

DONG WANG



INTEGRATING

CROP MODELLING

AND

REMOTE SENSING

TO DESIGN A NITROGEN MANAGEMENT SYSTEM
FOR SUSTAINABLE CROP PRODUCTION

Propositions

1. Optimisation of nitrogen management simultaneously improves crop productivity, farmer's income, and nitrogen use efficiency.
(this thesis)
2. Quantifying uncertainties of crop modelling and remote sensing makes crop growth predictions more certain.
(this thesis)
3. Interdisciplinary thinking accelerates research progress.
4. Unboxing artificial intelligence creates more fear than knowledge.
5. Curiosity- and mission-driven research are equally important.
6. Engaging with social media is the final step in applied science.
7. Unbridled technological evolution harms societal developments.
8. Education shapes society.

Propositions belonging to the thesis, entitled

Integrating crop modelling and remote sensing to design a nitrogen management system for sustainable crop production

Dong Wang

Wageningen, 24 April 2024

**Integrating crop modelling and remote sensing
to design a nitrogen management system
for sustainable crop production**

Dong Wang

Thesis committee

Promotor

Prof. Dr P.C. Struik
Professor of Crop Physiology
Wageningen University & Research

Co-promotors

Dr X. Yin
Senior scientist, Centre for Crop Systems Analysis
Wageningen University & Research

Dr L. Liang
Researcher
Shanghai Lankuaikui Technology Development Co. Ltd., Shanghai, China

Other members

Prof. Dr L.F.M. Marcelis, Wageningen University & Research
Dr M. van Oijen, Independent Scientist, Edinburgh, UK
Dr J.G.P.W. Clevers, Wageningen University & Research
Dr P.A.J. van Oort, Wageningen University & Research

This research was conducted under the auspices of the C.T. de Wit Graduate School for
Production Ecology and Resource Conservation

**Integrating crop modelling and remote sensing
to design a nitrogen management system
for sustainable crop production**

Dong Wang

Thesis

submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus
Prof. Dr C. Kroeze,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Wednesday 24 April 2024
at 11 a.m. in Omnia Auditorium.

Dong Wang

Integrating crop modelling and remote sensing to design a nitrogen management system for sustainable crop production, 246 pages.

PhD thesis, Wageningen University, Wageningen, NL (2024)
With references, with summary in English

ISBN: 978-94-6469-882-4

DOI: <https://doi.org/10.18174/646128>

Abstract

Dong Wang, 2024, *Integrating crop modelling and remote sensing to design a nitrogen management system for sustainable crop production*. PhD thesis, Wageningen University & Research, Wageningen, The Netherlands. 246 pp.

Growing global food demand requires increased crop production to meet food security. Nitrogen (N) fertilisation has significantly contributed to increasing regional and global crop production. Severe environmental issues have been aggravated, however, as the N loss has increased from the blindly increasing N fertilisation for more crop production. To ensure a food-secure future, sustainable crop production is expected for attaining a high yield and high N use efficiency while reducing environmental cost. Decision tools for precise N fertilisation are required to simultaneously achieve those goals by matching crop N demand and N application in the field. However, obstacles lie in varied crop N needs and field N supply under changing weather conditions, soil status and management practices. Due to the inevitably spatial and temporal variations in crop growth, compared with *in situ* crop N needs, N rates can be easily under- or over-recommended. This dissertation aims to design a smart N management system that determines the optimal N rates from the temporally simulated and spatially monitored crop N needs by making use of the complementarities of crop modelling and remote sensing.

As crop growth and production are primarily determined by photosynthesis, the mechanistic photosynthesis-driven crop model GECROS was adopted. Its key photosynthetic parameters and their relationship with specific leaf N content (SLN) were investigated for four major field crops, two C₃ crops (rice and wheat) and two C₄ crops (maize and sorghum). Measured data were collected from experiments with different N treatments in the greenhouse. Photosynthetic capacity linearly increased with SLN and the regressed linear relationships for C₃ crops differed significantly from those for C₄ crops. The linear slope of Rubisco carboxylation capacity versus SLN for C₃ crops was significantly steeper than that of C₄ crops. Although estimated photosynthetic parameters increased with SLN, the difference among crops tended to be insignificant in the relationship between photosynthetic parameters and SLN.

Given the significant role of leaf N in affecting photosynthetic parameters, the diagnosis of canopy and leaf N status for major field crops was further investigated based on Unmanned Aerial Vehicle-acquired hyperspectral images in the field. Among five machine learning algorithms (referred to as nonparametric regression algorithms), the Partial Least Squares

Regression and Support Vector Regression models performed better than the others. Combined use of hyperspectral features, canopy reflectance, vegetation indices and texture information, improved the model performance. Regarding different predicting pathways, the prediction of leaf N traits could be improved by identifying indirect predicting pathways, although they might become much uncertain due to the introduced more regression processes.

Uncertainties of remote sensing predictions and crop model simulations were quantified to improve the forecasting of crop status by a developed Bayesian methodology. Within this method, a Markov Chain Monte Carlo approach and a Gaussian Process Regression model were adopted to obtain crop model simulations and remote sensing predictions, respectively, together with their corresponding uncertainties. The data assimilation method of Ensemble Kalman Filter was applied to update crop growth status based on the quantified uncertainties. The simulated leaf traits, leaf weight, leaf N content and leaf area index, from crop model GECROS were directly updated by the remotely sensed ones to agree better with the measurements. Those of carbon and N status of aboveground plant and grain were updated indirectly with the improved performance of forecasting. A better performance of the data assimilation system was achieved by the estimated uncertainties, compared with the ones when the pre-assumed uncertainty was used alone or when it was combined with the inflation factor being introduced into Ensemble Kalman Filter.

Based on the improved monitoring and forecasting of temporal and spatial variations of crop growth, an integrated remote sensing-crop model method by data assimilation was proposed to optimise N management *in situ*. A crop model-based N optimisation method was also investigated to temporally determine the optimal N management. The optimised in-season topdressing N of the proposed methods performed well in improving yield, profit and N use efficiency simultaneously, based on the simulated crop growth with actual weather data. The integrated crop model-remote sensing method suits better for smart fertiliser management.

Above findings suggest the merits of the designed N management system in achieving multi-objectives by connecting leaf photosynthesis, and enhancing monitoring and forecasting of crop growth. Thus, crop modelling- and remote sensing-based N management system can promisingly achieve sustainable crop production under smart farming.

Keywords: Major crop, leaf photosynthesis, leaf nitrogen trait, crop modelling, remote sensing, machine learning, uncertainty quantification, nitrogen optimisation

Table of contents

Chapter 1	General introduction	1
Chapter 2	Estimating photosynthetic parameter values of rice, wheat, maize and sorghum to enable smart crop cultivation	21
Chapter 3	Estimating leaf and canopy nitrogen contents in major field crops across the growing season from hyperspectral images using nonparametric regression	65
Chapter 4	Enhancing crop growth forecasting by incorporating estimated uncertainties for time series hyperspectral image data and crop model GECROS into Ensemble Kalman Filter	115
Chapter 5	Developing remote sensing- and crop model-based methods to optimise nitrogen management in rice fields	169
Chapter 6	General discussion	213
	Summary	233
	Acknowledgements	237
	List of publications	241
	PE&RC Training and Education Statement	243
	Curriculum Vitae	245
	Funding	246

Chapter 1

General introduction

1.1 Background

1.1.1 Challenges of sustainable crop production

Global agriculture faces enormous challenges regarding food security, environmental degradation and climate change (Foley et al., 2011; Zhang et al., 2015). A 100–110% increase in global food production from 2005 to 2050 is required to meet the growing demand (Tilman et al., 2011). However, the food insecurity level is still on the rise since 2014, especially in Asia (FAO, 2022). Also, severe environmental issues, like global warming (IPCC, 2021), water eutrophication (Diaz and Rosenberg, 2008) and inhalable particulate matter formation (Lelieveld et al., 2015), are aggravated by agriculture. For example, the nitrogen (N) loss from the regionally or globally increased N application (Fig. 1.1a) leads to adverse impacts on human health (Wang and Lu, 2020).

The increased availability of N fertiliser has enabled a strong growth of global crop production (Evans, 1997), and remained predominantly supporting major field crop production (Fig. 1.1a). For instance, compared to the year 2000, the worldwide agricultural consumption of chemical fertilisers in 2020 increased by ca. 48.7%, and ca. 49.4% of this increase was from N (FAO, 2022). Yet, N fertiliser has also been identified as the major contributor to economic loss (Wang and Zhao, 2019). Taking China for example, while making efforts to ensure food security, the estimated economic cost, as a result of damages in environment and human health, is equivalent to 7 to 10% of Chinese agricultural gross domestic product (Norse and Ju, 2015).

Seventeen Sustainable Development Goals in the United Nations' 2030 Agenda have been outlined in 2015 for Sustainable Development, including ending poverty and hunger, achieving food and nutrition security, and promoting sustainable agriculture. The improvement of N use efficiency is crucial for addressing the challenges and meeting the goals of sustainable crop production (Zhang et al., 2015). Thus, for a sustainable food-secure future, crop production is required to attain a high yield and high efficiency while reducing environmental cost (Cui et al., 2018). To accomplish this, both socio-economic factors and technological innovations should be taken into consideration (Zhang et al., 2015).

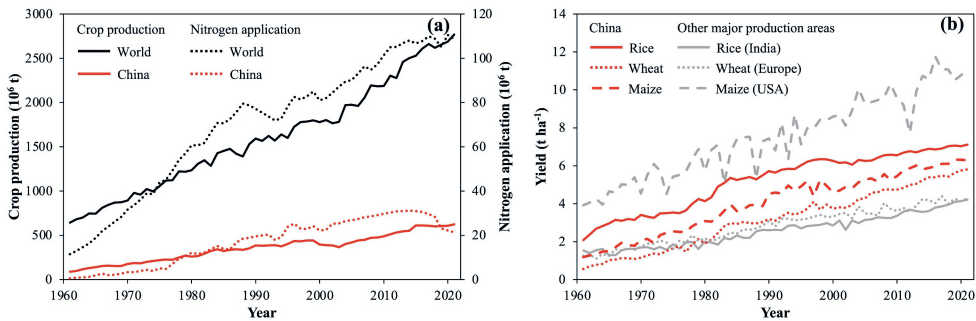


Fig. 1.1. Crop production and N use in the world and in China from 1961 to 2021. (a) Total amount of crop production for major field crops (rice, wheat, and maize) and agricultural use of N in the world and in China. (b) Yields of major field crops in China and those in other major production areas in the world. All data were collected from <https://www.fao.org/faostat/en/#data>.

1.1.2 Current crop management strategies in China

Achieving food security has consistently been a significant challenge for China (Lu et al., 2015). Food insecurity particularly tends to be worsened by its dominant smallholder farming and current land tenure system, the Household Responsibility System (Cui et al., 2018; Miao et al., 2011). As Chinese farmer cannot buy, sell or inherit the farmland, but only rent or use it, most farmers lack a long-term interest in improving soil fertility for the sake of ensuring sustainable agricultural production (Miao et al., 2011). Short-sighted decision-making and irresponsible use of land resources have become a problem (Gao et al., 2006). For instance, in pursuit of high profit by maximising yield (Fig. 1.1b), N application rate of major field crops, rice, wheat and maize, in China has risen up to 305 kg N ha⁻¹, compared to 74 kg N ha⁻¹ worldwide (Cui et al., 2018). However, yield stagnation of major crops has been widely observed for the past decade, especially in China, which occurs across 79%, 56% and 31% of cropland areas of rice, wheat and maize, respectively (Ray et al., 2012). Given the reduced potential of yield increase and the increased risk of N loss, there is an urgent need for sustainable agricultural production in China. This can be achieved by balancing crop demand and N application (Miao et al., 2011). To make field N optimisation accessible to smallholder farmers, a region-specific, yet fixed N rate without any soil and crop testing, is recommended by the government to avoid significant overdosing or N deficiency (Cui et al., 2013). Due to the considerable temporal and spatial variation of crop growth, the recommended regional optimal N rate can be either under- or over-applied, compared with *in situ* crop N needs, and thus should be further improved to become site-specific. Although considerable efforts have been

made to develop decision tools for precise fertilisation, innovative technologies are still required to be further developed in pursuit of simultaneously improving productivity, profitability and N use efficiency (Wu and Ma, 2015).

1.1.3 Opportunities for smart crop management

Smart crop management is crucial, not only for food security but also for raising income of smallholder farmers, whose livelihood depends on agricultural production, in many parts of the world (Lipper et al., 2014). To meet the challenge of sustainable agricultural production, the best hope lies in the genetic improvement of cultivars like those selected in the Green Revolution of the twentieth century and the required crop management for these cultivars. As field management affects both crop production and environmental outcomes, much attention has been paid to the effective acquisition and utilisation of “agricultural big data” via information technologies for improving crop management (Basso and Antle, 2020). Since the mid-1980s, information-based field management systems, or so-called precision agriculture techniques, have emerged (Pierce and Nowak, 1999). Due to the changing weather conditions, variable soil properties and their interactions with crop growth (Cassman et al., 2002), there are spatial and temporal variations in crop growth and crop needs. The development and combination of remote sensing, ground sensors and the Internet of Things, which can connect things simultaneously via the internet, enable to design the optimal field management to improve profitability and environmental sustainability (Gebbers and Adamchuk, 2010; Walter et al., 2017). For example, a smart crop N-management strategy that determines the optimal N rate dynamically during the growing season based on soil status, crop growth status and (future) weather conditions is expected to synchronise N application with crop N demand for sustainable agricultural production.

1.2 A review of smart nitrogen management

1.2.1 Field nitrogen management and dynamic crop modelling

Traditionally, optimum amounts of N application can be determined from curves of crop yield responses to the resource concerned, also called production functions (Redman and Allen, 1954). For decades, the focus of optimisation studies has been put on economic factors, like prices of fertiliser and grain (Tembo et al., 2008). Field experimentation is conducted in order to collect necessary data to establish the production function, e.g., the relationship between

yield and fertiliser application (Brorsen and Richter, 2012). The obtained function can be used to determine the amounts of the fertilisation needed to maximise yield, economic profit, or fertiliser use efficiency, respectively (Goulding et al., 2008). This approach can be extended to optimal resource allocation between two or more production units (e.g., fields or farms), given that production functions differ between units, for example, as a result of different soil fertility (Raun et al., 2002; Tembo et al., 2008). However, it will be a tremendous and laborious job to conduct such field experiments for establishing production functions of every unit under every climatic and edaphic condition. Also, production functions based on field experiments are static, while field crop growth is dynamic, constantly responding to growth environmental variables and varying from year to year. Dynamic crop growth simulation models (thereafter, crop models) can help to simulate crop yields if they are well calibrated and validated using experimental data (Whisler et al., 1986).

Dynamic crop models predict crop yield by quantifying nonlinear responses of many intermediate physiological processes to environmental variables like soil and weather variables (Loomis et al., 1979). The models, therefore, allow for the interactions of component processes related to time-dependent phenological, morphological, and physiological characteristics to be simulated on a daily basis. Many crop models predict crop yield, relying on model mechanisms for an accurate dynamic prediction of the primary production process – photosynthesis (Lawlor, 1995), while other physiological processes like leaf area expansion, transpiration and nutrient uptake also have a strong impact on predicted yields (Boote et al., 2013; Sinclair and Jamieson, 2006). All these physiological processes interact, in feedforward and/or feedback manners. For example, in early crop models, leaf area index (LAI) was merely considered as carbon limited, but a positive feedback loop is unavoidable: leaf weight → leaf area → photosynthesis → plant growth → leaf weight, and such a positive feedback loop generally makes predicted yield extremely sensitive to small errors in LAI-related parameters (Penning de Vries et al., 1989). Since both leaf expansion and leaf senescence are highly related to N (Sinclair and De Wit, 1976), an N-limited mechanism of leaf area development was later introduced in the photosynthesis-driven crop model GECROS (Yin and van Laar, 2005) and the combined simulation of carbon- and N-limited LAI can overcome the problem caused by aforementioned positive feedback (Yin et al., 2000). Moreover, as 75% of N in mature leaves is present in chloroplasts and 84% of N in chloroplasts is associated with photosynthesis-related proteins and molecules (Evans and Clarke, 2019), key photosynthetic parameters are known to correlate with specific leaf N content (SLN) (e.g., Evans, 1983). Thus, the quantified relationship

between photosynthetic parameters and leaf N status is of great importance in simulating crop photosynthetic capacity, as well as supporting leaf area growth and biomass accumulation (Sinclair and Horie, 1989).

To meet the N needs of crop leaf development and other organ formation, soil and fertiliser N availabilities are both necessitated. Indigenous N supply simulated by crop model normally follows soil N processes including N mineralisation and immobilisation and N losses via leaching, denitrification and volatilisation, and their interactions with weather conditions like temperature and precipitation, and soil physical and chemical properties (e.g., Yin and van Laar, 2005). With all the crop and soil related N processes, crop models are empowered to simulate crop growth under different N managements and optimise N management in the field by synchronising N fertiliser application with crop N demand (Cassman et al., 2002; Scharf, 2015). For instance, Cammarano et al. (2021) demonstrated that the optimum total application amount of N fertiliser for reducing N leaching and improving grain yield and quality can be determined based on the calibrated crop model DSSAT. However, uncertainties arise from model process descriptions, parameterisations, and required forcing inputs (Marin et al., 2017). Even with the best available crop models, difficulties exist in accurately simulating many intermediate crop characteristics and soil nutrient availabilities for crop uptake. Processes in crop model associated with N management are still of substantial uncertainties (Scharf, 2015). Especially, considering their interactions, the spatial and temporal variations in soil N supply and crop N uptake are further complicated (Jin et al., 2017). Due to the lack of *in situ* information of soil properties, crop status and management strategies, applying crop models in varied croplands remains limited (Huang et al., 2016).

1.2.2 Use of sensor technology to in situ monitor crop growth status

Because information-based technological facilities are available, instant and accurate *in situ* data of crop growth and environmental conditions makes farming processes data-driven (Wolfert et al., 2017). Instead of data collection through destructive sampling, the acquisition of *in situ* canopy state variables can be derived from spectral imaging captured by remote or proximal sensing techniques based on the spectroscopy (El-Shikha et al., 2007; Sims and Gamon, 2002; Viña et al., 2011), which, at present, emerge as a key element for the observation of in-field temporal and spatial variations in crop growth (Walter et al., 2017). Vegetation parts, like leaves, tassels, and panicles, absorbing or scattering solar radiation by chemical constituents or structural elements, determine the canopy reflectance and the shape of the

overall light spectrum (380–2500 nm) (Jacquemoud and Baret, 1990). The absorption of solar radiation in the visible (380–700 nm) spectrum is mainly due to chlorophyll and other pigments, while absorption in the near (800–1300 nm) and shortwave (1300–2500 nm) infrared spectra is influenced by the protein and cellulose (Fourty and Baret, 1997). Scattering, including both reflection and transmission, depends on the wavelength and angle of incidence (Bousquet et al., 2005; Jacquemoud and Ustin, 2001). For instance, canopy architecture, like LAI and the distribution and uniformity of vegetated and non-vegetated areas, mainly accounts for the magnitude and direction variation of the reflected signal (Kuusk, 1995; Verhoef, 1984).

Relying on the causality between canopy reflectance and biophysical characteristics, the key canopy traits can be retrieved in real-time based on the developed sensor technology. The deployed platform of sensors differentiates between proximal sensing (ground-based platforms, such as hand-held devices or vehicles) and remote sensing (air- or space-based platforms, such as Unmanned Aerial Vehicles (UAVs) and satellites). With respect to monitoring crop N status, canopy leaf-N content (N_{leaves}) has received particular attention in the field of remote and proximal sensing, due to its significance in determining crop growth (Sinclair and Horie, 1989). By mathematically combining specific multispectral or hyperspectral bands, spectral vegetation indices (VIs) are able to minimise soil, atmospheric and sun-target-sensor geometry effects, thus enhance the information contained in spectral data and finally extract variability characterised by differential crop canopy status (Moulin, 1999). There are several vegetation indices proposed as surrogates for N_{leaves} estimation, such as Nitrogen Reflectance Index (Bausch and Duke, 1996) and Double-peak Canopy Nitrogen Index (Chen et al., 2010). Although these indices are widely used, there are many more spectral bands available in sensing images, especially for hyperspectral data, which may help improve the estimation of N_{leaves} and better understand the crop canopy status. For instance, numerous studies derived VI from the near-infrared region and the red-edge region located around 730 nm, which appears to have a consistent relationship with crop N status (Berger et al., 2020; Clevers and Gitelson, 2013). Consequently, greater attention has been devoted to machine learning methods, due to its capability to process a large number of inputs and handle nonlinear tasks (Chlingaryan et al., 2018).

In order to estimate N_{leaves} more accurately via remote and proximal sensing data, several machine learning methods have been explored. Artificial Neural Networks have been generally employed in remote sensing to predict vegetation parameters (Kaul et al., 2005). When analysing high-dimensional data such as hyperspectral imaging data, Support Vector

Regression has the potential to dissolve overfitting problems (Shao et al., 2009) and Random Forest Regression suits best in reducing the redundancy (Abdel-Rahman et al., 2012). Higher estimation accuracy yielded from machine learning algorithms has been demonstrated from the comparative studies of crop N monitoring, albeit mainly for N concentration predictions (Berger et al., 2020). Thus, the comparison of machine learning algorithms on predicting N_{leaves} is expected to be further explored, especially given that N_{leaves} is more useful and suitable to describe the optimal N status for N management in the field (Baret et al., 2007).

The use of remote sensing in precision crop management dates back to the 1980's (Moran et al., 1997). Exploiting hyperspectral remote sensing for in-season variate N management is identified as “perhaps of greatest interest”, compared with other applications of hyperspectral remote sensing, in precision agriculture (Mulla, 2013). Instead of N_{leaves} , proxies, like Sufficiency Index and Response index, are mainly developed based on VIs to guide field management by characterising the degree of crop N deficiency (Holland and Schepers, 2010; Raun et al., 2002). As most of those methods are empirical and the impacts of changing environmental conditions on crop growth are more likely to be ignored, the optimised N management in the field might be largely different (Zhang et al., 2021). Long-term experimental data that reflects the interactions of soil-crop-weather continuum is anticipated to solve this, but this data is hard to obtain (Cammarano et al., 2021). Although there is progress in combining soil, weather and management data with remotely sensed VIs, the simulated yield response to N topdressing might fail due to the unreasonable prediction from trained machine learning models (Wang et al., 2021). Consequently, utilising remote sensing data solely is considered as insufficient and it is a necessary to incorporate complementary information that describes the crop system (Weiss et al., 2020).

1.2.3 Field management system and the integration of crop modelling and remote sensing

There are intrinsic quantitative relationships between crop, environment and management practices in crop model, while its missing spatial variation of crop growth status can be provided from the in-season remote sensing predictions. For N fertilisation recommendation, a decision-support tool based on the crop model and remote sensing methods has been developed to support crop management (Basso et al., 2001). A uniform management practice, thus, is applied in the relatively homogeneous area (Mulla, 2013). The delineation of the management zone is so far a critically uncertain step of this method (Jin et al., 2017). To delineate site-specific

management zones, several approaches have been proposed, for example based on remotely sensed spectral information (Basso et al., 2001; Zhang et al., 2010), soil properties (Fleming et al., 2004), multi-year's yield maps (Diker et al., 2004), or a combination of them processed by various clustering algorithms (Shaddad et al., 2015; Song et al., 2009). However, there is still a lack of an efficient method fitting all situations (Derby et al., 2007) and there is a discrepancy in the derived management zones between those derived from satellite remote sensing and those from soil maps (Jin et al., 2017). A comprehensive and adaptive system is expected to have a better use of the crop model simulations and remote sensing predictions to optimise N management in the field under the variable weather and soil conditions.

To fully utilise crop modelling and remote sensing, opportunities to improve crop model estimates have been highlighted by assimilating remote sensing data, for example in recent reviews that focused on providing an overview of crop model development, remote sensing technology, data assimilation algorithms, and source of uncertainty (Dlamini et al., 2023; Huang et al., 2019; Jin et al., 2018). Various crop models have been developed in recent decades, yet the increase in quality hardly matches the growth in number (Yin et al., 2021). Compared with the commonly used crop models WOFOST, DSSAT and APSIM (Jin et al., 2018), the crop model GECROS is hardly used for data assimilation, even though it was developed on a sound physiological basis. Model application in assimilating with in-season remote sensing observations needs to be extended, in which the discrepancies between model simulations and actual observations are expected to be reduced. While conducting data assimilation, most studies still used the satellite data like MODIS and Landsat, which is limited by the coarse spatial resolution and the susceptibility to weather. The mismatch between remote sensing data and agricultural landscape might be caused by the relatively coarse resolution in either space or time (Dlamini et al., 2023). As proximal sensing is labour-intensive and time-consuming over large areas, UAV-based remote sensing potentially represents a good compromise for determining the field variability on smallholder's farm and on cloudy days (Argento et al., 2021). The UAV-acquired remote sensing images are relatively low-cost, with high spatial resolution and user-preferred temporal resolution (Dlamini et al., 2023). The assimilation of high-resolution remote sensing data into crop model with robust mechanisms is expected to lead to a more detailed spatial characterisation of accurate crop growth and yield estimates (Kasampalis et al., 2018).

With existing challenges on data assimilation (Jin et al., 2018), the current assimilation of remote sensing observations and crop model simulations is mainly aimed at improving the

accuracy of in-season crop growth and end-of-season yield (Ines et al., 2013). Its further application in the field management optimisation should be investigated. To this end, a robust dynamic crop model linked with remote sensing data into a decision support system needs to be designed to support farmers to manage fertilisation in real-time. It is expected that such an integrated approach enables crop management to be made on a more accurate quantitative basis (Wolfert et al., 2017) and supports smart farming (Muangprathub et al., 2019; Nawandar and Satpute, 2019).

1.3 Research objectives

There has been an increasing gap between growing food demand and steady food production in China, and minimising this gap requires multiple efforts. In addition, to improve environmental sustainability and farmers' livelihood, not only higher yield but also higher N use efficiency and economic profit should be considered. Thus, the overall objective of this study is to design a system for smart N management in the wake of modernising agricultural production in China. The general hypothesis is that optimum N management can be determined to meet temporally and spatially varied crop needs yet to achieve the multi-objectives simultaneously in terms of productivity, economic profit and sustainability. To this end, a state-of-the-art generic, photosynthesis-driven crop model developed in Wageningen, GECROS (Yin and Struik, 2017; Yin and van Laar, 2005), will be used as an engine, and real-time remote sensing will be explored to complement the GECROS model in the system. To achieve this objective, experimental research was designed that follows:

- 1) Estimation of crucial photosynthetic parameters in the GECROS model for common field crops.
- 2) Assessment of the prediction of canopy and leaf N from UAV-acquired hyperspectral images using machine learning algorithms.
- 3) Integration of real-time remote sensing, dynamic crop model and data assimilation to develop a system of predicting crop growth and yield.
- 4) Use of the above system to investigate smart N management methods for maximising yield, profit or resource use efficiency.

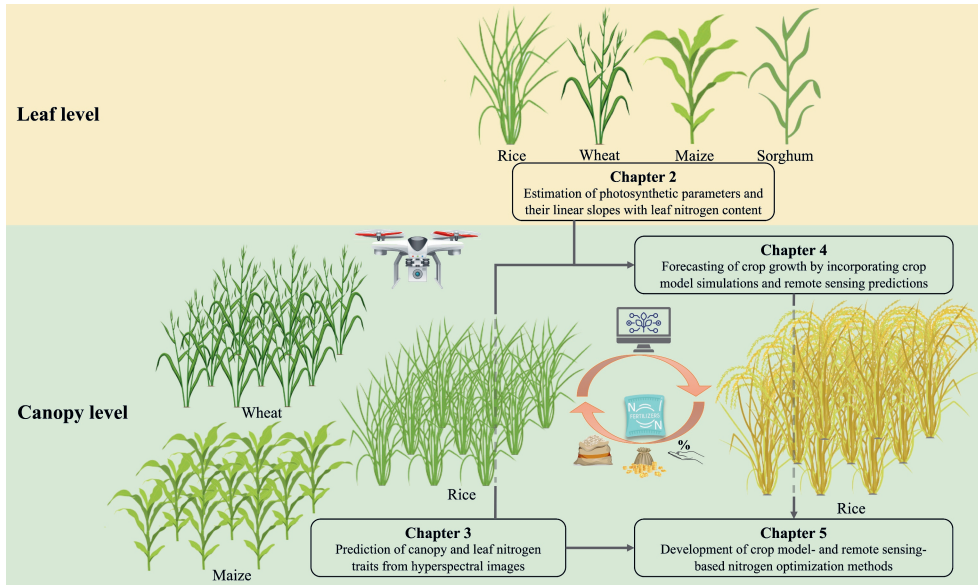


Fig. 1.2. Outline of the thesis.

1.4 Outline of the thesis

This thesis is composed of six chapters. This chapter (Chapter 1) forms the general introduction, in which the research objectives are identified based on a problem analysis and literature review. Four research chapters (Chapter 2 to Chapter 5) and a general discussion (Chapter 6) are arranged and structured as illustrated in Fig. 1.2.

Given that the crop model GECROS used in this study simulates crop growth as being photosynthesis driven, **Chapter 2** aims to establish the relationships between key photosynthetic parameters and SLN. Photosynthetic parameters, including CO_2 -diffusion parameters and photosynthetic capacity parameters, and their linear slopes with SLN are estimated in two C_3 crops, rice and wheat, and two C_4 crops, maize and sorghum, grown under three N levels. The applications of the tight associations between photosynthesis parameters and SLN in crop modelling and remote sensing are discussed in view of decision making on N management in the field in the context of smart farming.

In **Chapter 3**, canopy and leaf N traits are predicted from remote sensing images for the site- and time-specific monitor of crop N status. Different combinations of extracted features from the collected UAV-based hyperspectral images are applied to predict crop N traits for rice,

wheat and maize in field experiments in China with different N rates, using five machine-learning algorithms (or referred to as nonparametric regression algorithms). The importance of monitoring crop N status in time is highlighted, especially regarding linking it with crop models for better prediction of crop growth and well-grounded guidance of *in situ* N management in the field.

To effectively integrate remote sensing predictions and crop model simulations, a new data assimilation framework with quantified system uncertainties is developed in **Chapter 4**. Taking rice as a case study, remotely sensed in-season canopy leaf traits, including leaf weight, N_{leaves} and LAI, are directly updated in the crop model GECROS for the better simulation of crop growth and yield formation. The contribution of the developed method is well established, with respect to uncertainty quantification in data assimilation community and accurate forecasting of crop carbon and N status for smart farming.

From the above studies, smart N management based on crop modelling and remote sensing is expected to be provided for sustainable agricultural production. Two new N optimisation methods – the crop model method using historical weather data, and an integrated remote sensing-crop model method by data assimilation – are proposed in **Chapter 5** and compared with three reference methods: farmer practice optimised by the yield response curve, and the Sufficiency Index- or Response Index-based remote sensing methods. The potential to simultaneously improve the productivity, profit and N use efficiency of those two new proposed methods will be investigated in our field experiments of rice.

Chapter 6 summarises the main findings of this thesis, discussing the essential aspects of the utilisation of crop modelling and remote sensing in smart N management. In this chapter, the contributed innovation and potential application of the designed real-time smart N management system are exhibited. The prospective for the improvement of the designed system is emphasised in view of smart farming for the achievement of sustainable agricultural production.

References

- Abdel-Rahman, E. M., Ahmed, F. B., & Ismail, R. (2012). Random forest regression and spectral band selection for estimating sugarcane leaf nitrogen concentration using EO-1 Hyperion hyperspectral data. *International Journal of Remote Sensing*, 34, 712-728. doi:10.1080/01431161.2012.713142
- Argento, F., Anken, T., Abt, F., Vogelsanger, E., Walter, A., & Liebisch, F. (2021). Site-specific nitrogen management in winter wheat supported by low-altitude remote sensing and soil data. *Precision Agriculture*, 22, 364-386. doi:10.1007/s11119-020-09733-3
- Baret, F., Houles, V., & Guerif, M. (2007). Quantification of plant stress using remote sensing observations and crop models: the case of nitrogen management. *Journal of Experimental Botany*, 58, 869-880. doi:10.1093/jxb/erl231
- Basso, B., & Antle, J. (2020). Digital agriculture to design sustainable agricultural systems. *Nature Sustainability*, 3, 254-256. doi:10.1038/s41893-020-0510-0
- Basso, B., Ritchie, J. T., Pierce, F. J., Braga, R. P., & Jones, J. W. (2001). Spatial validation of crop models for precision agriculture. *Agricultural Systems*, 68, 97-112.
- Bausch, W. C., & Duke, H. R. (1996). Remote sensing of plant nitrogen status in corn. *Transactions of the ASAE*, 39, 1869-1875.
- Berger, K., Verrelst, J., Féret, J.-B., Wang, Z., Woche, M., Strathmann, M., Danner, M., Mauser, W., & Hank, T. (2020). Crop nitrogen monitoring: Recent progress and principal developments in the context of imaging spectroscopy missions. *Remote Sensing of Environment*, 242, 111758. doi:10.1016/j.rse.2020.111758
- Boote, K. J., Jones, J. W., White, J. W., Asseng, S., & Lizaso, J. I. (2013). Putting mechanisms into crop production models. *Plant, Cell and Environment*, 36, 1658-1672. doi:10.1111/pce.12119
- Bousquet, L., Lachérade, S., Jacquemoud, S., & Moya, I. (2005). Leaf BRDF measurements and model for specular and diffuse components differentiation. *Remote Sensing of Environment*, 98, 201-211. doi:10.1016/j.rse.2005.07.005
- Brorsen, B. W., & Richter, F. G. C. (2012). Experimental designs for estimating plateau-type production functions and economically optimal input levels. *Journal of Productivity Analysis*, 38, 45-52.
- Cammarano, D., Basso, B., Holland, J., Gianinetti, A., Baronchelli, M., & Ronga, D. (2021). Modeling spatial and temporal optimal N fertilizer rates to reduce nitrate leaching while improving grain yield and quality in malting barley. *Computers and Electronics in Agriculture*, 182. doi:10.1016/j.compag.2021.105997
- Cassman, K. G., Dobermann, A., & Walters, D. T. (2002). Agroecosystems, nitrogen-use efficiency, and nitrogen management. *AMBIO: A Journal of the Human Environment*, 31, 132-140.
- Chen, P., Haboudane, D., Tremblay, N., Wang, J., Vigneault, P., & Li, B. (2010). New spectral indicator assessing the efficiency of crop nitrogen treatment in corn and wheat. *Remote Sensing of Environment*, 114, 1987-1997. doi:10.1016/j.rse.2010.04.006
- Chlingaryan, A., Sukkarieh, S., & Whelan, B. (2018). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *Computers and Electronics in Agriculture*, 151, 61-69. doi:10.1016/j.compag.2018.05.012

- Clevers, J. G. P. W., & Gitelson, A. A. (2013). Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on Sentinel-2 and -3. *International Journal of Applied Earth Observation and Geoinformation*, 23, 344-351. doi:10.1016/j.jag.2012.10.008
- Cui, Z., Chen, X., & Zhang, F. (2013). Development of regional nitrogen rate guidelines for intensive cropping systems in China. *Agronomy Journal*, 105, 1411-1416. doi:10.2134/agronj2012.0398
- Cui, Z., Zhang, H., Chen, X., Zhang, C., Ma, W., Huang, C., Zhang, W., Mi, G., Miao, Y., Li, X., Gao, Q., Yang, J., Wang, Z., Ye, Y., Guo, S., Lu, J., Huang, J., Lv, S., Sun, Y., Liu, Y., Peng, X., Ren, J., Li, S., Deng, X., Shi, X., Zhang, Q., Yang, Z., Tang, L., Wei, C., Jia, L., Zhang, J., He, M., Tong, Y., Tang, Q., Zhong, X., Liu, Z., Cao, N., Kou, C., Ying, H., Yin, Y., Jiao, X., Zhang, Q., Fan, M., Jiang, R., Zhang, F., & Dou, Z. (2018). Pursuing sustainable productivity with millions of smallholder farmers. *Nature*, 555, 363-366. doi:10.1038/nature25785
- Derby, N. E., Casey, F. X. M., & Franzen, D. W. (2007). Comparison of nitrogen management zone delineation methods for corn grain yield. *Agronomy Journal*, 99, 405-414. doi:10.2134/agronj2006.0027
- Diaz, R. J., & Rosenberg, R. (2008). Spreading dead zones and consequences for marine ecosystems. *Science*, 321, 926-929. doi:10.1126/science.1156401
- Diker, K., Heermann, D. F., & Brodahl, M. K. (2004). Frequency analysis of yield for delineating yield response zones. *Precision Agriculture*, 5, 435-444.
- Dlamini, L., Crespo, O., van Dam, J., & Kooistra, L. (2023). A global systematic review of improving crop model estimations by assimilating remote sensing data: implications for small-scale agricultural systems. *Remote Sensing*, 15, 4066. doi:10.3390/rs15164066
- El-Shikha, D. M., Waller, P., Hunsaker, D., Clarke, T., & Barnes, E. (2007). Ground-based remote sensing for assessing water and nitrogen status of broccoli. *Agricultural Water Management*, 92, 183-193. doi:10.1016/j.agwat.2007.05.020
- Evans, J. R. (1983). Nitrogen and photosynthesis in the flag leaf of wheat (*Triticum aestivum* L.). *Plant Physiology*, 72, 297-302.
- Evans, J. R., & Clarke, V. C. (2019). The nitrogen cost of photosynthesis. *Journal of Experimental Botany*, 70, 7-15. doi:10.1093/jxb/ery366
- Evans, L. T. (1997). Adapting and improving crops: the endless task. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 352, 901-906.
- FAO. (2022). *World Food and Agriculture – Statistical Yearbook 2022*. Rome, Italy. doi:10.4060/cc2211en
- Fleming, K. L., Heermann, D. F., & Westfall, D. G. (2004). Evaluating soil color with farmer input and apparent soil electrical conductivity for management zone delineation. *Agronomy Journal*, 96, 1581-1587.
- Foley, J. A., Ramankutty, N., Brauman, K. A., Cassidy, E. S., Gerber, J. S., Johnston, M., Mueller, N. D., O'Connell, C., Ray, D. K., West, P. C., Balzer, C., Bennett, E. M., Carpenter, S. R., Hill, J., Monfreda, C., Polasky, S., Rockstrom, J., Sheehan, J., Siebert, S., Tilman, D., & Zaks, D. P. (2011). Solutions for a cultivated planet. *Nature*, 478, 337-342. doi:10.1038/nature10452

