and canopy reflectance, which resulted in two types of synthetic datasets. The concept of a ‘synthetic dataset’ is relative to a real dataset collected from practical environments (e.g. field experiments). Here the ‘synthetic dataset’ was generated by physical models including APSIM and PROSAIL. According to research objectives, these synthetic datasets were reconstructed and used in a series of analyses defined below to investigate the effects of application of observation geometry, wavelength range, and retrieval methods on estimation accuracy of target crop traits. Finally, the prediction models were applied to real multispectral images collected in wheat plots contrasting in canopy structure, to evaluate the practical performance of this ‘synthetic data-derived’ predictive model.

Models and their integration

APSIM Next Generation (https://www.apsim.info/apsim-next-generation/) is the new version of APSIM, which is an evolution of the classic version 7.10 (Holzworth et al., 2018). APSIM Next Generation or APSIM is driven by major processes in crop physiology and interactions with environment factors and management practices, and is widely used to simulate crop development and growth (e.g. LAI or dry weight of organ parts) during the growth season at a daily time step. APSIM has been validated in many regions around the world for wheat (e.g. Chen et al., 2017). APSIM is not a structural model, so the crop variables such as biomass and leaf area per unit ground area are state variables computed each day as single variables, and do not have any 2D or 3D structural attributes. The model does utilize geometrical theory to determine how the relationship between LAI and light extinction coefficient (a crop trait) affects interception of incoming radiation each day, including the effects of plant row spacing, but APSIM Next Generation does not account for direct versus diffuse light conditions or change in radiation over the day (Holzworth et al., 2018). PROSAIL is the combination of PROSPECT (a leaf optical property model) and SAIL (a canopy bidirectional reflectance model). PROSAIL links the spectral variation of canopy reflectance (mainly related to leaf biochemical contents) with its directional variation (primarily related to canopy architecture and soil/vegetation contrast), which is key to simultaneously estimating canopy biophysical variables such as leaf chlorophyll content and LAI (Jacquemoud et al., 2009). The current version of the PROSAIL model, PROSAIL-D, is available for download (http://tele-detection.ipgp.jussieu.fr/prosail/, retrieved on 18 July 2022). Input and output variables of this PROSAIL model are presented in Table 1. The 14 input parameters can be divided into four categories: leaf properties (leaf mesophyll structure parameter (Ns), leaf water content per leaf area or leaf equivalent water thickness (Cw), leaf dry matter content per leaf area (Cmi), leaf Chl a and b content per leaf area (Cab), leaf carotenoid content per leaf area (Car), leaf anthocyanin content per leaf area (Cant), and leaf brown pigment concentration (Cbrown)), background soil properties (rsol), canopy architecture (LAI, leaf inclination distribution function (LIDF), and hot spot size parameter (hspos)), and solar-object-sensor observation geometry (solar zenith angle (SZA), viewing zenith angle (VZA), and relative azimuth angle (RAA)). PROSAIL can output directional canopy reflectance, which is also represented using canopy reflectance or reflectance without explicit specification in the following sections. For further details, refer to the original papers about the PROSPECT (Jacquemoud and Baret, 1990) and SAIL model (Verhoef, 1984).

The coupling of APSIM and PROSAIL is realized by passing output variables of APSIM to PROSAIL as input variables. This permits the coupling model to estimate canopy reflectance from 400 nm to 2500 nm in 1 nm intervals at defined observation conditions (determined by latitude, day of year, and time of day) given that required parameters are specified. The transformation of variables is based on a series of equations (Table 2) and details are provided in Supplementary Protocol S1.
The synthetic datasets

Four sites were chosen to represent diverse soil and climate environments across the Australian wheatbelt (Table 3). Simulations were run with historical weather records from 2000 to 2019, typical soil conditions with best initial soil water in the long term, and local management practices (i.e. fertilization, Table 3; Chenu et al., 2013). For each site in each year, nine cultivars varying in maturity type (Young, spring/very-fast; Gauntlet, spring/fast; Ellison, spring/mid-fast; Wills, spring/mid; Lancer, spring/slow; Forrest, spring/very-slow; Longsword, winter/fast; Kittyhawk, winter/mid; and Manning, winter/slow) and three sowing dates (1 May, 29 May, and 26 June) were selected to characterize different wheat growth patterns (Table 3). In total, 2160 seasons were simulated by APSIM. Daily outputs of major crop traits (Table 2) were sampled fortnightly from emergence to physical maturity, resulting in 25201 unique records; these APSIM output records were converted into PROSAIL input records using the variable transformation formulas in Table 2, which were used to determine the range and density distribution of each PROSAIL input variable.

The synthetic dataset was a set of data pairs constructed from PROSAIL input parameters and corresponding canopy reflectance. Based on the combination mode of PROSAIL input parameters, the synthetic datasets were assigned two names: rdata (combined randomly) and cdata (combined based on constraints). The rdata synthetic dataset was generated in the traditional way used in previous model inversion studies (e.g. Atzberger, 2004; Baret et al., 2007), which used the ranges and distributions of the input parameters converted from APSIM outputs but allowed PROSAIL to be run using samples from full parameter space for any combination of inputs. The second synthetic dataset with biological
Table 1 Description of input parameters and output of PROSAIL

| Variable | Unit | Description |
|----------|------|-------------|
| Input    |      |             |
| Ns       | Unitless | Leaf mesophyll structure parameter, relates to the cellular arrangement within the leaf. |
| Cw       | g cm\(^{-2}\) or cm | Leaf water content (g cm\(^{-2}\)) or leaf equivalent water thickness (cm) |
| Cm       | g cm\(^{-2}\) | Leaf dry matter content per leaf area |
| Cab      | µg cm\(^{-2}\) | Leaf Chl a and b content per leaf area |
| Car      | µg cm\(^{-2}\) | Leaf carotenoid content per leaf area |
| Cant     | µg cm\(^{-2}\) | Leaf anthocyanin content per leaf area |
| Cbrown   | Unitless | Leaf brown pigment concentration |
| rsoil    | Unitless | Reflectance of soil as a libertarian surface. It is usually adjusted by a soil brightness factor: assol (to be multiplied with single rsoil spectrum) or psoil (scaling factor between the two model-implemented rsoil spectra—wet versus dry). |
| LAI      | m\(^{2}\) m\(^{-2}\) | Leaf area index |
| LIDF     | – | Leaf inclination distribution function. There are two methods provided to calculate LIDF: use two parameters (LIDFa and LIDFb) or a single parameter (ALA for average leaf angle, degree) |
| hspot    | m m\(^{-1}\) | Hot spot size parameter. It is primarily designed to correct the canopy reflection regarding bidirectional effects. Hot spot effect is the case where a spot displays maximum reflectivity and appears brighter than surroundings because of no visible shadows at the hot spot position where the sensor is in direct alignment between the sun and the ground target. |
| SZA      | ° | Solar zenith angle (from vertical) |
| VZA      | ° | Viewing (or observing) zenith angle (from vertical) |
| RAA      | ° | Relative azimuth angle. It is equal to viewing azimuth minus solar azimuth angle. |
| Output   |      |             |
| resv     | Unitless | Directional canopy reflectance |

Constraints (cdata) used only input data converted from APSIM output variables to explore a subspace of input parameters (i.e. limited by the APSIM biology) to run PROSAIL. Both rdata and cdata have 25 201 samples. Taking rdata as a table with fixed row–column order, the rdata can be considered as breaking the potential associations among parameters within rows (Supplementary Fig. S1).