- Fourty, T., & Baret, F. (1997). Vegetation water and dry matter contents estimated from top-of-the-atmosphere reflectance data: A simulation study. *Remote Sensing of Environment*, 61, 34-45.
- Gao, C., Sun, B., & Zhang, T. (2006). Sustainable nutrient management in Chinese agriculture: Challenges and perspective. *Pedosphere*, 16, 253-263.
- Gebbers, R., & Adamchuk, V. I. (2010). Precision agriculture and food security. *Science*, 327, 828-831.
- Goulding, K., Jarvis, S., & Whitmore, A. (2008). Optimizing nutrient management for farm systems. *Philosophical Transaction of the Royal Society B*, 363, 667-680. doi:10.1098/rstb.2007.2177
- Holland, K. H., & Schepers, J. S. (2010). Derivation of a variable rate nitrogen application model for in-season fertilization of corn. *Agronomy Journal*, 102, 1415-1424. doi:10.2134/agronj2010.0015
- Huang, J., Gómez-Dans, J. L., Huang, H., Ma, H., Wu, Q., Lewis, P. E., Liang, S., Chen, Z., Xue, J.-H., Wu, Y., Zhao, F., Wang, J., & Xie, X. (2019). Assimilation of remote sensing into crop growth models: Current status and perspectives. *Agricultural and Forest Meteorology*, 276-277, 107609. doi:10.1016/j.agrformet.2019.06.008
- Huang, J. X., Sedano, F., Huang, Y. B., Ma, H. Y., Li, X. L., Liang, S. L., Tian, L. Y., Zhang, X. D., Fan, J. L., & Wu, W. B. (2016). Assimilating a synthetic Kalman filter leaf area index series into the WOFOST model to improve regional winter wheat yield estimation. *Agricultural and Forest Meteorology*, 216, 188-202. doi:10.1016/j.agrformet.2015.10.013
- Ines, A. V. M., Das, N. N., Hansen, J. W., & Njoku, E. G. (2013). Assimilation of remotely sensed soil moisture and vegetation with a crop simulation model for maize yield prediction. *Remote Sensing of Environment*, 138, 149-164. doi:10.1016/j.rse.2013.07.018
- IPCC. (2021). *Climate Change 2021: The Physical Science Basis*. Cambridge, The United Kingdom and New York, The United States: Cambridge University Press.
- Jacquemoud, S., & Baret, F. (1990). PROSPECT: A model of leaf optical properties spectra. *Remote Sensing of Environment*, 34, 75-91.
- Jacquemoud, S., & Ustin, S. L. (2001). *Leaf optical properties: A state of the art*. Paper presented at the 8th International Symposium of Physical Measurements & Signatures in Remote Sensing, CNES, Aussois, France.
- Jin, X., Kumar, L., Li, Z., Feng, H., Xu, X., Yang, G., & Wang, J. (2018). A review of data assimilation of remote sensing and crop models. *European Journal of Agronomy*, 92, 141-152. doi:10.1016/j.eja.2017.11.002
- Jin, Z., Prasad, R., Shriver, J., & Zhuang, Q. (2017). Crop model- and satellite imagery-based recommendation tool for variable rate N fertilizer application for the US Corn system. *Precision Agriculture*, 18, 779-800. doi:10.1007/s11119-016-9488-z
- Kasampalis, D., Alexandridis, T., Deva, C., Challinor, A., Moshou, D., & Zalidis, G. (2018). Contribution of Remote Sensing on Crop Models: A Review. *Journal of Imaging*, 4, 52. doi:10.3390/jimaging4040052
- Kaul, M., Hill, R. L., & Walthall, C. (2005). Artificial neural networks for corn and soybean yield prediction. *Agricultural Systems*, 85, 1-18. doi:10.1016/j.agsy.2004.07.009

- Kuusk, A. (1995). A fast, invertible canopy reflectance model. *Remote Sensing of Environment*, 51, 342-350.
- Lawlor, D. W. (1995). Photosynthesis, productivity and environment. *Journal of Experimental Botany*, 46, 1449-1461.
- Lelieveld, J., Evans, J. S., Fnais, M., Giannadaki, D., & Pozzer, A. (2015). The contribution of outdoor air pollution sources to premature mortality on a global scale. *Nature*, 525, 367-371. doi:10.1038/nature15371
- Lipper, L., Thornton, P., Campbell, B. M., Baedeker, T., Braimoh, A., Bwalya, M., Caron, P., Cattaneo, A., Garrity, D., Henry, K., Hottle, R., Jackson, L., Jarvis, A., Kossam, F., Mann, W., McCarthy, N., Meybeck, A., Neufeldt, H., Remington, T., Sen, P. T., Sessa, R., Shula, R., Tibu, A., & Torquebiau, E. F. (2014). Climate-smart agriculture for food security. *Nature Climate Change*, 4, 1068-1072. doi:10.1038/nclimate2437
- Loomis, R. S., Rabbinge, R., & Ng, E. (1979). Explanatory models in crop physiology. *Annual Review of Plant Physiology*, 30, 339-367.
- Lu, Y., Jenkins, A., Ferrier, R. C., Bailey, M., Gordon, I. J., Song, S., Huang, J., Jia, S., Zhang, F., Liu, X., Feng, Z., & Zhang, Z. (2015). Addressing China's grand challenge of achieving food security while ensuring environmental sustainability. *Science Advances*, 1, e1400039. doi:10.1126/sciadv.1400039
- Marin, F., Jones, J. W., & Boote, K. J. (2017). A stochastic method for crop models: including uncertainty in a sugarcane model. *Agronomy Journal*, 109, 483-495. doi:10.2134/agronj2016.02.0103
- Miao, Y., Stewart, B. A., & Zhang, F. (2011). Long-term experiments for sustainable nutrient management in China. A review. *Agronomy for Sustainable Development*, 31, 397-414. doi:10.1051/agro/2010034
- Moran, M. S., Inoue, Y., & Barnes, E. M. (1997). Opportunities and limitations for image-based remote sensing in precision crop management. *Remote Sensing of Environment*, 61, 319-346.
- Moulin, S. (1999). Impacts of model parameter uncertainties on crop reflectance estimates: A regional case study on wheat. *International Journal of Remote Sensing*, 20, 213-218. doi:10.1080/014311699213730
- Muangprathub, J., Boonnam, N., Kajornkasirat, S., Lekbangpong, N., Wanichsombat, A., & Nillaor, P. (2019). IoT and agriculture data analysis for smart farm. *Computers and Electronics in Agriculture*, 156, 467-474. doi:10.1016/j.compag.2018.12.011
- Mulla, D. J. (2013). Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosystems Engineering*, 114, 358-371. doi:10.1016/j.biosystemseng.2012.08.009
- Nawandar, N. K., & Satpute, V. R. (2019). IoT based low cost and intelligent module for smart irrigation system. *Computers and Electronics in Agriculture*, 162, 979-990. doi:10.1016/j.compag.2019.05.027
- Norse, D., & Ju, X. (2015). Environmental costs of China's food security. *Agriculture, Ecosystems & Environment*, 209, 5-14. doi:10.1016/j.agee.2015.02.014

- Penning de Vries, F. W. T., Jansen, D. M., ten Berge, H. F. M., & Bakema, A. (1989). *Simulation of ecophysiological processes of growth in several annual crops*. Wageningen, The Netherlands: Pudoc and Los Baños, Philippines: International Rice Research Institute.
- Pierce, F. J., & Nowak, P. (1999). Aspects of precision agriculture. *Advances in Agronomy*, 67, 1-85.
- Raun, W. R., Solie, J. B., Johnson, G. V., Stone, M. L., Mullen, R. W., Freeman, K. W., Thomason, W. E., & Lukina, E. V. (2002). Improving nitrogen use efficiency in cereal grain production with optical sensing and variable rate application. *Agronomy Journal*, 94, 815-820.
- Ray, D. K., Ramankutty, N., Mueller, N. D., West, P. C., & Foley, J. A. (2012). Recent patterns of crop yield growth and stagnation. *Nature Communications*, 3, 1293. doi:10.1038/ncomms2296
- Redman, J. C., & Allen, S. Q. (1954). Some interrelationships of economic and agronomic concepts. *Journal of Farm Economics*, 36, 453-465.
- Scharf, P. C. (2015). Managing nitrogen. In P. C. Scharf (Eds.), *Managing nitrogen in crop production* (pp. 25-76). Madison, WI, USA: American Society of Agronomy, Inc., Crop Science Society of America, Inc., and Soil Science Society of America, Inc.
- Shaddad, S. M., Madrau, S., Castrignanò, A., & Mouazen, A. M. (2015). Data fusion techniques for delineation of site-specific management zones in a field in UK. *Precision Agriculture*, 17, 200-217. doi:10.1007/s11119-015-9417-6
- Shao, Y., Zhao, C., Bao, Y., & He, Y. (2009). Quantification of nitrogen status in rice by least squares support vector machines and reflectance spectroscopy. *Food and Bioprocess Technology*, 5, 100-107. doi:10.1007/s11947-009-0267-y
- Sims, D. A., & Gamon, J. A. (2002). Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sensing of Environment*, 81, 337-354.
- Sinclair, T. R., & De Wit, C. T. (1976). Analysis of the Carbon and Nitrogen Limitations to Soybean Yield. *Agronomy Journal*, 68, 319-324.
- Sinclair, T. R., & Horie, T. (1989). Leaf nitrogen, photosynthesis, and crop radiation use efficiency: A review. *Crop Science*, 29, 90-98.
- Sinclair, T. R., & Jamieson, P. D. (2006). Grain number, wheat yield, and bottling beer: An analysis. *Field Crops Research*, 98, 60-67. doi:10.1016/j.fcr.2005.12.006
- Song, X., Wang, J., Huang, W., Liu, L., Yan, G., & Pu, R. (2009). The delineation of agricultural management zones with high resolution remotely sensed data. *Precision Agriculture*, 10, 471-487. doi:10.1007/s11119-009-9108-2
- Tembo, G., Brorsen, B. W., Epplin, F. M., & Tostão, E. (2008). Crop input response functions with stochastic plateaus. *American Journal of Agricultural Economics*, 90, 424-434.
- Tilman, D., Balzer, C., Hill, J., & Befort, B. L. (2011). Global food demand and the sustainable intensification of agriculture. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 20260-20264. doi:10.1073/pnas.1116437108
- Verhoef, W. (1984). Light scattering by leaf layers with application to canopy reflectance modeling: The SAIL model. *Remote Sensing of Environment*, 16, 125-141.

- Viña, A., Gitelson, A. A., Nguy-Robertson, A. L., & Peng, Y. (2011). Comparison of different vegetation indices for the remote assessment of green leaf area index of crops. *Remote Sensing of Environment*, 115, 3468-3478. doi:10.1016/j.rse.2011.08.010
- Walter, A., Finger, R., Huber, R., & Buchmann, N. (2017). Smart farming is key to developing sustainable agriculture. *Proceedings of the National Academy of Sciences of the United States of America*, 114, 6148-6150. doi:10.1073/pnas.1707462114
- Wang, X., Miao, Y., Dong, R., Zha, H., Xia, T., Chen, Z., Kusnierek, K., Mi, G., Sun, H., & Li, M. (2021). Machine learning-based in-season nitrogen status diagnosis and side-dress nitrogen recommendation for corn. *European Journal of Agronomy*, 123, 126193. doi:10.1016/j.eja.2020.126193
- Wang, Y., & Lu, Y. (2020). Evaluating the potential health and economic effects of nitrogen fertilizer application in grain production systems of China. *Journal of Cleaner Production*, 264, 121635. doi:10.1016/j.jclepro.2020.121635
- Wang, Y., & Zhao, G. (2019). Life cycle assessment of potential pollutant-induced human capital loss caused by different agricultural production systems in Beijing, China. *Journal of Cleaner Production*, 240, 118141. doi:10.1016/j.jclepro.2019.118141
- Weiss, M., Jacob, F., & Duveiller, G. (2020). Remote sensing for agricultural applications: A meta-review. *Remote Sensing of Environment*, 236, 111402. doi:10.1016/j.rse.2019.111402
- Whisler, F. D., Acock, B., Baker, D. N., Fye, R. E., Hodges, H. F., Lambert, J. R., Lemmon, H. E., McKinion, J. M., & Reddy, V. R. (1986). Crop simulation models in agronomic systems. *Advances in Agronomy*, 40, 141-208.
- Wolfert, S., Ge, L., Verdouw, C., & Bogaardt, M. J. (2017). Big data in smart farming – A review. *Agricultural Systems*, 153, 69-80. doi:10.1016/j.agsy.2017.01.023
- Wu, W., & Ma, B. (2015). Integrated nutrient management (INM) for sustaining crop productivity and reducing environmental impact: A review. *Science of the Total Environment*, 512-513, 415-427. doi:10.1016/j.scitotenv.2014.12.101
- Yin, X., Schapendonk, A. H. C. M., Kropff, M. J., van Oijen, M., & Bindraban, P. S. (2000). A generic equation for nitrogen-limited leaf area index and its application in crop growth models for predicting leaf senescence. *Annals of Botany*, 85, 579-585.
- Yin, X., & Struik, P. C. (2017). Can increased leaf photosynthesis be converted into higher crop mass production? A simulation study for rice using the crop model GECROS. *Journal of Experimental Botany*, 68, 2345-2360. doi:10.1093/jxb/erx085
- Yin, X., Struik, P. C., & Goudriaan, J. (2021). On the needs for combining physiological principles and mathematics to improve crop models. *Field Crops Research*, 271, 108254. doi:10.1016/j.fcr.2021.108254
- Yin, X., & van Laar, H. H. (2005). *Crop systems dynamics: an ecophysiological simulation model for genotype-by-environment interactions*. Wageningen, The Netherlands: Wageningen Academic Publishers.

- Zhang, J., Wang, W., Krienke, B., Cao, Q., Zhu, Y., Cao, W., & Liu, X. (2021). In-season variable rate nitrogen recommendation for wheat precision production supported by fixed-wing UAV imagery. *Precision Agriculture*, 23, 830-853. doi:10.1007/s11119-021-09863-2
- Zhang, X., Davidson, E. A., Mauzerall, D. L., Searchinger, T. D., Dumas, P., & Shen, Y. (2015). Managing nitrogen for sustainable development. *Nature*, 528, 51-59. doi:10.1038/nature15743
- Zhang, X., Shi, L., Jia, X., Seielstad, G., & Helgason, C. (2010). Zone mapping application for precision-farming: A decision support tool for variable rate application. *Precision Agriculture*, 11, 103-114. doi:10.1007/s11119-009-9130-4

Chapter 2

Estimating photosynthetic parameter values of rice, wheat, maize and sorghum to enable smart crop cultivation

Dong Wang^{1,2}, Winda Rianti^{1,3}, Fabián Gálvez^{1,4}, Peter E. L. van der Putten¹,
Paul C. Struik¹ and Xinyou Yin¹

¹ Centre for Crop Systems Analysis, Department of Plant Sciences, Wageningen University & Research, 6700 AK Wageningen, The Netherlands

² Shanghai Lankuaikei Technology Development Co. Ltd., No. 888 Huanhu West 2nd Road, Pudong New District, Shanghai, China

³ Faculty of Agriculture, Universitas Singaperbangsa Karawang, Karawang, Jawa Barat 41361, Indonesia

⁴ Faculty of Life Sciences, Polytechnic Superior School of Litoral (ESPOL), Campus Gustavo Galindo Velasco, Km 30.5 Perimetral Road, Guayaquil, Ecuador

Abstract

Crop models can support the design of smart crop management practices. The Farquhar-von Caemmerer-Berry (FvCB) model is increasingly being used in these models for quantifying leaf photosynthesis. Nitrogen (N) is required for many functional machineries of photosynthesis, thus relationships between FvCB-model parameters and specific leaf N content (SLN) should be established. We conducted combined gas exchange and chlorophyll fluorescence measurements on fully expanded leaves of two C₃ crops, rice (*Oryza sativa*) and wheat (*Triticum aestivum*), and two C₄ crops, maize (*Zea mays*) and sorghum (*Sorghum bicolor*), grown under three N levels. Photosynthetic parameters were estimated and linear relationships between these parameters and SLN were quantified in both C₃ and C₄ crop types. The efficiency of converting incident light into linear electron transport for C₃ crops or into ATP production for C₄ crops showed a weak increase with SLN. The maximum electron transport rate (J_{\max}) for C₃ crops or the maximum ATP production rate ($J_{\max,atp}$) for C₄ crops significantly increased with SLN. The increase in Rubisco carboxylation capacity (V_{\max}) with SLN was significantly higher in C₃ than in C₄ crops. Triose phosphate utilisation for C₃ crops and PEP carboxylation capacity ($V_{p\max}$) for C₄ crops increased significantly with SLN as well. Except for J_{\max} at 21% O₂ and V_{\max} of C₃ crops, there was no significant difference among crops in the relationship between estimated photosynthetic parameters and SLN. The tight associations of photosynthesis parameters with SLN were discussed in view of decision making on N management in the context of smart farming.

Keywords: C₃, C₄, crop modelling, decision support, leaf nitrogen content, photosynthesis model, photosynthesis parameters

2.1 Introduction

Breeding semi-dwarf crop cultivars in the 1960s that could benefit from an increase in inputs of fertiliser, irrigation water and pesticides has resulted in the first Green Revolution (Bailey-Serres et al., 2019; Pingali, 2012). Achieving further food security to meet demands for ever growing populations in a sustainable way requires high yields and efficient resource use, a great challenge for modern agriculture. Smart farming is key to fulfilling this challenge in sustainable agriculture (Walter et al., 2017). Smart farming entails smart crop cultivation tailored to specific crops grown under specific conditions.

Smart crop cultivation requires in-season adaptive management, based on crop yield forecasting that is conditional on the status of crop, water and nutrients in the cropping systems. Together with the development of unmanned aerial vehicles, the exploitation of remote and proximal sensing apparatus has become a key to sophisticated *in situ* smart crop cultivation in open fields (Walter et al., 2017). For instance, site-specific crop growth and nitrogen (N) status have been retrieved from multi- or hyper-spectral images recently (Hank et al., 2019). However, there is a huge phenotypic gap between retrieved-N and crop yield. Crop models have long been developed based on crop physiological principles and used to predict crop growth and yield in response to various environmental variables (including N supply), and thus may serve as a tool to bridge the gap. As photosynthesis is the primary physiological process in crop production (Lawlor, 1995) and is very sensitive to environmental conditions and crop management, an accurate and robust modelling of photosynthesis or related parameters is a prerequisite for crop models in predicting crop growth and yield. Ambient weather conditions, including solar radiation, air temperature, humidity and CO₂ concentration, have a direct impact on crop photosynthesis. Agronomic management practices, such as fertilisation can also affect the crop's photosynthetic performance, because they affect canopy development and the photosynthetic proteins' contents, and thus light capture and gas exchange (McDowell, 2011). Similarly, rainfall and irrigation determine soil water balance, thereby indirectly affecting crop photosynthesis. Crop protection also plays a significant role in maintaining crop photosynthesis and growth.

Based on their approach of simulating the process of photosynthesis, crop models can be classified into three categories (Li et al., 2015): models based on the canopy radiation use efficiency (RUE) (Monteith, 1977), models utilising the light-response curve of single-leaf photosynthesis and integrating it to the canopy level (de Wit, 1978), and models describing the

photosynthetic process via the biochemical model of Farquhar et al. (1980) (“the FvCB model”, hereafter) and upscaling it from leaf to canopy. The RUE concept is used in many crop models (Boote et al., 1996) and works fairly well under many conditions (Sinclair and Rawlins, 1993). However, RUE is a composite parameter that depends not only on species-related biosynthesis products and photosynthesis type (Kiniry et al., 1989), but also on leaf physiological status (e.g., N content) and environmental factors (e.g., temperature, radiation, and water stress intensity) affecting leaf photosynthetic rate (Sinclair and Horie, 1989; Sinclair and Muchow, 1999). Without the specific prediction of photosynthetic processes describing an effective RUE, models based on RUE are hardly convincing under nonoptimal conditions (Boote et al., 1996), because these crop models lack appropriate algorithms to describe the effects of complex interactions between multiple environmental and physiological factors on RUE (Yin et al., 2021b). As for the simple light response models, although the photosynthesis can be upscaled from leaf to canopy, the interaction effects of multiple factors on leaf-photosynthesis parameters have not been modelled adequately (Yin et al., 2021b). The FvCB model is based on the understanding of the major biochemical component processes of photosynthesis. Its further developments and extensions have been reviewed recently by Yin et al. (2021a). Because of its mechanistic level and simplicity, the FvCB model has been widely used not only for analysing leaf biochemistry and quantifying leaf photosynthesis responses under different environmental conditions (Bernacchi et al., 2003; Bernacchi et al., 2001), but also for upscaling to project productivities of canopies and ecosystems and to model global carbon fluxes responding to environmental changes (Rogers, 2014).

The FvCB model predicts the net leaf photosynthetic rate (A) of C_3 species as the minimum of the Rubisco-limited (A_c) and electron (e^-) transport-limited (A_j) rate of photosynthesis (Farquhar et al., 1980). The model for both A_c and A_j accounts for the CO_2 released through photorespiration, in which oxygenation of ribulose biphosphate is innately linked to its carboxylation by Rubisco and potentially reduces the rate of photosynthesis by over 20% (Ehleringer et al., 1991). Later, it was modified by adding a third limitation set by triose phosphate utilisation (TPU), in which A may be ceiled by the rate as determined by the export of triose phosphate (A_p) (Herold, 1980). Maximum rates of these three limited processes are determined by Rubisco carboxylation capacity (V_{cmax}), e^- transport capacity (J_{max}), and rate of TPU (T_p), respectively.

Compared with C_3 photosynthesis, C_4 photosynthesis is more complicated because of its operation of the additional CO_2 concentrating mechanism (CCM) cycle, also called C_4 cycle,

prior to the C_3 cycle as occurring in C_3 photosynthesis. In C_4 photosynthesis, CO_2 is firstly converted into bicarbonate which is fixed by phosphoenolpyruvate (PEP) carboxylase (PEPc) into C_4 -acids in thin-walled mesophyll cells. Then these C_4 -acids are transported to thick-walled bundle sheath cells, in which C_4 -acids are decarboxylated. The released CO_2 is re-fixed via Rubisco in the C_3 cycle. As PEPc has a high affinity to bicarbonate and a likely higher maximal velocity than Rubisco, the C_4 cycle runs faster than the C_3 cycle. This results in CO_2 concentrations in bundle sheath cells that are 10- to 20-fold higher than in the mesophyll cells. Therefore, photorespiration is suppressed in C_4 photosynthesis and this improves the carboxylation efficiency of Rubisco (Osmond et al., 1982). An FvCB-type C_4 photosynthesis model was developed (von Caemmerer and Furbank, 1999), assuming that reactions of both C_4 and C_3 cycles are limited by either enzyme activity or e^- transport. Yin et al. (2011b) revised this C_4 model, considering that either cycle can be limited by enzyme activity or by e^- transport, and thus there are four possible limitations in the model for C_4 photosynthesis.

Gas exchange measurements of photosynthetic CO_2 -response curves have been widely used to estimate parameters of the FvCB model (Dubois et al., 2007; Sharkey et al., 2007). Combined gas exchange and chlorophyll fluorescence measurements have also been used to estimate additional photosynthetic parameters, such as mesophyll conductance (g_m) (Evans and von Caemmerer, 1996), utilising the fact that photosystem II (PSII) e^- transport efficiency could be assessed by chlorophyll fluorescence measurement (Genty et al., 1989). Using combined gas exchange and chlorophyll fluorescence measurements of both CO_2 - and light-response curves, Yin et al. (2009) and Yin et al. (2011b) described methods to estimate photosynthetic parameters for C_3 and C_4 photosynthesis, respectively. There are many reports on using these methods to estimate photosynthetic parameters within an individual species. However, estimating these photosynthetic parameters simultaneously for major C_3 and C_4 crops to better support yield forecasting of these crops is rare.

Key photosynthetic parameters such as V_{cmax} and J_{max} are known to correlate with specific leaf N content (SLN) (Evans and Clarke, 2019; Harley et al., 1992b; Niinemets and Tenhunen, 1997). This is because N is a key-element in various photosynthetic machineries, such as the light harvesting complexes, the e^- transport systems, the CO_2 fixation enzymes and other enzymes (Laisk et al., 2002). Within chloroplasts, 84% of N is associated with photosynthesis-related proteins and molecules and 75% of N in mature leaves is present in chloroplasts (Evans and Clarke, 2019).

In this study, photosynthetic parameters of the FvCB models will be estimated by combined gas exchange and chlorophyll fluorescence measurements for four major crops (C₃: rice and wheat, C₄: maize and sorghum) under three different N levels. Our objectives are: 1) to quantify photosynthetic parameters for these C₃ and C₄ crop species; 2) to quantify the relationships of these parameters with SLN; and 3) to assess if these relationships differ between different crop species. These relationships are discussed in the context of crop modelling in order to better support N management in smart production of major crops.

2.2 Materials and methods

2.2.1 Plant materials and growth conditions

The experiment was conducted as a split-plot design (crop species as the whole-plot factor, N treatment as the split-plot factor), with four replicates. The selected C₃ crops were rice (cv. IR64) and wheat (cv. Paragon), and the C₄ crops were maize (cv. P8057) and sorghum (cv. CSM63E). Seeds were sown in 7-L pots with 5.2 L soil. Soil was prepared by mixing sandy soil and perlite substrate in equal volumes for wheat, maize and sorghum while for rice, perlite substrate was substituted by quartz sand. Plants were grown under well-watered conditions in a greenhouse of UNIFARM, Wageningen University & Research, the Netherlands. Natural light was automatically supplemented by an artificial light source (600 W HPS Hortilux Schröder Lamps, Monster, Netherlands), which was switched on when global solar radiation outside the greenhouse dropped below 400 W m⁻² and switched off when it exceeded 500 W m⁻². The photoperiod was 12 h d⁻¹. The average temperature was 25-27 °C during the day and 22-23 °C during the night, and the relative humidity was 60-80%. Nitrogen was supplied weekly through a nutrient solution and three N levels (low N: 50 mg N; middle N: 300 mg N; and high N: 1000 mg N, per pot) were applied resulting in three different SLN levels. Seeds were sown for four consecutive weeks to create four replicates of plants so that the workload for measurements (see below) could be spread over weeks; one leaf was tagged per replicate for conducting measurements.

2.2.2 Gas exchange and chlorophyll fluorescence measurements

Measurements were conducted on selected leaves (the 6-8th for rice, the 6-7th for wheat, the 5th for maize and the 7-8th for sorghum, counted from below), on average 9 days after full expansion of the leaves. Simultaneous gas exchange and chlorophyll fluorescence

measurements were conducted over the same leaf area, using an open gas exchange system Li-Cor 6400XT and an integrated fluorescence chamber head (Li-Cor, Lincoln, Nebraska, USA). During the measurements, the leaf temperature was set at 25 °C and the leaf-to-air vapour pressure difference was controlled within a range of 0.6 to 1.9 kPa.

First, incident irradiance (I_{inc}) response curves of net photosynthetic rate (A) at the 21% O₂ level were assessed. Photon flux density (at a red:blue ratio of 90%:10%) was decreased stepwise: 1750, 1000, 500, 200, 150, 100, 70, and 45 $\mu\text{mol m}^{-2} \text{s}^{-1}$ while keeping ambient CO₂ (C_a) at 400 $\mu\text{mol mol}^{-1}$. To obtain a CO₂ response curve of A at 21% O₂, C_a was first decreased stepwise: 400, 250, 150, 100, and 65 $\mu\text{mol mol}^{-1}$ and then increased stepwise: 400, 600, 1000, and 1500 $\mu\text{mol mol}^{-1}$ while keeping I_{inc} at a saturating level, 1750 $\mu\text{mol m}^{-2} \text{s}^{-1}$, evidenced by the above observed I_{inc} response curves at the 21% O₂ level. In order to convert chlorophyll fluorescence-based apparent PSII photochemical efficiency into linear e⁻ transport for C₃ crops and into ATP production rate for C₄ crops (see below), an extra I_{inc} response curve was measured, in which photon flux densities were in a decreasing series: 350, 200, 150, 100, 70, and 45 $\mu\text{mol m}^{-2} \text{s}^{-1}$ while keeping C_a at 1000 $\mu\text{mol mol}^{-1}$ combined with 2% O₂ to ensure a non-photorespiratory condition. All gas exchange data were measured at a constant flow rate (300 $\mu\text{mol s}^{-1}$).

At each step in I_{inc} or C_a , the steady-state fluorescence yield (F_s) was recorded once A reached the steady state (taking 5 minutes on average to reach). Afterwards, the multiphase flash was applied to estimate the maximum fluorescence yield (F_m') (Loriaux et al., 2013). The photon flux density was increased from 2500 to ca 8450 $\mu\text{mol m}^{-2} \text{s}^{-1}$ (Phase 1), then attenuated by 40% (Phase 2) and finally increased back to 8450 $\mu\text{mol m}^{-2} \text{s}^{-1}$ (Phase 3). F_m' was estimated as the intercept of the linear regression of fluorescence yield in Phase 2 against the reciprocal of flash intensity. The apparent quantum efficiency of PSII e⁻ transport (Φ_2) was calculated as $\Delta F/F_m' = (F_m' - F_s)/F_m'$ (Genty et al., 1989).

All gas exchange data and the CO₂ level of substomatal cavity (C_i) were corrected for CO₂ leakage in the chamber based on apparent CO₂ response curves using heat-killed leaves (Flexas et al., 2007), measured with the same flow rate (300 $\mu\text{mol s}^{-1}$).

2.2.3 Leaf nitrogen content assessment

A leaf portion from the leaf position used for measuring photosynthesis was excised for N content measurements. For rice and wheat, a 5-6-cm long part of a leaf was cut down to measure

specific leaf area. For maize and sorghum, a 2-cm² leaf disk was punched. All leaf materials were weighed after drying at 70 °C to constant weight. Then, total N concentration was analysed using the Micro-Dumas combustion method. Leaf N content on the area basis (SLN) was derived from these data.

2.2.4 C_3 model

The calculation of A_c and A_j in the FvCB model is proposed as (Farquhar et al., 1980):

$$A = \frac{(C_c - \Gamma^*)x_1}{C_c + x_2} - R_d \quad (2.1)$$

where for A_c , $x_1 = V_{\text{cmax}}$ and $x_2 = K_{\text{mC}}(1 + O_c/K_{\text{mO}})$; for A_j , $x_1 = J/4$ and $x_2 = 2\Gamma^*$. In the model, C_c is the CO₂ level at the carboxylation site of Rubisco and Γ^* is the CO₂ compensation point in the absence of day respiration (R_d), defined as $0.5O_c/S_{\text{c/o}}$, where O_c is the O₂ level in chloroplasts and $S_{\text{c/o}}$ is the relative CO₂/O₂ specificity factor for Rubisco (von Caemmerer et al., 1994). K_{mC} and K_{mO} are Michaelis-Menten constants of Rubisco for CO₂ and O₂, respectively (Table 2.1). J is the linear e⁻ transport rate.

If the TPU limitation occurs, A can be simplified as

$$A_p = 3T_p - R_d \quad (2.2)$$

Overall, the C_3 model predicts A as the minimum of three limited rates of CO₂ assimilation:

$$A = \min(A_c, A_j, A_p) \quad (2.3)$$

Eqn (2.1) uses C_c as input, but C_c is generally unknown. This is commonly solved by combining Eqn (2.1) and the Fick's law of gas diffusion $C_c = C_i - A/g_m$, and the obtained solution for A can be expressed as a quadratic function of C_i (von Caemmerer, 2000). This procedure assumes that the mesophyll conductance g_m is constant (i.e., independent of light or CO₂ levels at the given temperature) based on early reports (e.g., Loreto et al., 1992). However, our result based on the variable J method of Harley et al. (1992a) showed that g_m was variable, varying with C_i and light levels (see Results). Thus, we used a simplified version of the model (Yin et al., 2009), in which a dimensionless parameter δ is introduced dealing with variable g_m (see Supplementary Appendix A) in such a way that g_m values at various C_i and light levels are emergent properties of the parameter δ (see later).

Table 2.1 Pre-set model parameters with their definitions and values.

Parameter	Definition	C ₃ species		C ₄ species	
		Value	Reference	Value	Reference
K_{mC}	Michaelis-Menten constant of Rubisco for CO ₂ ($\mu\text{mol mol}^{-1}$)	291	Cousins et al. (2010)	485	Cousins et al. (2010)
K_{mO}	Michaelis-Menten constant of Rubisco for O ₂ (mmol mol^{-1})	194	Cousins et al. (2010)	146	Cousins et al. (2010)
K_P	Michaelis-Menten constant of PEPc for CO ₂ ($\mu\text{mol mol}^{-1}$)	-		40	Leegood and von Caemmerer (1989)
$S_{c/o}$	Relative CO ₂ /O ₂ specificity of Rubisco ($\text{mmol } \mu\text{mol}^{-1}$)	3.022	Cousins et al. (2010)	2.862	Cousins et al. (2010)
α	Fraction of O ₂ evolution in bundle sheath cells (-)	-		0.1	Chapman et al. (1980)
x	Fraction of ATP used for CCM (-)	-		0.4	von Caemmerer and Furbank (1999)
n_b	Base leaf nitrogen, at and below which leaf photosynthesis is zero (g N m^{-2})	-		0.24	Yin et al. (2011b)

2.2.5 C_4 model

For C_4 species, A is written as (von Caemmerer and Furbank, 1999):

$$A = V_p - L - R_m \quad (2.4)$$

where V_p is the rate of PEP carboxylation in the mesophyll cells, L is the CO_2 leakage rate from the bundle sheath to the mesophyll, and R_m is the mitochondrial respiration rate occurring in mesophyll cells, considered as $R_d/2$ (von Caemmerer and Furbank, 1999), where R_d is leaf day respiration, the same as defined for the C_3 model.

The standard C_4 model assumes that C_4 acid decarboxylation in bundle sheath cells is not limited and occurs at the same rate as V_p . If V_p is PEP carboxylase limited, V_p is described as:

$$V_p = \frac{C_m V_{pmax}}{C_m + K_p} \quad (2.5a)$$

where V_{pmax} is the carboxylation capacity of PEPc; K_p is the Michaelis-Menten constant of PEPc for CO_2 ; C_m is the CO_2 concentration of mesophyll cells, described as $C_m = C_i - A/g_m$, following the Fick's law of gas diffusion. So far, there is no strong evidence reported for variable g_m in C_4 species; so, we treated g_m as a constant, in contrast with the C_3 case. If V_p is e^- transport limited, V_p is expressed as:

$$V_p = x J_{atp} / 2 \quad (2.5b)$$

where x is the partition fraction of ATP to the C_4 cycle and J_{atp} is the ATP production rate driven by e^- transport. Note that the constant 2 in Eqn (2.5b) applies to the C_4 subtype to which maize and sorghum belong; this coefficient may need to change if the model is applied to other subtypes (Yin et al., 2021a).

The term L in Eqn (2.4) is given by:

$$L = g_{bs}(C_c - C_m) \quad (2.6)$$

in which g_{bs} is the bundle sheath conductance and C_c is the CO_2 concentration at the carboxylation site of Rubisco in bundle sheath cells.

As for the Rubisco carboxylation and oxygenation reacted in the bundle sheath cells, A can be written, like in the C_3 photosynthetic model, as:

$$A = \frac{(C_c - \gamma^* O_c) x_1}{C_c + x_2 O_c + x_3} - R_d \quad (2.7)$$

where γ^* is defined as $0.5/S_{c/o}$, O_c is the O_2 level at the carboxylation sites of Rubisco in the bundle sheath cells, described as $O_c = \alpha A / (0.047 g_{bs}) + O_i$, where α is the fraction of O_2 evolution happening in bundle sheath cells (von Caemmerer and Furbank, 1999). When enzymatic

activity limits, $x_1 = V_{\text{cmax}}$, $x_2 = K_{\text{mC}}/K_{\text{mO}}$ and $x_3 = K_{\text{mC}}$. When A is limited by e^- transport, $x_1 = (1 - x)J_{\text{atp}}/3$, $x_2 = 7\gamma/3$ and $x_3 = 0$ (Yin et al., 2011b).

Consequently, as the reactions related to both C_3 and C_4 cycles can be limited by either enzyme activity or e^- transport, the four-limitation C_4 model is described as:

$$A = \min(A_{\text{EE}}, A_{\text{ET}}, A_{\text{TE}}, A_{\text{TT}}) \quad (2.8)$$

where A_{EE} represents A when both C_3 and C_4 cycles are limited by enzyme actions, A_{ET} is A when the C_4 cycle is limited enzymatically and the C_3 cycle is limited by e^- transport, A_{TE} is A when the C_4 cycle is limited by e^- transport and the C_3 cycle is limited enzymatically, and A_{TT} means A when both cycles are limited by e^- transport. Mathematical solutions to A_{EE} , A_{ET} , A_{TE} , and A_{TT} are from Yin et al. (2011b) and are given in Supplementary Appendix B.

2.2.6 Estimation steps of model parameters

The procedures of Yin et al. (2009) and Yin et al. (2011b) are adopted to estimate C_3 and C_4 photosynthetic parameters, respectively.

2.2.6.1 Estimating calibration factor (s) and day respiration (R_d)

To convert chlorophyll fluorescence-based data for PSII e^- transport efficiency into linear e^- transport rate $J(C_3)$ or to ATP production rate $J_{\text{atp}}(C_4)$ as required by their respective models (see Eqn (2.1) and Eqn (2.7)), it is necessary to obtain a calibration factor based on measurements under non-photorespiratory condition (Valentini et al., 1995; Bernacchi et al., 2003). We followed the procedure of Yin et al. (2009) and Yin et al. (2011b) for C_3 and C_4 models, respectively, for this calibration. Unlike other calibration methods, their procedure simultaneously provides an estimate of R_d , because the calibration factor s and day respiration R_d were estimated as the slope and intercept, respectively, of the linear regression of A against $I_{\text{inc}}\Phi_2/4$ for C_3 species (Yin et al., 2009) and of A against $I_{\text{inc}}\Phi_2/3$ for C_4 species (Yin et al., 2011b), using data under the non-photorespiratory condition. Different stoichiometric coefficients (4 vs 3) were used to agree with the A_j submodel, which was formulated in terms of e^- and ATP demands by the C_3 cycle for C_3 and C_4 models, respectively (see the e^- transport limited part of Eqn (2.1) and Eqn (2.7)). To distinguish them, the slope factor was denoted as s_{c3} and s_{c4} for C_3 and C_4 species, respectively. The selected data for this linear regression was from combined measurements of gas exchange and chlorophyll fluorescence within the e^- transport-limited range (i.e., with I_{inc} between 20 and 200 $\mu\text{mol m}^{-2} \text{s}^{-1}$) under non-

photorespiratory conditions (keeping C_a at $1000 \mu\text{mol mol}^{-1}$ combined with 2% O_2 , which is supposed to suppress photorespiration).

2.2.6.2 Estimating PSII electron transport efficiency under limiting light (Φ_{2LL})

Φ_{2LL} was estimated by fitting the light response curve of Φ_2 to the equation (Yin et al., 2009):

$$\Phi_2 = \frac{\alpha_{2LL} I_{\text{inc}} \beta + J_{2\text{max}} - \sqrt{(\alpha_{2LL} I_{\text{inc}} \beta + J_{2\text{max}})^2 - 4\theta_2 J_{2\text{max}} \alpha_{2LL} I_{\text{inc}} \beta}}{2\theta_2 \alpha_{2LL} I_{\text{inc}} \beta / \Phi_{2LL}} \quad (2.9)$$

where α_{2LL} is the quantum efficiency of PSII e^- transport under limiting light on the combined PSI- and PSII-absorbed light basis and given as $\Phi_{2LL}(1 - f_{\text{cyc}})/[\Phi_{2LL}/\Phi_{1LL} + (1 - f_{\text{cyc}})]$, in which f_{cyc} is the fraction of e^- at PSI that follow cyclic e^- transport around PSI and Φ_{1LL} is the quantum efficiency of PSI e^- transport under limiting light on the PSI-absorbed light basis. β is the leaf absorbance by photosynthetic pigments, $J_{2\text{max}}$ is the maximum rate of all e^- transport through PSII under saturated light (J_2) and θ_2 is the convexity factor for response of J_2 to absorbed photon flux density (I_{abs}). Since θ_2 and Φ_{2LL} are not affected by the various values of f_{cyc} or Φ_{1LL} , arbitrary yet physiologically relevant values of f_{cyc} (e.g., 0) or Φ_{1LL} (e.g., 1.0) can be used to estimate Φ_{2LL} (Yin et al., 2009). Thus, with measured I_{inc} , β and Φ_2 , Φ_{2LL} , θ_2 and $J_{2\text{max}}$ can be fitted as output. As estimated θ_2 and $J_{2\text{max}}$ are not used in the FvCB model, only Φ_{2LL} will be discussed later.

2.2.6.3 Calculating electron transport parameters

Once s and Φ_{2LL} are known, for C_3 species, the efficiency of converting incident irradiance into linear e^- transport (κ_{2LL}) can be calculated as $s_{c3}\Phi_{2LL}$ (Yin et al., 2009). Following a standard model, the light response curve of the calculated J can be described by:

$$J = [\kappa_{2LL} I_{\text{inc}} + J_{\text{max}} - \sqrt{(\kappa_{2LL} I_{\text{inc}} + J_{\text{max}})^2 - 4\theta J_{\text{max}} \kappa_{2LL} I_{\text{inc}}}] / (2\theta) \quad (2.10a)$$

where J_{max} is the light-saturated maximum value of linear e^- transport rate, and q is a curvature factor. Values of J_{max} and q can be estimated by fitting Eqn (2.10a) to the data for the light response curve of the calculated J ($= s_{c3} I_{\text{inc}} \Phi_2$).

Compared with C_3 species, energy supply is more complex in C_4 species because of the coordinated functioning of mesophyll and bundle sheath cells. Nonetheless, as energy is highly likely to be shared between the two types of cells (Kanai and Edwards, 1999), modelling energy supply or production rate as a whole is pragmatic and sufficient for our analysis on the basis of measurements of photosynthetic quantum yield and PSII photochemical efficiency conducted

on the whole leaf (von Caemmerer and Furbank, 1999; Yin and Struik, 2012). This is achieved using a factor, x , for energy partitioning to the C_4 cycle, and $(1 - x)$ for partitioning to the C_3 cycle (see Eqn (2.5b) and Eqn (2.7), respectively). Thus, for C_4 species, J_{atp} can be calculated as $J_{\text{atp}} = s_{c4} I_{\text{inc}} \Phi_2 / (1 - x)$ (Yin et al., 2011b). Alike with Eqn (2.10a), the light response curve of J_{atp} can be described as (Yin et al., 2011b):

$$J_{\text{atp}} = \left[\kappa_{2\text{LL},\text{atp}} I_{\text{inc}} + J_{\text{max},\text{atp}} - \sqrt{(\kappa_{2\text{LL},\text{atp}} I_{\text{inc}} + J_{\text{max},\text{atp}})^2 - 4\theta J_{\text{max},\text{atp}} \kappa_{2\text{LL},\text{atp}} I_{\text{inc}}} \right] / (2\theta) \quad (2.10b)$$

where $\kappa_{2\text{LL},\text{atp}}$, the efficiency of converting incident light into ATP under limiting light, was calculated by $s_{c4} \Phi_{2\text{LL}} / (1 - x)$. Thus, the light-saturated maximum value of ATP production rate, $J_{\text{max},\text{atp}}$, can be estimated from fitting Eqn (2.10b) to the light response curve of the calculated J_{atp} .

2.2.6.4 Estimating T_p and δ

Since prediction of A_c is affected by possible uncertainties of K_{mC} and K_{mO} (see sensitivity analysis later), g_m was estimated after embedding $J = s_{c3} I_{\text{inc}} \Phi_2$ into the A_j model (see Eqns (A2.1) and (A2.2) in Supplementary Appendix A) for C_3 species (Yin et al., 2009). Hence, the parameter δ in the g_m model and T_p were parameterised together using measured data under limiting light ($I_{\text{inc}} < 200 \mu\text{mol m}^{-2} \text{s}^{-1}$) of the A - I_{inc} curves and higher ambient CO_2 ($C_a > 400 \mu\text{mol mol}^{-1}$) of the A - C_i curves where A is expected to be limited either by e^- transport or by TPU.

2.2.6.5 Estimating V_{cmax} and other parameters

In our study, in order to have a better fit of measured data, J_{max} and θ for C_3 species were estimated together with V_{cmax} by fitting the complete C_3 model (see Supplementary Appendix A) combined with Eqn (2.10a) based on the whole dataset, using estimated R_d , $\kappa_{2\text{LL}}$, δ and T_p as input. As J_{max} differed between non-photorespiratory and photorespiratory conditions, the dummy variable method (Yin et al., 2009) was used for the different values of J_{max} under 21% and 2% O_2 conditions.

For C_4 species, with estimated R_d and s_{c4} as input, the complete C_4 model (see Supplementary Appendix B) combined with $J_{\text{atp}} = s_{c4} I_{\text{inc}} \Phi_2 / (1 - x)$ was used for estimating g_{bs} , V_{cmax} and V_{pmax} simultaneously from fitting to the whole dataset. As g_m is less important than

g_{bs} in C_4 species, in order to reduce over-fitting, an overall estimate for g_m was made and g_{bs} was estimated based on the proven linear relationship between g_{bs} and $(SLN - n_b)$, where n_b is the base leaf N, at which leaf photosynthesis reduces to zero (Retta et al., 2016; Yin et al., 2011b).

2.2.6.6 Calculating the dependence of g_m in C_3 species in response to C_i and I_{inc}

Once all parameters were obtained, the dependence of g_m on light and CO_2 in C_3 species was back-calculated based on the following equation by Yin et al. (2009):

$$g_m = A + \delta(A + R_d)/(C_i - \Gamma_*) \quad (2.11)$$

where A is modelled net CO_2 assimilation rate for specific light and CO_2 levels.

2.2.7 Model input constants and statistical analyses

As many parameters of the FvCB model and its C_4 equivalents are conserved, we set these conserved parameters to constant values (Table 2.1). The definitions of estimated model parameters are listed in Table 2.2.

Simple linear regression was carried out by Microsoft Excel. Non-linear regression was performed in SAS using the GAUSS method in PROC NLIN (SAS Institute Inc, Cary, NC, USA). Data for individual replicates were pooled to obtain robust parameter estimates for each treatment.

The difference between C_3 and C_4 species in SLN was evaluated by the use of the Student's t test in Excel. Photosynthetic parameters were linearly regressed against SLN, for which only the estimates of the parameter for N treatments were used. Analysis of covariance (ANCOVA) was conducted in R (R Core Team, 2021) to examine any significant differences in the regression equations between crops.

Table 2.2 Model parameters to be estimated.

Parameter	Definition
g_{bs}	Bundle sheath diffusion conductance ($\text{mmol m}^{-2} \text{s}^{-1}$)
g_m	Mesophyll diffusion conductance ($\text{mol m}^{-2} \text{s}^{-1}$)
J_{\max}	Maximum rate of linear e^- transport for C_3 species under saturated light ($\mu\text{mol } e^- \text{m}^{-2} \text{s}^{-1}$)
$J_{\max,atp}$	Maximum chloroplastic ATP production rate for C_4 species under saturated light ($\mu\text{mol ATP m}^{-2} \text{s}^{-1}$)
R_d	Day respiration ($\mu\text{mol CO}_2 \text{m}^{-2} \text{s}^{-1}$)
s_{c3}	A lumped parameter as the calibration factor for C_3 species ($\text{mol CO}_2 (\text{mol CO}_2)^{-1}$)
s_{c4}	A lumped parameter used to estimate R_d for C_4 species ($\text{mol ATP (mol } e^-)^{-1}$)
T_p	Rate of triose phosphate export from the chloroplast ($\mu\text{mol phosphate m}^{-2} \text{s}^{-1}$)
$V_{c\max}$	Maximum rate of Rubisco activity-limited carboxylation ($\mu\text{mol CO}_2 \text{m}^{-2} \text{s}^{-1}$)
$V_{p\max}$	Maximum rate of PEPC activity-limited carboxylation ($\mu\text{mol CO}_2 \text{m}^{-2} \text{s}^{-1}$)
χ_{gbs}	Linear slope of bundle-sheath conductance versus ($\text{SLN} - n_b$) ($\text{mmol CO}_2 \text{g}^{-1} \text{N s}^{-1}$)
$\chi_{j\max}$	Linear slope of J_{\max} versus SLN ($\mu\text{mol } e^- \text{g}^{-1} \text{N s}^{-1}$)
$\chi_{j\max,atp}$	Linear slope of $J_{\max,atp}$ versus SLN ($\mu\text{mol ATP g}^{-1} \text{N s}^{-1}$)
χ_{Tp}	Linear slope of T_p versus SLN ($\mu\text{mol phosphate g}^{-1} \text{N s}^{-1}$)
$\chi_{v\max}$	Linear slope of $V_{c\max}$ versus SLN ($\mu\text{mol CO}_2 \text{g}^{-1} \text{N s}^{-1}$)
$\chi_{vp\max}$	Linear slope of $V_{p\max}$ versus SLN ($\mu\text{mol CO}_2 \text{g}^{-1} \text{N s}^{-1}$)
δ	Parameter to quantify the variable g_{ms} , representing the carboxylation to mesophyll resistance ratio (-)
κ_{2LL}	Conversion efficiency of incident light into linear e^- transport for C_3 species at the strictly limiting light level ($\text{mol } e^- (\text{mol photon})^{-1}$)
$\kappa_{2LL,atp}$	Conversion efficiency of incident light into ATP production for C_4 species at the strictly limiting light level ($\text{mol ATP (mol photon)}^{-1}$)
θ	Convexity factor for response of e^- transport or of ATP production to incident light (-)
ϕ_{2LL}	ϕ_2 at the strictly limiting light level (mol mol^{-1})

2.3 Results

2.3.1 Effects of nitrogen supply on photosynthetic capacity A_{\max}

Measured A while keeping incident I_{inc} at $1750 \mu\text{mol m}^{-2} \text{s}^{-1}$ and C_a at $400 \mu\text{mol mol}^{-1}$ in A - I_{inc} curves were recorded as A_{\max} . When O_2 level decreased from 21% to 2%, A_{\max} increased in rice and wheat due to suppressed photorespiration; but this increase was very small in maize and sorghum as a result of the operation of CCM in C_4 species (Fig. 2.1). Under the same leaf N conditions, A_{\max} of maize was higher than that of sorghum and there was a significant difference in the intercept of their regression equations for the A_{\max} -SLN linear plot ($p < 0.05$). Moreover, the linear regression equation of A_{\max} versus SLN for C_4 species differed significantly from that for C_3 species under both O_2 conditions in terms of either slope or intercept ($p < 0.05$) (Table S2.1). The increasing trends of A_{\max} of rice and wheat were similar when SLN ranged from 0.59 to 1.44 g N m^{-2} , but SLN of wheat plants of the high-N treatment reached 2.21 g N m^{-2} and A_{\max} at this high SLN tended to level off (Fig. 2.1a).

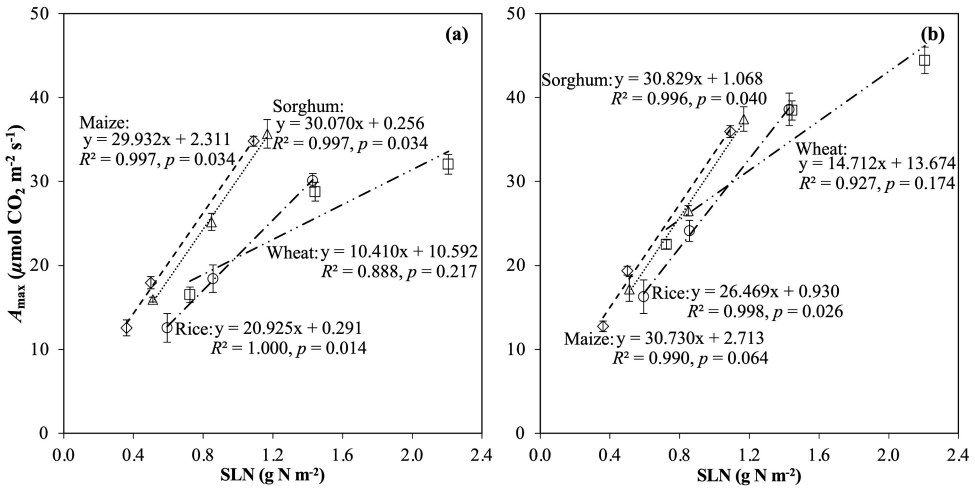


Fig. 2.1 Measured A_{\max} (maximum rate of light saturated assimilation at ambient CO_2 level) in relation to SLN (specific leaf nitrogen content) for rice (circles), wheat (squares), maize (diamonds) and sorghum (triangles) under conditions of 21% (a) and 2% O_2 (b). Data points are shown with bars for standard errors of the means.

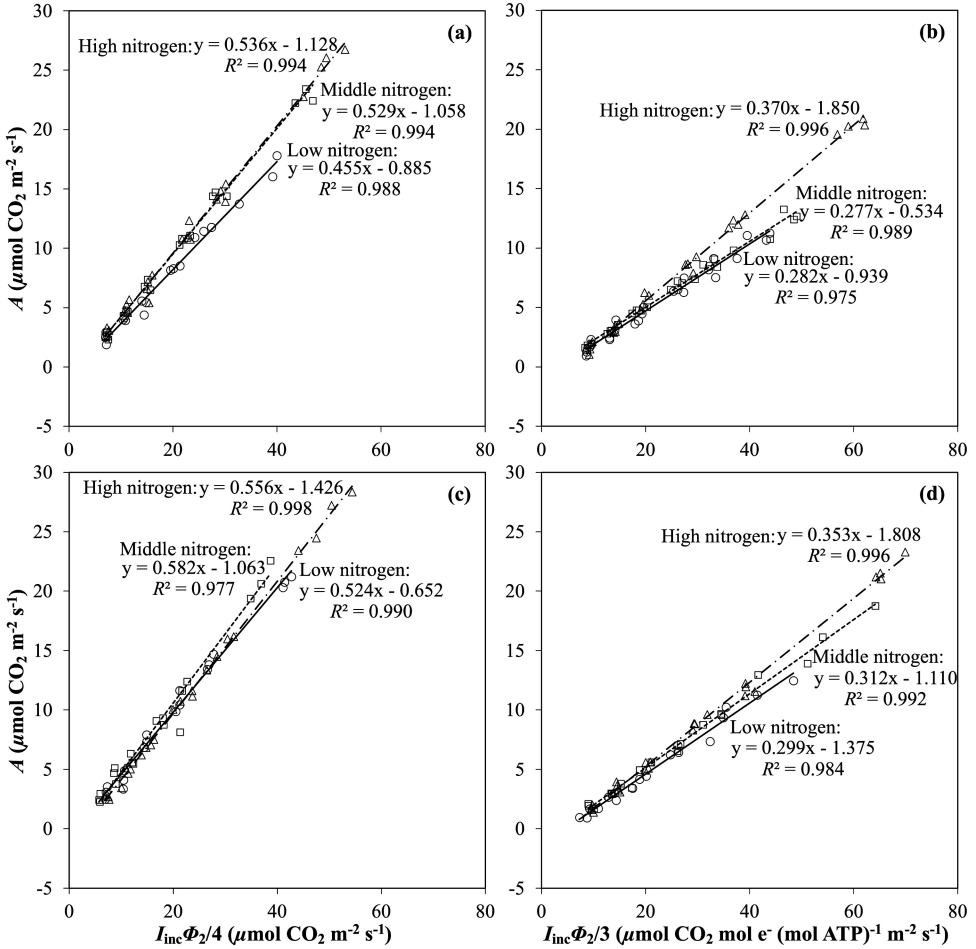


Fig. 2.2 Linear regression of assimilation rate A against $I_{\text{inc}} \Phi_2/4$ of rice (a), wheat (c), and of A against $I_{\text{inc}} \Phi_2/3$ of maize (b) and sorghum (d) under low- (open circles), middle- (open squares) and high- (open triangles) nitrogen supply, respectively. Measurements were conducted under non-photorespiratory condition. Each data point reflects measurement of individual replicate. The intercept of the regression represents the estimation of day respiration (R_d), and the slope represents the estimation of the calibration factor (i.e., s_{c3} for rice and wheat, and s_{c4} for maize and sorghum; see the text).

2.3.2 Estimations of calibration factor and day respiration R_d

Calibration factor and R_d were estimated by linear regression (Fig. 2.2). Estimated values of s_{c3} were in the range of 0.455 to 0.582 mol CO₂ (mol CO₂)⁻¹ and s_{c4} varied from 0.277 to 0.370 mol ATP (mol e⁻)⁻¹. The linear relationship between s_{c3} of C₃ crops and SLN was non-significant ($p > 0.05$) (Fig. 2.3a). Likewise, s_{c4} increased with SLN for C₄ crops, but non-significantly ($p > 0.05$) (Fig. 2.3a). The estimated R_d increased with SLN, but non-significantly in both C₃ and C₄ crops ($p > 0.05$) (Fig. 2.3b). Moreover, there was no significant ($p > 0.05$) difference in these regressed linear equations between rice and wheat, nor between maize and sorghum.

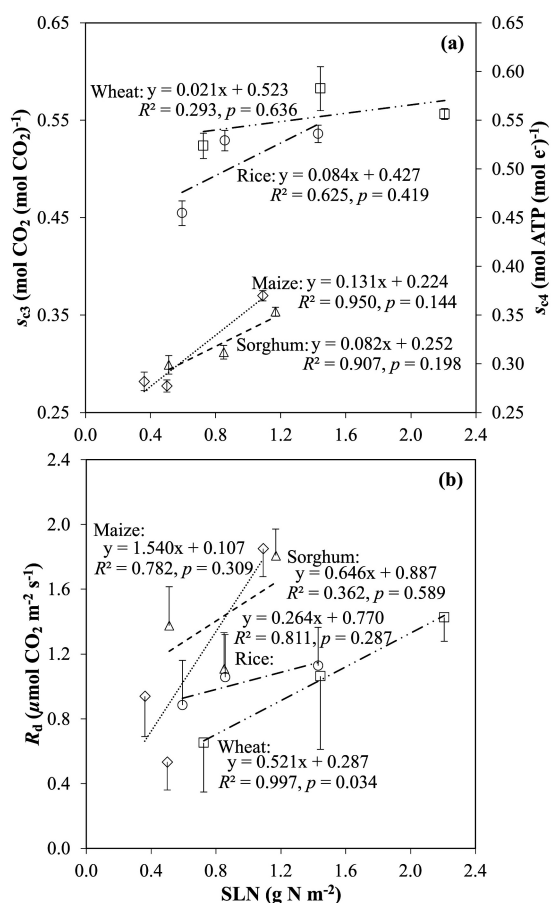


Fig. 2.3 Calibration factor for C₃ crops s_{c3} or for C₄ crops s_{c4} (a) and day respiration R_d (b) in relation to specific leaf nitrogen content (SLN). Rice (circles), wheat (squares), maize (diamonds) and sorghum (triangles). The vertical bars represent the standard errors of the estimates.

Table 2.3 Estimated (standard error in brackets if applicable) of ϕ_{2LL} and κ_{2LL} or $\kappa_{2LL,atp}$. Estimates are made separately for different O₂ conditions.

Parameter	Low nitrogen	Mid nitrogen	High nitrogen	Low nitrogen	Mid nitrogen	High nitrogen
	Wheat					
SLN (g N m ⁻²)	0.593 (0.032)	0.856 (0.038)	1.429 (0.068)	0.724 (0.054)	1.444 (0.115)	2.208 (0.068)
ϕ_{2LL} (mol mol ⁻¹)	21% O ₂	0.645 (0.026)	0.681 (0.010)	0.683 (0.006)	0.631 (0.012)	0.699 (0.013)
	2% O ₂	0.626 (0.040)	0.671 (0.015)	0.669 (0.009)	0.601 (0.018)	0.660 (0.013)
κ_{2LL}^* [mol e ⁻ (mol photon) ⁻¹]	21% O ₂	0.293	0.360	0.366	0.323	0.389
	2% O ₂	0.285	0.355	0.359	0.367	0.367
	Maize					
	Sorghum					
SLN (g N m ⁻²)	0.361 (0.018)	0.500 (0.051)	1.090 (0.068)	0.511 (0.042)	0.849 (0.063)	1.168 (0.055)
ϕ_{2LL} (mol mol ⁻¹)	21% O ₂	0.593 (0.016)	0.605 (0.007)	0.601 (0.003)	0.531 (0.007)	0.649 (0.016)
	2% O ₂	0.581 (0.019)	0.591 (0.009)	0.586 (0.005)	0.489 (0.008)	0.618 (0.020)
$\kappa_{2LL,atp}^*$ [mol ATP (mol photon) ⁻¹]	21% O ₂	0.279	0.279	0.371	0.276	0.382
	2% O ₂	0.273	0.273	0.362	0.254	0.364

* Since κ_{2LL} and $\kappa_{2LL,atp}$ were calculated as $s_{c3}\phi_{2LL}$ and $s_{c4}\phi_{2LL}/(1 - x)$ (see materials and methods) for C₃ and C₄ crops, respectively, standard errors are not available for these variables.

2.3.3 Estimations of photosynthetic efficiency parameters

Estimated Φ_{2LL} ranged from 0.49 to 0.70 mol mol⁻¹ and was higher in C₃ species than in C₄ species under both 21% and 2% O₂ levels (Table 2.3). The increase of Φ_{2LL} with SLN for both C₃ and C₄ species was non-significant ($p > 0.05$) under any of the two O₂ levels (Table S2.1). Φ_{2LL} changed slightly with increasing O₂ level from 2% to 21% and the ratio of estimated value of Φ_{2LL} under 21% and 2% O₂ levels varied from 1.02 to 1.07 among the four species.

The estimated κ_{2LL} values of the C₃ species were in the range of 0.285 to 0.389 mol e⁻ (mol photon)⁻¹ and $\kappa_{2LL,atp}$ of C₄ species varied from 0.254 to 0.382 mol ATP (mol photon)⁻¹. As κ_{2LL} and $\kappa_{2LL,atp}$ were calculated from calibration factor (s_{c3} and s_{c4}) and Φ_{2LL} (which both varied slightly with SLN), it follows that κ_{2LL} and $\kappa_{2LL,atp}$ increased with increasing SLN (Table 2.3).

2.3.4 Estimation of CO₂-diffusion parameters

For rice and wheat, we first calculated g_m using the commonly used method of Harley et al. (1992a), which showed that g_m varied with C_i and I_{inc} and was zero at the zero irradiance (Fig. S2.1). We thus incorporated the variable- g_m formula into the FvCB equation for A_j to estimate the parameter δ (see Supplementary Appendix A). The calculated g_m at A_{max} from using this variable- g_m algorithm, i.e., Eqn (2.11), had a good linear relationship with SLN as well (Fig. 2.4a). If SLN was at the same level, g_m of rice tended to be higher than g_m of wheat (Fig. 2.4a).

With respect to C₄ species, we were not able to judge whether or not CO₂-diffusion parameters are variable, so an overall g_m was estimated from fitting the C₄ model (Supplementary Appendix B) to measurements. The estimated g_m was 2.99 (s.e. 0.65) and 1.87 (s.e. 0.28) mol m⁻² s⁻¹ for maize and sorghum, respectively, which were - not surprisingly - higher than the values of the C₃ species. g_{bs} was estimated under different N supplies and both the estimated g_{bs} and its response slope to SLN were lower in maize than in sorghum (Fig. 2.4b).

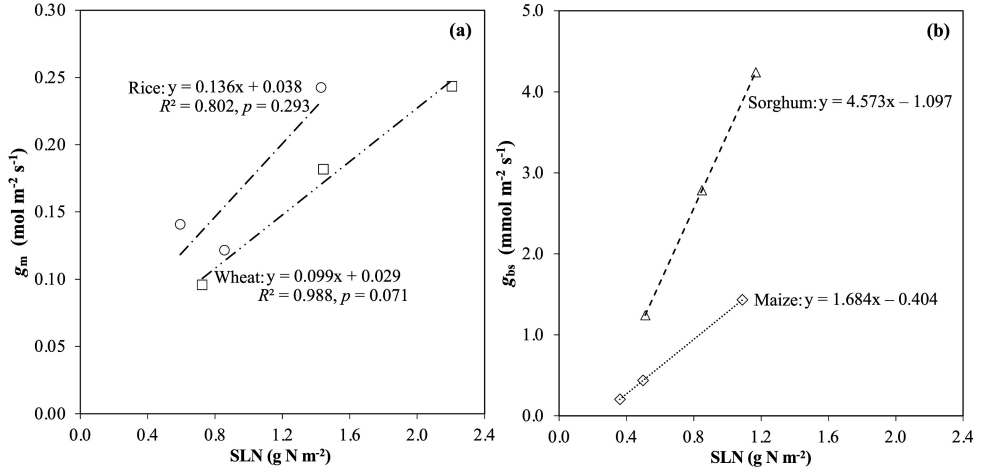


Fig. 2.4 Mesophyll conductance g_m (a) of rice (circles) and wheat (squares) and bundle-sheath conductance g_{bs} (b) of maize (diamonds) and sorghum (triangles) in relation to specific leaf nitrogen content (SLN). In Panel a, g_m represents the value of g_m when A reached A_{\max} (i.e., keeping I_{inc} at $1750 \mu\text{mol m}^{-2} \text{s}^{-1}$ and C_a at $400 \mu\text{mol mol}^{-1}$). The estimated g_{bs} of maize and sorghum were calculated from the equation, $g_{bs} = \chi_{gbs}(\text{SLN} - n_b)$, where the linear slope χ_{gbs} was estimated, together with other parameters, from fitting the full C_4 model to the whole data of each crop (see the text); thus, there are no R^2 and p values for the equation in panel b.

2.3.5 Estimations of photosynthetic capacity parameters and their linear slopes with leaf nitrogen content

There were significant and positive linear relationships between estimated photosynthetic capacity parameters and SLN (Table S2.1). Within C_3 species, the regression slope of J_{\max} versus SLN ($\chi_{J_{\max}}$) of rice was significantly ($p < 0.05$) higher than that of wheat, while the difference of the intercept was non-significant ($p > 0.05$), as reflected by the higher values of estimated J_{\max} of wheat than rice under low N input but lower values for wheat under high N input (Fig. 2.5a). For the C_4 crops, $J_{\max, \text{atp}}$ of maize was higher than that of sorghum, although there was no significant difference ($p > 0.05$) between regressed linear equations of $J_{\max, \text{atp}}$ versus SLN in both crops.

The estimate of V_{\max} increased with SLN in both C_3 and C_4 species (Fig. 2.5b) and the equation of V_{\max} versus SLN for C_3 species was significantly ($p < 0.01$) different from that for C_4 species in the intercept, but not in the slope ($\chi_{V_{\max}}$) ($p > 0.05$). Moreover, $\chi_{V_{\max}}$ of rice was significantly ($p < 0.05$) higher than that of wheat. However, the response of V_{\max} to SLN was

weak in C₄ species and there was no significant ($p > 0.05$) difference between maize and sorghum (Fig. 2.5b).

The estimate of T_p in C₃ species (Fig. 2.5c) and that of V_{pmax} in C₄ species (Fig. 2.5d) increased with SLN. But the trend of T_p with SLN did not differ significantly ($p > 0.05$) between rice and wheat, nor did that of V_{pmax} with SLN between maize and sorghum.

Together with above mentioned photosynthetic capacity parameters in C₃ and C₄ species, the convexity factor θ was estimated. The overall estimated θ was 0.84 (0.02) and 0.81 (0.02) for rice and wheat, respectively, while its value was up to 0.98 (0.01) and 0.99 (0.01) for maize and sorghum, respectively (results not shown).

2.3.6 Modelled CO₂ and irradiance responses of net assimilation rate

The modelled versus measured $A-C_i$ and $A-I_{inc}$ curves across crop species \times N supply combinations are shown in Figure S2.2. A increased with N supply and such an effect tended to be enlarged when suppressing O₂ level from 21% to 2%. Meanwhile, modelled $A-C_i$ and $A-I_{inc}$ curves were slightly underestimated in low and middle N supply treatments when I_{inc} was higher than 500 $\mu\text{mol m}^{-2} \text{s}^{-1}$ or C_a was kept higher than 400 $\mu\text{mol mol}^{-1}$, but overestimated when there was high N input (Fig. S2.2). Notably, both in maize and sorghum, the modelled A with response to C_i fluctuated under low and middle N input condition, as a result of fluctuations in the measured Φ_2 that were used as input to calculate J_{atp} for model fitting. Nonetheless, our results showed that overall, both C₃ and C₄ photosynthetic models fitted data well ($R^2 > 0.98$).

2.3.7 Sensitivity analysis

Figure S2.3 shows the relative changes of the estimated J_{max} and V_{cmax} for C₃ species with up to $\pm 50\%$ change in input parameter values. As expected, the enzyme-related pre-set parameters, like K_{mC} , K_{mO} and $S_{c/o}$ hardly had any influence on the estimates of J_{max} (Fig. S2.3a-c). V_{cmax} seemed to linearly change with K_{mC} (Fig. S2.3a). However, the linear change of V_{cmax} with K_{mO} and $S_{c/o}$ only existed when their values were lower than the pre-set values (Fig. S2.3b, c). Regarding the sensitivity to the previous estimated parameters, V_{cmax} and J_{max} hardly changed with R_d (Fig. S2.3d) while κ_{2LL} and T_p tended to have a slight influence on the estimates of V_{cmax} and J_{max} and δ changed the estimates of V_{cmax} significantly (Fig. S2.3e-g).

For C₄ species, the changes in pre-set and estimated parameters were mostly reflected in changes in χ_{gbs} (Fig. S2.4). Except for α , most of these parameters can have a noticeable

influence on χ_{gbs} (Fig. S2.4). K_p and R_d can cause the notable change of g_m (Fig. S2.4c, g). The change in V_{cmax} was slightly affected by K_{mC} , K_{mO} and $S_{\text{c/o}}$ (Fig. S2.4a, b, d). Compared with other parameters, V_{pmax} was more likely to be influenced by K_p (Fig. S2.4c). n_b showed, not surprisingly, a systematic influence on the estimates of χ_{gbs} and the relationship between n_b and χ_{gbs} tended to be positively linear (Fig. S2.4f).

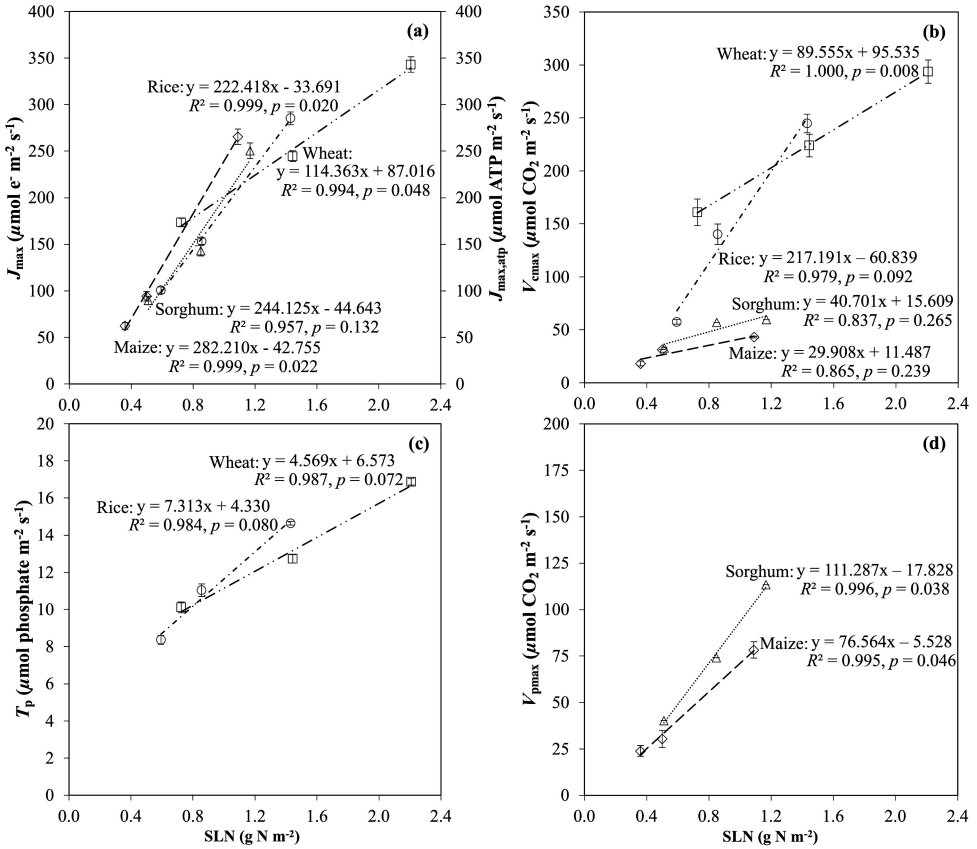


Fig. 2.5 Estimated maximum rate of linear electron transport J_{max} (for rice and wheat) or maximum rate of ATP production $J_{\text{max,atp}}$ (for maize and sorghum) (a), maximum rate of Rubisco carboxylation V_{cmax} (b), rate of triose phosphate utilisation T_p (for rice and wheat) (c), and maximum rate of PEP carboxylation V_{pmax} (for maize and sorghum) (d) in relation to specific leaf nitrogen content (SLN) for rice (circles), wheat (squares), maize (diamonds) and sorghum (triangles) under 21% O₂ condition. Rice and wheat had different estimates of J_{max} under 21% and 2% O₂ levels (see the text), but in order to make them comparable with maize and sorghum, only J_{max} under 21% O₂ level is shown. Data points are shown with standard errors of the estimates.

2.4 Discussion

Major cereal crops (rice, wheat, maize, and sorghum) grow on a global cultivation area of almost 700 million hectares, and supply approximately 50% of the world's caloric intake (Singer et al., 2019). In order to support smart farming, an accurate prediction of growth of major crops is essential. Photosynthesis is the primary process of crop growth and responds to variations in multiple environmental variables and management manipulations (in particular, N supply). The model of Farquhar et al. (1980) has been proven to be essential for modelling leaf photosynthesis in response to interactions of multiple physiological and environmental variables, and can be used as the basic model for upscaling to canopy and crop scales (Gu et al., 2014; Yin and Struik, 2017). However, reports about estimating and comparing photosynthetic parameters for these major crops together are quite rare, as most studies mainly focus on a single crop (e.g., Qian et al., 2012; Retta et al., 2016; Yin et al., 2009). Here, in our study, the primary photosynthetic parameters for four major crops, two C_3 species (rice and wheat) and two C_4 species (maize and sorghum) were quantified together. Although our estimated parameter values were subject to uncertainties in input constants of the FvCB model (Figs. S2.3 and S2.4), our results could serve as a referenced dataset in related research.

2.4.1 Effects of leaf nitrogen content on photosynthetic capacity

Photosynthetic capacity A_{\max} has been found to be highly, positively correlated with SLN (Yoshida and Coronel, 1976), and R^2 for the linear regression is up to 0.80 (Field and Mooney, 1983). Furthermore, N related proxies have also been found to linearly increase with A_{\max} under ambient or enhanced CO_2 condition, like leaf chlorophyll content ($R^2 > 0.95$) (Wong, 1979). Our results show that with an increase in SLN, the linear slopes of A_{\max} to SLN were ranging from 14.71 to 30.83 $\mu\text{mol } CO_2 \text{ g}^{-1} \text{ N s}^{-1}$ and, except for wheat, R^2 values were higher than 0.99 (Fig. 2.1). The steeper slope of A_{\max} with respect to SLN in C_4 species is in line with earlier reports (Anten et al., 1995; Byrd et al., 1992; Sage and Percy, 1987). This is basically because of the higher efficiency of carboxylation of Rubisco associated with the CCM in the C_4 photosynthetic pathway (Black Jr, 1973; Schmitt and Edwards, 1981), as evidenced by the increased slope values for C_3 crops when O_2 was decreased to 2% (Fig. 2.1b).