The main difference between rdata and cdata is that the value of input parameters can be combined randomly in rdata, while in cdata the input parameter combinations are constrained to wheat growth patterns estimated using APSIM. In fact, some traits (e.g. LAI, Cab, Cm, and Cw) of crops are directly or indirectly associated with each other, meaning that they are not independent and cannot be randomly or arbitrarily combined. For example, the total leaf area gradually increases and reaches its maximum value at the flag leaf during the wheat-growing season; the total leaf water content reduces along with more senescent leaves (Supplementary Fig. S2). This means that the rdata, generated in a random combination way without consideration of biological constraints among traits, may include those ‘unrealistic’ combinations that cannot exist in field conditions.

The design of simulation experiments

The major objective of these simulation experiments is to compare the theoretical estimation accuracy of target crop traits (i.e. LAI, Cab, Cm, and Cw) based on two synthetic datasets without (rdata) and with (cdata) biological constraints in terms of application of two observation geometries, five wavelength ranges, and three retrieval methods. Reflectance in a specific wavelength range and angles related to observation geometry were used as predictive variables of the predictive model to estimate crop traits with varying retrieval methods. In total, 60 simulation experiments were conducted (2 × 2 × 3 × 2 = 24) (see below for details).

Because the canopy reflectance is co-determined by the canopy status as well as the observation geometry (i.e. VZA, SZA, and RAA), some scholars considered taking these angles together with reflectance as predictive variables to increase estimation accuracy (e.g. Bacour et al., 2006; Baret et al., 2007; Dhakar et al., 2019). The solar zenith and azimuth angles change during the compositing period of images from a drone during a flight, and the variance of angles within a flight increases with increasing flight time, but calculation of these angles is tedious. Thus, two ways of angle utilization (i.e. ‘na’, with angles in model input; ‘na’, without angles in model input) were considered here to analyse the effects on estimation accuracy of target crop traits. In this research, these angles only included SZA and RAA as a nadir view was assumed here (VZA=0).

Five sets of wavelength ranges were considered to suit different types of instrumentations and applications: four sets for narrow-band reflectance at 400–2500 nm [visible to shortwave infrared (VNIR)]; 400–750 nm (visible, VIS), 1000–2500 nm (SWIR), and 400–1000 nm [visible to near infrared (VNIR)], and one set for broad-band reflectance at 400–900 nm [visible to near infrared (multispectral) (mVNIR)] (Fig. 2). The narrow-band reflectance mimics virtual hyperspectral cameras with very high spectral resolution of 1 nm at defined ranges. The broad-band reflectance mimics a commercial multispectral camera taken as an example (MicaSense Rededge, https://www.micasense.com). Five broad-band reflectances of the multispectral camera were resampled from the simulated 1 nm reflectance values based on the spectral response coefficients (provided by MicaSense; Supplementary Fig. S3) in ranges of 463–485 nm (blue), 550–570 nm (green), 663–673 (red), 820–860 nm (NIR), and 712–722 (red-edge). The band reflectance can be calculated as follows:

$$
\rho_i = \frac{\sum_{j=1}^{\lambda_m} q_{ijk} \times \rho_j}{\sum_{j=1}^{\lambda_m} q_{ijk}}
$$

where \(\rho_i\) represents simulated band reflectance for the ith band of the camera; \(\rho_j\) represents the simulated hyperspectral reflectance for the jth wavelength; \(q_{ijk}\) represents the response coefficient of the ith band of the camera for the jth wavelength; and \(\lambda_{min}\) and \(\lambda_{max}\) represent the lower and upper band wavelength limits for the ith band, respectively.

Three retrieval methods were selected, namely look-up table (LUT), random forest (RF), and feed-forward neural network (FFNN) as described below.
Table 2 Variable transformation from APSIM output to PROSAIL input

| APSIM output variable | Transformation formula | PROSAIL input variable |
|-----------------------|------------------------|------------------------|
| LAI\textsubscript{total}, LDW | NS=(0.9×SLA+0.025)/(SLA−0.1) (Jacquemoud and Baret, 1990) where SLA=10×LAI\textsubscript{total}/LDW | Ns |
| Z\textsubscript{w}, LAI\textsubscript{total}, LAI\textsubscript{dead} | \( C_w = \frac{-0.000196 - Z_w + 0.0298}{0.0223} \exp(-1.90 \cdot f\textsubscript{dead}) \quad (f\textsubscript{dead} = 0) \) \( f\textsubscript{dead}=LAI\textsubscript{dead}/LAI\textsubscript{total} \) | Cw |
| LDW, LAI\textsubscript{total} | Cm=10^{-4}×LDW/LAI\textsubscript{total} | Cm |
| CNC, LAI\textsubscript{total} | Cab=26×LNC (Yang et al., 2015) where LNC=CNC/LAI\textsubscript{total} | Cab |
| / | Car=0.216×Cab (Yang et al., 2015) | Car |
| / | Fixed C\textsubscript{ant} to 0 | C\textsubscript{ant} |
| / | Fixed C\textsubscript{brown} to 0 | C\textsubscript{brown} |
| / | Fixed ps\textsubscript{oil} to 1 | ps\textsubscript{oil} |
| LAI\textsubscript{total} | LAI=LAI\textsubscript{total} | LAI |
| / | Fixed AL\textsubscript{A} to 50° | AL\textsubscript{A} |
| LAI\textsubscript{total} | h\textsubscript{spot}=a/LAI\textsubscript{total} (Verhoef and Bach, 2003) (a is an empirical parameter and is set as 0.5) | h\textsubscript{spot} |
| L, DOY | \cos(SZA)=\sin(L)\sin(δ)+\cos(L)\cos(δ)\cos(h) \text{ given } δ=23.45\sin[360/365(284+DOY)] \text{ and } h=15(AST−12) | SZA |
| / | Fixed V\textsubscript{ZA} to 0 | V\textsubscript{ZA} |
| L, DOY | RAA=S\textsubscript{AA}−V\textsubscript{AA} \text{ given } V\textsubscript{AA}=0 (when V\textsubscript{ZA}=0) | RAA |
| / | Fixed V\textsubscript{ZA} to 0 | V\textsubscript{ZA} |

‘/’ denotes no output for equivalent input; SLA (cm\(^2\)/mg\(^{-1}\)), specific leaf area, that is leaf area per unit leaf dry weight; LAI\textsubscript{total} (m\(^2\)/m\(^{-2}\)), total leaf area index; LDW (g/m\(^{-2}\)), leaf dry weight per unit planting area; Z\textsubscript{w}, decimal Zadok score for the growth stage; f\textsubscript{dead} the fraction of dead leaves; LAI\textsubscript{dead} (m\(^2\)/m\(^{-2}\)), leaf area index of senesced or dead leaves; LNC (g/m\(^{-2}\)), leaf nitrogen content per unit leaf area; CNC (g/m\(^{-2}\)), canopy nitrogen content per unit planting area; L (°), local latitude; δ (°), solar declination angle; DOY, day of year; h (°), hour angle; AST (h), apparent solar time for phenotyping crops and here is set at three levels: 10.00, 12.00, and 14.00 h.

Table 3 Information of genotype, environment, and management used for simulation to represent the Australian wheatbelt

| Cropping area | Site | Latitude | Longitude | Soil classification | PAWC at sowing (mm) | Nitrogen (kg ha\(^{-1}\)) |
|---------------|------|----------|-----------|--------------------|---------------------|-------------------------|
| West          | Merredin | -31.48   | 118.28    | Shallow loamy duplex | 86                  | 20-20-30                |
| South-east    | Yanco  | -34.61   | 146.42    | Brown sodosol      | 191                 | 40-40-40                |
| East          | Narrabri | -30.34   | 149.76    | Grey vertosol      | 218                 | 130-0-0                 |
| East          | Gatton | -27.54   | 152.33    | Black vertosol     | 225                 | 20-30-0                 |

Plant-available water content (PAWC) at sowing is indicated for each soil at its best level of initial soil water (referring to Chenu et al., 2013). Applied nitrogen is represented by ‘x-y-z’: x, applied at sowing; y, z, applied at stage of “beginning of stem elongation” and “flag leaf just visible”, respectively.