2.4.2 Estimated biochemical parameters of photosynthetic capacity and their relationships with nitrogen

SLN significantly ($p < 0.001$) differed between C_3 and C_4 species. For example, SLN of rice was 1.6 times the value of maize (Table 2.3). V_{cmax} , J_{max} and T_p increased with increasing SLN (Harley et al., 1992b). Our results confirmed the linear increase of these photosynthetic parameters with SLN in rice and wheat (Fig. 2.5a-c). Values of V_{cmax} of rice and wheat tended to be higher than those of maize and sorghum at the same SLN (Fig. 2.5b), which may indicate that a lower amount of Rubisco is required in C_4 species for achieving the same rate of CO_2 assimilation as in C_3 species. The values of estimated J_{max} of rice and wheat varied from 73.3 to 343.1 $\mu\text{mol e}^- \text{m}^{-2} \text{s}^{-1}$ under 21% and 2% O_2 conditions, higher than those of previous studies (Li et al., 2009; Yin et al., 2009). The SLN of wheat was significantly ($p < 0.01$) higher than that of rice, and this was compensated for by the significantly higher estimated $\chi_{J_{\text{max}}}$ under 21% O_2 conditions in rice than in wheat. The lower $\chi_{J_{\text{max}}}$ of wheat was in line with the small increment of A_{max} under high N condition (Fig. 2.1), which indicates that the increasing trend of A with SLN would be ceased while plant growth is no longer limited by N supply (Hirose, 1984). For example, more than one third of the ability of carboxylation by Rubisco in wheat leaves would be inactivated once carboxylation exceeds 155 $\mu\text{mol m}^{-2} \text{s}^{-1}$, although more N tends to be accumulated (Evans, 1983).

Estimated values of V_{cmax} for C_4 species ranged from 18.3 to 59.6 $\mu\text{mol CO}_2 \text{m}^{-2} \text{s}^{-1}$ (Fig. 2.5b) and V_{pmax} varied from 24.0 to 113.4 $\mu\text{mol CO}_2 \text{m}^{-2} \text{s}^{-1}$ (Fig. 2.5d), in line with earlier reports (Ghannoum et al., 2000) and both of them varied in proportion to SLN (Retta et al., 2016; Yin et al., 2011b). The estimated V_{pmax} also tended to have a higher value than V_{cmax} and the slopes of the regressed equation of V_{pmax} on SLN ($\chi_{V_{\text{pmax}}}$) were 2.6 and 2.7 times values of $\chi_{V_{\text{cmax}}}$ in maize and sorghum, respectively (Fig. 2.5d). The higher V_{pmax} than V_{cmax} has also been observed in sugarcane when varying N supply ($V_{\text{pmax}} = 1.3 V_{\text{cmax}}$) (Tofanello et al., 2021). Higher values of V_{pmax} than V_{cmax} , combined with a lower Michaelis-Menten constant value of PEPc than of Rubisco for CO_2 (Table 2.1), guarantees the faster C_4 cycle than the C_3 cycle, thereby, effectively sustaining the CCM in C_4 photosynthesis.

2.4.3 Estimations of R_d and its relationship with nitrogen

Environmental conditions also affect respiration. For instance, increased respiration rate of rice tends to be the primary cause of its yield losses under high temperature, while the photosynthetic rate hardly changed (Li et al., 2021). With respect to N response, earlier research has reported that leaf respiration occurring in the dark (R_{dk}) scales with SLN in several trees and shrubs (Ryan, 1995). Reich et al. (2008) demonstrated the strong relationship between R_{dk} and leaf N by utilising a database containing 287 species. It has also been reported that the estimated R_d of maize generally increases with SLN (Retta et al., 2016). In our study, we also found R_d increased with N application and there was a linear relationship between R_d of C_3 and C_4 species and SLN, although the correlation was weaker for the C_4 species than for the C_3 species (Fig. 2.3b). Meanwhile, in line with the results in earlier reports that the estimated R_d of maize was higher than those of rice and wheat (Griffin and Turnbull, 2013; Yin et al., 2011a), we also found that the estimated R_d of sorghum, a common C_4 species, was also higher than the estimated R_d of rice and wheat (Fig. 2.3b). This is in contrary to the expectation that a higher observed SLN is supposed to lead to higher respiration costs (Mooney and Gulmon, 1982) as protein turnover represents a considerable expenditure of energy and thus associates with a significant proportion of leaf respiration (Penning de Vries, 1975). The reasons for higher R_d in C_4 than in C_3 crops need further studies to elucidate.

2.4.4 Estimation of g_m and g_{bs}

g_m may limit A by 20% in C_3 species (Warren, 2008). The analysis of our data using the variable J method (Harley et al., 1992a) confirmed that g_m of rice and wheat varied with C_i and I_{inc} (Fig. S2.1). The estimated g_m was lower than $0.8 \text{ mol m}^{-2} \text{ s}^{-1}$ (Fig. S2.1), in line with the results of von Caemmerer and Evans (2015) derived from the gas exchange and carbon isotope discrimination measurements. g_m of rice tended to be higher than that of wheat under different N supplies (Fig. S2.1, 2.4a), compared with different g_m between the two crops reported by Ouyang et al., (2017). It has been found that there is a positive linear correlation between g_m and A_{max} ($R^2 = 0.84$) (von Caemmerer and Evans, 2015). We also found that g_m under the condition where A_{max} was measured tended to linearly increase with SLN (Fig. 2.4a). C_4 species has a higher g_m than C_3 species, because g_m in C_4 species arises from only the mesophyll cell-wall and plasma membrane resistance whereas g_m in C_3 species additionally includes resistance components from chloroplasts (Evans, 1996). In our estimates, g_m for maize and sorghum were

2.99 (0.65) and 1.87 (0.28) mol m⁻² s⁻¹, respectively, similar to the estimates of Yin et al. (2011b) for maize.

Compared with g_m , g_{bs} is crucial in C₄ photosynthesis as it determines the efficiency of the CCM (Kromdijk et al., 2014) and tends to be more likely affected by pre-set model constants (Fig. S2.4). Under N deficient conditions, a low g_{bs} helps maintain high CO₂ concentration at Rubisco sites in compensation for the low PEPc activity (Tofanello et al., 2021). Our results also showed that estimated g_{bs} of maize and sorghum based on the assumed linear relationship of g_{bs} with SLN fitted data well, in line with the early reports about maize (Retta et al., 2016; Yin et al., 2011b). Like the observed difference of g_m in different C₃ species, g_{bs} varied between species as well and the estimated g_{bs} of sorghum was notably higher than the estimates for maize (Fig. 2.4b). A recent study showed that the increase in g_{bs} with SLN can hardly be explained by leaf anatomical traits, but is more likely associated with plasmodesmata density and membrane permeability (Retta et al., 2016). This is in line with the significantly enhanced expression of the plasma membrane intrinsic protein in *Cleome gynandra* (C₄) (Brautigam et al., 2011), which could confer higher permeability to CO₂ (Weber and von Caemmerer, 2010).

2.4.5 Implications of nitrogen response of photosynthetic parameters for smart crop management

Given the above-discussed crucial roles of N in affecting photosynthetic parameters and also considering the N effect on other aspects of crop growth, N management is central to precision cultivation of crops. Nitrogen management can be optimised with the diagnosis of crop N status. With well estimated photosynthetic parameters, crop models can predict real-time course of crop growth, thereby, potentially being able to support real-time crop N management.

Photosynthetic parameters may be estimated directly from exploring modern technologies. J_{max} has been directly retrieved from hyperspectral reflectance for quantifying seasonal and stressful photosynthetic changes (Lawson et al., 2020). Large seasonal and spatial variations in V_{cmax} have also been observed, especially for diverse crop rotation systems, based on the mapped global canopy V_{cmax} from 11 years of satellite chlorophyll fluorescence records (He et al., 2019). For instance, a higher V_{cmax} of winter wheat was observed in spring than V_{cmax} of maize in the summer in Shandong, China (He et al., 2019), in line with the differences of V_{cmax} for wheat and maize in our estimates (Fig. 2.5b). Nonetheless, until now, studies about directly estimating other photosynthetic parameters, such as T_p and V_{pmax} , or in other major field crops

are scarce. However, methods to retrieve crop N status have been developed previously at different temporal and spatial resolutions from remote and proximal sensing images (Berger et al., 2020). By means of retrieving crop N, these key photosynthetic parameters can be indirectly retrieved from hyperspectral imagery based on our established correlation between photosynthetic parameters and SLN.

With the predicted photosynthetic parameters, photosynthetic capacity at the leaf level can be estimated. After up-scaling to the canopy photosynthesis, together with other growth and development processes, crop growth and its responses to weather, soil and field management can be estimated quantitatively by crop models. Correspondingly, according to actual crop growth status and predicted crop growth, the decisions of the adaptation of N fertiliser to the target yield or to environmental objectives can be made before the date of actual N applications. Thus, the identified relationship between estimated photosynthetic parameters and SLN in our study can serve for large scale photosynthetic parameter estimation by means of N retrieving from remote and proximal sensing images and be supportive the optimisation of N management in smart crop cultivation. This is an area of our ongoing investigation.

2.5 Concluding remarks

We estimated photosynthetic parameters for four major crops, two C_3 species (rice and wheat) and two C_4 species (maize and sorghum), as well as their relationships with SLN. Our results showed that the estimates changed between C_3 and C_4 types, as well as between crops within each type. The significant relationships of these parameters with SLN can be used in crop models for the simulation of photosynthetic process from the leaf to the canopy level. This, combined with the modelling of other growth processes, can help to bridge the gap between crop yield and retrieved crop N status from widely available hyperspectral data, and, therefore, may effectively support smart crop N management.

List of abbreviations

ANCOVA, Analysis of covariance; A_{\max} , maximum rate of light saturated assimilation at ambient CO_2 level; C_a , ambient CO_2 level; C_c , the CO_2 level at the carboxylating sites of Rubisco; C_i , the CO_2 level of substomatal cavity; CCM, CO_2 -concentrating mechanism; F_m' , the maximum fluorescence yield; F_s , the steady-state fluorescence yield; FvCB model, the model of Farquhar, von Caemmerer & Berry (1980); SLN, specific leaf nitrogen content (g N

m⁻²); PEP, phosphoenolpyruvate; PEPc, PEP carboxylase; PSII, Photosystem II; RUE: radiation use efficiency; TPU, triose phosphate utilisation.

References

- Anten, N. P. R., Schieving, F., & Werger, M. J. A. (1995). Patterns of light and nitrogen distribution in relation to whole canopy carbon gain in C₃ and C₄ mono- and dicotyledonous species. *Oecologia*, 101, 504-513.
- Bailey-Serres, J., Parker, J. E., Ainsworth, E. A., Oldroyd, G. E. D., & Schroeder, J. I. (2019). Genetic strategies for improving crop yields. *Nature*, 575, 109-118. doi:10.1038/s41586-019-1679-0
- Berger, K., Verrelst, J., Féret, J.-B., Wang, Z., Woche, M., Strathmann, M., Danner, M., Mauser, W., & Hank, T. (2020). Crop nitrogen monitoring: Recent progress and principal developments in the context of imaging spectroscopy missions. *Remote Sensing of Environment*, 242, 111758. doi:10.1016/j.rse.2020.111758
- Bernacchi, C. J., Pimentel, C., & Long, S. P. (2003). In vivo temperature response functions of parameters required to model RuBP-limited photosynthesis. *Plant, Cell and Environment*, 26, 1419-1430.
- Bernacchi, C. J., Singaas, E. L., Pimentel, C., Portis Jr., A. R., & Long, S. P. (2001). Improved temperature response functions for models of Rubisco-limited photosynthesis. *Plant, Cell and Environment*, 25, 253-259.
- Black Jr., C. C. (1973). Photosynthetic carbon fixation in relation to net CO₂ uptake. *Annual Review of Plant Physiology*, 24, 253-286.
- Boote, K. J., Jones, J. W., & Pickering, N. B. (1996). Potential uses and limitations of crop models. *Agronomy Journal*, 88, 704-716.
- Brautigam, A., Kajala, K., Wullenweber, J., Sommer, M., Gagneul, D., Weber, K. L., Carr, K. M., Gowik, U., Mass, J., Lercher, M. J., Westhoff, P., Hibberd, J. M., & Weber, A. P. (2011). An mRNA blueprint for C₄ photosynthesis derived from comparative transcriptomics of closely related C₃ and C₄ species. *Plant Physiology*, 155, 142-156. doi:10.1104/pp.110.159442
- Byrd, G. T., Sage, R. F., & Brown, R. H. (1992). A comparison of dark respiration between C₃ and C₄ plants. *Plant Physiology*, 100, 191-198.
- Chapman, K. S. R., Berry, J. A., & Hatch, M. D. (1980). Photosynthetic metabolism in bundle sheath cells of the C₄ species *Zea mays*: sources of ATP and NADPH and the contribution of photosystem II. *Archives of Biochemistry and Biophysics*, 202, 330-341.
- Cousins, A. B., Ghannoum, O., von Caemmerer S., & Badger, M. R. (2010). Simultaneous determination of Rubisco carboxylase and oxygenase kinetic parameters in *Triticum aestivum* and *Zea mays* using membrane inlet mass spectrometry. *Plant, Cell and Environment*, 33, 444-452. doi:10.1111/j.1365-3040.2009.02095.x
- de Wit, C. T. (1978). *Simulation of assimilation, respiration and transpiration of crops*. Wageningen, The Netherlands: Pudoc.
- Dubois, J.-J. B., Fiscus, E. L., Booker, F. L., Flowers, M. D., & Reid, C. D. (2007). Optimizing the statistical estimation of the parameters of the Farquhar–von Caemmerer–Berry model of photosynthesis. *New Phytologist*, 176, 402-414. doi:10.1111/j.1469-8137.2007.02182.x

- Ehleringer, J. R., Sage, R. F., Flanagan, L. B., & Pearcy, R. W. (1991). Climate change and the evolution of C₄ photosynthesis. *Trends in Ecology and Evolution*, 6, 95-99.
- Evans, J. R. (1983). Nitrogen and photosynthesis in the flag leaf of wheat (*Triticum aestivum* L.). *Plant Physiology*, 72, 297-302.
- Evans, J. R. (1996). Developmental constraints on photosynthesis: effects of light and nutrition. In N. R. Baker (Eds.), *Photosynthesis and the Environment* (pp. 281-304). Dordrecht, The Netherlands: Kluwer.
- Evans, J. R., & Clarke, V. C. (2019). The nitrogen cost of photosynthesis. *Journal of Experimental Botany*, 70, 7-15. doi:10.1093/jxb/ery366
- Evans, J. R., & von Caemmerer, S. (1996). Carbon dioxide diffusion inside leaves. *Plant Physiology*, 110, 339-346.
- Farquhar, G. D., von Caemmerer, S., & Berry, J. A. (1980). A biochemical model of photosynthetic CO₂ assimilation in leaves of C₃ species. *Planta*, 149, 78-90.
- Field, C., & Mooney, H. A. (1983). Leaf age and seasonal effects on light, water, and nitrogen use efficiency in a California shrub. *Oecologia*, 56, 348-355.
- Flexas, J., Diaz-Espejo, A., Berry, J. A., Cifre, J., Galmes, J., Kaldenhoff, R., Medrano, H., & Ribas-Carbo, M. (2007). Analysis of leakage in IRGA's leaf chambers of open gas exchange systems: quantification and its effects in photosynthesis parameterization. *Journal of Experimental Botany*, 58, 1533-1543. doi:10.1093/jxb/erm027
- Genty, B., Briantais, J.-M., & Baker, N. R. (1989). The relationship between the quantum yield of photosynthetic electron transport and quenching of chlorophyll fluorescence. *Biochimica et Biophysica Acta*, 990, 87-92.
- Ghannoum, O., von Caemmerer, S., Ziska, L. H., & Conroy, J. P. (2000). The growth response of C₄ plants to rising atmospheric CO₂ partial pressure: a reassessment. *Plant, Cell and Environment*, 23, 931-942.
- Griffin, K. L., & Turnbull, M. H. (2013). Light saturated RuBP oxygenation by Rubisco is a robust predictor of light inhibition of respiration in *Triticum aestivum* L. *Plant Biology*, 15, 769-775. doi:10.1111/j.1438-8677.2012.00703.x
- Gu, J., Yin, X., Stomph, T. J., & Struik, P. C. (2014). Can exploiting natural genetic variation in leaf photosynthesis contribute to increasing rice productivity? A simulation analysis. *Plant Cell and Environment*, 37, 22-34. doi:10.1111/pce.12173
- Hank, T. B., Berger, K., Bach, H., Clevers, J. G. P. W., Gitelson, A., Zarco-Tejada, P., & Mauser, W. (2019). Spaceborne imaging spectroscopy for sustainable agriculture: contributions and challenges. *Surveys in Geophysics*, 40, 515-551. doi:10.1007/s10712-018-9492-0
- Harley, P. C., Loreto, F., Di Marco, G., & Sharkey, T. D. (1992a). Theoretical considerations when estimating the mesophyll conductance to CO₂ flux by analysis of the response of photosynthesis to CO₂. *Plant Physiology*, 98, 1429-1436.
- Harley, P. C., Thomas, R. B., Reynolds, J. F., & Strain, B. R. (1992b). Modelling photosynthesis of cotton grown in elevated CO₂. *Plant Cell and Environment*, 15, 271-282.

- He, L., Chen, J. M., Liu, J., Zheng, T., Wang, R., Joiner, J., Chou, S., Chen, B., Liu, Y., Liu, R., & Rogers, C. (2019). Diverse photosynthetic capacity of global ecosystems mapped by satellite chlorophyll fluorescence measurements. *Remote Sensing of Environment*, 232, 111344. doi:10.1016/j.rse.2019.111344
- Herold, A. (1980). Regulation of photosynthesis by sink activity-the missing link. *New Phytologist*, 86, 131-144.
- Hirose, T. (1984). Nitrogen use efficiency in growth of *Polygonum cuspidatum* Sieb. et Zucc. *Annals of Botany*, 54, 695-704.
- Kanai, R., & Edwards, G. E. (1999). The biochemistry of C₄ photosynthesis. In R. F. Sage, & R. K. Monson (Eds.), *C₄ plant biology* (pp. 49-87). Toronto, Canada: Academic Press.
- Kiniry, J. R., Jones, C. A., O'Toole, J. C., Blanchet, R., Cabelguenne, M., & Spanel, D. A. (1989). Radiation-use efficiency in biomass accumulation prior to grain-filling for five grain-crop species. *Field Crops Research*, 20, 51-64.
- Kromdijk, J., Ubierna, N., Cousins, A. B., & Griffiths, H. (2014). Bundle-sheath leakiness in C₄ photosynthesis: a careful balancing act between CO₂ concentration and assimilation. *Journal of Experimental Botany*, 65, 3443-3457. doi:10.1093/jxb/eru157
- Laisk, A., Oja, V., Rasulov, B., Ramma, H., Eichelmann, H., Kasparova, I., Pettai, H., Padu, E., & Vapaavuori, E. (2002). A computer-operated routine of gas exchange and optical measurements to diagnose photosynthetic apparatus in leaves. *Plant Cell and Environment*, 25, 923-943. doi:10.1046/j.1365-3040.2002.00873.x
- Lawlor, D. W. (1995). Photosynthesis, productivity and environment. *Journal of Experimental Botany*, 46, 1449-1461.
- Lawson, T., Bernacchi, C., Driever, S., Pederson, T., Dracup, E., Guan, K., Ainsworth, E., Montes, C. M., Serbin, S., Wu, J., Fu, P., & Meacham-Hensold, K. (2020). Plot-level rapid screening for photosynthetic parameters using proximal hyperspectral imaging. *Journal of Experimental Botany*, 71, 2312-2328. doi:10.1093/jxb/eraa068
- Leegood, R.C., & von Caemmerer, S. (1989). Some relationships between contents of photosynthetic intermediates and the rate of photosynthetic carbon assimilation in leaves of *Zea mays* L. *Planta*, 178, 258-266.
- Li, G., Chen, T., Feng, B., Peng, S., Tao, L., & Fu, G. (2021). Respiration, rather than photosynthesis, determines rice yield loss under moderate high-temperature conditions. *Frontiers in Plant Science*, 12, 678653. doi:10.3389/fpls.2021.678653
- Li, T., Hasegawa, T., Yin, X., Zhu, Y., Boote, K., Adam, M., Bregaglio, S., Buis, S., Confalonieri, R., Fumoto, T., Gaydon, D., Marcaida, M., 3rd, Nakagawa, H., Oriol, P., Ruane, A. C., Ruget, F., Singh, B., Singh, U., Tang, L., Tao, F., Wilkens, P., Yoshida, H., Zhang, Z., & Bouman, B. (2015). Uncertainties in predicting rice yield by current crop models under a wide range of climatic conditions. *Global Change Biology*, 21, 1328-1341. doi:10.1111/gcb.12758
- Li, Y., Gao, Y., Xu, X., Shen, Q., & Guo, S. (2009). Light-saturated photosynthetic rate in high-nitrogen rice (*Oryza sativa* L.) leaves is related to chloroplastic CO₂ concentration. *Journal of Experimental Botany*, 60, 2351-2360. doi:10.1093/jxb/erp127

- Loreto, F., Harley, P. C., Di Marco, G., Sharkey, T. D. (1992). Estimation of mesophyll conductance to CO₂ flux by three different methods. *Plant Physiology*, 98, 1437-1443.
- Loriaux, S. D., Avenson, T. J., Welles, J. M., McDermitt, D. K., Eckles, R. D., Riensche, B., & Genty, B. (2013). Closing in on maximum yield of chlorophyll fluorescence using a single multiphase flash of sub-saturating intensity. *Plant Cell and Environment*, 36, 1755-1770. doi:10.1111/pce.12115
- McDowell, N. G. (2011). Mechanisms linking drought, hydraulics, carbon metabolism, and vegetation mortality. *Plant Physiology*, 155, 1051-1059. doi:10.1104/pp.110.170704
- Monteith, J. L. (1977). Climate and the efficiency of crop production in Britain. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 281, 277-294.
- Mooney, H. A., & Gulmon, S. L. (1982). Constraints on leaf structure and function in reference to herbivory. *BioScience*, 32, 198-206.
- Niinemets, Ü., & Tenhunen, J. D. (1997). A model separating leaf structural and physiological effects on carbon gain along light gradients for the shade-tolerant species *Acer saccharum*. *Plant Cell and Environment*, 20, 845-866.
- Osmond, C. B., Winter, K., & Ziegler, H. (1982). Physiological plant ecology II. In O. L. Lange, P. S. Nobel, C. B. Osmond, & H. Ziegler (Eds.), *Encyclopedia of plant physiology* (pp. 479-547). Berlin, Heidelberg, Germany and New York, The United States: Springer-Verlag.
- Ouyang, W., Struik, P. C., Yin, X., & Yang, J. (2017). Stomatal conductance, mesophyll conductance, and transpiration efficiency in relation to leaf anatomy in rice and wheat genotypes under drought. *Journal of Experimental Botany*, 68, 5191-5205. doi:10.1093/jxb/erx314
- Penning de Vries, F. W. T. (1975). The cost of maintenance processes in plant cells. *Annals of Botany*, 39, 77-92.
- Pingali, P. L. (2012). Green revolution: impacts, limits, and the path ahead. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 12302-12308. doi:10.1073/pnas.0912953109
- Qian, T., Elings, A., Dieleman, J. A., Gort, G., & Marcelis, L. F. M. (2012). Estimation of photosynthesis parameters for a modified Farquhar-von Caemmerer-Berry model using simultaneous estimation method and nonlinear mixed effects model. *Environmental and Experimental Botany*, 82, 66-73. doi:10.1016/j.envexpbot.2012.03.014
- R Core Team. (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.
- Reich, P. B., Tjoelker, M. G., Pregitzer, K. S., Wright, I. J., Oleksyn, J., & Machado, J. L. (2008). Scaling of respiration to nitrogen in leaves, stems and roots of higher land plants. *Ecology Letter*, 11, 793-801. doi:10.1111/j.1461-0248.2008.01185.x
- Retta, M., Yin, X., van der Putten, P. E., Cantre, D., Berghuijs, H. N., Ho, Q. T., Verboven, P., Struik, P. C., & Nicolai, B. M. (2016). Impact of anatomical traits of maize (*Zea mays* L.) leaf as affected by nitrogen supply and leaf age on bundle sheath conductance. *Plant Science*, 252, 205-214. doi:10.1016/j.plantsci.2016.07.013
- Rogers, A. (2014). The use and misuse of $V_{c,max}$ in Earth System Models. *Photosynthetic Research*, 119, 15-29. doi:10.1007/s11220-013-9818-1

- Ryan, M. G. (1995). Foliar maintenance respiration of subalpine and boreal trees and shrubs in relation to nitrogen content. *Plant Cell and Environment*, 18, 765-772.
- Sage, R. F., & Pearcy, R. W. (1987). The nitrogen use efficiency of C₃ and C₄ plants: II. Leaf nitrogen effects on the gas exchange characteristics of *Chenopodium album* (L.) and *Amaranthus retroflexus* (L.). *Plant Physiology*, 84, 959-963.
- Schmitt, M. R., & Edwards, G. E. (1981). Photosynthetic capacity and nitrogen use efficiency of maize, wheat, and rice: a comparison between C₃ and C₄ photosynthesis. *Journal of Experimental Botany*, 32, 459-466.
- Sharkey, T. D., Bernacchi, C. J., Farquhar, G. D., & Singsaas, E. L. (2007). Fitting photosynthetic carbon dioxide response curves for C₃ leaves. *Plant Cell and Environment*, 30, 1035-1040. doi:10.1111/j.1365-3040.2007.01710.x
- Sinclair, T. R., & Horie, T. (1989). Leaf nitrogen, photosynthesis, and crop radiation use efficiency: A review. *Crop Science*, 29, 90-98.
- Sinclair, T. R., & Muchow, R. C. (1999). Radiation use efficiency. *Advances in Agronomy*, 65, 215-265. doi:10.1016/s0065-2113(08)60914-1
- Sinclair, T. R., & Rawlins, S. L. (1993). Inter-seasonal variation in soybean and maize yields under global environmental change. *Agronomy Journal*, 85, 406-409.
- Singer, S. D., Foroud, N. A., & Laurie, J. D. (2019). Molecular improvement of grain: Target traits for a changing world. In P. Ferranti, E. M. Berry, J. R. Anderson (Eds.), *Encyclopedia of Food Security and Sustainability* (pp. 545-555). Oxford, The United Kingdom: Elsevier.
- Tofanello, V. R., Andrade, L. M., Flores-Borges, D. N. A., Kiyota, E., Mayer, J. L. S., Creste, S., Machado, E. C., Yin, X., Struik, P. C., & Ribeiro, R. V. (2021). Role of bundle sheath conductance in sustaining photosynthesis competence in sugarcane plants under nitrogen deficiency. *Photosynthetic Research*, 149, 275-287. doi:10.1007/s11120-021-00848-w
- Valentini, R., Epron, D., de Angelis, P., Matteucci, G., & Dreyer, E. (1995). *In situ* estimation of net CO₂ assimilation, photosynthetic electron flow and photorespiration in Turkey oak (*Q. cerris* L.) leaves: diurnal cycles under different levels of water supply. *Plant Cell and Environment*, 18, 631-640.
- von Caemmerer, S. (2000). *Biochemical models of leaf photosynthesis*. Melbourne, Australia: CSIRO Publishing.
- von Caemmerer, S., & Evans, J. R. (2015). Temperature responses of mesophyll conductance differ greatly between species. *Plant Cell and Environment*, 38, 629-637. doi:10.1111/pce.12449
- von Caemmerer, S., Evans, J. R., Hudson, G. S., & Andrews, T. J. (1994). The kinetics of ribulose-1,5-bisphosphate carboxylase/oxygenase in vivo inferred from measurements of photosynthesis in leaves of transgenic tobacco. *Planta*, 195, 88-97.
- von Caemmerer, S., & Furbank, R. T. (1999). Modelling C₄ photosynthesis. In R. F. Sage, & R. K. Monson (Eds.), *C₄ plant biology* (pp. 173-211). Toronto, Canada: Academic Press.
- Walter, A., Finger, R., Huber, R., & Buchmann, N. (2017). Opinion: Smart farming is key to developing sustainable agriculture. *Proceedings of the National Academy of Sciences of the United States of America*, 114, 6148-6150. doi:10.1073/pnas.1707462114

- Warren, C. R. (2008). Does growth temperature affect the temperature responses of photosynthesis and internal conductance to CO₂? A test with *Eucalyptus regnans*. *Tree Physiology*, 28, 11-19. doi:10.1093/treephys/28.1.11
- Weber, A. P. M., & von Caemmerer, S. (2010). Plastid transport and metabolism of C₃ and C₄ plants—comparative analysis and possible biotechnological exploitation. *Current Opinion in Plant Biology*, 13, 256-264. doi:10.1016/j.pbi.2010.01.007
- Wong, S. C. (1979). Elevated atmospheric partial pressure of CO₂ and plant growth. I. Interactions of nitrogen nutrition and photosynthetic capacity in C₃ and C₄ plants. *Oecologia*, 44, 68-74.
- Yin, X., Busch, F. A., Struik, P. C., & Sharkey, T. D. (2021a). Evolution of a biochemical model of steady-state photosynthesis. *Plant Cell and Environment*, 44, 2811-2837. doi:10.1111/pce.14070
- Yin, X., & Struik, P. C. 2012. Mathematical review of the energy transduction stoichiometries of C₄ leaf photosynthesis under limiting light. *Plant Cell and Environment*, 35, 1299-1312. doi:10.1111/j.1365-3040.2012.02490.x
- Yin, X., & Struik, P. C. (2017). Can increased leaf photosynthesis be converted into higher crop mass production? A simulation study for rice using the crop model GECROS. *Journal of Experimental Botany*, 68, 2345-2360. doi:10.1093/jxb/erx085
- Yin, X., Struik, P. C., & Goudriaan, J. (2021b). On the needs for combining physiological principles and mathematics to improve crop models. *Field Crops Research*, 271, 108254. doi:10.1016/j.fcr.2021.108254
- Yin, X., Struik, P. C., Romero, P., Harbinson, J., Evers, J. B., van der Putten, P. E. L., & Vos, J. (2009). Using combined measurements of gas exchange and chlorophyll fluorescence to estimate parameters of a biochemical C₃ photosynthesis model: a critical appraisal and a new integrated approach applied to leaves in a wheat (*Triticum aestivum*) canopy. *Plant Cell and Environment*, 32, 448-464. doi:10.1111/j.1365-3040.2009.01934.x
- Yin, X., Sun, Z., Struik, P. C., & Gu, J. (2011a). Evaluating a new method to estimate the rate of leaf respiration in the light by analysis of combined gas exchange and chlorophyll fluorescence measurements. *Journal of Experimental Botany*, 62, 3489-3499. doi:10.1093/jxb/err038
- Yin, X., Sun, Z., Struik, P. C., van der Putten, P. E. L., van Ieperen, W., & Harbinson, J. (2011b). Using a biochemical C₄ photosynthesis model and combined gas exchange and chlorophyll fluorescence measurements to estimate bundle-sheath conductance of maize leaves differing in age and nitrogen content. *Plant Cell and Environment*, 34, 2183-2199. doi:10.1111/j.1365-3040.2011.02414.x
- Yoshida, S., & Coronel, V. (1976). Nitrogen nutrition, leaf resistance, and leaf photosynthetic rate of the rice plant. *Soil Science and Plant Nutrition*, 22, 207-211.

Supplementary Appendix A in Chapter 2

g_m model and the solution of A_c and A_j for C_3 species

Given that mesophyll diffusion conductance was identified as zero if light approaches zero (see Results), we used a simplified version of the variable- g_m model of Yin et al. (2009) without the terms for residual mesophyll diffusion conductance:

$$g_m = \delta(A + R_d)/(C_c - \Gamma_*) \quad (\text{A2.1})$$

With this simplified form, parameter δ represents the carboxylation to mesophyll resistance ratio (Yin et al., 2020).

Combining Eqn (A2.1) with Eqn (2.1) and replacing C_c with $(C_i - A_c/g_m)$ or $(C_i - A_j/g_m)$, correspondingly, and then solving the quadratic equation for A_c and A_j gives:

$$A_c \text{ or } A_j = (-b - \sqrt{b^2 - 4ac})/(2a) \quad (\text{A2.2})$$

where $a = x_2 + \Gamma_* + \delta(C_i + x_2)$

$$b = -\{(x_2 + \Gamma_*)(x_1 - R_d) + \delta(C_i + x_2)(x_1 - R_d) + \delta[x_1(C_i - \Gamma_*) - R_d(C_i + x_2)]\}$$

$$c = \delta(x_1 - R_d)[x_1(C_i - \Gamma_*) - R_d(C_i + x_2)]$$

where x_1 and x_2 are defined in Eqn (2.1) in the main text.

Supplementary Appendix B in Chapter 2

The solution of A_{EE} , A_{ET} , A_{TE} , and A_{TT} for C_4 species

Combining Eqns (2.4-2.7) in the main text can solve the value A for each of these limitation combinations.

(1) If V_p is limited by the activity of PEPc (see Eqn (2.5a)), A_{EE} and A_{ET} can be solved as one of the roots of a standard cubic equation, which is suitable for calculating either A_{EE} or A_{ET} under any combinations of C_i , I_{inc} and O_i :

$$A_{EE} \text{ or } A_{ET} = -2\sqrt{Q} \cos(\psi/3) - p/3 \quad (B2.1)$$

where $Q = (p^2 - 3q)/9$

$$\psi = \cos^{-1} \left(U/\sqrt{Q^3} \right)$$

$$U = (2p^3 - 9pq + 27r)/54$$

in which, $p = m/(g_m + g_{bs} - x_2 g_m \alpha / 0.047)$

$$q = n/(g_m + g_{bs} - x_2 g_m \alpha / 0.047)$$

$$r = o/(g_m + g_{bs} - x_2 g_m \alpha / 0.047)$$

and m , n and o are expressed as

$$m = d - (x_3 + x_2 O_i) g_m g_{bs} + (R_d - x_1)(g_m + g_{bs}) - (x_1 \gamma_* g_m + x_2 R_d g_m - x_2 k / g_{bs}) \alpha / 0.047$$

$$n = f + (x_3 + x_2 O_i) k + d(R_d - x_1) - g_m g_{bs} [x_1 \gamma_* O_i + R_d (x_3 + x_2 O_i)] \\ + (x_1 \gamma_* + x_2 R_d) k \alpha / (0.047 g_{bs})$$

$$o = R_d [f + (x_3 + x_2 O_i) k] - x_1 (f - k \gamma_* O_i)$$

and d , f and k are expressed as

$$d = g_m [R_m - V_{pmax} - C_i (g_m + 2g_{bs}) - K_p (g_m + g_{bs})]$$

$$f = g_m^2 [C_i V_{pmax} + (C_i + K_p)(g_{bs} C_i - R_m)]$$

$$k = g_m^2 g_{bs} (C_i + K_p)$$

(2) If V_p is limited by electron transport (see Eqn (2.5b)), A_{TE} and A_{TT} can be solved as one of the roots of a generated quadratic equation:

$$A_{TE} \text{ or } A_{TT} = (-b + \sqrt{b^2 - 4ac}) / (2a) \quad (B2.2)$$

where $a = x_2 g_m \alpha / 0.047 - g_m - g_{bs}$

$$b = g_m (C_i g_{bs} + V_p - R_m) + (x_3 + x_2 O_i) g_m g_{bs} + (x_1 \gamma_* + x_2 R_d) g_m \alpha / 0.047 \\ + (g_m + g_{bs})(x_1 - R_d)$$

$$c = -g_m (C_i g_{bs} + V_p - R_m)(x_1 - R_d) + g_m g_{bs} [x_1 \gamma_* O_i + R_d (x_3 + x_2 O_i)]$$

Supplementary Table in Chapter 2

Table S2.1 The linear regression equations between estimated photosynthetic parameters and specific leaf nitrogen content for C₃ and C₄ crops when data for C₃ (rice and wheat) or for C₄ species (maize and sorghum) are pooled for regression. The linear regressions were conducted separately for different O₂ conditions, if applicable.

Parameter		C ₃ species	C ₄ species
A_{\max}	21% O ₂	$y^{**} = 12.587x + 7.854$	$y^{***} = 29.002x + 2.036$
	2% O ₂	$y^{**} = 17.645x + 9.396$	$y^{***} = 29.976x + 2.495$
s_{c3} or s_{c4}		$y^{ns} = 0.049x + 0.471$	$y^{**} = 0.106x + 0.236$
R_d		$y^* = 0.374x + 0.583$	$y^{ns} = 1.216x + 0.362$
Φ_{2LL}	21% O ₂	$y^{ns} = 0.032x + 0.620$	$y^{ns} = 0.029x + 0.574$
	2% O ₂	$y^{ns} = 0.030x + 0.595$	$y^{ns} = 0.016x + 0.560$
κ_{2LL} or $\kappa_{2LL,atp}^1$	21% O ₂	$y^* = 0.049x + 0.290$	$y^* = 0.123x + 0.222$
	2% O ₂	$y^{ns} = 0.047x + 0.278$	$y^{ns} = 0.113x + 0.217$
g_m		$y^* = 0.089x + 0.064$	NU
g_{bs}		NA	$y^{ns} = 3.656x - 1.005$
J_{\max} or $J_{\max,atp}^1$	21% O ₂	$y^{**} = 142.8x + 44.1$	$y^{**} = 250.6x - 36.2$
	2% O ₂	$y^{**} = 90.3x + 36.4$	$y^{***} = 248.9x - 37.8$
V_{cmax}		$y^{**} = 128.7x + 31.2$	$y^* = 40.7x + 10.1$
T_p		$y^{**} = 4.956x + 6.300$	NA
V_{pmax}		NA	$y^{**} = 98.3x - 13.2$

NA, not applicable; NU, not used here, as g_m considered as the constant in C₄ crops in this study.

¹ s_{c3} , κ_{2LL} and s_{c4} for C₃ crops; s_{c4} , $\kappa_{2LL,atp}$ and $J_{\max,atp}$ for C₄ crops.

^{ns}, *, **, *** represent the significance level of regressed linear equations ($p > 0.05$, $p < 0.05$, $p < 0.01$, $p < 0.001$, respectively).