The implementation of retrieval methods

Based on different combinations of two synthetic datasets, five wavelength ranges, and two types of angle utilization considered, there are 20 (2 × 5 × 2 = 20) different simulation experiments that need to be conducted with each retrieval method (Table 4). Experiments 1–10 and 11–20 are constructed from the same subset of rdata and cdata, respectively, with predictive variables varying in wavelength range and angle utilization. For example, there are 353 predictive variables (351 narrow-band reflectance variables at 400–750 nm plus two angle variables including SZA and RAA) for Experiment 1, while Experiment 10 has five predictive variables (five broad-band reflectance variables, i.e. blue, green, red, NIR, and red-edge). The response variables for all experiments are the four crop traits: LAI, Cab, Cm, and Cw.

Both predictive variables and response variables were normalized with the zero-mean normalization approach to prevent any scaling issue. No noise or bias was added to synthetic data. For each experiment, 75% of samples of the entire dataset were randomly selected as the training set to generate the LUT or train RF and FFNN models. The remaining 25% of samples were used as the test set to evaluate the theoretical performance of estimation accuracy. A 4-fold cross-validation...
was conducted with similar results in cross-validation (Supplementary Fig. S4).

Look-up table
The LUT is a statistical method based on global searching which was generated here with the training set. The estimation of target variables is achieved by minimizing a cost function that measures the difference in predictive variables between the tested sample (from the test set) and records in the LUT. A cost function based on root mean square error (RMSE) was used to find the best solutions of estimated variables:

$$RMSE = \frac{1}{m} \sum_{i=1}^{m} (V_i - V_{i,LUT})^2$$

where m represents the number of predictive variables (here it is reflectance or angles); $V_i$ represents the ith predictive variable of a sample from the test set, while $V_{i,LUT}$ represents the ith predictive variable of a record stored in LUT. For estimation of each test sample, the entire LUT was sorted based on RMSE, and the mean of the target traits corresponding to the best solutions leading to minimum RMSE were considered as the final solution of estimated variables (i.e. LAI, Cab, Cm, and Cw). Referring to the setting of the number of best solutions ($k$) in previous studies (e.g. Combal et al., 2003; Quan et al., 2017), we used the mean of the best 10 solutions ($k=10$) instead of the best one solution ($k=1$) to alleviate the ill-posed problem. Increasing $k$ from 1 to 10 apparently improved the estimation accuracy (Supplementary Fig. S5) but increasing $k$ from 10 to 50 produced no further improvement (data not shown), which was consistent with reported results for $k$ selection in Quan et al. (2017).

Random forest
RF is an ensemble method based on decision trees, belonging to the scope of machine learning. The final prediction is the mean of predictions of multiple decision trees (base learners), trained on different subsets of the same training dataset, in order to overcome the overfitting by an individual base learner (Breiman, 2001). The literature suggests three key hyperparameters (e.g. Lu and He, 2019; Shah Fig. 2. Selected wavelength ranges in the visible to infrared range. There are a few strong absorption peaks at wavelength around 970, 1200, 1450, 1930, and 2500 nm for water (Ma et al., 2019), and around 450 nm and 680 nm for chlorophyll (Féret et al., 2017).

Table 4

| No. | Model input | Dataset type | No. | Model input | Dataset type |
|-----|-------------|--------------|-----|-------------|--------------|
| 1   | VNSWIR      | wa           | rdata | VNSWIR      | wa           | cdata        |
| 2   | VIS         | wa           | rdata | VIS          | wa           | cdata        |
| 3   | SWIR        | wa           | rdata | SWIR         | wa           | cdata        |
| 4   | VNIR        | wa           | rdata | VNIR         | wa           | cdata        |
| 5   | mVNIR       | wa           | rdata | mVNIR        | wa           | cdata        |
| 6   | VNSWIR      | na           | rdata | VNSWIR       | na           | cdata        |
| 7   | VIS         | na           | rdata | VIS          | na           | cdata        |
| 8   | SWIR        | na           | rdata | SWIR         | na           | cdata        |
| 9   | VNIR        | na           | rdata | VNIR         | na           | cdata        |
| 10  | mVNIR       | na           | rdata | mVNIR        | na           | cdata        |

Model input represents the predictive variables of look-up table, random forest model, or feedforward neural network model. Angle utilization: ‘wa’, with angles; ‘na’ without angles.
et al., 2019), namely the number of trees or base learners (n_estimators), the number of features to consider in finding the best split (max_features), and the minimum number of samples required at a leaf node (min_samples_leaf). max_features was set as \(\log_2(n_{\text{features}})\) where \(n_{\text{features}}\) represents the number of predictive variables to add the attribute disturbance and increase the interindividual differences of base learners (Breiman, 2001).

Several values were evaluated to find the optimal settings, namely, 5, 50, 100, 200, 300, 400, 500, and 1000 for n_estimators, and 1, 5, 10, and 15 for min_samples_leaf. The mean squared error was used for evaluation of model performance during training. n_estimators and min_samples_leaf were set to 200 and 1, respectively, through independent evaluation for each experiment, which provided the optimal performance over the corresponding validation set (i.e. a subset of the training set) for all experiments. The RF model was implemented in Python Version 3.7.2 using the scikit-learn 0.24.2 (https://scikit-learn.org).

Feed-forward neural network

FFNNs, also called deep feedforward networks, or multilayer perceptrons, are the quintessential deep learning models (Goodfellow et al., 2016). An FFNN consists of an input layer, an output layer, and one or more hidden layers. According to the universal approximation theorem (Hornik et al., 1989), an FFNN can approximate an arbitrary function even with only one hidden layer that is sufficiently wide. However, simply increasing the number of neurons can easily lead to overparameterization, and previous studies showed that increasing hidden layers typically resulted in better generalization (e.g. Lin et al., 2014; Zhang et al., 2017). Five key hyperparameters (governing the training process and the architecture) need to be determined before training an FFNN, namely the activation function (transform data), the optimizer (specify how the training proceeds), the learning rate (control the magnitude of weights update during training), the unit number, and the hidden layer number.

To find the optimal model structure, three common values of learning rate (i.e. 0.1, 0.01, and 0.001), 10 activation functions, eight optimizers, five levels of hidden layer number from shallow (1) to deep (5), and eight levels of unit number from narrow (2^1) to wide (2^8) were tested. The mean squared error was used as the loss function for evaluation of model performance during training and the maximum training epoch was up to 1000 with an early stopping strategy to prevent overfitting. The configurations of the five parameters (i.e. with 256 units in each of three hidden layers, 0.001 as learning rate, ‘relu’ as activation, ‘Adamax’ as optimizer) were obtained through independent evaluation for each experiment, which provided the optimal and robust performance over the corresponding validation set (i.e. a subset of the training set) for all experiments. The FFNN model was implemented using Keras in TensorFlow 2.3.0 (https://www.tensorflow.org/).

Validation using the field experiment

In order to demonstrate the applicability of adding biological constraints to synthetic datasets in practice, the final predictive models were tested on the spectral imagery from a field experiment. Due to lack of in situ reference data, only models trained on synthetic datasets in the range of the mVNIR were validated on real multispectral images to predict LAI, Cm, and Cw. However, quadrat harvest and UAV phenotyping did not always occur on the same day. Missing LAI measurements from quadrat harvest were interpolated with a fitted piecewise function based on all observed LAIs from quadrat harvests across the growth season (Chen et al., 2022), while missing Cm and Cw values were interpolated with a linear function. The distributions of crop traits at each UAV phenotyping date between plant emergence and flag leaf are presented in Fig. 3.