Supplementary Figures in Chapter 2

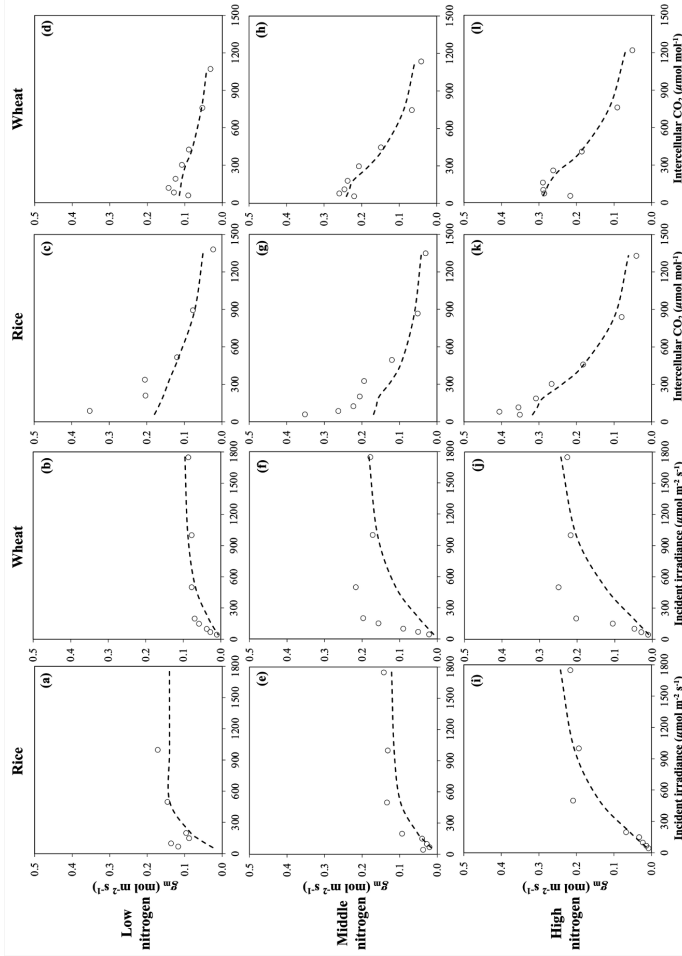


Fig. S2.1 The response of estimated mesophyll conductance g_m to incident irradiance and intercellular CO_2 of rice and wheat with different levels of nitrogen supply. g_m represented by circles are calculated using the variable J method (Harley et al., 1992), following the equation: $g_m = A / \{C_i - F_s [J_s + 8(A + R_d)] / [J_s + 4(A + R_d)]\}$, based on the average A , C_i and $J_s = s_{c3} I_{inc} \Delta F / F_i$ of four replicated leaves and obtained at 21% O_2 . The dashed lines represent the calculated g_m according to Eqn (2.11) using parameters fitted to the model, in which estimated δ under low, middle and high nitrogen supply was 2.063, 0.786 and 0.846 for rice, and 0.463, 0.700 and 0.640 for wheat, respectively.

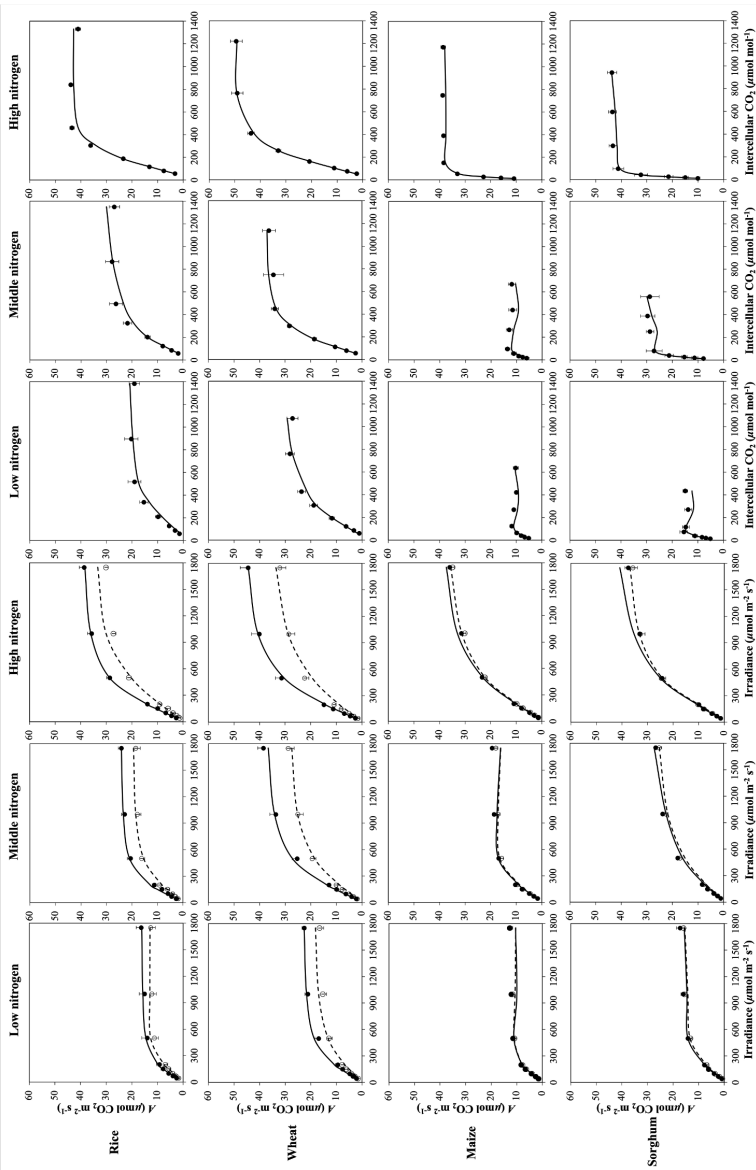


Fig. S2.2 Measured (points) and modelled (curves) irradiance and CO₂ response curves of net CO₂ assimilation rate A , at 2% (filled circles and full lines) and 21% (open circles and dash lines) of different nitrogen supplies. Data points are shown as means (with standard errors) of three or four replicate leaves. Curves are drawn from model fitted values.

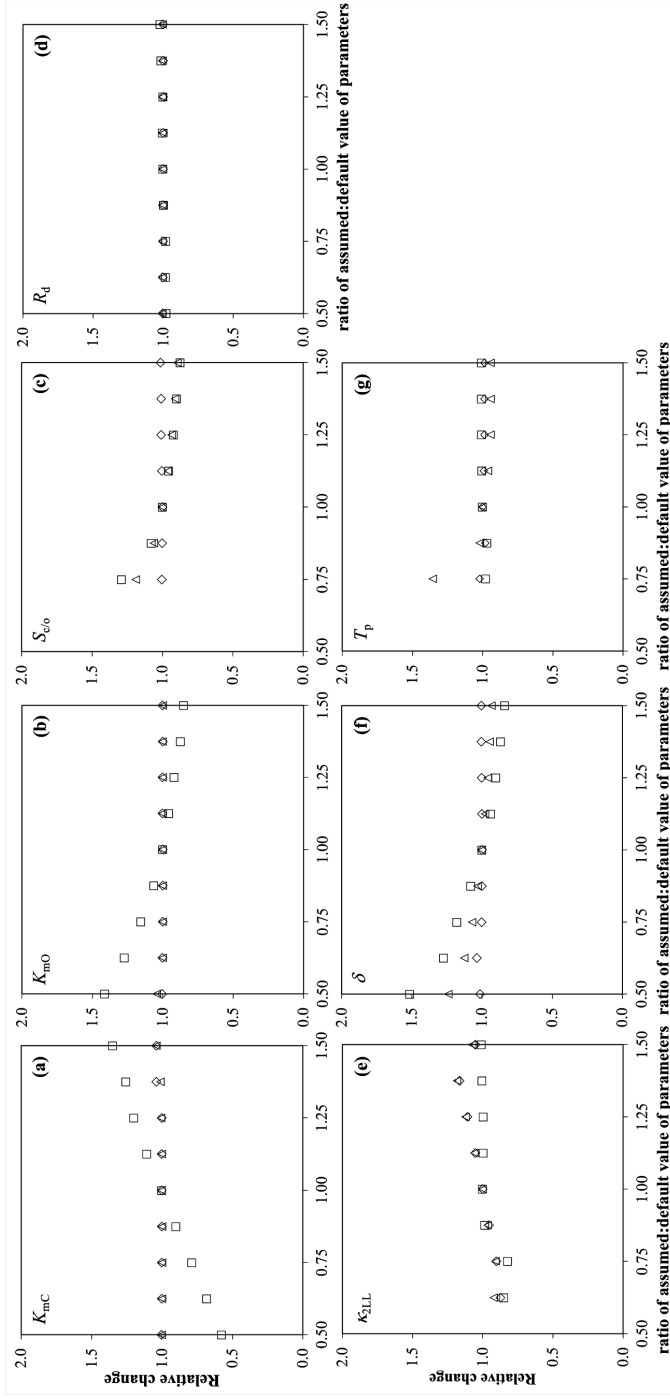


Fig. S2.3 Relative changes in the estimated maximum rate of Rubisco carboxylation V_{cmax} (open squares), maximum rate of linear electron transport J_{max} at 2% O_2 (open diamonds) and J_{max} at 21% O_2 (open triangles) in C_3 crops in response to a change of pre-set model parameters, K_{mC} (a), K_{mO} (b) and S_{eO} (c) and estimated parameters, R_d (d), K_{2LL} (e), δ (f) and T_p (g). The relative changes shown in this figure were the average value of that of rice and wheat.

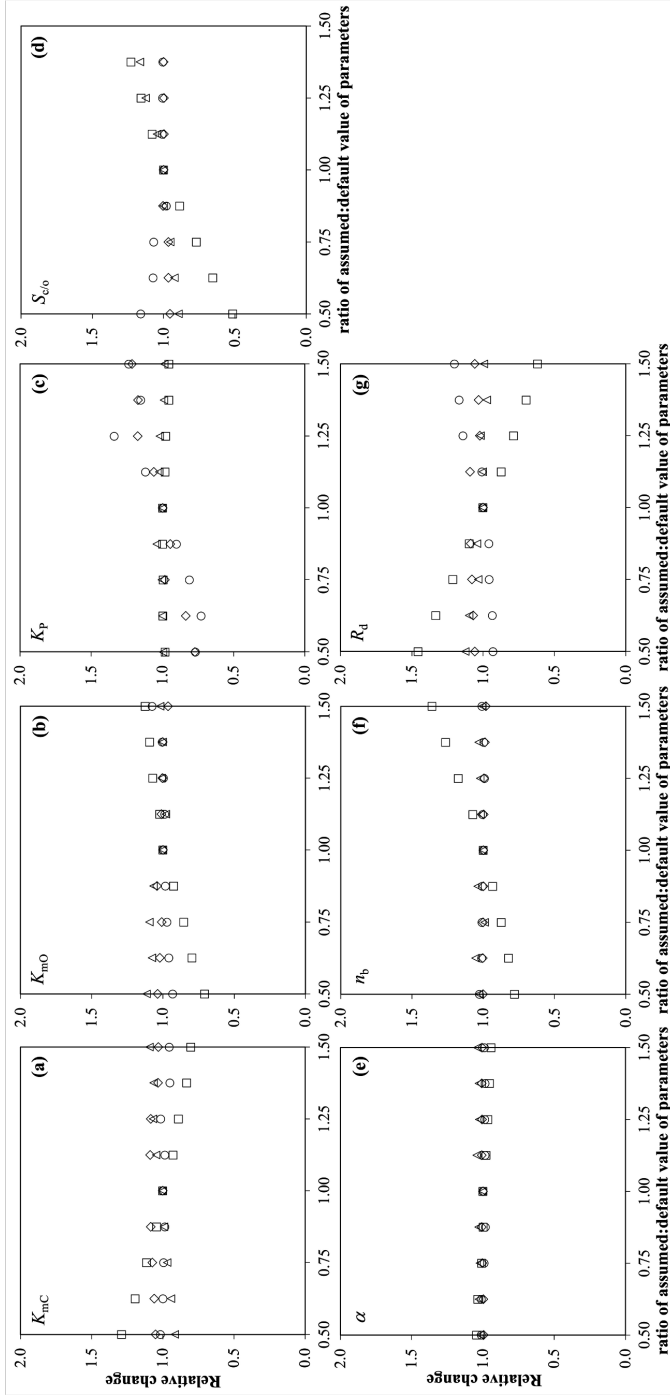


Fig. S2.4 Relative changes in the estimated mesophyll conductance g_m (open circles), the slope of the bundle sheath conductance-leaf nitrogen content linearity χ_{obs} (open squares), maximum rate of PEP carboxylation V_{max} (open diamonds) and maximum rate of Rubisco carboxylation V_{cmax} (open triangles) in C_4 crops in response to a change in pre-set model parameters, K_{mC} (a), K_{mO} (b), K_P (c), S_{cO} (d), α (e) and n_b (f), and estimated parameter R_d (g). The relative changes shown in this figure were the average value of that of maize and sorghum.

References

- Harley, P. C., Loreto, F., Di Marco, G., & Sharkey, T. D. (1992). Theoretical considerations when estimating the mesophyll conductance to CO₂ flux by analysis of the response of photosynthesis to CO₂. *Plant Physiology*, 98, 1429-1436.
- Yin, X., Struik, P. C., Romero, P., Harbinson, J., Evers, J. B., van der Putten, P. E. L., & Vos, J. (2009). Using combined measurements of gas exchange and chlorophyll fluorescence to estimate parameters of a biochemical C₃ photosynthesis model: a critical appraisal and a new integrated approach applied to leaves in a wheat (*Triticum aestivum*) canopy. *Plant Cell and Environment*, 32, 448-464. doi:10.1111/j.1365-3040.2009.01934.x
- Yin, X., van der Putten, P. E. L., Belay, D., & Struik, P. C. (2020). Using photorespiratory oxygen response to analyse leaf mesophyll resistance. *Photosynthetic Research*, 144, 85-99. doi:10.1007/s11120-020-00716-z

Chapter 3

Estimating leaf and canopy nitrogen contents in major field crops across the growing season from hyperspectral images using nonparametric regression

Dong Wang^{1,2}, Paul C. Struik¹, Lei Liang² and Xinyou Yin¹

¹ Centre for Crop Systems Analysis, Department of Plant Sciences, Wageningen University & Research, 6700 AK Wageningen, The Netherlands

² Shanghai Lankuaikei Technology Development Co. Ltd., No. 888 Huanhu West 2nd Road, Pudong New District, Shanghai, China

Abstract

Estimating leaf nitrogen (N) status is crucial for site- and time-specific crop N management, and can be accomplished more routinely than ever before with the advent of hyperspectral imaging techniques. We conducted field experiments with different nitrogen supply for rice, wheat and maize, in China, in which three types of hyperspectral features were extracted, including canopy reflectance (Ref), vegetation indices (VIs), and texture information (Tex). These features as well as crop development stage (DS) were applied to estimate crop N parameters, using five nonparametric regression algorithms: Partial Least Squares Regression (PLSR), Support Vector Regression (SVR), Random Forest Regression, Deep Neural Network, and Convolution Neural Network. The performance of PLSR and SVR models was more robust than that of the others and could be improved by incorporating the combined feature set RefVIsTex, although there was no further improvement when also incorporating DS. The prediction of the mass-based leaf N trait, leaf N concentration (LNC), was better than that of the area-based trait, specific leaf N (SLN). The models also predicted specific leaf area (SLA) better than its reciprocal, specific leaf weight. Values of SLN were better predicted via an indirect method (predicted via SLA; denoted as SLN_{sla}) than via the direct method (SLN_{dir}). However, when upscaled to canopy, the predicted canopy leaf-N content (N_{leaves}) using SLN_{dir} agreed better with measured N_{leaves} than that using SLN_{sla} , and even better than the direct prediction $N_{leaves,dir}$ in rice and maize. These results were discussed in view of coupling the predicted leaf and canopy N traits with dynamic crop growth models that can be used for optimising field N management in sustainable agricultural production.

Keywords: Major crop, nonparametric regression algorithm, hyperspectral image, leaf nitrogen trait, crop nitrogen management

3.1 Introduction

The security and sustainability of modern agricultural production are at risk. While crop production has increased 50% since 2000, global agricultural land area decreased by 2% and the overall fertiliser use was 40% higher in 2018 than in 2000, while 52% of the increase was nitrogen (N) (FAO, 2020). Fertilisation with N is needed as it strongly influences crop growth, production and quality (Beeckman et al., 2018; Lemaire et al., 2008) while soils often cannot supply enough N to meet the demands of modern varieties to achieve high yields. Crop N uptake, the N absorption by the roots, depends on soil N availability and crop N demand. Crop demand changes during the vegetative and reproductive phases, while N movement and reallocating within the crop occur as well (Ohyama, 2010). Based on the understanding of these and other physiological processes, numerous dynamic crop models have been developed (e.g., Kropff et al., 1994; Yin and Struik, 2017), which can simulate responses of crop growth to N fertilisation. In general, N fertilisation can increase crop yield. However, in view of the cost of N fertilisers and the diminishing return of productivity (yield gain per unit N-fertiliser diminishes with increasing application of N), the amount of fertilisers applied to crops may be excessive (Schröder et al., 2000; Skiba, 2014). Moreover, severe environmental issues have arisen because of N losses, through volatilisation (Davidson, 2009), leaching (Padilla et al., 2018) and runoff (Zeng et al., 2021). Thus, it is necessary to guide farmers to perform a sustainable site- and time-specific N-fertiliser management (Weiss et al., 2020). To this end, monitoring crop N status in time and linking this with crop models to predict crop growth more accurately (Jin et al., 2018) may allow the development of effective strategies for smart crop N management.

Monitoring crop N status via destructive sampling and biophysical chemical analyses of crop N is costly and time consuming. Instead, remote sensing has been developed as a powerful tool to monitor crop N status. Although multispectral satellite images tend to be temporally frequent, hyperspectral satellite images suit better for crop phenotyping because of the higher spectral resolution (Marshall and Thenkabail, 2015). However, hyperspectral satellite images cannot be acquired in time: delivery always takes weeks, while crop N status rapidly changes especially during the vegetative phase, like from tillering to panicle-initiation for rice, during which fertilisers are applied mostly. Consequently, hyperspectral sensors equipped on an Unmanned Aerial Vehicle (UAV) serve with a high versatility for meeting the requirements on spectral, spatial, and temporal resolution (Homolová et al., 2013). Of the studies on predicting

N from hyperspectral images in an agricultural context, more than two thirds are about rice, wheat and maize (Berger et al., 2020), mirroring the large agricultural areas around the globe cropped with these crops (Khoury et al., 2014).

For site-specific field N management, the monitoring of the aboveground leaf-N content in a canopy (N_{leaves}) is crucial as it determines canopy photosynthetic capacity and thus crop productivity (Peng et al., 1995). In crop models, canopy photosynthesis is calculated as the integration of photosynthetic rates of individual leaves in a canopy, while leaf photosynthesis is often related to leaf N status (Boote et al., 1998; Kropff et al., 1994; Yin & van Laar 2005). To parameterise this relationship, data for leaf photosynthesis as measured by gas exchange are used, which are expressed on a leaf-area basis (Evans, 1983). It follows that leaf-N parameters used to model leaf photosynthesis are also expressed on a leaf area basis. Indeed, specific leaf N content (SLN, g N m^{-2}) has been widely used as a typical leaf-N parameter in crop models to calculate leaf photosynthesis rate (Evans, 1983; Kropff et al., 1994; Wang et al., 2022; Yin et al., 2009). Thus, predicting crop N status (e.g., based on SLN) from UAV-based hyperspectral imaging and linking this prediction with crop modelling can guide farmers for better crop N management.

Investigating SLN is also supported by the expectation that the light reflected or transmitted by a leaf is considered to interact within the whole leaf thickness in spectral domains of absorption, and thus, the reflectance and transmittance are supposed to be more directly sensitive to the variation in the area-based content of absorbing constituents, like SLN, but rather less to the mass-based concentration, leaf N concentration (LNC, g g^{-1}) (Baret and Fourty, 1997). However, it is surprising that the estimation of SLN is still rare (Berger et al., 2020) whereas there are many studies regarding the prediction of LNC from remote sensing images (e.g., Moharana and Dutta, 2016; Raj et al., 2021). Very few studies have compared predictions of SLN and LNC. For instance, Ecartot et al. (2013) showed that the prediction of LNC in wheat leaf samples collected from a broad range of genotypes (wild and cultivated forms) had better accuracy than that of SLN. However, Li et al. (2018) found that SLN in rice and wheat had better correlation than LNC with the determined features in the shortwave infrared region (1000 to 2500 nm). Given that these few studies already showed inconsistencies, predictions of SLN and LNC should be further evaluated and compared, especially across various crop species.

On the other hand, SLN can be calculated from LNC, using the leaf thickness parameter specific leaf weight (SLW, g m^{-2}) (Baret and Fourty, 1997), or its reciprocal, specific leaf area

(SLA, $\text{m}^2 \text{g}^{-1}$). This implies that, while SLN can be directly predicted, it can also be indirectly estimated based on the prediction of LNC and SLW or SLA. Note that SLA is a parameter that can inherently cause differences in photosynthetic N use efficiency across species (Poorter and Evans, 1998); therefore, predicting SLA may generate additional information. While it has been reported that SLN in wheat was better predicted by the indirect approach than by the direct method (Ecarnot et al., 2013), a comparison of performance of indirect and direct approaches in predicting SLN is yet to be made for more species. Similarly, the performance of indirectly predicted N_{leaves} by upscaling SLN and LNC, compared with that of the directly predicted N_{leaves} , should be investigated as well. Kattenborn et al. (2019) indicated that the area-based indicators of leaf pigment are supposed to be more suitable for upscaling from leaf to canopy level, since upscaling area-based leaf N only requires an estimate of the canopy state variable leaf area index (LAI), as is done in crop models (Kropff et al., 1994; Yin and Struik, 2017). Nevertheless, the prediction of area- and mass-based leaf N for upscaling from leaf to canopy level need to be further quantified.

Methodologies relating to N prediction from remote sensing have been developed since the pioneering work by Thomas and Oerther (1972). Generally, predicting methods can be classified into four categories: parametric regression methods like methods based on vegetation indices (VIs), nonparametric regression methods including linear and nonlinear regression algorithms, physical methods like inversion of radiative transfer models, and hybrid regression methods by combining physically based methods with nonparametric regression methods (Verrelst et al., 2019). Nonparametric regression methods, which explore the direct link between target traits and the given spectral data, are receiving more attention than the other three methods in the recent decade (Berger et al., 2020). Although regression algorithms like Artificial Neural Networks (ANN) have been employed in recent studies (Liu et al., 2016; Yao et al., 2015), deep learning (DL)-based regression methods, such as Deep Neural Network (DNN) and Convolution Neural Network (CNN), are still under-utilised for crop N prediction based on canopy-level hyperspectral reflectance (Fu et al., 2020). Given that DNN has shown more potential than other regression methods in combining different types of features, like canopy spectral, VIs and texture information, for predicting soybean yield (Maimaitijiang et al., 2020), predicting leaf and canopy N by DL-based regression algorithms should be explored.

Spectral and textural features are fundamental pattern elements in imagery interpretation. The full spectrum has been widely used in chemometrics (Atzberger et al., 2010), while the spectrum within visible (400 to 700 nm) and near infrared (NIR, 800 to 1300 nm) regions are

frequently reported for crop N estimation (Homolová et al., 2013). For instance, as a chlorophyll molecule contains four N atoms, there is an association between N supply and reflectance of the visible and also the so-called red-edge, located between visible and NIR and defined as the position of the sharp change in leaf reflectance between 680 to 750 nm (e.g. Horler et al., 1983; Berger et al., 2020). Serbin et al. (2012) suggested that visible and NIR spectra data could be promising raw data to establish the robust calibration that allows to infer some leaf parameters such as LNC and its temporal variation. Spectral information has also been widely used in numerous VIs. While spectral features describe the variations in various bands of an electromagnetic spectrum, textural features contain the information about the spatial characteristics of canopy architecture within a band (Colombo, 2003; Haralick et al., 1973). Fusing canopy texture information together with spectral features is gaining momentum in phenotyping crop traits such as LAI (Liu et al., 2021), biomass (Zheng et al., 2018b), and yield (Maimaitijiang et al., 2020). Zheng et al. (2018a) showed that LNC in rice was better predicted after incorporating texture information into VIs by using stepwise multiple linear regression. Thus, texture features should also be taken into consideration in predicting SLN and N_{leaves} . Especially, taking advantage of high spatial resolution of the UAV-acquired hyperspectral images (the ground pixel size is usually less than a meter), texture information is supposed to be one type of the important features in site-specific field N management.

In this study, the prediction of area- and mass-based leaf N traits from hyperspectral images will be conducted by combining canopy spectral, VIs and texture information by nonparametric regression algorithms for three major crops (rice, wheat, maize). Our objectives are: 1) to explore the utilisation of the combinations of different feature types and nonparametric regression algorithms in the prediction of leaf N traits at leaf level, SLN and LNC, and at canopy level, N_{leaves} , in different crop species; 2) to compare the performance in estimating N_{leaves} between directly predicted N_{leaves} and its values upscaled from SLN and LNC; and 3) to identify the most robust nonparametric regression algorithms and the most effective pathways (direct vs. indirect) that can be applied in obtaining crop N information of major crops. Achieving these objectives would help monitor crop N status that can be linked with crop models for more accurately predicting crop growth and guide N-fertiliser management in the context of sustainable precision agriculture.

3.2 Materials and methods

3.2.1 Experimental design

Field experiments with different N rates from deficient to excessive were conducted in China: rice experiment was in Chongming, Shanghai, while maize and winter wheat experiments were in Luohe and Zhoukou, respectively, Henan province (Fig. 3.1). Treatments for each experiment were arranged as a randomised block design. There were four replications (four blocks) in the rice experiment, and within each block, six N rates, varying from 0 to 320 kg N ha⁻¹, were applied to rice (cv. Nanjing 46), planted on 4 June 2020 with a row spacing of 20 cm and a within-row plant spacing of 2-3 cm. The plot size was 6 × 30 m. Experiments of maize (cv. Zhengdan 958) and winter wheat (cv. Bainong 4199) were designed with eight blocks, and in each block seven N rates, varying from 0 to 400 kg N ha⁻¹, were applied. Each plot of maize was 10 × 20 m and each plot of wheat was 9 × 18 m (Fig. 3.1). Maize was sown on 9 June 2020 with 60 cm row spacing and 25 cm plant spacing in the row. Winter wheat was planted after harvesting maize on 24 October 2020 with a row spacing of 21 cm and a within-row plant spacing of 1-2 cm. Urea N fertiliser (N, 46%) was split-applied according to the rate and the time schedule described in Table S3.1. Based on the local practice, for each experiment, sufficient phosphate (112.5, 120 and 75 kg P₂O₅ ha⁻¹ for rice, maize and winter wheat, respectively) and potash (112.5, 120 and 105 kg K₂O ha⁻¹ for rice, maize and winter wheat) fertilisers were applied to prevent phosphorus and potassium deficiencies. Irrigation, and pest, disease and weed control were the same for all treatments within a crop species and followed local standard practices. Figure 3.2 presents a flow chart detailing the overall procedure used in this study. The acronyms of the targeted leaf traits at the leaf and canopy level are summarised in Table 3.1.

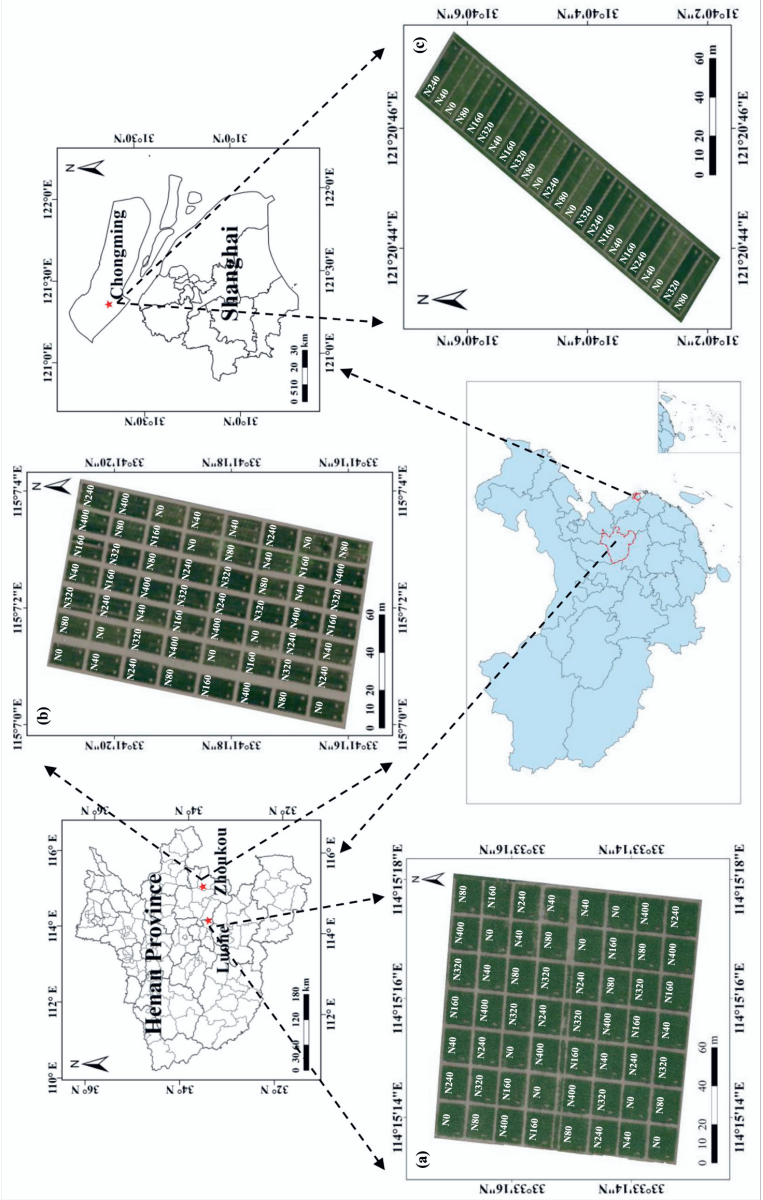


Fig. 3.1 Study sites of the field experiments of maize (a), wheat (b) and rice (c). The true colour images (Red: 638 nm, Green: 550 nm, Blue: 470 nm) at the stem-elongating stage are illustrated here. N0, N40, N80, N160, N240, N320 and N400 denote different nitrogen (N) application rates (see Table S3.1 for details).

Table 3.1 List of acronyms of leaf traits at the leaf and canopy level with their definitions or calculations, and units.

Acronym	Definition or calculation	Unit
LAI	Leaf area index	$\text{m}^2 \text{ leaf } (\text{m}^2 \text{ ground})^{-1}$
LNC	Leaf nitrogen concentration	$\text{g N } (\text{g dry matter})^{-1}$
SLA	Specific leaf area	$\text{m}^2 \text{ leaf } (\text{g dry matter})^{-1}$
SLN	Specific leaf nitrogen content	$\text{g N } (\text{m}^2 \text{ leaf})^{-1}$
SLN_{dir}	SLN directly predicted from hyperspectral images	$\text{g N } (\text{m}^2 \text{ leaf})^{-1}$
SLN_{sla}	SLN indirectly predicted as the predicted LNC divided by the predicted SLA	$\text{g N } (\text{m}^2 \text{ leaf})^{-1}$
SLW	Specific leaf weight ($= 1/\text{SLA}$)	$\text{g dry matter } (\text{m}^2 \text{ leaf})^{-1}$
N_{leaves}	Leaf nitrogen content at the canopy level	$\text{g N } (\text{m}^2 \text{ ground})^{-1}$
$N_{\text{leaves,dir}}$	N_{leaves} directly predicted from hyperspectral images	$\text{g N } (\text{m}^2 \text{ ground})^{-1}$
$N_{\text{leaves,SLNdir}}$	N_{leaves} indirectly predicted as the product of SLN_{dir} and the predicted LAI	$\text{g N } (\text{m}^2 \text{ ground})^{-1}$
$N_{\text{leaves,SLNsla}}$	N_{leaves} indirectly predicted as the product of SLN_{sla} and the predicted LAI	$\text{g N } (\text{m}^2 \text{ ground})^{-1}$

3.2.2 Data acquisition

3.2.2.1 Canopy leaf trait measurements

Crops were destructively sampled from all experimental plots on six dates (Table 3.2). At each sampling date, aboveground plants were cut within an area of 2.25, 1.0, and 2.4 m² in each plot for rice, wheat and maize, respectively. The total fresh samples were weighted immediately and ca. 20% of the biomass was used as subsamples to be dissected into component plant parts, including green leaves, stems, and grains. Yellow leaves were not included. LAI was measured on the fresh green leaves using a LI-3100C Area Meter (Li-Cor, Lincoln, NE, USA). Then, green leaves were weighed after being oven-dried at 70 °C to constant weight. All green leaf samples were mixed and ground, and then stored for chemical testing. LNC, expressed on the basis of leaf dry weight, was determined by the Micro-Keldjahl method. N_{leaves} was calculated as the product of LNC and leaf weight per unit ground area. SLW was calculated as the ratio of leaf weight to LAI and SLA was calculated as 1/SLW. SLN was calculated as the product of LNC and SLW. Data records of wheat at maturity were excluded in this study, as there were no green leaves remaining at that time.

3.2.2.2 Canopy reflectance measurements

Canopy reflectance data were obtained before each field destructive sampling by a Cubert S185 hyperspectral snapshot camera (Cubert GmbH, Ulm, Baden-Württemberg, Germany). The hyperspectral camera was equipped with a DJI Ronin-MX three-axis gimbal stabiliser and installed on a DJI M600 PRO hexacopter with global positioning system (GPS) and inertial navigation system modules (DJI, Shenzhen, China). The gimbal stabiliser ensured that the view angle of the camera was constant and acquired the nadir images. The hyperspectral camera captures 125 spectral bands in the range of 450-950 nm with a sampling interval of 4 nm. The light is split into two beams after passing through the camera lens. 80% of the entire light entering the hyperspectral camera creates a 50×50 pixels hyperspectral cube with a 12-bit dynamic range and the remaining 20% enters the panchromatic camera to create a panchromatic image with a resolution of 1000×1000 pixels (e.g., Aasen et al., 2015; Lu et al., 2020). After pan-sharpening the hyperspectral cube to the resolution of panchromatic image by the Cube-Pilot software (Cubert GmbH, Ulm, Baden-Württemberg, Germany), hyperspectral images with a size of 1000×1000 pixels can be generated.

Table 3.2 Sampling dates and the number of destructive field samples at corresponding growth stages of rice, maize and wheat.

Growth stage	Rice	No. of samples	Maize	No. of samples	Winter wheat	No. of samples
Tillering stage/ Early stem- elongating stage	22 July	24	7 July	19	22 March	56
Stem-elongating stage	15 August	24	20 July	19	8 April	56
Flowering stage	8 September	24	4 August	19	27 April	56
Grain-filling stage	21 September	24	25 August	19	10 May	56
Grain-filling stage	10 October	24	7 September	19	25 May	56
Maturity	30 October	24	23 September	18	4 June	-
Total no. of samples		144		113		280

The flight campaigns were performed between 10 a.m. and 2 p.m. in consistent weather conditions (cloudless or heavy cloud) to minimise changes in illumination. The hyperspectral camera was calibrated by a white and black board before flying. The flight height was 60 m, resulting in a spatial resolution of ca. 2 cm, and the flying speed was fixed at 6 m s⁻¹. The hyperspectral images were created automatically at a sampling time interval of 1 ms and the forward and side overlaps were set to 80% and 60%, respectively, under the control of a connected microcomputer. The hyperspectral images were orthographically stitched by the Agisoft PhotoScan software (Agisoft LLC, St. Petersburg, Russia) based on GPS coordinates or image textures. Finally, the region of interest was drawn from the sampling area on the obtained hyperspectral ortho-images. The mean spectrum of the destructive sampling area was used as the corresponding reflectance of each sample. The bands beyond 902 nm and below 470 nm were dropped in the later data analysis, due to the low spectral quality (Lu et al., 2020).

3.2.3 Features collections for leaf traits prediction

3.2.3.1 Canopy spectral features and vegetation indices

The raw bands from hyperspectral images were used as canopy spectral features. Additionally, a set of VIs was selected for predicting leaf traits. In our study, 37 narrowband hyperspectral VIs were collected, in which 6 were directly related to N, 16 were related to chlorophyll, and 15 were related to canopy traits like biomass, coverage or LAI (Table S3.2). One more red-edge position was added in the list of VIs (Table S3.2), which was regressed by the inverted Gaussian method (Miller et al., 1990). The wavelengths with the sampling interval of 4 nm in our study were linearly interpolated to 1 nm to match the needs of calculation of hyperspectral VIs.

3.2.3.2 Canopy texture information extraction

Texture information from each hyperspectral band was extracted by the commonly used grey level co-occurrence matrix algorithm (GLCM) (Haralick et al., 1973). Eight GLCM-based texture features, including mean, variance, homogeneity, contrast, dissimilarity, entropy, second moment and correlation, were computed using the ENVI 5.2 software (Exelis Visual Information Solutions, Boulder, Colorado, USA). To avoid complexity of computation, only four bands, 470, 550, 638, and 850 nm (for blue, green, red, and NIR bands, respectively), were selected here for canopy texture information extraction.

3.2.3.3 Crop development stage

As leaf traits were measured at various dates during crop growth and these traits varied in time, we considered to include crop development stage as a co-predictor in nonparametric regression algorithms. Development stage was simulated by the crop growth simulation model GECROS (Yin and Struik, 2017), in which the development stage at seedling emergence, flowering and maturity were denoted as dimensionless values 0.0, 1.0 and 2.0, respectively, and the effect of temperature fluctuations on rate of crop development was taken into account.