The multispectral camera used in this study was a MicaSense Red-Edge camera (https://www.micasense.com). The multispectral data were captured from a UAV-based phenotyping platform during 10.00 h to 14.00 h on three defined dates (Fig. 3), with a flight height of 20 m resulting in a ground sample distance of ~1.3 cm. Raw multispectral images were processed in Pix4Dmapper software (https://www.pix4d.com/product/pix4dmapper-photogrammetry-software) to generate the calibrated reflectance of each band for the whole field. A normalized difference vegetation index (NDVI) map was computed from red and NIR bands based on the formula: NDVI=(NIR−R)/NIR+R. The NDVI map was used to generate the vegetation–background binary map based on threshold classification. The threshold was empirically set as 0.5 (tillering stage), 0.65 (stem elongation stage), and 0.75 (flag leaf stage). These threshold values can effectively classify soil pixels as background and avoid classifying green leaves as background, especially when plants are small (Supplementary Fig. S6). Using the binary map as a mask, the value of background pixels in the original reflectance map was replaced with the corresponding soil reflectance used in synthetic data, resulting in a new reflectance map named ‘background-corrected reflectance map’. Each background-corrected reflectance map for the entire field was then segmented into individual plots according to the experimental design. Marginal areas from adjacent plots and plot gaps as well as harvest areas were trimmed from the plot images. The pixel-scale reflectance from the background-corrected reflectance map was averaged to generate the plot-scale reflectance that was used in predictive models to predict crop traits. Finally, the predicted plot value for LAI, Cm, and Cw was compared with observed LAI, Cm, and Cw, respectively, to calculate the related metrics for performance evaluation. Further details about image processing can be found in the study of Chen et al. (2022).

Results and discussion

In the synthetic data study, compared with models using rdata, the models using cdata significantly improved the estimation accuracy of four response variables or crop traits (i.e. LAI, Cab, Cm, and Cw), although the magnitude of improvement depended on the application of angle utilization, wavelength ranges, and retrieval methods (Supplementary Fig. S7). The theoretical effects of these three factors are discussed in detail below. Additionally, models trained on synthetic datasets in the range of the mVNIR were tested on real multispectral data to predict LAI, Cm, and Cw.

Slight effects of utilization of angles related to observation geometry

The relative root mean squared error (RRMSE) was compared for four crop traits (i.e. LAI, Cab, Cm, and Cw) predicted for models with and without angles (i.e. SZA and RAA) related to irrigation and fertilization regimes, and each block was split into small plots of ~14 m^2 (2 × 7 m), each with seven rows and a 25 cm row spacing. Quadrat harvests comprising 0.5 m length of four inner rows (~0.5 m^2) were taken in 84 plots (comprising seven genotypes, four water–nitrogen treatments, and three replicates) to conduct ground measurements of LAI, Cm, and Cw. However, quadrat harvest and UAV phenotyping did not always occur on the same day. Missing LAI measurements from quadrat harvest were interpolated with a fitted piecewise function based on all observed LAIs from quadrat harvests across the growth season (Chen et al., 2022), while missing Cm and Cw values were interpolated with a linear function. The distributions of crop traits at each UAV phenotyping date between plant emergence and flag leaf are presented in Fig. 3.
observation geometry as model inputs (Table 4). Utilization of observation geometry in the models only caused slight changes of RRMSE to estimate crop traits (i.e. –4.9% to 1.9%, –0.9% to 11.3%, –2.0% to 3%, and –4.2% to 2.3% for LAI, Cab, Cm, and Cw, respectively, in the test dataset), whichever retrieval method and wavelength range was used (Fig. 4). Changes in RRMSE ranged from –4.9% to 11.3% in the models trained on rdata and from –1.5% to 0.5% in the models trained on cdata. The minor improvements in the latter were caused by the better performance of models trained on cdata (see below for details). The utilization of observation geometry resulted in slight improvements of estimation accuracy when retrieving with FFNN but no improvements with LUT or RF (Supplementary Fig. S7), which agreed with methodologies in the literature; for example, these angles have been used in neural network-based methods (e.g. Bacour et al., 2006; Dhakar et al., 2019), but rarely used in non-neural network-based methods including LUT and RF (e.g. Shah et al., 2019; Upreti et al., 2019). In this context, it was concluded that it would be reasonable to omit these angles from model inputs in the subsequent analyses.

**FFNN outperformed RF and LUT in joint estimation of multiple crop traits**

Compared with LUT and RF, on average, the use of FFNN reduced RRMSE by 1.9% (LAI), 2.7% (Cab), 5.5% (Cm), and 4.2% (Cw) for rdata and within 1% for the four traits for cdata (Fig. 5). The better performance from FFNN is thought to be due to superior performance of neural networks in approximating the function between inputs and outputs and optimizing the mathematical constraints among multiple

![Figure 3](image-url)  
**Fig. 3.** Distribution of observed values of crop traits (LAI, Cm, and Cw) at three UAV phenotyping dates corresponding to tillering stage, start of stem elongation stage, and flag leaf stage, respectively (i.e. 18, 40, and 59 d after sowing).

![Figure 4](image-url)  
**Fig. 4.** Comparison of relative root mean squared error (RRMSE, %) of four crop traits (i.e. LAI, Cab, Cm, and Cw) predicted on synthetic test datasets with different retrieval methods and wavelength ranges with or without angles as model input. cdata and rdata denote the synthetic dataset with and without biological constraints, respectively. Both data types had 30 independent data points, 10 for each method with various model inputs (Table 4).
outputs, given the use of a universal approximation theorem (Hornik et al., 1989). The estimation of crop traits for rdata was mainly based on the correlation between reflectance and traits, while the mathematical constraints, learnt during FFNN training, provided supplemental information to improve estimation accuracy. This finding was consistent with previous studies reporting that multitask deep learning models can exploit correlations among traits to improve prediction accuracy for green area index and the fraction of absorbed photosynthetically active radiation (e.g. Danner et al., 2021) as well as physiological traits related to photosynthesis (Fur banker et al., 2021). The information carried in reflectance across the entire range from 400 nm to 2500 nm was insufficient to predict Cm but sufficient to predict LAI, Cab, and Cw based on sensitivity analysis results (Supplementary Fig. S8). This conclusion was consistent with research for similar parameter ranges (Danner et al. 2019), which suggested that the averaged reduction of RRMSE in using FFNN instead of RF or LUT for estimation of Cm was larger than that for LAI, Cab, and Cw. Since biological constraints were more effective than mathematical constraints learnt during training FFNN for joint estimation (see below for details), the use of cdata appeared to weaken the superiority of FFNN.

**Significant effects of chosen subsets of wavelength ranges**

The effective wavelength ranges for prediction varied for crop traits. For estimation of LAI with rdata, the RRMSE for each wavelength range was ordered as: SWIR (7.7%)<VNSWIR (8.7%)<VNIR (8.9%)<VIS (10.4%)<mVNIR (11.8%) (Fig. 6A). SWIR was not suitable to retrieve Cab for rdata (RRMSE=60.1%), while the RRMSE was ordered as: VIS (7.3%)<VNIR (7.8%)<VNSWIR (7.9%)<mVNIR (10.9%) for the other four ranges (Fig. 6B). The RRMSE of Cm estimation for rdata was smaller in VNSWIR/VNIR/SWIR (6.0–6.7%) than in VIS (12.7%) and mVNIR (15.1%) (Fig. 6C). The best wavelength range for Cw estimation for rdata was SWIR (5.7%), followed by VNSWIR (8.0%) and VNIR (11.2%), while mVNIR and VIS were not suitable (RRMSE>50%) (Fig. 6D). The suitability of wavelength range for prediction of specific crop traits was supported by sensitivity analysis results (Supplementary Fig. S4) and confirmed with previous reported results (e.g. Baret et al., 1998; Gitelson, 2004; Ustin et al., 2009; Féret et al., 2017).

Compared with models trained on rdata, the models trained on cdata significantly improved estimation accuracy, resulting in substantial reduction of RRMSE by 5.9–9.0% for LAI, 4.4–49.2% for Cab, 4.0–12.5% for Cm, and 4.4–43.7% for Cw (Fig. 6). However, for models trained on cdata, the SWIR, VNSWIR, and VNIR models still resulted in smaller RRMSE to VIS or mVNIR models for estimation of LAI, Cm, and Cw, while VIS, VNIR, VNSWIR, and mVNIR models resulted in smaller RRMSE than the SWIR model for estimation of Cab.

The models trained on cdata also resulted in relatively accurate estimation of a specific crop trait from a given wavelength range which was unsuitable for estimation of the corresponding crop trait for rdata. For example, models using cdata enabled Cw to be retrieved from VIS (RRMSE=6.9%) and mVNIR (RRMSE=8.1%) (Fig. 7A, B), enabled Cab to be retrieved from SWIR (RRMSE=10.9%) (Fig. 7C), and significantly improved Cm estimation from VIS (RRMSE=1.8%) and mVNIR (RRMSE=2.6%) (Fig. 7D, E). Under these conditions, the estimation of Cw, Cab, or Cm was mainly based on prior biological constraints between crop traits defined in APSIM instead of the correlations between crop traits and radiometric information. Overall, the models trained on cdata could achieve accurate joint estimation in VNIR, with an RRMSE of 1.8, 3.4, 2.1, and 2.8%, respectively, for LAI, Cab, Cm, and Cw (Fig. 6).

![Fig. 5](image_url) Comparison of relative root mean squared error (RRMSE, %) of four crop traits (i.e. LAI, Cab, Cm, and Cw) predicted on synthetic test datasets with feedforward neural network (FFNN) and look-up table (LUT) or random forest (RF). cdata and rdata denote the synthetic dataset with and without biological constraints, respectively. Both data types had 10 independent data points: five for FFNN versus LUT and the other five for FFNN versus RF (without angles in model input, Table 4).
Application of the FFNN model on real UAV-based multispectral images

The established FFNN models trained over synthetic datasets in the range of mVNIR were applied on MicaSense multispectral images captured from a UAV-based phenotyping platform to evaluate models’ practical performance with ground references of crop traits (i.e. LAI, Cm, and Cw) between the plant emergence and flag leaf stages (Fig. 8; Table 5). In the comparison of observed values against predictions, the FFNN model trained on rdata showed a good agreement for LAI (RMSE=0.36 m² m⁻²); however, the model trained on cdata resulted in less accurate estimation of LAI (RMSE=0.54 m² m⁻²), with an increasing overestimation for higher observed LAI (Fig. 8A; Table 5). The model trained on rdata failed to predict Cw from multispectral data, with a near horizontal slope of 0.01 between predictions and observations (Fig. 8C; Table 5), due to the lack of effective spectral information in the range of mVNIR for Cw prediction as mentioned above. In contrast, the model trained on cdata enable Cw to be predicted with a slope of 0.51 (Fig. 8C; Table 5), in which situation the prediction was based on prior biological constraints between crop traits expressed in cdata. No matter whether the models were trained on rdata or cdata, poor estimation of Cm (Fig. 8B) occurred, with systematic overestimation to leverage the estimation of other traits, as the estimation of Cm was mainly based on its correlation with other traits described above. Despite the fact that practical performance was less accurate than the theoretical performance due to uncertainties in simulations and measurements, the model trained on cdata achieved good estimation of LAI (RMSE=0.54 m² m⁻²) and Cw (RMSE=0.0028 g cm⁻²) on real multispectral data (Table 5).

Potentials and limitations

The estimation accuracy of crop traits via deep learning models was substantially improved when biological constraints were introduced to the synthetic dataset as the inference/estimation of crop traits was not solely based on correlations between crop traits and radiometric information, but was also based on biological constraints among simulated crop traits. Crop traits are not independent but are explicitly or implicitly associated with each other, so they cannot be randomly combined during sampling (Supplementary Fig. S1). The ill-posed problem was effectively alleviated by the introduction of biological constraints in two ways. (i) A higher quality training dataset was provided. Those combinations of parameter values that were unrealistic to crop growth patterns defined in APSIM were omitted from the entire parameter space, resulting in a more realistic subset (i.e. cdata). This higher quality training dataset resulted in more appropriate solutions to be found in searching LUT and a better approximation function between predictive and
response variables in training of RF and FFNN. (ii) More effective constraints among response variables (i.e. crop traits) were learned in FFNN training. Given the fully connected nature of FFNN, the constraints among response variables would be learned to reduce possible combinations of values of response variables corresponding to given predictive variables. The constraints were totally mathematically based for rdata but biologically based for cdata such that crop traits within training data are constrained to APSIM’s biophysical model of trait interactions. Our study indicated that these biological constraints were able to reduce the dependence on effective information carried in spectra (Figs 6, 7) to achieve accurate estimation of crop traits. The FFNN model trained on cdata enabled the prediction of leaf water content (Cw) from real multispectral images in the visible to NIR range (i.e. mVNIR) (Table 5).

The accurate estimation presented in the synthetic data was obtained based on assumptions and hypotheses (resulting in optimistic situations rather than actual cases for variable retrieval) and might not be achieved from experimental data due to model and measurement uncertainties (Table 5). In this research, we fixed the ALA (average leaf angle) to an overall mean value for generating synthetic datasets as it is difficult to determine ALA accurately, which conflicts with the fact that the variable ALA was reported to significantly affect canopy reflectance (e.g. Verrelst and Rivera, 2017; Danneker et al., 2019). LAI and ALA are not independent during the inversion process. Assigning a fixed value to ALA can stabilize the inversion process in circumstance where constraints are not clear (Jacquemoud et al., 1995); that is probably the reason why ALA was fixed in some studies (e.g. Thorp et al., 2012; Guo et al., 2019). The relationships for leaf mesophyll structure parameter (Ns) and specific leaf area (SLA) and for Cab and leaf nitrogen content (LNC) in this study were found under specific situations and they are not so consistent across varying

Fig. 7. Known values and values of crop traits predicted with the feedforward neural network (FFNN) over the synthetic test dataset. Leaf water content (Cw, g cm\(^{-2}\)) was predicted from VIS (A) and mVNIR (B); leaf chlorophyll content (Cab, µg cm\(^{-2}\)) was predicted from SWIR (C); leaf dry matter content (Cm, g cm\(^{-2}\)) was predicted from VIS (D) and mVNIR (E). See Fig. 2 for details of wavelengths. cdata and rdata denote the synthetic dataset with and without biological constraints, respectively. Raw data have been aggregated to 15 bins based on the full range of x- and y-axes. Completed results predicted with LUT, RF, and FFNN from varying wavelength ranges for four crop traits (i.e. LAI, Cab, Cm, and Cw) can be found in Supplementary Figs S9, S10, and S11, respectively.
environmental or/and genotypic conditions. However, the models trained on cdata effectively avoided unreasonable solutions being found during LUT searching or RF and FFNN training, resulting in improvement of prediction accuracy (Supplementary Figs S8–S10). Despite the limiting situations represented in the experimental dataset, in the best situation, models trained on experimental data generally achieved more accurate estimation of crop traits than those trained on synthetic data (without biological constraints) when testing on the same subset of experimental data (Camacho et al., 2021), which supports the effectiveness of biological constraints that are implicitly included in experimental data.