3.2.4 Nonparametric regression algorithms

Commonly used linear nonparametric regression algorithms like Partial Least Squares Regression (PLSR), two machine learning algorithms including Support Vector Regression (SVR) and Random Forest Regression (RFR), and two DL-based algorithms DNN and CNN were implemented in the canopy leaf traits predictions (Fig. 3.2). The prepared feature information, including canopy reflectance (Ref), hyperspectral VIs, texture information (Tex), and development stage (DS), was selectively assembled and then compressed into a one-dimensional vector as the feature set (Table 3.3). Random assignment of instances was used for splitting the training and testing data for each model, in which 75% data at each growing stage was randomly selected and pooled in the training dataset and the remaining 25% was used for independent testing. Thus, the full-growth-stage models were built on the training set and validated on the testing set. All data analysis and model training were conducted in the Python environment. Regression models were implemented relying on the package of scikit-learn (0.24.1) (Pedregosa et al., 2011) and TensorFlow (2.5) (Abadi et al., 2015) in Python. Table S3.3 gives the specific properties and numerical settings of the hyperparameters for each regression algorithms. *k*-fold cross validated root mean square error (*RMSE*) was minimised to tune the pivotal hyperparameters in machine learning models.

Table 3.3 Description of the combinations of feature sets from hyperspectral images and crop model simulations.

Acronym	Feature type	Feature number
Ref	Canopy reflectance	109
RefVIs	Canopy reflectance + Vegetation indices	146
RefVIsTex	Canopy reflectance + Vegetation indices + Texture information	178
All	Canopy reflectance + Vegetation indices + Texture information + Development stage	179

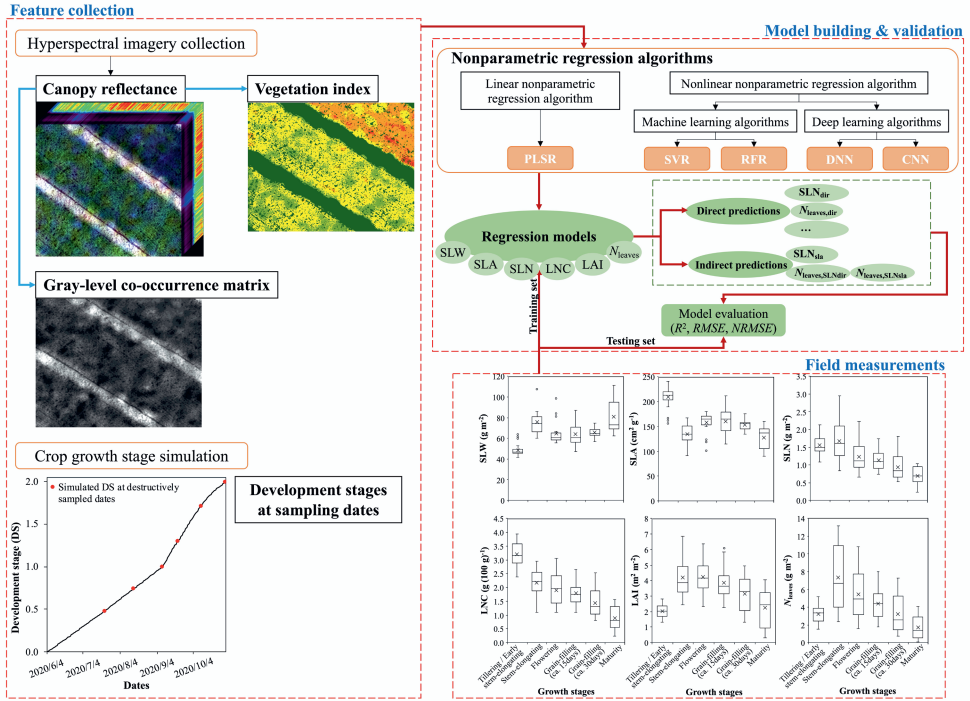


Fig. 3.2 The workflow diagram of data acquisition, feature collection, and data analysis in this study. Nonparametric regression algorithms are: PLSR = Partial Least Squares Regression, SVR = Support Vector Regression, RFR = Random Forest Regression, DNN = Deep Neural Network, CNN = Convolution Neural Network. Crop traits are: SLW = specific leaf weight, SLA = specific leaf area, SLN = specific leaf nitrogen, LNC = leaf nitrogen concentration, LAI = leaf area index, N_{leaves} = canopy leaf-nitrogen content. $N_{leaves,dir}$ and SLN_{dir} represent the direct predictions of N_{leaves} and SLN, respectively. Within the indirect predictions, SLN_{sla} was calculated as the predicted LNC divided by the predicted SLA, $N_{leaves,SLNdir}$ was derived from predicted LAI and SLN_{dir} , and $N_{leaves,SLNsla}$ was derived from predicted LAI and SLN_{sla} (indirectly predicted SLN via SLA) (see also Table 3.1).

3.2.4.1 Partial Least Squares Regression (PLSR)

PLSR is characterised by the ability to perform regression modelling even when the independent variables have severe multiple correlations. PLSR also allows regression modelling even when there are fewer sample points than given variables, thereby allowing all original independent variables being included in the final model (Geladi and Kowalski, 1986). The only hyperparameter in PLSR optimised in this study was the number of potential components (“n_components”, Table S3.3).

3.2.4.2 Support Vector Regression (SVR)

SVR has unique advantages in the case of small samples as well as nonlinear and high-dimensional inputs (Drucker et al., 1997). It improves the predictability by mapping low-dimensional samples into high-dimensional space and thus the non-linear inputs can be linearly separated by using a kernel function. SVR is efficient in finding sample information to achieve the best compromise between model complexity and learning ability.

3.2.4.3 Random Forest Regression (RFR)

RFR is a robust ensemble learning technique able to handle large numbers of input variables. RFR combines a large set of decision trees (Breiman, 2001). Together with the number of trees, tree traits and the number of input features can be optimised to get the best model (Table S3.3).

3.2.4.4 Artificial Neural Networks (ANNs)

Normally, an ANN is composed of an input layer, several hidden layers and an output layer and each layer contains a number of neurons. In this study, the fully connected feedforward DNN is based on simply increasing the depth (number of layers) of ANNs (Längkvist et al., 2014) (Fig. S3.1a). Fully connected 1-D CNN with several layers of convolution and pooling was designed as well (Fig. S3.1b). The stacking multiple layers are able to extract higher-order feature information (Längkvist et al., 2014). The optimum composition of hidden layers was determined by the minimum value of *RMSE* in validation dataset, while the number of neurons in each layer was asserted beforehand. Herein, the selected activation function for hidden layers was the Rectified Linear Unit with the form of $f(z) = \max(0, z)$ (LeCun et al., 2015). The weights of model neurons were initialised by the initialisation method of He et al. (2015) and then updated by applying the gradient descent algorithm of the Adam optimiser. A reduced learning rate on plateau and early stopping were implemented as well to prevent overfitting.

3.2.5 Model performance evaluation

The performances of all models were evaluated by the results in the testing sets (unless specified), with the coefficient of determination (R^2) and the *RMSE* with the following equation:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3.1)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (3.2)$$

where \hat{y}_i is a predicted value, y_i is a measured value, \bar{y} is the average value of measured values across the whole growing season, and n is the number of measurements.

Additionally, the normalised *RMSE* (*NRMSE*) was applied here as the metric of model performance evaluation in the needs of the comparison among crop species or among leaf traits, calculated by:

$$NRMSE = RMSE / \bar{y} \quad (3.3)$$

3.3 Results

3.3.1 Canopy spectral reflectance and canopy and leaf nitrogen traits

Differences in canopy reflectance were observed across growth stages in the three crop species and the change tended to be similar in rice and wheat (Fig. 3.3a-c). Values of SLN in rice and wheat became lower with the progress of the growth stage, but SLN in maize even increased at maturity (Fig. 3.3d), which might be caused by the rapid decline in LAI associated with senescence of the upper leaves while the cob leaves remained green (Fig. S3.2). Values of LNC decreased from the tillering stage of rice and wheat or the early stem-elongating stage of maize onwards (Fig. 3.3e). Measured N_{leaves} in both rice and wheat reached the peak at the stem-elongating stage and then decreased until maturity, while the highest values of N_{leaves} in maize occurred at flowering (Fig. 3.3f). Unlike in rice and wheat, there was no significant difference in measured SLN, LNC and N_{leaves} between treatments in maize (results not shown), presumably because the residual N supply from soil was very high.

3.3.2 Correlograms of collected features with respect to leaf and canopy nitrogen traits

Among the three crops, the correlation coefficients (r) of N_{leaves} with canopy reflectance changed in a similar trend and slightly decreased when wavelength was between 680 nm and 710 nm, which belongs to the red-edge region (Fig. S3.3). SLN and LNC in rice showed weaker correlations than N_{leaves} with canopy reflectance in the NIR region and the r ranged from 0.15 to 0.27 and -0.06 to 0.11, respectively (Fig. S3.3a). Within the collected VIs, the highest values of r for SLN, LNC and N_{leaves} , with VIs were up to 0.77, 0.86 and 0.82 in rice, and 0.72, 0.93 and 0.95 in wheat (Fig. S3.4a-f). In maize, LNC displayed a weak correlation with collected VIs and the r ranged from -0.37 to 0.40 (Fig. S3.4g-i).

Regarding correlations with the texture information, leaf N traits were more correlated with the features extracted from blue, green and red bands than those from NIR in rice (Fig. S3.5a-d), while the higher correlations tended to be achieved from the bands of red and NIR in wheat and maize (Fig. S3.5e-l). The LNC in maize was highly correlated with the texture information from the NIR band and the highest value of r was up to 0.81 (Fig. S3.5l).

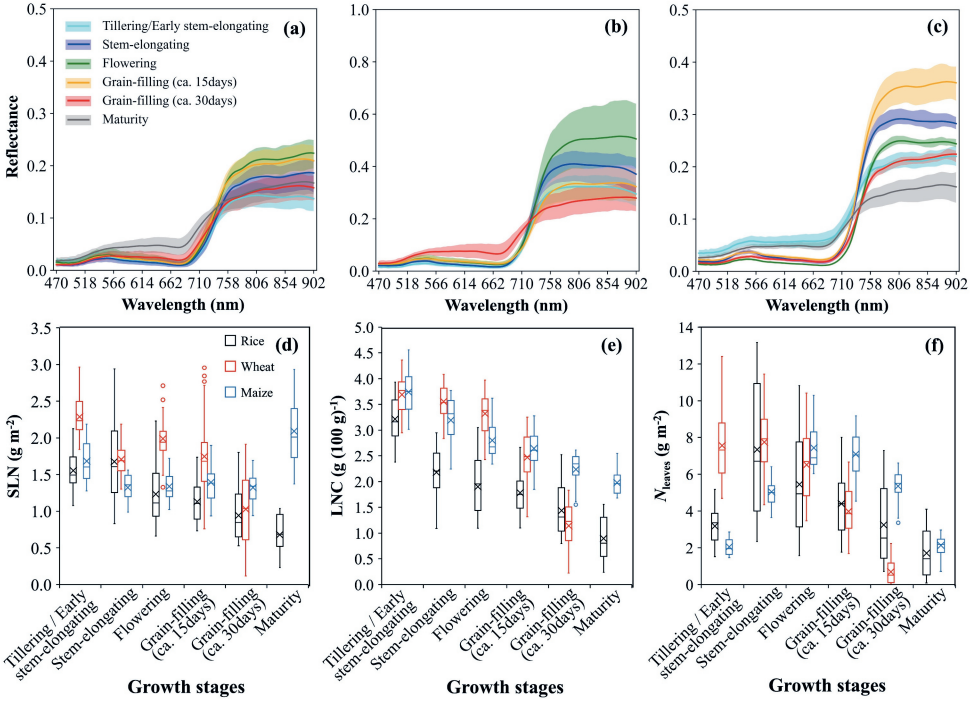


Fig. 3.3 Range of the collected canopy reflectance of rice (a), wheat (b) and maize (c), and the data distribution of destructively sampled specific leaf nitrogen (SLN) (d), leaf nitrogen concentration (LNC) (e), and canopy leaf-nitrogen content (N_{leaves}) (f) at different stages during the whole growing season.

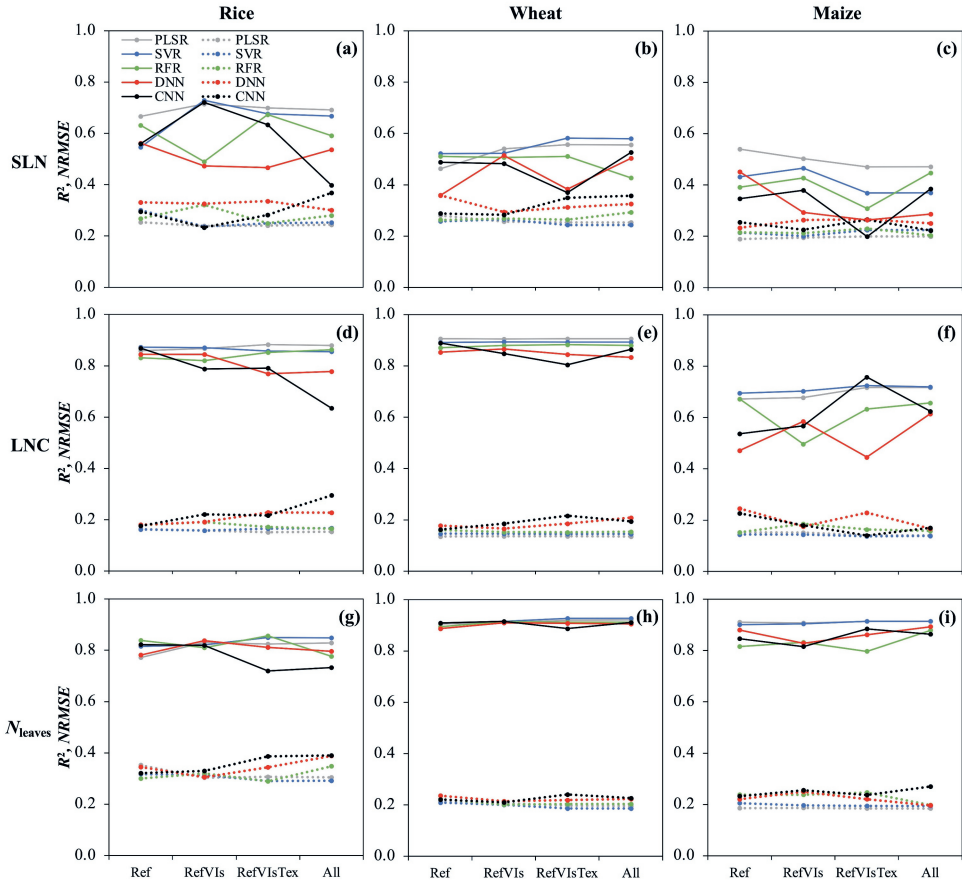


Fig. 3.4 The prediction performance, R^2 (full lines) and $NRMSE$ (dotted lines), in the testing dataset of different regression models in directly predicted specific leaf nitrogen (SLN) (a-c), leaf nitrogen concentration (LNC) (d-f) and canopy leaf-nitrogen content (N_{leaves}) (g-i) in rice, wheat and maize with various sets of input feature types. Nonparametric regression algorithms: PLSR = Partial Least Squares Regression, SVR = Support Vector Regression, RFR = Random Forest Regression, DNN = Deep Neural Network, CNN = Convolution Neural Network. Acronyms in the x-axis for four sets of feature types are defined in Table 3.3.

3.3.3 Performance of nonparametric regression algorithms

For SLN and LNC, PLSR and SVR models tended to improve their performance in rice and wheat when incorporating more features and the predicted LNC agreed with its measurements better than the predicted SLN among three crop species (Figs 3.4a-f, S3.6). For SLN in rice and wheat from PLSR and SVR, $NRMSE$ decreased after incorporating VIs and Tex into the feature set of Ref, whereas $NRMSE$ in maize tended to increase (Fig. 3.4a-c). For LNC, the performance

of PLSR and SVR hardly changed in wheat while incorporating different feature types, but improved slightly in maize (Fig. 3.4e-f). For the predicted SLN and LNC, RFR tended to perform better than DNN among three crop species, albeit worse than PLSR and SVR (Fig. 3.4a-f). Although there was a chance that CNN performed better than other regression algorithms (Fig. 3.4a, f), its performance lacked the consistency and tended to become worse when incorporating more feature types into the model (Fig. 3.4b-e).

For the predictions of N_{leaves} , PLSR and SVR always performed better than others and their performance tended to be consistently improved in the testing dataset when incorporating different types of features among three crop species (Fig. 3.4g-i). In detail, *NRMSE* of predicted N_{leaves} in rice from PLSR decreased from 0.352 to 0.303 after incorporating VIs into the feature set of Ref and hardly changed thereafter, while that of SVR further decreased to 0.291 after incorporating the feature of Tex (Fig. 3.4g). Also in wheat and maize, no increment of model performance was obtained while further incorporating the DS (Fig. 3.4h-i). Due to the increased complexity of their algorithms, RFR, DNN and CNN models were more likely to be overfitted in the training set (Fig. S3.7g-i) and their performance for N_{leaves} of the testing set in different crops tended to be highly variable while incorporating different types of features (Fig. 3.4g-i). For instance, after incorporating VIs into the feature set of Ref, *NRMSE* of predicted N_{leaves} in rice from DNN decreased from 0.345 to 0.305 and then increased to 0.388 after further incorporating the DS (Fig. 3.4g).

3.3.4 Performance in modelling canopy and leaf nitrogen by direct and indirect predictions

The best performing regression algorithm for predicting SLW and SLA changed among crop species, and in general, the predicted SLA agreed with its measurements better than SLW (Figs. 3.5, S3.8). The incorporation of feature sets also improved the predictions of SLW and SLA by PLSR and SVR in three crops (Fig. 3.5). Among different regression algorithms, PLSR and SVR performed better than others in rice (Fig. 3.5a, d), while DNN and CNN performed better in wheat (Fig. 3.5b, e). In maize, the predictions of SLW from RFR agreed with the measurements better, while PLSR and SVR performed better in the predictions of SLA (Fig. 3.5c, f).

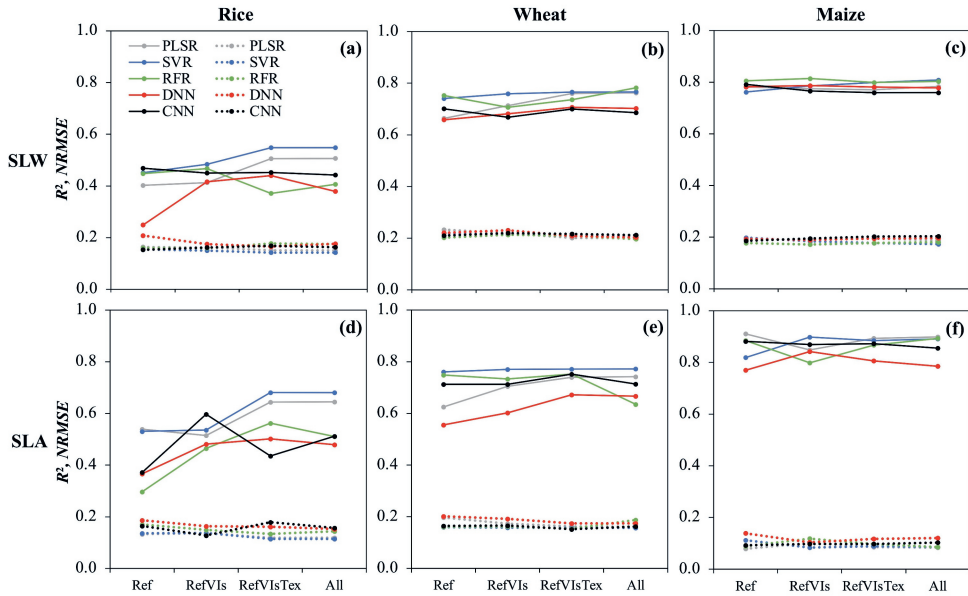


Fig. 3.5 The prediction performance, R^2 (full lines) and $NRMSE$ (dotted lines), in the testing dataset of different regression models in directly predicted specific leaf weight (SLW) (a-c) and specific leaf area (SLA) (d-f) in rice, wheat and maize with various sets of input feature types. Nonparametric regression algorithms: PLSR = Partial Least Squares Regression, SVR = Support Vector Regression, RFR = Random Forest Regression, DNN = Deep Neural Network, CNN = Convolution Neural Network. Acronyms in the x-axis for four sets of feature types are defined in Table 3.3.

As SLA was better predicted than SLW, the indirect predictions of SLN from SLA, SLN_{sla} , are only presented here (Table 3.4). Compared with the directly predicted SLN (SLN_{dir}), SLN_{sla} tended to agree better with the measurements, in which $NRMSE$ decreased from 0.241, 0.254, and 0.189 to 0.240, 0.244 and 0.183, while R^2 increased from 0.699, 0.557 and 0.540 to 0.725, 0.583 and 0.611, for rice, wheat and maize, respectively. Both direct and indirect predictions from PLSR were better than those from SVR. Regarding the utilised feature sets, in rice and wheat, both predicted SLN agreed better with measurements while using the feature set of RefVIsTex. However, in maize, better performance was achieved when using the feature set of Ref only.

To upscale the predicted SLN to N_{leaves} , LAI was predicted well using the feature set of Ref or RefVIsTex while applying the regression algorithms of PLSR and SVR (Figs S3.9-S3.10). The indirect predictions of $N_{\text{leaves,SLN}_{\text{dir}}}$ and $N_{\text{leaves,SLN}_{\text{sla}}}$ were estimated from SLN_{dir} and SLN_{sla} , respectively. Using more features from Ref to RefVIsTex universally improved directly or indirectly predicted N_{leaves} , except for the slight decrease in the indirect predictions from SVR in wheat and maize (Fig. 3.6). Therefore, for further analysis, we only show the predictions from applying the feature set of RefVIsTex. Within the comparison between direct and indirect predictions, $N_{\text{leaves,dir}}$ in wheat agreed with the measurements better than others, while in rice and maize N_{leaves} tended to be better predicted via the indirect method ($N_{\text{leaves,SLN}_{\text{dir}}}$). For instance, compared with $N_{\text{leaves,dir}}$ from PLSR, *NRMSE* of $N_{\text{leaves,SLN}_{\text{dir}}}$ in rice decreased from 0.307 to 0.286 (Fig. 3.6a). Between the two indirect predictions, the predicted $N_{\text{leaves,SLN}_{\text{dir}}}$ generally had a lower *NRMSE* in the three crop species than the predicted $N_{\text{leaves,SLN}_{\text{sla}}}$ (Fig. 3.6).

Table 3.4 The prediction performance in the testing dataset of directly and indirectly predicted specific leaf nitrogen (SLN) in rice, wheat and maize using feature sets of Ref and RefVIsTex (see Table 3.3 for their definition) and applying the regression algorithms of Partial Least Squares Regression (PLSR) and Support Vector Regression (SVR) ^a.

Method	Feature set		Rice		Wheat		Maize	
			SLN_{dir}	SLN_{sla}	SLN_{dir}	SLN_{sla}	SLN_{dir}	SLN_{sla}
PLSR	Ref	R^2	0.666	0.659	0.462	0.564	0.540	0.611
		<i>NRMSE</i>	0.254	0.266	0.281	0.251	0.189	0.183
	RefVIsTex	R^2	0.699	0.725	0.557	0.583	0.471	0.440
		<i>NRMSE</i>	0.241	0.240	0.254	0.244	0.200	0.207
SVR	Ref	R^2	0.546	0.626	0.522	0.533	0.432	0.582
		<i>NRMSE</i>	0.301	0.276	0.258	0.256	0.215	0.183
	RefVIsTex	R^2	0.677	0.661	0.582	0.563	0.369	0.481
		<i>NRMSE</i>	0.250	0.262	0.243	0.246	0.224	0.197

^a For each set of analysis, there is a directly predicted SLN (SLN_{dir}) and an indirectly predicted SLN (SLN_{sla}), the SLN calculated as the predicted leaf nitrogen concentration divided by the predicted specific leaf area (see also Table 3.1). The regression algorithms for the prediction of SLA were specified as SVR, Convolution Neural Network, and PLSR for rice, wheat and maize, respectively, as these algorithms yielded the best performance for the three crops. For the better predicted SLN in the direct vs indirect predictions' comparison, its corresponding R^2 and *NRMSE* values are given in bold.

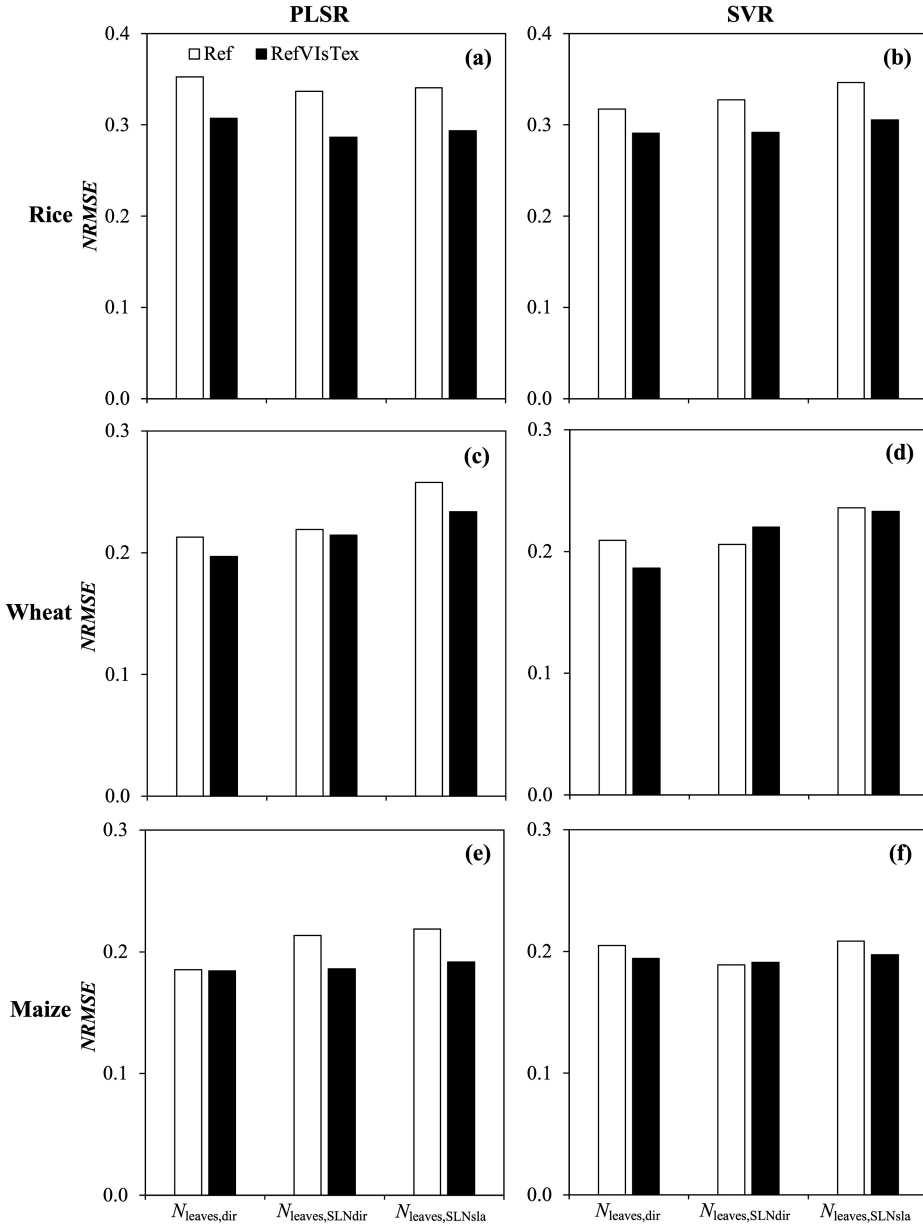


Fig. 3.6 The prediction performance in the testing dataset of directly and indirectly predicted canopy leaf-nitrogen content (N_{leaves}) in rice (a, b), wheat (c, d) and maize (e, f) using feature sets of Ref and RefVIsTex (see Table 3.3 for their definition) and regression algorithms of Partial Least Squares Regression (PLSR) and Support Vector Regression (SVR). The directly predicted N_{leaves} was labelled as $N_{leaves,dir}$. The indirectly predicted $N_{leaves,SLNdir}$ and $N_{leaves,SLNsla}$ were upscaled from directly predicted specific leaf nitrogen (SLN) SLN_{dir} and from indirectly predicted SLN_{sla} , respectively (also see Table 3.1).

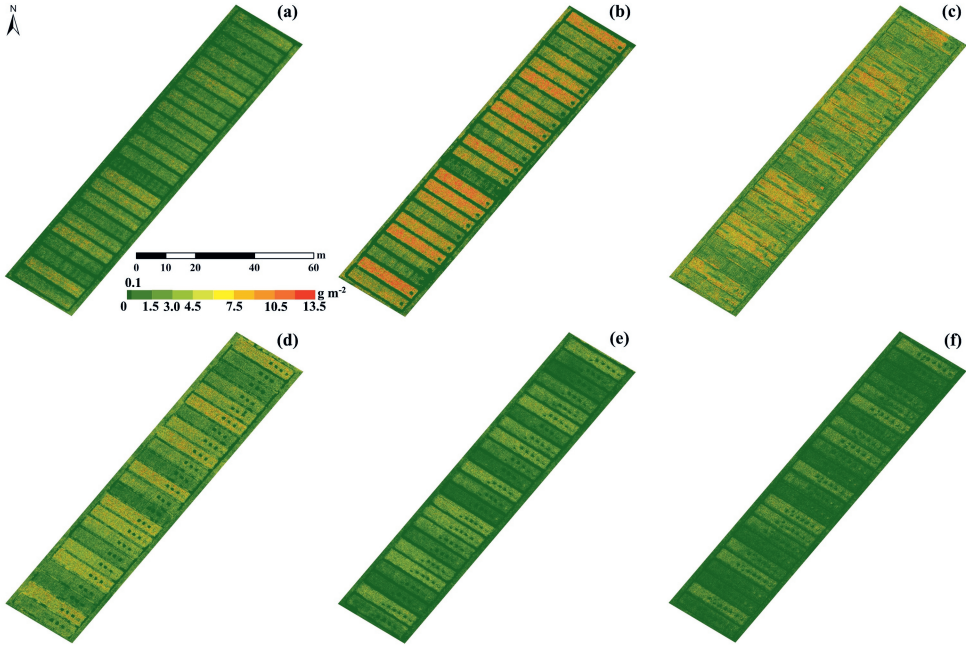


Fig. 3.7 The predicted canopy leaf-nitrogen content (N_{leaves}) from its indirect predictions $N_{\text{leaves,SLNdir}}$ (see Table 3.1 for its definition) at the experimental area of rice across the whole growing season using feature set of RefVIsTex (see Table 3.3 for its definition) and regression algorithm of Partial Least Squares Regression. Predicted $N_{\text{leaves,SLNdir}}$ at the stage of tillering (a), stem-elongating (b), flowering (c), grain-filling (ca. 15 days) (d), grain-filling (ca. 30 days) (e), and maturity (f) are shown.

3.3.5 Predicted canopy leaf-nitrogen content along stages of the whole season

The predicted N_{leaves} values in rice, wheat and maize were mapped in the whole experimental area and illustrated for different growing stages (Figs 3.7, S3.11-S3.12). Taking rice as an example, in line with the field measurements (Fig. 3.3f), the highest N_{leaves} occurred at stem elongation and remarkable differences were shown between experimental plots (Fig. 3.7b). Such differences among treatments gradually narrowed after stem elongation (Fig. 3.7b-f). A similar trend was observed in mapped N_{leaves} of wheat (Fig. S3.11), in agreement with its field measurements (Fig. 3.3f). In maize, the differences in N_{leaves} among growing stages were predicted well (Fig. S3.12). The hardly observed differences of predicted N_{leaves} between experimental plots also agreed with the fact that there was no significant difference in measured N_{leaves} between treatments (Figs 3.3f, S3.12). Moreover, the differences in crop growth status within an experimental plot was also noticeable (Figs 3.7, S3.11-S3.12).

3.4 Discussion

Leaf and canopy N contents across growing stages for three major field crops were predicted from UAV based hyperspectral image features, by adopting different nonparametric regression algorithms and predicting pathways (direct vs. indirect). The mapped N_{leaves} across successive growth stages demonstrates the potential utilisation of the prediction models.

3.4.1 Comparison of direct and indirect predictions

As SLA and SLW are reciprocals of each other, the difference between them has been hardly noticed. But there are still some debates about predicting them from remote sensing images. For instance, SLW could be predicted well and the reported R^2 values were ca. 0.90 (Ecarnot et al., 2013; Silva-Perez et al., 2018). However, Ali et al. (2017) indicated that SLA tended to be more correlated with leaf reflectance than SLW, whereas Hussain et al. (2020) showed that the predictions of SLW matched its measurements better than the predictions of SLA did. With crop growth, our data range of SLW tended to become wider than that of SLA in three crop species (results not shown) whereas the differences of canopy reflectance between treatments tended to be diminished (Fig. 3.3a-c), which might contribute to the better predicted SLA than that of SLW (Figs 3.5, S3.8).

Both SLN and LNC are common leaf level indicators to monitor leaf N. As the components in whole leaf thickness accounts for the light reflection or absorption, SLN is supposed to be better predicted than LNC (Baret and Fourty, 1997). However, our results showed that the predicted LNC agreed with its measurements better than the predicted SLN did (Figs 3.4a-f, S3.6), in line with the previous results of Ecarnot et al. (2013) and Li et al. (2018) using measured raw reflectance spectra. Under N shortage, leaf weight increases because of accumulation of starch and cell wall material (McDonald et al., 1986; Terry et al., 1983) and thus SLA or SLW varies significantly (Jones and Hesketh, 1980). As leaf expansion is limited, the change of leaf N status is reflected less in SLN than in LNC. Moreover, as SLN is normally determined from LNC and SLA, meaning that it requires more measurement steps to determine SLN than to determine LNC, data for SLN must contain more measurement noise than for LNC. Due to the better predicted LNC and SLA, the indirectly predicted SLN_{sla} tended to match measurements than directly predicted SLN_{dir} (Table 3.4). However, when upscaling to canopy level, the indirectly predicted N_{leaves} from SLN_{sla} became worse than that from SLN_{dir} (Fig. 3.6). As SLN_{sla} was estimated from predicted LNC and SLA, this result supported the assertion by

Kattenborn et al. (2019) that area-based leaf traits, like SLN, are more suitable for upscaling from leaf to canopy level than mass-based leaf traits, like LNC, as less regression error was introduced. The indirect predictions of $N_{\text{leaves,SLN}}$ in rice and maize tended to agree with measurements than the directly predicted $N_{\text{leaves,dir}}$ (Fig. 3.6). Even though the indirect methods achieved better prediction in our results (Table 3.4, Fig. 3.6), as more regression processes were introduced, the results might be more uncertain (Malenovsky et al., 2019). Thus, the uncertainty of the regression processes should be further investigated.

3.4.2 Contribution of feature types in canopy and leaf nitrogen prediction

Fusing different types of features contributes to plant N traits prediction. Besides the raw bands, VIs have also been considered as supplementary information serving as canopy spectral features (Maimaitijiang et al., 2020). There were tight correlations between VIs and leaf N traits (Fig. S3.4). However, with Ref features used as the baseline, adding VIs to the model did not always yield better prediction (Fig. 3.4), which might be caused by the introduced noisy information from VIs and/or the correlations between Ref and VIs features. Due to the different changing rate in crop biomass accumulation or N uptake between before and after heading, the predictability using hyperspectral VIs could be significantly influenced (Li et al., 2010; Yu et al., 2013). On the other hand, as VIs are normally derived from several bands (Table S3.2), the information provided from VIs might be limited compared with that from Ref in the full spectrum (Wang et al., 2021). From the extracted canopy texture features, the additional information associated with spatial canopy architecture and structure characteristics is potentially provided (Colombo, 2003). The fusion of canopy texture information benefits the predictions not only of the canopy traits like LAI and biomass (Liu et al., 2021; Zheng et al., 2018b), but also of crop N status like leaf chlorophyll and plant aboveground N content (Qiao et al., 2020; Zheng et al., 2018a). Similar to these previous studies, our results also showed that fusing canopy texture information contributed to the prediction of leaf N traits, especially for the regression models based on PLSR and SVR (Table 3.4, Figs 3.4, 3.6).

Although both measured canopy reflectance and leaf N traits changed across growth stages (Fig. 3.3), the performance of the regressing models in predicting leaf N traits hardly improved after directly incorporating DS into the models (Fig. 3.4). The feature of DS might be utilised in a more proper way. For instance, Li et al. (2022) found that the coefficient values (slope and intercept) of ordinary least squares regression of aboveground biomass in wheat with VIs had a strong relationship with phenological stages. With the calibrated slopes and intercepts at

different stages, aboveground biomass can be well estimated over the entire growing season (Li et al., 2022). Moreover, the appropriate digital indicator of crop growth stages is of importance as well. The simulated DS in this study is only one way of quantifying development stages, which is calculated from hourly temperatures, using a bell-shaped temperature response function capturing base, optimum and ceiling temperatures (Yin and van Laar, 2005). It is similar to growth stages with the Zadoks scale, using a numeric system between 0 and 100 representing stages from sowing to harvest (Zadoks et al., 1974). Li et al. (2022) showed that compared with the thermal-based digital indicator of growth stages, growth stages with the Zadoks scale performed better in the prediction of aboveground biomass of wheat.

3.4.3 Characteristics of regression algorithms in canopy and leaf nitrogen prediction

The performance of five popular nonparametric regression algorithms were evaluated for N traits in three major field crops in this study. Among these methods, PLSR used the least number of hyperparameters and SVR next to it, compared with RFR, DNN and CNN that have many hyperparameters (Table S3.3) and thus, in principle, should have the more flexibility in the prediction of leaf N traits. However, our results showed that PLSR and SVR tended to be more stable and PLSR was likely to achieve the best model performance in the prediction of N_{leaves} , as well as SLN and LNC (Figs 3.4, S3.7), in line with the results in the prediction of LNC in wheat (Yao et al., 2015). This might be caused by the limited size of our dataset, as PLSR is known to be more effective than others, when the number of samples is smaller than the number of variables (Geladi and Kowalski, 1986). With nearly 1000 samples, DNN based yield prediction models systematically performed better than PLSR and SVR based models while increasing feature types (Maimaitijiang et al., 2020). Moreover, with more than 7000 samples from different years and locations, CNN provided higher accuracy in LNC prediction in grass (Pullanagari et al., 2021). Thus, as the numerous combinations of hyperparameters in RFR and of hidden layers and neuro numbers in DNN and CNN have already been trialled in this study (Table S3.3, Fig. S3.1), increasing the number of samples might be the essential for improving the predictability of DNN and CNN.