The major limitation factor is the gap between observed and simulated spectra under given conditions due to simplification in RTMs, when the predictive model, completely trained on a synthetic dataset, is applied to retrieve crop traits from experimental data in practice (e.g. Danner et al., 2019). Considering crop growth patterns, the synthetic dataset with biological constraints should better represent the actual situations than that without biological constraints. Previous studies provided evidence that a model completely trained on a synthetic dataset (without biological constraints) could achieve acceptable estimation accuracy of crop traits (e.g. LAI, Cab, etc.) on experimental data (e.g. Dhakar et al., 2019; Upreti et al., 2019). For example, Danner et al (2021) reported that an FFNN model trained over synthetic data (without constraints, with artificial noise) was able to predict crop traits of multiple crops from 242 hyperspectral reflectance bands in the range of VNSWIR with an RRMSE of 9% (LAI) and 5% (Cab) on synthetic test data (our study: 9% for LAI and

Table 5 Theoretical (testing over synthetic data) and practical performance (testing over experimental data) of feed-forward neural network (FFNN) models trained over a synthetic dataset

| Test set        | Model | Crop trait | r   | RMSE  | RRMSE | Regression function |
|-----------------|-------|------------|-----|-------|-------|---------------------|
| Synthetic data  | rdata | LAI        | 0.99| 0.17  | 11.79 | y=0.98xx+0.04       |
| (n=6298)        | cdata | LAI        | 1   | 0.04  | 2.83  | y=x                 |
|                 | rdata | Cm         | 0.74| 0.00084| 15.06 | y=0.44xx+0.003      |
|                 | cdata | Cm         | 0.99| 0.00013| 2.59  | y=0.99xx            |
|                 | rdata | Cw         | 0.64| 0.0093 | 50.46 | y=0.0049xx+0.019    |
|                 | cdata | Cw         | 0.95| 0.0015 | 8.08  | y=0.99xx+0.001      |
|                 | rdata | LAI        | 0.96| 0.36  | 24.99 | y=1.03xx+0.09       |
|                 | cdata | LAI        | 0.96| 0.54  | 37.94 | y=1.22xx+0.11       |
|                 | rdata | Cm         | 0.35| 0.0018 | 44.55 | y=0.37xx+0.001      |
|                 | cdata | Cm         | 0.23| 0.0017 | 41.91 | y=0.08xx+0.01       |
|                 | rdata | Cw         | 0.10| 0.0060 | 24.80 | y=0.01xx+0.02       |
|                 | cdata | Cw         | 0.69| 0.0028 | 11.34 | y=0.51xx+0.01       |

Evaluation metrics include correlation coefficient (r), root mean squared error (RMSE), relative root mean squared error (RRMSE), and regression function. ‘cdata’ and ‘rdata’ represents the synthetic dataset with and without biological constraints, respectively. The unit of RMSE for leaf area index (LAI), leaf dry matter content (Cm), and leaf water content (Cw) is m² m⁻², g cm⁻², and g cm⁻², respectively. For the regression function, ‘x’ and ‘y’ in the regression function represent the known or observed value and its prediction, respectively.

The practical performance of the FFNN model trained over cdata is formatted in bold.
8% for Cab in VNSWIR with 2101 bands) and of 28% (LAI) and 19% (Cab) on experimental data. It is reasonable to expect that a model trained on a novel synthetic dataset with biological constraints will result in satisfactory estimation on experimental data. The estimation accuracy of Cab was reported to be improved by using a matrix-based LUT to consider the relationship between LAI and Cab, reducing RMSE from 10.5–22.9 μg cm\(^{-2}\) to 7.9–13.8 μg cm\(^{-2}\) (M. Xu et al., 2019). Different from the study of M. Xu et al. (2019), the framework of integrating APSIM and PROSAIL proposed here demonstrated a more systematic way to add biological constraints throughout the growing season and across environments. However, there is no guarantee that the model trained on cdata can always outperform the model trained on rdata when it is used on experimental data (Supplementary Fig. S12). The likely reason is that the biological constraints, defined in APSIM, may not sufficiently describe the actual situations of wheat development and growth, and errors in the characterization of environment and genotype. Hence, we suggest testing this proposed concept framework on real datasets captured with varying sensors from more diverse environments. Additionally, for a specific application in practice, the APSIM model can be calibrated with phenoology and biomass information in the conventional way before being used to inform PROSAIL input parameters. This calibrated APSIM and PROSAIL should be able to generate improved synthetic data with reflectance spectra highly correlated to the real ones captured in the field, which in turn will result in more accurate estimation of crop traits predicted with models trained on the improved synthetic data.

Fine hyperspectral reflectance at 1 nm spectral resolution theoretically provided large capacity to retrieve crop traits depending on the range of wavelength (Fig. 6). Some commercial systems already provide fine hyperspectral reflectance, such as Compact Airborne Spectrographic Imager (408–947@7.5 nm, imaging sensor) (e.g. Haboudane et al., 2004), Cubert FireFlyEYE V185 (450–950 nm@4 nm, imaging sensor) (e.g. Zhu et al., 2019), and ASD FieldSpec 4 Hi_Res (350–2500 nm@1 nm) (e.g. Shah et al., 2019, point sensor). However, fine reflectance is still hard and expensive to obtain from current commercial sensors. We also mimicked a common multispectral camera with five broad-band reflectance in blue, green, red, NIR, and red-edge regions (i.e. mVNIR). The use of mVNIR could also reveal advantages of using a deep learning model (i.e. FFNN) and adding biological constraints (i.e. cdata) to retrieve crop traits from reflectance in theory (Fig. 6) and in practice (Fig. 8). Regardless of uncertainties/ unrealities of physical models and possible error propagation in variable transformation between models, superior solutions can be expected in the future as these models (i.e. PROSAIL and APSIM) have been evolving to better describe the reality (e.g. Verhoeff, 1984; Jacquemoud and Baret, 1990; Fêret et al., 2017; Holsworth et al., 2014, 2018), and their validity for addressing practical problems (e.g. crop trait retrieval or crop growth simulation) has been widely reported (e.g. Jacquemoud et al., 2009; Chenu et al., 2017; Berger et al., 2018). As it is foreseeable that many innovative applications can be achieved from the integration of models, this will further stimulate the improvement of current versions towards more powerful models in the future.

**Conclusion**

This research demonstrated a conceptual framework of integrating APSIM and PROSAIL to predict crop traits, and the advantages of adding biological constraints in synthetic training data for joint estimation of multiple crop traits (i.e. LAI, Cab, Cm, and Cw). Our research indicated that angles related to observation geometry had slight effects on estimation accuracy, and FFNN achieved more accurate results in estimation of multiple crop traits than LUT and RF. Adding biological constraints among related variables significantly improved the estimation accuracy of crop traits. The models trained on a synthetic dataset with biological constraints enabled the estimation of crop traits not only based on correlations between crop traits and radiometric information, but also based on associations among crop traits themselves, which allowed Cab and Cw to be estimated from reflectance in SWIR and VIS or mVNIR, respectively. Although the good estimations presented in theory might not be completely reproduced in practical applications due to measurement and model uncertainties, the validation of an established FFNN model on real UAV-based multispectral images in the mVNIR range suggested the effectiveness of adding biological constraints in training synthetic datasets for prediction of multiple crop traits.

**Supplementary data**

The following supplementary data are available at *JXB* online. Protocol S1. APSIM–PROSAIL coupling model. Fig. S1. Concept map of combinations of PROSAIL input parameters. Fig. S2. Measured and predicted leaf water content for total leaves. Fig. S3. Spectral response coefficient for each band of the MicaSense RedEdge camera. Fig. S4. Cross-validation results for each band of the MicaSense RedEdge camera. Fig. S5. Comparison of the number of best solutions on LUT predictions. Fig. S6. Schematics of NDVI threshold classification results. Fig. S7. Estimation accuracy from LUT, RF, or FFNN on the synthetic test dataset. Fig. S8. Global sensitivity analysis. Fig. S9. Known values against predictions with LUT on the synthetic test dataset. Fig. S10. Known values against predictions with RF on the synthetic test dataset.
Fig. S11. Known values against predictions with FFNN on the synthetic test dataset.