3.5 Conclusions

We evaluated the prediction of area- and mass-based leaf N traits using different regression algorithms and feature sets extracted from hyperspectral images for three major crops, rice,

wheat and maize. Our results showed that for the prediction of N traits at leaf level, SLN and LNC, and at canopy level N_{leaves} using the limited size of the dataset, regression algorithms of PLSR and SVR tended to perform better than others, and the incorporation of different types of features, like canopy spectral, VIs and texture information, helped improve model performance. The regression algorithm of PLSR performed slightly better than SVR in the prediction of N_{leaves} . Predicting N_{leaves} in an indirect way by upscaling SLN contributed to the further improvement of prediction of N_{leaves} , compared with the directly predicted N_{leaves} , although the regression process might be more uncertain. These results provide useful information for the next-step work that will couple the real-time accurately retrieved leaf N traits into dynamic crop models so as to predict the real-time growth status more accurately, given that canopy photosynthesis is directly linked with crop N status and primarily determines crop growth rates. This would better enable crop models as a decision-making tool to guide farmers for in situ field N management targeting the maximal yield or resource use efficiency.

Acknowledgement

We acknowledge the financial support from the High-tech Industry and Scientific and Technological Innovation Project of Lin-gang Special Area, Shanghai (grant number: SH-LG-GK-2020-02-19).

References

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jozefowicz, R., Jia, Y., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Schuster, M., Monga, R., Moore, S., Murray, D., Olah, C., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., & Zheng, X. (2015). TensorFlow: A System for Large-Scale Machine Learning. Software available from tensorflow.org.
- Ali, A. M., Darvishzadeh, R., Skidmore, A. K., & van Duren, I. (2017). Specific leaf area estimation from leaf and canopy reflectance through optimization and validation of vegetation indices. *Agricultural and Forest Meteorology*, 236, 162-174. doi:10.1016/j.agrformet.2017.01.015
- Aasen, H., Burkart, A., Bolten, A., & Bareth, G. (2015). Generating 3D hyperspectral information with lightweight UAV snapshot cameras for vegetation monitoring: From camera calibration to quality assurance. *ISPRS Journal of Photogrammetry and Remote Sensing*, 108, 245-259. doi:10.1016/j.isprsjprs.2015.08.002
- Atzberger, C., Guérif, M., Baret, F., & Werner, W. (2010). Comparative analysis of three chemometric techniques for the spectroradiometric assessment of canopy chlorophyll content in winter wheat. *Computers and Electronics in Agriculture*, 73, 165-173. doi:10.1016/j.compag.2010.05.006
- Baret, F., & Fourty, T. (1997). Radiometric estimates of nitrogen status of leaves and canopies. In G. Lemaire (Eds.), *Diagnosis of nitrogen status in crops* (pp. 201-227). Heidelberg, Germany: Springer-Verlag.
- Beeckman, F., Motte, H., & Beeckman, T. (2018). Nitrification in agricultural soils: impact, actors and mitigation. *Current Opinion in Biotechnology*, 50, 166-173.
- Berger, K., Verrelst, J., Féret, J.-B., Wang, Z., Woche, M., Strathmann, M., Danner, M., Mauser, W., & Hank, T. (2020). Crop nitrogen monitoring: Recent progress and principal developments in the context of imaging spectroscopy missions. *Remote Sensing of Environment*, 242, 111758. doi:10.1016/j.rse.2020.111758
- Boote, K. J., Jones, J. W., Hoogenboom, G., & Pickering, N. B. (1998). The CROPGRO model for grain legumes. In G. Y. Tsuji, G. Hoogenboom, & P. K. Thornton (Eds.), *Understanding options for agricultural production* (pp. 99-128). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Breiman, L. (2001). Random forests. *Machine learning*, 45, 5-32.
- Colombo, R. (2003). Retrieval of leaf area index in different vegetation types using high resolution satellite data. *Remote Sensing of Environment*, 86, 120-131. doi:10.1016/s0034-4257(03)00094-4
- Davidson, E. A. (2009). The contribution of manure and fertilizer nitrogen to atmospheric nitrous oxide since 1860. *Nature Geoscience*, 2, 659-662.
- Drucker, H., Burges, C. J., Kaufman, L., Smola, A., & Vapnik, V. (1997). Support vector regression machines. In M. Mozer, J. Jordan, & T. Petsche (Eds.), *Neural information processing systems 9* (pp. 155-161). Cambridge, Massachusetts, The United States: MIT Press.

- Ecarnot, M., Compan, F., & Roumet, P. (2013). Assessing leaf nitrogen content and leaf mass per unit area of wheat in the field throughout plant cycle with a portable spectrometer. *Field Crops Research*, 140, 44-50. doi:10.1016/j.fcr.2012.10.013
- Evans, J. R. (1983). Nitrogen and photosynthesis in the flag leaf of wheat (*Triticum aestivum* L.). *Plant Physiology*, 72, 297-302.
- FAO. (2020). *World Food and Agriculture - Statistical Yearbook 2020*. Rome, Italy.
- Fu, Y., Yang, G., Li, Z., Li, H., Li, Z., Xu, X., Song, X., Zhang, Y., Duan, D., Zhao, C., & Chen, L. (2020). Progress of hyperspectral data processing and modelling for cereal crop nitrogen monitoring. *Computers and Electronics in Agriculture*, 172, 105321. doi:10.1016/j.compag.2020.105321
- Geladi, P., & Kowalski, B. R. (1986). Partial least-squares regression: a tutorial. *Analytica Chimica Acta*, 185, 1-17.
- Haralick, R. M., Shanmugam, K., & Dinstein, I. H. (1973). Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, 610-621.
- He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision* (pp. 1026-1034).
- Homolová, L., Malenovský, Z., Clevers, J. G. P. W., García-Santos, G., & Schaepman, M. E. (2013). Review of optical-based remote sensing for plant trait mapping. *Ecological Complexity*, 15, 1-16. doi:10.1016/j.ecocom.2013.06.003
- Horler, D. N. H., Dockray, M., & Barber, J. (1983). The red edge of plant leaf reflectance. *International Journal of Remote Sensing*, 4, 273-288.
- Hussain, S., Gao, K., Din, M., Gao, Y., Shi, Z., & Wang, S. (2020). Assessment of UAV-Onboard Multispectral Sensor for Non-Destructive Site-Specific Rapeseed Crop Phenotype Variable at Different Phenological Stages and Resolutions. *Remote Sensing*, 12, 397. doi:10.3390/rs12030397
- Jin, X., Kumar, L., Li, Z., Feng, H., Xu, X., Yang, G., & Wang, J. (2018). A review of data assimilation of remote sensing and crop models. *European Journal of Agronomy*, 92, 141-152. doi:10.1016/j.eja.2017.11.002
- Jones, J. W., & Hesketh, J. D. (1980). Predicting leaf expansion. In J. D. Hesketh & J. W. Jones (Eds.), *Predicting Photosynthesis for Ecosystem Models. Volume II* (pp. 85-122). Boca Raton, Florida, The United States: CRC Press.
- Kattenborn, T., Schiefer, F., Zarco-Tejada, P., & Schmidtlein, S. (2019). Advantages of retrieving pigment content [$\mu\text{g}/\text{cm}^2$] versus concentration [%] from canopy reflectance. *Remote Sensing of Environment*, 230, 111195. doi:10.1016/j.rse.2019.05.014
- Khoury, C. K., Bjorkman, A. D., Dempewolf, H., Ramirez-Villegas, J., Guarino, L., Jarvis, A., Rieseberg, L. H., & Struik, P. C. (2014). Increasing homogeneity in global food supplies and the implications for food security. *Proceedings of the National Academy of Sciences of the United States of America*, 111, 4001-4006. doi:10.1073/pnas.1313490111

- Kropff, M. J., Van Laar, H. H., & Matthews, R. B. (1994). *ORYZA1: An ecophysiological model for irrigated rice production*. Wageningen, The Netherlands: DLO-Research Institute for Agrobiology and Soil Fertility.
- Långkvist, M., Karlsson, L., & Loutfi, A. (2014). A review of unsupervised feature learning and deep learning for time-series modeling. *Pattern Recognition Letters*, 42, 11-24. doi:10.1016/j.patrec.2014.01.008
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521, 436-444. doi:10.1038/nature14539
- Lemaire, G., Jeuffroy, M.-H., & Gastal, F. (2008). Diagnosis tool for plant and crop N status in vegetative stage: Theory and practices for crop N management. *European Journal of Agronomy*, 28, 614-624. doi:10.1016/j.eja.2008.01.005
- Li, D., Wang, X., Zheng, H., Zhou, K., Yao, X., Tian, Y., Zhu, Y., Cao, W., & Cheng, T. (2018). Estimation of area- and mass-based leaf nitrogen contents of wheat and rice crops from water-removed spectra using continuous wavelet analysis. *Plant Methods*, 14, 76. doi:10.1186/s13007-018-0344-1
- Li, F., Miao, Y., Hennig, S. D., Gnyp, M. L., Chen, X., Jia, L., & Bareth, G. (2010). Evaluating hyperspectral vegetation indices for estimating nitrogen concentration of winter wheat at different growth stages. *Precision Agriculture*, 11, 335-357. doi:10.1007/s11119-010-9165-6
- Li, Z., Zhao, Y., Taylor, J., Gaulton, R., Jin, X., Song, X., Li, Z., Meng, Y., Chen, P., Feng, H., Wang, C., Guo, W., Xu, X., Chen, L., & Yang, G. (2022). Comparison and transferability of thermal, temporal and phenological-based in-season predictions of above-ground biomass in wheat crops from proximal crop reflectance data. *Remote Sensing of Environment*, 273, 112967. doi:10.1016/j.rse.2022.112967
- Liu, H., Zhu, H., & Wang, P. (2016). Quantitative modelling for leaf nitrogen content of winter wheat using UAV-based hyperspectral data. *International Journal of Remote Sensing*, 38, 2117-2134. doi:10.1080/01431161.2016.1253899
- Liu, S., Jin, X., Nie, C., Wang, S., Yu, X., Cheng, M., Shao, M., Wang, Z., Tuohuti, N., Bai, Y., & Liu, Y. (2021). Estimating leaf area index using unmanned aerial vehicle data: shallow vs. deep machine learning algorithms. *Plant Physiology*, 187, 1551-1576. doi:10.1093/plphys/kiab322
- Lu, J., Li, W., Yu, M., Zhang, X., Ma, Y., Su, X., Yao, X., Cheng, T., Zhu, Y., Cao, W., & Tian, Y. (2020). Estimation of rice plant potassium accumulation based on non-negative matrix factorization using hyperspectral reflectance. *Precision Agriculture*, 22, 51-74. doi:10.1007/s11119-020-09729-z
- Maimaitijiang, M., Sagan, V., Sidike, P., Hartling, S., Esposito, F., & Fritschi, F. B. (2020). Soybean yield prediction from UAV using multimodal data fusion and deep learning. *Remote Sensing of Environment*, 237, 111599. doi:10.1016/j.rse.2019.111599
- Malenovský, Z., Homolová, L., Lukeš, P., Buddenbaum, H., Verrelst, J., Alonso, L., Schaepman, M. E., Lauret, N., & Gastellu-Etchegorry, J.-P. (2019). Variability and uncertainty challenges in scaling imaging spectroscopy retrievals and validations from leaves up to vegetation canopies. *Surveys in Geophysics*, 40, 631-656. doi:10.1007/s10712-019-09534-y

- Marshall, M., & Thenkabail, P. (2015). Advantage of hyperspectral EO-1 Hyperion over multispectral IKONOS, GeoEye-1, WorldView-2, Landsat ETM+, and MODIS vegetation indices in crop biomass estimation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 108, 205-218. doi:10.1016/j.isprsjprs.2015.08.001
- McDonald, A. J. S., Lohammar, T., & Ericsson, A. (1986). Growth response to step-decrease in nutrient availability in small birch (*Betula pendula* Roth). *Plant, Cell & Environment*, 9, 427-432.
- Miller, J. R., Hare, E. W., & Wu, J. (1990). Quantitative characterization of the vegetation red edge reflectance 1. An inverted-Gaussian reflectance model. *International Journal of Remote Sensing*, 11, 1755-1773. doi:10.1080/01431169008955128
- Moharana, S., & Dutta, S. (2016). Spatial variability of chlorophyll and nitrogen content of rice from hyperspectral imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 122, 17-29. doi:10.1016/j.isprsjprs.2016.09.002
- Ohyama, T. (2010). Nitrogen as a major essential element of plants. In T. Ohyama, & K. Sueyoshi (Eds.), *Nitrogen Assimilation in Plants* (pp. 1–18). Kerala, India: Research Signpost.
- Padilla, F. M., Gallardo, M., & Manzano-Agugliaro, F. (2018). Global trends in nitrate leaching research in the 1960–2017 period. *Science of the Total Environment*, 643, 400-413.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of machine Learning research*, 2825-2830.
- Peng, S., Cassman, K. G., & Kropff, M. J. (1995). Relationship between leaf photosynthesis and nitrogen content of field-grown rice in tropics. *Crop Science*, 35, 1627-1630.
- Poorter, H., & Evans, J. R. (1998). Photosynthetic nitrogen-use efficiency of species that differ inherently in specific leaf area. *Oecologia*, 116, 26-37.
- Pullanagari, R. R., Dehghan-Shoar, M., Yule, I. J., & Bhatia, N. (2021). Field spectroscopy of canopy nitrogen concentration in temperate grasslands using a convolutional neural network. *Remote Sensing of Environment*, 257, 112353. doi:10.1016/j.rse.2021.112353
- Qiao, L., Gao, D., Zhang, J., Li, M., Sun, H., & Ma, J. (2020). Dynamic Influence Elimination and Chlorophyll Content Diagnosis of Maize Using UAV Spectral Imagery. *Remote Sensing*, 12, 2650. doi:10.3390/rs12162650
- Raj, R., Walker, J. P., Pingale, R., Banoth, B. N., & Jagarlapudi, A. (2021). Leaf nitrogen content estimation using top-of-canopy airborne hyperspectral data. *International Journal of Applied Earth Observation and Geoinformation*, 104, 102584. doi:10.1016/j.jag.2021.102584
- Schröder, J. J., Neeteson, J. J., Oenema, O., & Struik, P. C. (2000). Does the crop or the soil indicate how to save nitrogen in maize production? Reviewing the state of the art. *Field Crops Research*, 66, 151-164.
- Serbin, S. P., Dillaway, D. N., Kruger, E. L., & Townsend, P. A. (2012). Leaf optical properties reflect variation in photosynthetic metabolism and its sensitivity to temperature. *Journal of Experimental Botany*, 63, 489-502. doi:10.1093/jxb/err294

- Silva-Perez, V., Molero, G., Serbin, S. P., Condon, A. G., Reynolds, M. P., Furbank, R. T., & Evans, J. R. (2018). Hyperspectral reflectance as a tool to measure biochemical and physiological traits in wheat. *Journal of Experimental Botany*, 69, 483-496. doi:10.1093/jxb/erx421
- Skiba, U. M. (2014). Nitrous oxide, climate change and agriculture. *CAB Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources*, 9, 10. doi:10.1079/pavsnr20149010
- Terry, N., Waldron, L. J., & Taylor, S. E. (1983). Environmental influences on leaf expansion. In J. E. Dale & F. L. Milthorpe (Eds.), *The growth and functioning of leaves* (pp. 179-205). Cambridge, The United Kingdom: Cambridge University Press.
- Thomas, J. R., & Oerther, G. F. (1972). Estimating nitrogen content of sweet pepper leaves by reflectance measurements. *Agronomy Journal*, 64, 11-13.
- Verrelst, J., Malenovsky, Z., Van der Tol, C., Camps-Valls, G., Gastellu-Etchegorry, J.-P., Lewis, P., North, P., & Moreno, J. (2019). Quantifying vegetation biophysical variables from imaging spectroscopy data: A review on retrieval methods. *Surveys in Geophysics*, 40, 589-629. doi:10.1007/s10712-018-9478-y
- Wang, D., Rianti, W., Gálvez, F., van der Putten, P. E. L., Struik, P. C., & Yin, X. (2022). Estimating photosynthetic parameter values of rice, wheat, maize and sorghum to enable smart crop cultivation. *Crop and Environment*, 1, 119-132. doi:10.1016/j.crope.2022.05.004
- Wang, L., Chen, S., Li, D., Wang, C., Jiang, H., Zheng, Q., & Peng, Z. (2021). Estimation of paddy rice nitrogen content and accumulation both at leaf and plant levels from UAV hyperspectral imagery. *Remote Sensing*, 13, 2956. doi:10.3390/rs13152956
- Weiss, M., Jacob, F., & Duveiller, G. (2020). Remote sensing for agricultural applications: A meta-review. *Remote Sensing of Environment*, 236, 111402.
- Yao, X., Huang, Y., Shang, G., Zhou, C., Cheng, T., Tian, Y., Cao, W., & Zhu, Y. (2015). Evaluation of six algorithms to monitor wheat leaf nitrogen concentration. *Remote Sensing*, 7, 14939-14966. doi:10.3390/rs71114939
- Yin, X., & Struik, P. C. (2017). Can increased leaf photosynthesis be converted into higher crop mass production? A simulation study for rice using the crop model GECROS. *Journal of Experimental Botany*, 68, 2345-2360. doi:10.1093/jxb/erx085
- Yin, X., Struik, P. C., Romero, P., Harbinson, J., Evers, J. B., PE, V. D. P., & Vos, J. (2009). Using combined measurements of gas exchange and chlorophyll fluorescence to estimate parameters of a biochemical C₃ photosynthesis model: a critical appraisal and a new integrated approach applied to leaves in a wheat (*Triticum aestivum*) canopy. *Plant, Cell and Environment*, 32, 448-464. doi:10.1111/j.1365-3040.2009.01934.x
- Yin, X., & van Laar, H. H. (2005). *Crop systems dynamics: an ecophysiological simulation model for genotype-by-environment interactions*. Wageningen, The Netherlands: Wageningen Academic Publishers.
- Yu, K., Li, F., Gnyp, M. L., Miao, Y., Bareth, G., & Chen, X. (2013). Remotely detecting canopy nitrogen concentration and uptake of paddy rice in the Northeast China Plain. *ISPRS Journal of Photogrammetry and Remote Sensing*, 78, 102-115. doi:10.1016/j.isprsjprs.2013.01.008

- Zadoks, J. C., Chang, T. T., & Konzak, C. F. (1974). A decimal code for the growth stages of cereals. *Weed Research*, 14, 415-421.
- Zeng, F., Zuo, Z., Mo, J., Chen, C., Yang, X., Wang, J., Wang, Y., Zhao, Z., Chen, T., Li, Y., Zhang, Z., Hu, Z., & Xu, H. (2021). Runoff losses in nitrogen and phosphorus from paddy and maize cropping systems: A field study in Dongjiang Basin, South China. *Frontiers in Plant Science*, 12, 675121. doi:10.3389/fpls.2021.675121
- Zheng, H., Cheng, T., Li, D., Yao, X., Tian, Y., Cao, W., & Zhu, Y. (2018a). Combining unmanned aerial vehicle (UAV)-based multispectral imagery and ground-based hyperspectral data for plant nitrogen concentration estimation in rice. *Frontiers in Plant Science*, 9, 936. doi:10.3389/fpls.2018.00936
- Zheng, H., Cheng, T., Zhou, M., Li, D., Yao, X., Tian, Y., Cao, W., & Zhu, Y. (2018b). Improved estimation of rice aboveground biomass combining textural and spectral analysis of UAV imagery. *Precision Agriculture*, 20, 611-629. doi:10.1007/s11119-018-9600-7

Supplementary Tables in Chapter 3

Table S3.1 Split-applied nitrogen (N) fertiliser rates at different growth stages for rice, maize and winter wheat.

Crop species	Application stage	N application rate (kg N ha ⁻¹)						
		0	40	80	160	240	320	400
Rice	Pre-planting	0	12	24	48	72	96	-
	At the beginning of tillering	0	8	16	32	48	64	-
	At the end of tillering	0	8	16	32	48	64	-
	At panicle initiation	0	12	24	48	72	96	-
Maize/winter wheat	Pre-planting	0	16	32	64	96	128	160
	At the beginning of stem-elongating	0	14	28	56	84	112	140
	At inflorescence initiation	0	10	20	40	60	80	100

Table S3.2 Collected narrowband hyperspectral vegetation indices in this study.

Vegetation index	Equation*	References
Red-edge position (REP)	$\lambda_0 + \sigma$ where λ_0 and σ are fitted from inverted Gaussian model ($refl_{\lambda} = refl_s - (refl_s - refl_0)\exp(\frac{(\lambda_0 - \lambda)^2}{2\sigma^2})$) using least squares method, in which $refl_s$ is the maximum or “shoulder” spectral reflectance, $refl_0$ and λ_0 are the minimum spectral reflectance and corresponding wavelength, and σ is the Gaussian function deviation.	Miller et al. (1990)
N related	Normalised Difference Vegetation Index (NDVI_b) $(refl_{573} - refl_{470}) / (refl_{573} + refl_{470})$ Double-peak Canopy Nitrogen index (DCNI) $(refl_{720} - refl_{700}) / (refl_{700} - refl_{670}) / (refl_{720} - refl_{670} + 0.03)$ 3 bands Vegetation Index (VI_3bands) $(refl_{850} - refl_{700} + 2refl_{470}) / (refl_{850} + refl_{700} - 2refl_{470})$ Normalised Difference Index of the Double-peak Areas (NDDA_ig) $\frac{(refl_{755} - refl_{REP}) - (refl_{REP} - refl_{680})}{(refl_{755} - refl_{REP}) + (refl_{REP} - refl_{680})}$ Ratio Index of the Double-peak Areas (RIDA_ig) $\frac{refl_{755} - refl_{REP}}{refl_{REP} - refl_{680}}$ MCARI/MTVI2 $MCARI = [(refl_{700} - refl_{670}) - 0.2(refl_{700} - refl_{550})](refl_{700} / refl_{670})$ $MTVI2 = \frac{1.5[1.2(refl_{800} - refl_{550}) - 2.5(refl_{670} - refl_{550})]}{[(2refl_{800} + 1)^2 - (6refl_{800} - 5refl_{670}) - 0.5]}^{1/2}$	Hansen and Schjoerring (2003) Chen et al. (2010) Wang et al. (2012) Feng et al. (2014) Feng et al. (2014) Eitel et al. (2007)

(To be continued)

Table S3.2 (Continued)

Vegetation index	Equation*	References
Chlorophyll related		
Pigment Specific Simple Ratio (PSSR_r)	$refl_{800}/refl_{635}$	Blackburn (1998)
Pigment Specific Simple Ratio (PSSR_b)	$refl_{800}/refl_{470}$	Blackburn (1998)
Chlorophyll Index (CI_re)	$refl_{750}/refl_{725} - 1$	Gitelson et al. (2005)
Modified Simple Ratio (mSR_re)	$\frac{refl_{750}/refl_{705} - 1}{(refl_{750}/refl_{705} + 1)^{1/2}}$	Wu et al. (2008)
Chlorophyll Vegetation Index (CVI)	$(refl_{850} * refl_{670})/refl_{550}^2$	Vincini et al. (2008)
Triangular Vegetation Index (TVI)	$0.5[120(refl_{750} - refl_{550}) - 200(refl_{670} - refl_{550})]$	Broge and Leblanc (2000)
Transformed Chlorophyll Absorption in Reflectance Index (TCARI)	$3[(refl_{700} - refl_{670}) - 0.2(refl_{700} - refl_{550})(refl_{700}/refl_{670})]$	Haboudane et al. (2002)
Modified Chlorophyll Absorption Ratio Index (MCARI)	$[(refl_{700} - refl_{670}) - 0.2(refl_{700} - refl_{550})(refl_{700}/refl_{670})]$	Daughtry et al. (2000)
TCARI/OSAVI	$\frac{refl_{550}[(refl_{700}/refl_{670}) - 0.2(refl_{700} - refl_{550})(refl_{700}/refl_{670})]}{3[(refl_{700} - refl_{670}) - 0.2(refl_{700} - refl_{550})(refl_{700}/refl_{670})]}$	Haboudane et al. (2002)
MCARI/OSAVI	$\frac{[(refl_{700} - refl_{670}) - 0.2(refl_{700} - refl_{550})(refl_{700}/refl_{670})]}{(1 + 0.16)(refl_{850} - refl_{670})(refl_{850} + refl_{670} + 0.16)}$	alike TCARI/OSAVI
Modified Chlorophyll Absorption Ratio Index (MCARI_re)	$[(refl_{750} - refl_{705}) - 0.2(refl_{750} - refl_{550})(refl_{750}/refl_{705})]$	Wu et al. (2008)
TCARI/OSAVI_re	$\frac{refl_{550}[(refl_{750}/refl_{705}) - 0.2(refl_{750} - refl_{550})(refl_{750}/refl_{705})]}{3[(refl_{750} - refl_{705}) - 0.2(refl_{750} - refl_{550})(refl_{750}/refl_{705})]}$	Wu et al. (2008)
MCARI/OSAVI_re	$\frac{[(refl_{750} - refl_{705}) - 0.2(refl_{750} - refl_{550})(refl_{750}/refl_{705})]}{(1 + 0.16)(refl_{750} - refl_{705})(refl_{750} + refl_{705} + 0.16)}$	Wu et al. (2008)

(To be continued)

Table S3.2 (Continued)

Vegetation index		Equation*	References
Chlorophyll related	Triangular Chlorophyll Index (TCI)	$1.2(refl_{700} - refl_{550}) - 1.5(refl_{670} - refl_{550})(refl_{700}/refl_{670})^{1/2}$	Haboudane et al. (2008)
	Triangular Greenness Index (TGI)	$-0.5[(670 - 480)(refl_{670} - refl_{550}) - (670 - 550)(refl_{670} - refl_{480})]$	Hunt et al. (2011)
Biomass, coverary, or LAI related	Inverted Red-edge Chlorophyll Index (IRECI)	$(refl_{783} - refl_{665})/(refl_{705}/refl_{740})$	Frampton et al. (2013)
	Optimised Soil-Adjusted Vegetation Index (OSAVI)	$(refl_{850} - refl_{670})/(refl_{850} + refl_{670} + L), L=0.16$	Rondeaux et al. (1996)
	Wide Dynamic Range Vegetation Index (WDRVI)	$(\alpha * refl_{850} - refl_{670})/(\alpha * refl_{850} + refl_{670}), \alpha = 0.2$	Gitelson (2004)
	Normalised Difference Vegetation Index (NDVI_bg)	$[(refl_{850} - (refl_{550} + refl_{470}))]/[(refl_{850} + (refl_{550} + refl_{470}))]$	Wang et al. (2007)
	Modified Triangular Vegetation Index2 (MTVI2)	$1.5[1.2(refl_{800} - refl_{550}) - 2.5(refl_{670} - refl_{550})]/[(2refl_{800} + 1)^2 - (6refl_{800} - 5refl_{670}) - 0.5]^{1/2}$	Haboudane et al. (2004)
	Modified Chlorophyll Absorption Ratio Index 2 (MCARI2)	$1.5[2.5(refl_{800} - refl_{670}) - 1.3(refl_{800} - refl_{550})]/[(2refl_{800} + 1)^2 - (6refl_{800} - 5refl_{670}) - 0.5]^{1/2}$	Haboudane et al. (2004)
	Normalised Difference Vegetation Index (NDVI_gr)	$(refl_{550} - refl_{670})/(refl_{550} + refl_{670})$	Tucker (1979)
	Renormalised Difference Vegetation Index (RDVI_r)	$\frac{refl_{850} - refl_{670}}{(refl_{850} + refl_{670})^{1/2}}$	Roujean and Breon (1995)
	Renormalised Difference Vegetation Index (RDVI_g)	$\frac{refl_{850} - refl_{550}}{(refl_{850} + refl_{550})^{1/2}}$	Roujean and Breon (1995)
	(To be continued)		

Table S3.2 (Continued)

Vegetation index	Equation*	References
Biomass, coverary, or LAI related		
Modified Wide Dynamic Range Vegetation Index (mWDRVI)	$\frac{a * refl_{850} - refl_{670}}{a * refl_{850} + refl_{670}} + \frac{1-a}{1+a}, \alpha=0.2$	Peng and Gitelson (2011)
Simple Ratio Index (SRID_re2)	$deriv_{719}/deriv_{703}$	Shaw et al. (1998)
Visible Index (VI_700)	$(refl_{700} - refl_{670})/(refl_{700} + refl_{670})$	Gitelson et al. (2002)
Visible Index (VI_green)	$(refl_{550} - refl_{670})/(refl_{550} + refl_{670})$	Gitelson et al. (2002)
Soil and Atmospherically Resistant Vegetation Index (SARVI)	$\frac{(refl_{850} - [refl_{670} - \gamma(refl_{470} - refl_{670})](1+C)}{refl_{850} + [refl_{670} - \gamma(refl_{470} - refl_{670})] + C}, C=0.5, \gamma=1$	Kaufman and Tanre (1992)
Visible Atmospherically Resistant Index (VARI_700)	$\frac{refl_{700} - 1.7refl_{670} + 0.7refl_{470}}{refl_{700} + 2.3refl_{670} - 1.3refl_{470}}$	Gitelson et al. (2002)
Visible Atmospherically Resistant Index (VARI_green)	$(refl_{550} - refl_{670})/(refl_{550} + refl_{670} - refl_{470})$	Gitelson et al. (2002)

* *refl* and *deriv* represent the reflectance and the first order derivative at each particular band, respectively.

Table S3.3 Hyperparameters in selected nonparametric regression algorithms.

Method ^a	Model parameters	Definition	Value
PLSR SVR	"n_components"	Number of components to keep.	Ranged from 1 to 20 and in intervals of 1
	"kernel"	The specified kernel function type.	"rbf", "linear", "poly"
	"C"	The regularisation parameter.	0.1, 1, 10
RFR	"gamma"	The kernel coefficient.	"scale", "auto", 0.1, 1, 10
	"degree"	Degree of the polynomial kernel function ("poly").	2, 3 (only applicable for "poly" kernel)
	"coef0"	Independent term in kernel function.	1 (only applicable for "poly" kernel)
	"n_estimator"	The number of trees in the forest.	1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50
	"max_depth"	The maximum depth of the tree.	Ranged from 2 to 20 and in intervals of 2
	"min_samples_leaf"	The minimum number of samples required to split an internal node.	Ranged from 2 to 10 and in intervals of 1
	"min_samples_split"	The minimum number of samples required to be at a leaf node.	Ranged from 2 to 10 and in intervals of 1
DNN & CNN	"max_features"	The number of features to consider when looking for the best split.	"auto", "sqrt", "log2"
	"filters"	The dimensionality of the output space.	8, 16, 32
	"kernel_size"	Specifying the height and width of the moving window.	3
	"kernel_initializer"	Initialiser for the kernel weights matrix.	"he_normal"
	"activation"	Specifying the used activation function.	"relu" (only applicable for hidden layer), "linear" (only applicable for output layer)
	"optimiser"	Specifying the selected algorithm for optimisation.	"adam"
	"loss"	Computing the metrics for model evaluation.	<i>RMSE</i>
	"batch_size"	The number of samples per batch of computation.	20
	"epochs"	Specifying the number of training iterations on a dataset.	1000

^a Nonparametric regression methods: PLSR = Partial Least Squares Regression, SVR = Support Vector Regression, RFR = Random Forest Regression, DNN = Deep Neural Network, CNN = Convolution Neural Network.

Supplementary Figures in Chapter 3

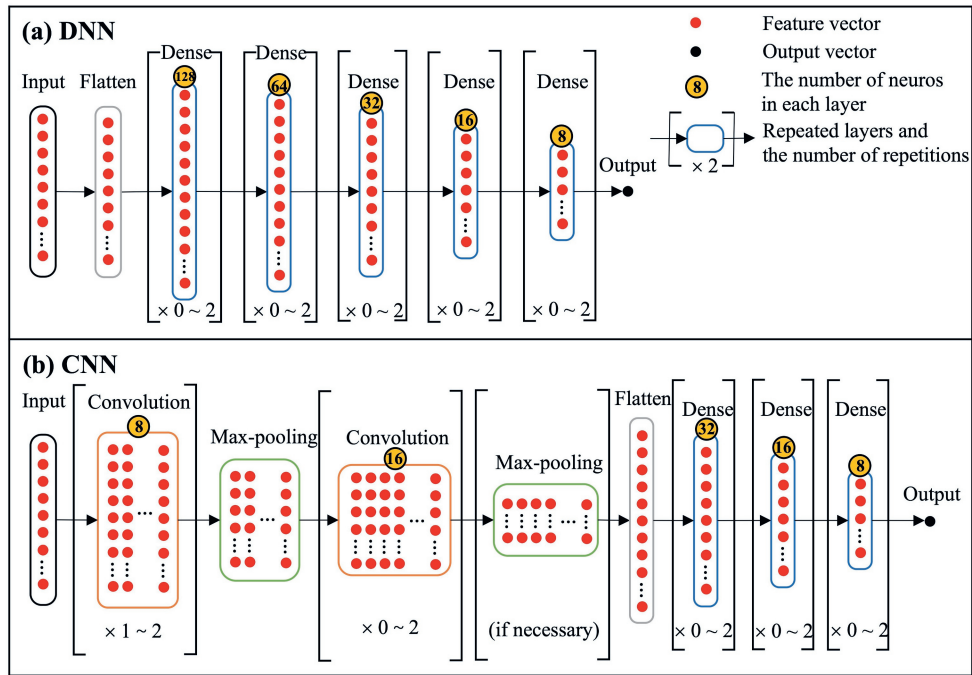


Fig. S3.1 A schematic illustration of deep learning-based neural network architectures of the full-connected deep neural network (DNN) (a) and convolution neural network (CNN) (b).

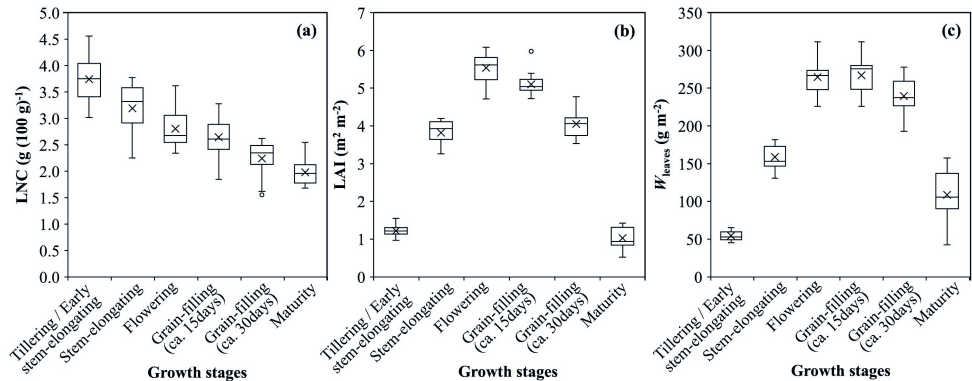


Fig. S3.2 The data distribution of destructively sampled leaf N concentration (LNC) (a), leaf area index (LAI) (b) and leaf weights (W_{leaves}) (c) in maize during the whole growing season.

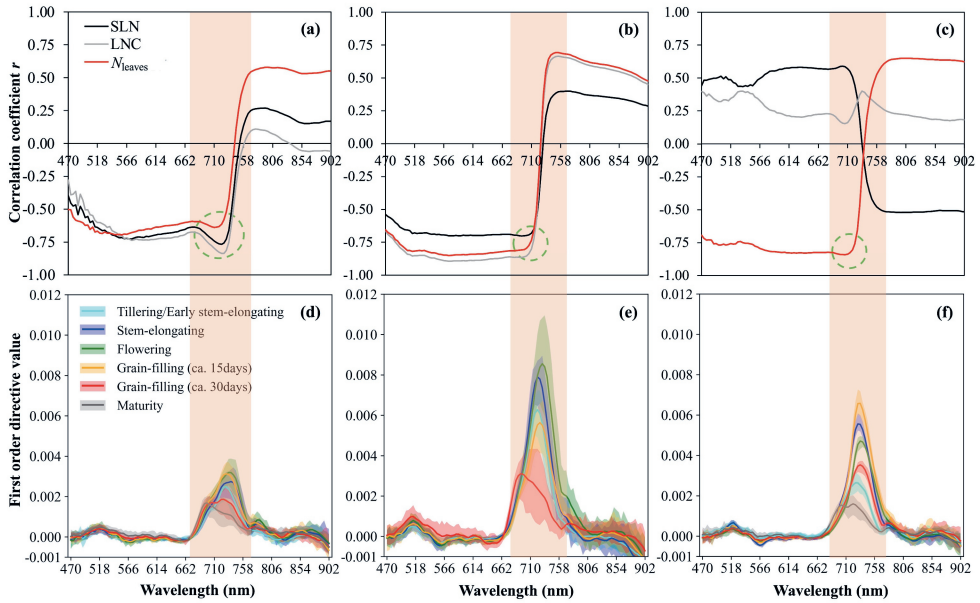


Fig. S3.3 The correlation coefficients (r) between specific leaf nitrogen (SLN), leaf nitrogen concentration (LNC) and canopy leaf-nitrogen content (N_{leaves}) in rice (a), wheat (b) and maize (c) with canopy reflectance. The orange bar represents the red-edge region in each crop species (d-f).