Fig. S12. Observed values against predictions with LUT, RF, and FFNN on real data.

Acknowledgements
QC received a scholarship co-funded by the China Scholarship Council and the University of Queensland and has been awarded a top-up scholarship by CSIRO as a visiting student.

Author contributions
QC: implementing the research and writing the draft; SC: helping in formulating research ideas; BZ: providing guidance for using the APSIM Next Generation model, and outlining the framework of this paper with QC; TC: providing technical advice on the application of the deep learning models. All authors engaged in reviewing and editing this manuscript.

Conflict of interest
The authors declare no conflict of interest.

Funding
The field experiment and ongoing research were supported by the Grains Research and Development Corporation [grant nos CSP00179 and UOQ2003-011RTX].

Data availability
The detailed description for the proposed concept framework to integrate APSIM and PROSAIL can be found in the supplementary data, and the iterative protocols are deposited at protocols.io: https://www.protocols.io/workspaces/integration-of-apsim-and-prosail/. Details (https://www.apsim.info/apsim-next-generation/) and source code (https://github.com/APSIMInitiative/ApsimX) of the APSIM Next Generation model used in this study are openly available online. Details and source codes supporting this work are available from the corresponding author upon request.

References
Atzberger C. 2004. Object-based retrieval of biophysical canopy variables using artificial neural nets and radiative transfer models. Remote Sensing of Environment 93, 53–67.

Bacour C, Baret F, Béal D, Weiss M, Pavageau K. 2006. Neural network estimation of LAI, IAPAR, ICover and LAI-Cab, from top of canopy MERIS reflectance data: principles and validation. Remote Sensing of Environment 105, 313–325.

Baret F, Buis S. 2008. Estimating canopy characteristics from remote sensing observations: review of methods and associated problems. In: Liang S, ed. Advances in land remote sensing: system, modeling, inversion and application. Dordrecht: Springer, 173–201.

Baret F, Guyot G, Begue A, Maurel P, Podaire A. 1998. Complementarity of middle-infrared with visible and near-infrared reflectance for monitoring wheat canopies. Remote Sensing of Environment 26, 213–225.

Baret F, Hagolle O, Geiger B, et al. 2007. LAI, IAPAR and ICover CYCLOPES global products derived from VEGETATION. Part 1: principles of the algorithm. Remote Sensing of Environment 110, 275–286.

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 Sensing 10, 85.

Breiman L. 2001. Random forests. Machine Learning 45, 5–32.

Camacho F, Fuster B, Li W, Weiss M, Ganguly S, Lacaze R, Baret F. 2021. Crop specific algorithms trained over ground measurements provide the best performance for GAI and IAPAR estimates from Landsat-8 observations. Remote Sensing of Environment 260, 112453.

Chapman SC, Zheng B, Andries B, Madec S, Solan B, George-jaeggli B, Hammer G, Jordan D. 2018. Visible, near infrared, and thermal spectral radiance on-board UAVs for high-throughput phenotyping of plant breeding trials. In: Thenkabail PS, Lyon JG, Huetel A, eds. Biophysical and biochemical characterization and plant species studies. Boca Raton, FL: CRC Press, 275–299.

Chen Q, Zheng B, Chenku H, Pu, Chapman SC. 2022. Unsupervised pilot-scale LAI phenotyping via UAV-based imaging, modelling and machine learning. Plant Phenomics 2022, 9768253.

Chenu K, Delhimfard R, Chapman SC. 2013. Large-scale characterization of drought pattern: A continent-wide modelling approach applied to the Australian wheatbelt—spatial and temporal trends. New Phytologist 198, 801–820.

Chenu K, Porter J, Martre P, Basso B, Chapman SC, Ewert F, Bind M, Asseng S. 2017. Contribution of crop models to adaptation in wheat. Trends in Plant Science 22, 472–490.

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 Sensing of Environment 84, 1–15.

Danner M, Berger K, Wocher M, Mauser W, Hank T. 2019. Fitted PROSAIL parameterization of leaf inclinations, water content and brown pigment content for winter wheat and maize canopies. Remote Sensing 11, 1150.

Danner M, Berger K, Wocher M, Mauser W, Hank T. 2021. Efficient RTM-based training of machine learning regression algorithms to quantify biophysical & biochemical traits of agricultural crops. ISPRS Journal of Photogrammetry and Remote Sensing 173, 278–296.

David E, Seroura M, Smith D, et al. 2021. Global wheat head detection 2021: an improved dataset for benchmarking wheat head detection methods. Plant Phenomics 2021, 9846158.

Dhakar R, Sehgal VK, Chakraborty D, Sahoo RN, Mukherjee J. 2019. Field scale wheat LAI retrieval from multispectral Sentinel 2A-MSI and LandSat 8-OLI imagery: effect of atmospheric correction, image resolutions and inversion techniques. Geocarto International 36, 2044–2064.

Dorigo WA, Zurita-Milla R, de Wit AJW, Brazile J, Singh R, Schaeppme ME. 2007. A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. International Journal of Applied Earth Observation and Geoinformation 9, 165–193.

Duan T, Zheng B, Guo W, Ninomiya S, Guo Y, Chapman SC. 2017. Comparison of ground cover estimates from experiment plots in cotton, sorghum and sugarcane based on images and ortho-mosaics captured by UAV. Functional Plant Biology 44, 169–183.

Ehsani A, Ghanian MA, Baxter I. 2015. Lights, camera, action: high-throughput plant phenotyping is ready for a close-up. Current Opinion in Plant Biology 24, 93–99.

Féret JB, Gitelson AA, Noble SD, Jacquemoud S. 2017. PROSPECT-D: towards modeling leaf optical properties through a complete lifecycle. Remote Sensing of Environment 193, 204–215.
Furbank RT, Silva-Perez V, Evans JR, Condon AG, Estavillo GM, He W, Newman S, Poiré R, Hall A, He Z. 2021. Wheat physiology predictor: predicting physiological traits in wheat from hyperspectral reflectance measurements using deep learning. Plant Methods 17, 1–15.

Gitelson A. 2004. Wide dynamic range vegetation index for remote quantification of biophysical characteristics of vegetation. Journal of Plant Physiology 161, 165–173.

Goodfellow I, Bengio Y, Courville A. 2016. Deep feedforward networks, deep learning. Cambridge, MA: The MIT Press.

Guo C, Tang Y, Lu J, Zhu Y, Cao W, Cheng T, Zhang L, Tian Y. 2019. Predicting wheat productivity: integrating time series of vegetation indices into crop modeling via sequential assimilation. Agricultural and Forest Meteorology 27, 69–80.

Haboudane D, Miller JR, Pattey E, Zarco-Tejada PJ, Strachan IB. 2004. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: modeling and validation in the context of precision agriculture. Remote Sensing of Environment 90, 337–352.

Holzworth DP, Huth NI, devil-PG, et al. 2014. APSIM—evolution towards a new generation of agricultural systems simulation. Environmental Modelling and Software 62, 327–350.

Holzworth DP, Huth NI, Fainges J, Brown H, Zurcher E, Gichot R, Verhoef W, Herrmann NI, Zheng B, Snow V. 2018. APSIM Next Generation: overcoming challenges in modernising a farming systems model. Environmental Modelling and Software 103, 43–51.

Hornik K, Stinchcombe M, White H. 1989. Multilayer feedforward networks are universal approximators. Neural Networks 2, 359–366.

Hu P, Chapman SC, Jin H, Guo Y, Zheng B. 2021. Comparison of modelling strategies to estimate phenotypic values from an unmanned aerial vehicle with spectral and temporal vegetation indexes. Remote Sensing 13, 28271–28219.

Jacquemoud S, Baret F. 1990. PROSPECT: a model of leaf optical properties spectra. Remote Sensing of Environment 34, 75–91.

Jacquemoud S, Baret F, Andrieu B, Danson FM, Jagged 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 Sensing of Environment 52, 163–172.

Jacquemoud S, Verhoef W, Baret F, Bacour C, Zarco-Tejada PJ, Asner GP, François C, Usin SL. 2009. PROSPECT + SAIL models: a review of use for vegetation characterization. Remote Sensing of Environment 113, S56–S66.

Jay S, Maupas F, Bendoula R, Gorretta N. 2017. Retrieving LAI, chlorophyll and nitrogen contents in sugar beet crops from multi-angular optical remote sensing: comparison of vegetation indices and PROSAIL inversion for field phenotyping. Field Crops Research 210, 33–46.

Jin X, Xu X, Song X, Li Z, Wang J, Guo W. 2013. Estimation of leaf water content in winter wheat using grey relational analysis—partial least squares modeling with hyperspectral data. Agronomy Journal 105, 1385–1392.

Li Z, Li Z, 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 data. Computers and Electronics in Agriculture 162, 174–182.

Lin M, Chen Q, Yan S. 2014. Network in network. In: 2nd International Conference on Learning and Representation (ICLR) 2014 - Conference Track Proceedings, 1–10.

Lu B, He Y. 2019. Evaluating empirical regression, machine learning, and radiative transfer modelling for estimating vegetation chlorophyll content using bi-seasonal hyperspectral images. Remote Sensing 11, 1979.

Lunagaria MM, Patel HR. 2019. Evaluation of PROSAIL inversion for retrieval of chlorophyll, leaf dry matter, leaf angle, and leaf area index of wheat using spectroradiometric measurements. International Journal of Remote Sensing 40, 8125–8145.

Ma J, Fan Y, Teng Q, Yang K, Huang S, Yang Y, Zhu H, Zang W, Zhu P. 2019. Estimation of high resolution corn vegetation water content based on airborne CASI/SASI hyperspectral data. In: IGARSS 2019 IEEE International Geoscience and Remote Sensing Symposium, 6685–6688.

Machwitz M, Giustarini L, Bossung C, Frantz D, Schlerf M, Lilienthal H, Wandera L, Matgen P, Hoffmann L, Udelhoven T. 2014. Enhanced biomass prediction by assimilating satellite data into a crop growth model. Environmental Modelling and Software 62, 437–453.

Madec S, Jin X, Lu H, de Solan B, Liu S, Duyfne F, Hertier E, Baret F. 2019. Ear density estimation from high resolution RGB imagery using deep learning technique. Agricultural and Forest Meteorology 264, 225–234.

Mahlein AK, Gerke EC, Steiner U, Dehne HW. 2012. Recent advances in sensing plant diseases for precision crop protection. European Journal of Plant Pathology 133, 197–209.

McCabe MF, Houborg R, Lucieer A. 2016. High-resolution sensing for precision agriculture: from Earth-observing satellites to unmanned aerial vehicles. In: Remote Sensing for Agriculture, Ecosystems, and Hydrology XVIII. 346–355. https://doi.org/10.1117/12.2241289

Quan X, He B, Yebra M, Yin C, Liao Z, Li X. 2017. Retrieval of forest fuel moisture content using a coupled radiative transfer model. Environmental Modelling Software 95, 290–302.

Roosjen PPJ, Brede B, Suomalainen JM, Bartholomeus HM, Kooistra L, Clevens JGPW. 2018. Improved estimation of leaf area index and leaf chlorophyll content of a potato crop using multi-angle spectral data—potential of unmanned aerial vehicle imagery. International Journal of Applied Earth Observation and Geoinformation 66, 14–26.

Shah SH, Angel Y, Houborg R, Ali S, McCabe MF. 2019. A random forest machine learning approach for the retrieval of leaf chlorophyll content in wheat. Remote Sensing 11, 920.

Sozzi M, Marinello F, Pezzuolo A, Sartori L. 2018. Benchmark of satellites image services for precision agricultural use. In: Proceedings of the European Conference on Agricultural Engineering (AgEng) 2018.

Thenkabail PS, Lyon JG, Huete A. 2019. Fifty years of advances in hyperspectral remote sensing of agriculture and vegetation—summary, insights, and highlights of volume III. In: Thenkabail PS, Lyon JG, Huete A, eds. Biophysical and biochemical characterization and plant species studies. Boca Raton, FL: CRC Press, 251–286.

Thorp KR, Wang G, West AL, Moran MS, Bronson KF, White JW, Mon J. 2012. Estimating crop biophysical properties from remote sensing data by inverting linked radiative transfer and ecophysiological models. Remote Sensing of Environment 124, 224–233.

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 Sensing 11, 481.

Ustin SL, Gitelson AA, Jacquemoud S, Schaepman M, Asner GP, Gamon JA, Zarco-Tejada P. 2009. Retrieval of foliar information about plant pigment systems from high resolution spectroscopy. Remote Sensing of Environment 113, S67–S77.

Verhoef W. 1984. Light scattering by leaf layers with application to canopy reflectance modeling; the SAIL model. Remote Sensing of Environment 16, 125–141.

Verhoef W, Bach H. 2003. Simulation of hyperspectral and directional radiance images using coupled biophysical and atmospheric radiative transfer models. Remote Sensing of Environment 87, 23–41.

Verrelst J, Camps-Valls G, Muñoz-Mari J, Rivera JP, Veroastraete F, Clevens JGPW, Moreno J. 2015. Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties—a review. ISPRS Journal of Photogrammetry and Remote Sensing 100, 273–290.

Verrelst J, Rivera JP. 2017. A global sensitivity analysis toolbox to quantify drivers of vegetation radiative transfer models. In: Petropoulos GP, Srivastav PL, eds. Sensitivity analysis in earth observation modelling. Amsterdam: Elsevier, 319–339.

Wan L, Zhu J, Du X, et al. 2021. A model for phenotyping crop fractional vegetation cover using imagery from unmanned aerial vehicles. Journal of Experimental Botany 72, 4691–4707.

Xu M, Liu R, Chen J, Liu Y, Shang R, Ju W, Wu C, Huang W. 2019. Retrieving leaf chlorophyll content using a matrix-based vegetation index combination approach. Remote Sensing of Environment 224, 60–73.
Xu XQ, Lu JS, Zhang N, Yang TC, He JY, Yao X, Cheng T, Zhu Y, Cao WX, Tian YC. 2019. Inversion of rice canopy chlorophyll content and leaf area index based on coupling of radiative transfer and Bayesian network models. ISPRS Journal of Photogrammetry and Remote Sensing 150, 185–196.

Yang G, Zhao C, Pu R, Feng H, Li Z, Li H, Sun C. 2015. Leaf nitrogen spectral reflectance model of winter wheat (Triticum aestivum) based on PROSPECT: simulation and inversion. Journal of Applied Remote Sensing 9, 095976.

Zhang Y, Chan W, Jaitly N. 2017. Very deep convolutional networks for end-to-end speech recognition. In: 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 4845–4849.

Zhang L, Niu Y, Zhang H, Han W, Li G, Tang J, Peng X. 2019. Maize canopy temperature extracted from UAV thermal and RGB imagery and its application in water stress monitoring. Frontiers in Plant Science 10, 1–18.

Zhu W, Sun Z, Huang Y, et al. 2019. Improving field-scale wheat LAI retrieval based on UAV remote-sensing observations and optimized VI-LUTs. Remote Sensing 11, 2456.