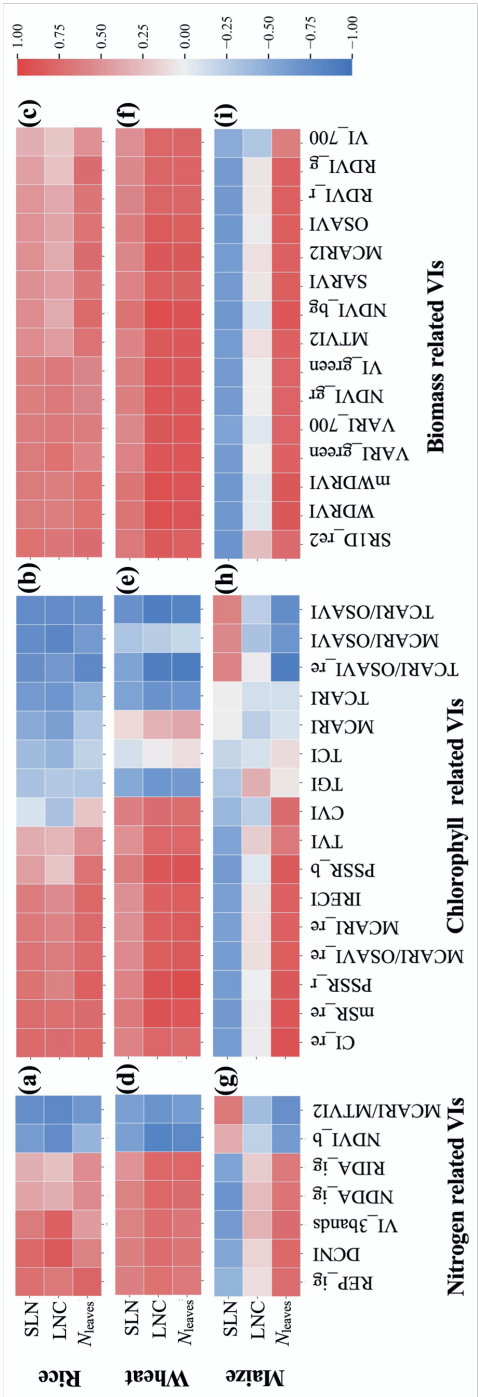


Fig. S3.4 The correlation coefficients (r) between specific leaf nitrogen (SLN), leaf nitrogen concentration (LNC) and canopy leaf-nitrogen content (N_{leaves}) of rice (a-c), wheat (d-f) and maize (g-i) with respect to collected nitrogen, chlorophyll, and biomass related vegetation indices. The definitions of vegetation indices (VIs) could be found in Table S3.2.

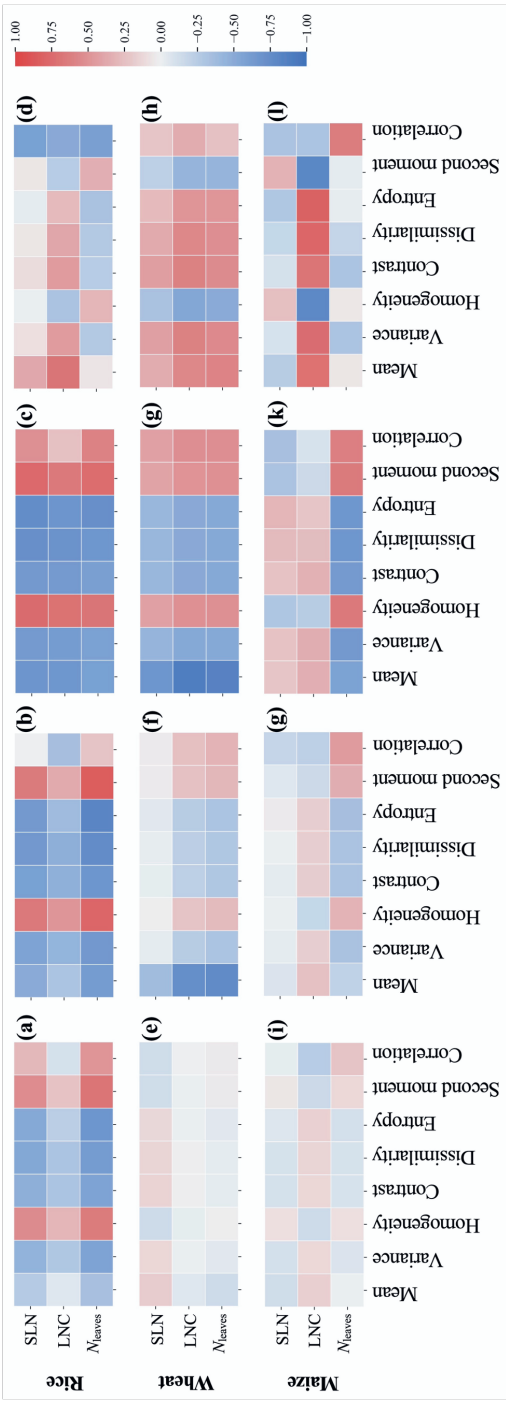


Fig. S3.5 The correlation coefficients (r) between specific leaf nitrogen (SLN), leaf nitrogen concentration (LNC) and canopy leaf nitrogen content (N_{leaves}) in rice (a-d), wheat (e-h) and maize (i-l) with respect to texture information extracted from bands of 470 (blue), 550 (green), 638 (red) and 850 (NIR) nm.

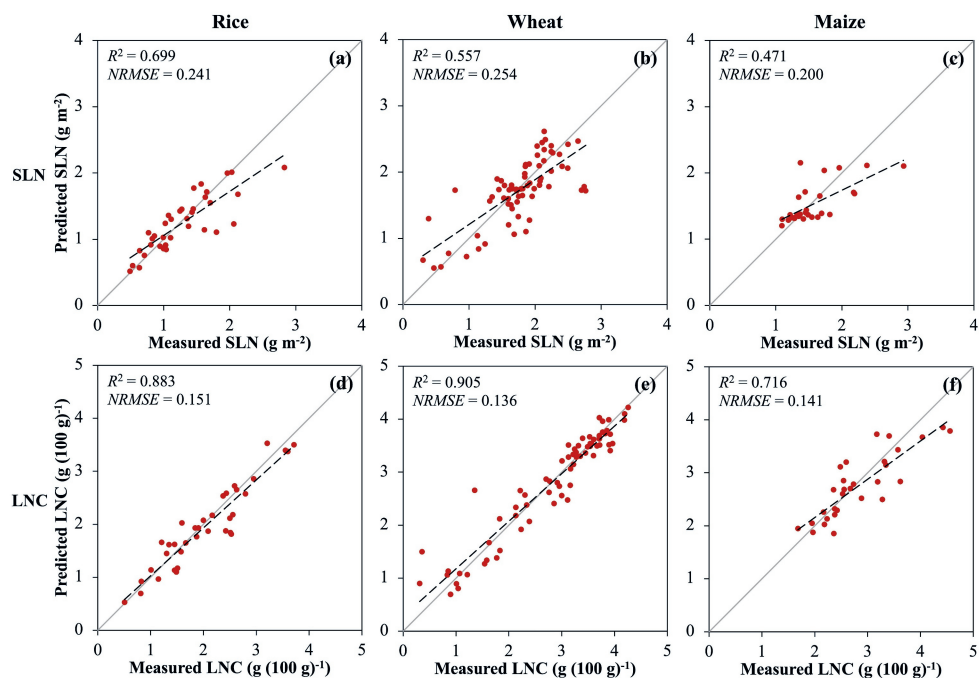


Fig. S3.6 The scatter plots for measured versus directly predicted specific leaf nitrogen (SLN) (a-c) and leaf nitrogen concentration (LNC) (d-f) in rice, wheat and maize using feature set of RefVIsTex (see Table 3.2 for its definition) from Partial Least Squares Regression algorithm in the testing dataset. The grey lines indicate a 1:1 relationship.

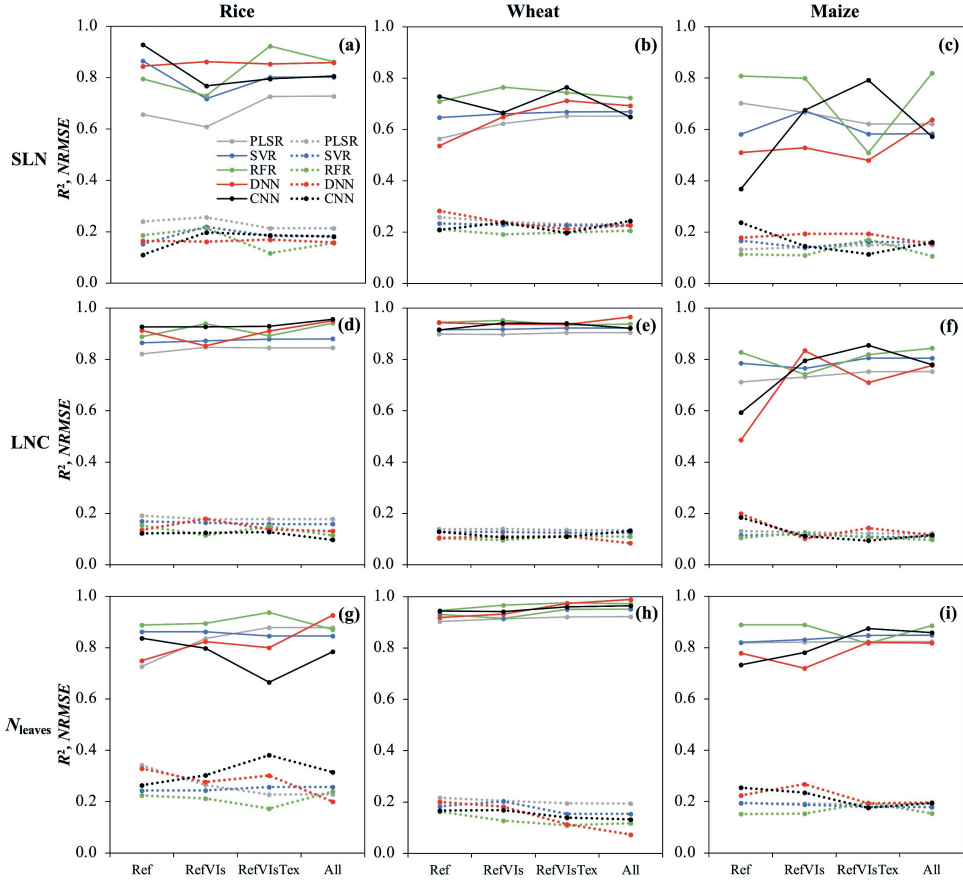


Fig. S3.7 The model performance, R^2 (full lines) and $NRMSE$ (dotted lines), in the training set of different regression algorithms in directly predicted specific leaf nitrogen (SLN) (a-c), leaf nitrogen concentration (LNC) (d-f) and canopy leaf-nitrogen content (N_{leaves}) (g-i) in rice, wheat and maize with various sets of input feature types. Nonparametric regression algorithms: PLSR = Partial Least Squares Regression, SVR = Support Vector Regression, RFR = Random Forest Regression, DNN = Deep Neural Network, CNN = Convolution Neural Network. Acronyms in the x-axis for four sets of feature types are defined in Table 3.2.

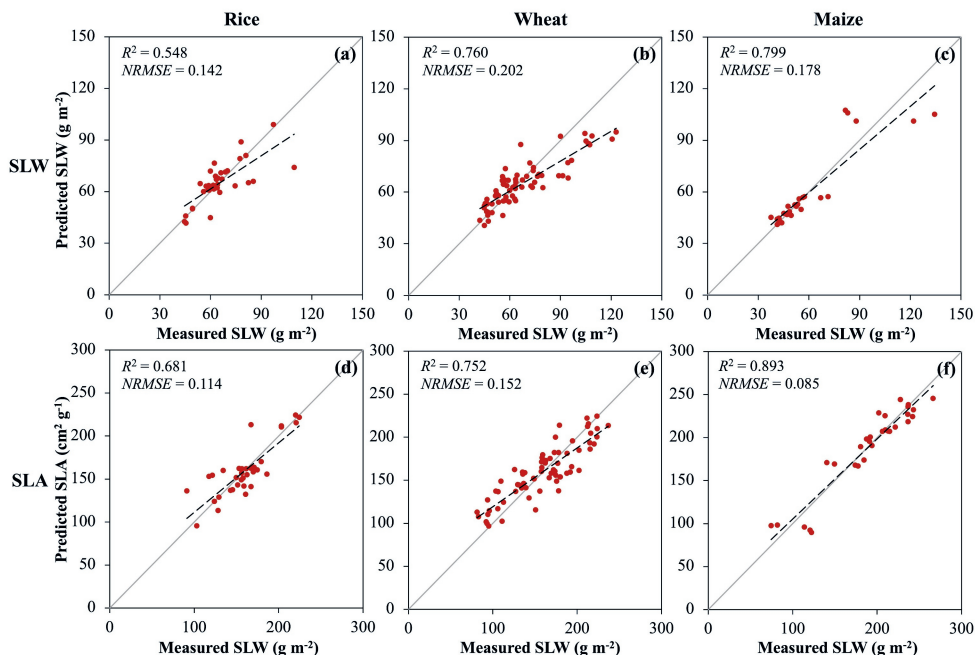


Fig. S3.8 The scatter plots for measured versus directly predicted specific leaf weight (SLW) (a-c) and specific leaf area (SLA) (d-f) in rice, wheat and maize in the testing dataset. The selected feature set was RefVIsTex (see Table 3.2 for its definition). The applications of the best performed regression algorithms of SLW and SLA were Support Vector Regression (SVR) (a) and SVR (d), Partial Least Squares Regression (PLSR) (b) and Convolution Neural Network (e), and Random Forest Regression (c) and PLSR (f) for rice, wheat and maize, respectively. The grey lines indicate a 1:1 relationship.

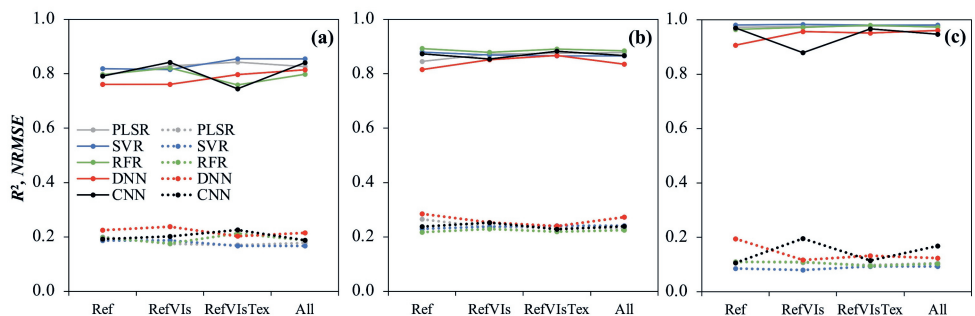


Fig. S3.9 The prediction performance, R^2 (full lines) and $NRMSE$ (dotted lines), in the testing dataset of different regression models in predicted leaf area index (LAI) in rice (a), wheat (b) and maize (c) with various sets of input feature types. Nonparametric regression algorithms: PLSR = Partial Least Squares Regression, SVR = Support Vector Regression, RFR = Random Forest Regression, DNN = Deep Neural Network, CNN = Convolution Neural Network. Acronyms in the x-axis for four sets of feature types are defined in Table 3.2.

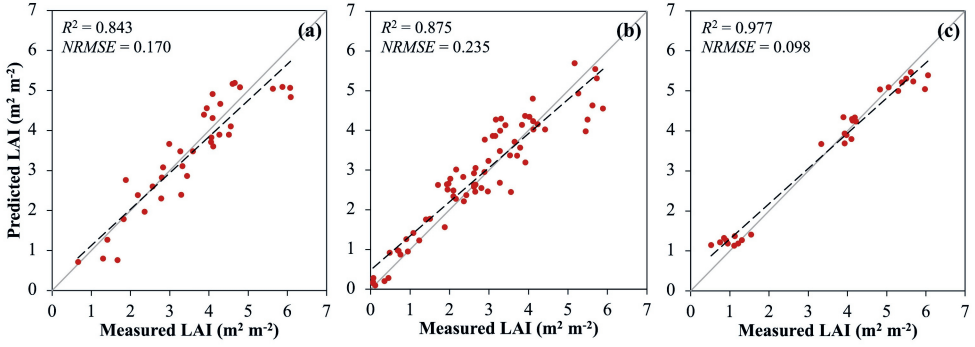


Fig. S3.10 The scatter plots for measured versus predicted leaf area index (LAI) in rice (a), wheat (b) and maize (c) in the testing dataset. The selected feature set was RefVIsTex (see Table 3.2 for its definition). The used regression algorithm was Partial Least Squares Regression. The grey lines indicate a 1:1 relationship.

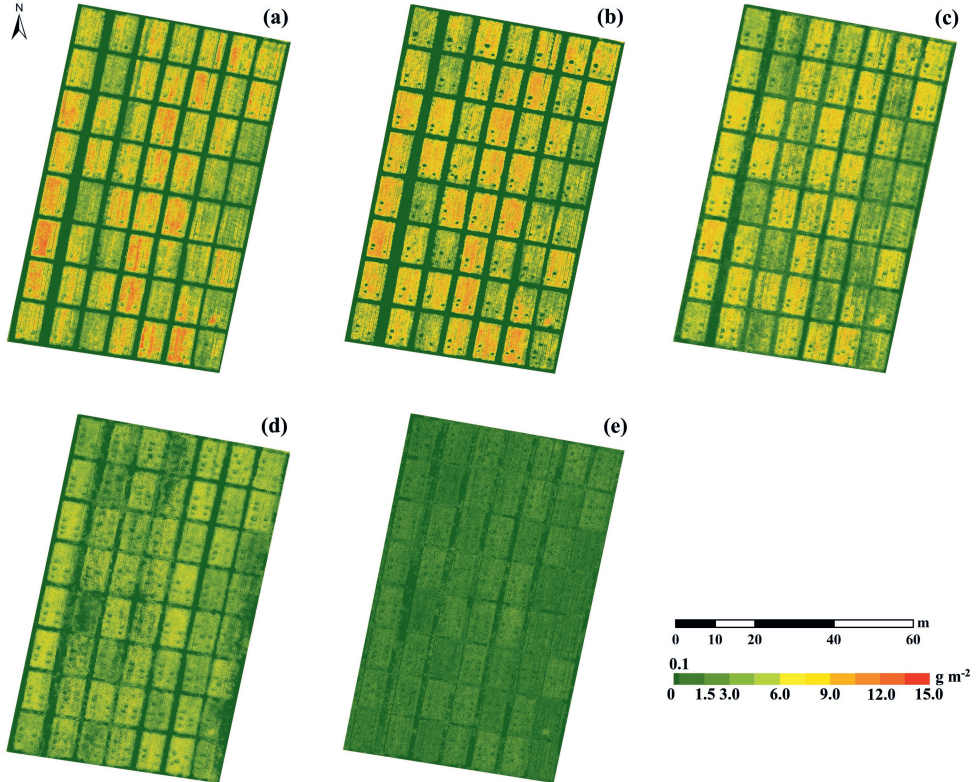


Fig. S3.11 The predicted canopy leaf-nitrogen content (N_{leaves}) from its direct predictions $N_{leaves,dir}$ (see Table 3.1 for its definition) at the experimental area of wheat across the whole growing season using feature set of RefVIsTex (see Table 3.2 for its definition) and regression algorithm of Partial Least Squares Regression. Predicted $N_{leaves,dir}$ at the stage of tillering (a), stem-elongating (b), flowering (c), grain-filling (ca. 15 days) (d), and grain-filling (ca. 30 days) (e) are shown successively.

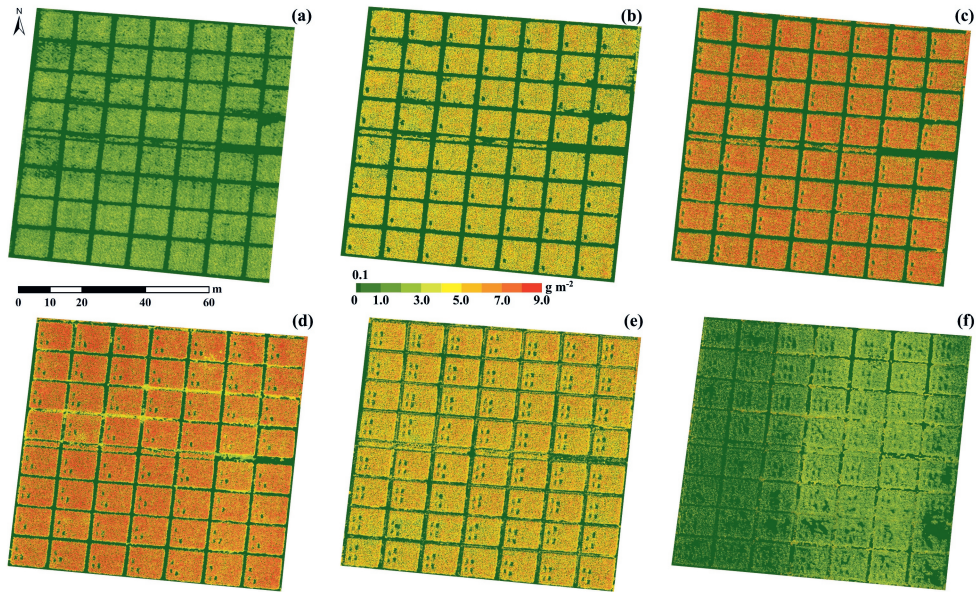


Fig. S3.12 The predicted canopy leaf-nitrogen content (N_{leaves}) from its indirect predictions $N_{\text{leaves,SLNdir}}$ (see Table 3.1 for its definition) at the experimental area of maize across the whole growing season using feature set of RefVIsTex (see Table 3.2 for its definition) and regression algorithm of Partial Least Squares Regression. Predicted $N_{\text{leaves,SLNdir}}$ at the stage of early stem-elongating (a), stem-elongating (b), flowering (c), grain-filling (ca. 15 days) (d), grain-filling (ca. 30 days) (e), and maturity (f) are shown successively.

References

- Blackburn, G. A. (1998). Quantifying chlorophylls and carotenoids at leaf and canopy scales: An evaluation of some hyperspectral approaches. *Remote Sensing of Environment*, 66, 273-285.
- Broge, N. H., & Leblanc, E. (2000). Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. *Remote Sensing of Environment*, 76, 156-172.
- Chen, P., Haboudane, D., Tremblay, N., Wang, J., Vigneault, P., & Li, B. (2010). New spectral indicator assessing the efficiency of crop nitrogen treatment in corn and wheat. *Remote Sensing of Environment*, 114, 1987-1997. doi:10.1016/j.rse.2010.04.006
- Daughtry, C. S. T., Walthall, C. L., Kim, M. S., De Colstoun, E. B., & McMurtrey III, J. E. (2000). Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. *Remote Sensing of Environment*, 74, 229-239.
- Eitel, J. U. H., Long, D. S., Gessler, P. E., & Smith, A. M. S. (2007). Using in-situ measurements to evaluate the new RapidEye™ satellite series for prediction of wheat nitrogen status. *International Journal of Remote Sensing*, 28, 4183-4190. doi:10.1080/01431160701422213
- Feng, W., Guo, B.-B., Wang, Z.-J., He, L., Song, X., Wang, Y.-H., & Guo, T.-C. (2014). Measuring leaf nitrogen concentration in winter wheat using double-peak spectral reflection remote sensing data. *Field Crops Research*, 159, 43-52. doi:10.1016/j.fcr.2014.01.010
- Frampton, W. J., Dash, J., Watmough, G., & Milton, E. J. (2013). Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 82, 83-92. doi:10.1016/j.isprsjprs.2013.04.007
- Gitelson, A. A. (2004). Wide dynamic range vegetation index for remote quantification of biophysical characteristics of vegetation. *Journal of plant physiology*, 161, 165-173.
- Gitelson, A. A., Kaufman, Y. J., Stark, R., & Rundquist, D. (2002). Novel algorithms for remote estimation of vegetation fraction. *Remote Sensing of Environment*, 80, 76-87.
- Gitelson, A. A., Vina, A., Ciganda, V., Rundquist, D. C., & Arkebauer, T. J. (2005). Remote estimation of canopy chlorophyll content in crops. *Geophysical Research Letters*, 32, L08403. doi:10.1029/2005GL022688, 2005
- Haboudane, D., Miller, J. R., Pattey, E., Zarco-Tejada, P. J., & Strachan, I. B. (2004). Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote Sensing of Environment*, 90, 337-352. doi:10.1016/j.rse.2003.12.013
- Haboudane, D., Miller, J. R., Tremblay, N., Zarco-Tejada, P. J., & Dextraze, L. (2002). Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. *Remote Sensing of Environment*, 81, 416-426.
- Haboudane, D., Tremblay, N., Miller, J. R., & Vigneault, P. (2008). Remote estimation of crop chlorophyll content using spectral indices derived from hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 46, 423-437.

- Hansen, P. M., & Schjoerring, J. K. (2003). Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. *Remote Sensing of Environment*, 86, 542-553. doi:10.1016/s0034-4257(03)00131-7
- Hunt, E. R., Daughtry, C. S. T., Eitel, J. U. H., & Long, D. S. (2011). Remote sensing leaf chlorophyll content using a visible band index. *Agronomy Journal*, 103, 1090-1099. doi:10.2134/agronj2010.0395
- Kaufman, Y. J., & Tanre, D. (1992). Atmospherically resistant vegetation index (ARVI) for EOS-MODIS. *IEEE Transactions on Geoscience and Remote Sensing*, 30, 261-270.
- Miller, J. R., Hare, E. W., & Wu, J. (1990). Quantitative characterization of the vegetation red edge reflectance 1. An inverted-Gaussian reflectance model. *International Journal of Remote Sensing*, 11, 1755-1773. doi:10.1080/01431169008955128
- Peng, Y., & Gitelson, A. A. (2011). Application of chlorophyll-related vegetation indices for remote estimation of maize productivity. *Agricultural and Forest Meteorology*, 151, 1267-1276. doi:10.1016/j.agrformet.2011.05.005
- Rondeaux, G., Steven, M., & Baret, F. (1996). Optimization of soil-adjusted vegetation indices. *Remote Sensing of Environment*, 55, 95-107.
- Roujean, J.-L., & Breon, F.-M. (1995). Estimating PAR absorbed by vegetation from bidirectional reflectance measurements. *Remote Sensing of Environment*, 51, 375-384.
- Shaw, D. T., Malthus, T. J., & Kupiec, J. A. (1998). High-spectral resolution data for monitoring Scots pine (*Pinus sylvestris* L.) regeneration. *International Journal of Remote Sensing*, 19, 2601-2608.
- Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8, 127-150. doi:10.1016/0034-4257(79)90013-0
- Vincini, M., Frazzi, E., & D'Alessio, P. (2008). A broad-band leaf chlorophyll vegetation index at the canopy scale. *Precision Agriculture*, 9, 303-319. doi:10.1007/s11119-008-9075-z
- Wang, F., Huang, J., Tang, Y., & Wang, X. (2007). New vegetation index and its application in estimating leaf area index of rice. *Rice Science*, 14, 195-203.
- Wang, W., Yao, X., Yao, X., Tian, Y., Liu, X., Ni, J., Cao, W., & Zhu, Y. (2012). Estimating leaf nitrogen concentration with three-band vegetation indices in rice and wheat. *Field Crops Research*, 129, 90-98. doi:10.1016/j.fcr.2012.01.014
- Wu, C., Niu, Z., Tang, Q., & Huang, W. (2008). Estimating chlorophyll content from hyperspectral vegetation indices: Modeling and validation. *Agricultural and Forest Meteorology*, 148, 1230-1241. doi:10.1016/j.agrformet.2008.03.005

Chapter 4

Enhancing crop growth forecasting by incorporating estimated uncertainties for time series hyperspectral image data and crop model GECROS into Ensemble Kalman Filter

Dong Wang^{1,2}, Paul C. Struik¹, Lei Liang² and Xinyou Yin¹

¹ Centre for Crop Systems Analysis, Department of Plant Sciences, Wageningen University & Research, 6700 AK Wageningen, The Netherlands

² Shanghai Lankuaikei Technology Development Co. Ltd., No. 888 Huanhu West 2nd Road, Pudong New District, Shanghai, China

Abstract

Accurate crop status forecasting benefits from assimilating remote sensing observations and crop model simulations. When conducting data assimilation (DA) using an Ensemble Kalman Filter (EnKF), arbitrary inflation factors are normally adopted to account for unspecified uncertainties, thus avoiding filter divergence. Here, we developed a Bayesian methodology in which the uncertainties were systematically quantified by combining disparate methods in one framework. Its applicability and performance with crop model GECROS using the EnKF framework were tested against data collected from two years of field experiments, in which aboveground biomass (W_{above}), grain weight (W_{grains}), aboveground nitrogen (N) content (N_{above}), grain N content (N_{grains}) and leaf traits like leaf dry weight, leaf N content and leaf area index were measured for rice. Using only the observations from the first year, the uncertain parameters in GECROS were calibrated by a Markov Chain Monte Carlo approach, while the parameters in the assumed error model that describes the uncertainties of crop model simulations were estimated simultaneously. The calibrated model parameters performed well in the validation year, except for the simulated leaf traits (Normalised Root Mean Squared Error ($NRMSE$) > 0.38). Remotely sensed leaf traits predicted by a Gaussian Process Regression (GPR) model were more accurate ($NRMSE < 0.34$), with uncertainties of the remote sensing observations estimated from the GPR model itself. Assimilating simulated and predicted leaf traits with their estimated uncertainties into EnKF prevented filter divergence, and the forecast accuracy of crop model improved in the validation year. Compared with simulation without assimilating in-season remote sensing observations, the $NRMSE$ of updated whole-season W_{above} and N_{above} both decreased from 0.37 to 0.20; and those of updated W_{grains} and N_{grains} at harvest decreased from 0.40 and 0.28 to 0.22 and 0.19, respectively. The developed method contributes to systematic uncertainty analysis in DA and accurate forecasting of in-season and end-of-season crop carbon and N status for smart farming.

Keywords: Crop growth, crop model, uncertainty quantification, Markov Chain Monte Carlo approach, Gaussian Process Regression, Ensemble Kalman Filter

4.1 Introduction

Accurate forecasting and mapping of spatiotemporal variabilities in aboveground biomass (W_{above}) and grain yield during the growing season are essential for informing farmers to perform field precision management under variable climatic conditions (Acevedo et al., 2020; Gao et al., 2017; Lobell and Azzari, 2017). Remote sensing science serves these purposes via connecting field measurements with sensor observations. For instance, crop yield forecasting based on regression models using field-measured yield and remote sensing features dates back to the 1970s (Idso et al., 1977). However, the poor exploration of model and data uncertainties has increasingly become a limitation for most remote sensing observations at high spatial resolution (Martínez-Ferrer et al., 2022). Although information like leaf area index (LAI, $\text{m}^2 \text{m}^{-2}$) and some weather variables can be incorporated into those regression methods to improve predictions (Johnson, 2014), the interactions within the continuum of atmosphere-crop-soil are widely overlooked. In this regard, it may be useful to explore dynamic crop models that have been developed since the 1960s to simulate crop growth and yield (e.g., de Wit, 1965; de Wit and Penning de Vries, 1985; Jones et al., 2003; Keating et al., 2003; Yin and van Laar, 2005) given that these models are based on in-depth understanding of crop physiological principles. However, as such crop models are generally developed and tested at the scale of a homogeneous plot, uncertainties are inevitable when applying them to heterogeneous farmers' fields. Uncertainties are also caused by incomplete knowledge of physiological processes, parameter values, meteorological conditions, soil properties and management practices (Hansen and Jones, 2000). For predictions using either remote sensing or crop modelling, all these uncertainties propagate, leading to bias in simulated in-season crop growth and end-of-season crop yield.

Combined utilisation of crop model simulations and remotely sensed observations is expected to produce a more accurate estimate than any of the two approaches on its own and attracts ever-increasing interests in smart farming (Houser et al., 2012; Jin et al., 2018). To this end, data assimilation (DA) methods have been developed (Jin et al., 2018). Monte Carlo-based Ensemble Kalman Filter (EnKF) (Evensen, 1994) is among the most popular methods for conducting DA (Carrassi et al., 2018), due to its simplicity, efficiency and adaptability to nonlinear and high-dimensional simulation models (Evensen, 2003; Kalnay et al., 2007). EnKF is an iterative procedure that keeps alternating between model forecasting and state updating. Each forecasting step produces an ensemble of different predictions that accounts for

uncertainty about model inputs, parameter values and model structure. Each updating step uses observations, weighted by measurement uncertainty, to correct the ensemble forecast. Sampling error can be minimised by using a large ensemble size (Whitaker and Hamill, 2012). However, inappropriately estimated system errors may lead to *filter divergence*, in which subsequent ensemble forecasts drift further from the truth (Anderson and Anderson, 1999; Jazwinski, 1970) and distributions of forecasted states become too narrow. Forecasting uncertainty is thus underestimated relative to observational uncertainty, making the observations essentially irrelevant. To alleviate filter divergence, additive or multiplicative inflation factors are commonly used (Huang et al., 2019). For instance, while assimilating remotely sensed soil moisture and LAI within the crop model DSSAT, a variant of EnKF, Ensemble Square Root Filter, was applied, in which fixed small inflation factors (1.05 for soil moisture and 1.50 for LAI) were included to prevent filter divergence (Ines et al., 2013). Instead of using fixed values, Kivi et al. (2022) adaptively estimated dynamic inflation factors to assimilate *in-situ* observed daily soil moisture for updating soil water and nitrogen (N) dynamics in the crop model APSIM. However, as inflation factors are not physically constrained, their application to complicated dynamic models with many different outputs is not straightforward (Ying and Zhang, 2015). Quantifying parameterisation errors in crop models and uncertainties of remotely sensed observations is indispensable when applying EnKF to achieve more accurate forecasts of crop growth status (Jin et al., 2018).

Parameter accuracy of a crop model significantly affects the performance of DA and yield forecasting (Kang and Özdoğan, 2019). To improve model parameter accuracy, various parameter inference methods have been developed and Bayesian approaches are becoming increasingly popular (e.g., Beven and Freer, 2001; Vrugt et al., 2009b). The interest in applying Bayesian approaches lies not only in inferring the most likely parameter values, but also in estimating their underlying posterior probability distribution functions (pdf) and even in estimating model structural error (Huang et al., 2019). Markov Chain Monte Carlo (MCMC) methods are typically used in these Bayesian approaches to link crop model simulations with observations. Based on data probability quantified by a likelihood function and, commonly, the Metropolis-Hasting search strategy (Hastings, 1970; Metropolis et al., 1953), the prior probability distribution for the parameters of the crop model and residual error model is updated to a posterior distribution conditioned by the information in the data. Normally, residual errors are assumed independent and identically distributed (i.i.d.), following a normal distribution with zero mean and constant variance (Box and Tiao, 1973). However, in-field observations

always have variable residuals throughout the growing season (Dumont et al., 2014). Thus, a likelihood function revised by observational variance was proposed for accounting for the heteroscedasticity in the crop model STICS (Dumont et al., 2014). A more generalised formal likelihood function based on a general error model was developed by Schoups and Vrugt (2010) for a hydrological model, which allows for the heteroscedasticity and non-Gaussian model residual errors. Their approach allows for diagnostic checking of residual error assumptions and does not require the i.i.d. assumption. As EnKF has been shown to be effective in cases with nonlinear or non-Gaussian errors (Han and Li, 2008; Katzfuss et al., 2016), investigating the applicability of integrating the calibrated uncertain parameters in a crop model with generalised error into the EnKF framework is in demand.

Errors in remote sensing data hamper the use of these data for predictions from nonparametric regression modelling, one of the most frequently used approaches to predict crop status from remote sensing data (Huang et al., 2019; Verrelst et al., 2019). Among the nonparametric models, the Gaussian Process Regression (GPR) model, developed within a Bayesian framework (Rasmussen and Williams, 2006), has been considered as a promising method, not only because of better prediction performance (Verrelst et al., 2012), but also because it quantifies predictive uncertainty (Berger et al., 2020a; Wang et al., 2019; Verrelst et al., 2019). Temporal and spatial transferability of GPR has been demonstrated by successfully transporting the GPR model to other images (Verrelst et al., 2013b). However, there is a need for comparison of DA from remote sensing (DA_{rs}) with DA from field measurements (DA_{fm}) (Huang et al., 2019). Due to destructive sampling, the sampling sites would not remain the same in ground observations but that inconsistency is normally neglected while conducting DA_{fm} . In contrast, in remote sensing predictions, although prediction errors always exist, temporal changes in crop growth can be predicted reasonably well. Thus, with those predominant merits of GPR, its performance upon further incorporating into the EnKF framework should be evaluated and compared with that of DA_{fm} .

Studies have been conducted to connect process-based simulations, field observations, and their uncertainties in order to have reliable forecasts using DA. In a hydrological modelling study, Vrugt et al. (2005) simultaneously estimated parameter uncertainties and structural errors as well as observational errors. In this approach, an inner EnKF loop for recursive state simulation and an outer global optimisation loop for posterior estimation were incorporated in simultaneous parameter estimation and data assimilation. However, even though model predictive ability is supposed to be enhanced by improved parameterisation, assessing model

structural and input errors may be hindered when parameter values are not fixed (Schoups and Vrugt, 2010). In an observing system simulation experiment that assimilated LAI and soil moisture data into the crop model SWAP, Hu et al. (2017) found that simultaneously updating parameters tended to worsen the performance of grain yield prediction when the uncertain parameters that directly determine biomass and grain formation were incorporated. A method for systematically quantifying uncertainties in the crop model simulation and remotely sensed observations from a separate Bayesian process and applying them into an EnKF framework is strongly needed for better forecasting of crop growth status. Such a method can help for the careful approximation and application of uncertainties in other DA algorithms or frameworks (Huang et al., 2019), and can be a potential reference of the desired model-data fusion framework for better Earth system forecasting (Gettelman et al., 2022).

The objective of this study was to develop a Bayesian methodology that combines disparate quantitative methods into one framework, i.e. incorporating the systematically analysed errors in crop model simulations and remote sensing observations into the data assimilation procedure of EnKF. We expect that this framework enhances the forecasting of the crop growth status. The methodology was validated in an actual case of a field experiment of rice. The crop model GECROS was selected for generating crop growth simulations, due to its generality and physiological robustness (Yin and Struik, 2017; Yin and van Laar, 2005) (see a brief description of the crop model GECROS in Supplementary Appendix A). Our specific objectives were: 1) to calibrate and validate GECROS under field conditions in China given the heteroscedastic and non-Gaussian residual error assumption; 2) to evaluate the performance of the GPR model for remote sensing prediction and its uncertainty estimation; 3) to assess the applicability of estimated uncertainties of the crop model simulations and the remote sensing observations in EnKF.

4.2 Materials and methods

Illustrations of study area and unmanned aerial vehicle-acquired time series hyperspectral images at experimental site are exhibited in Fig. 4.1. The framework of the developed method and its summarised description are shown in Fig. 4.2.

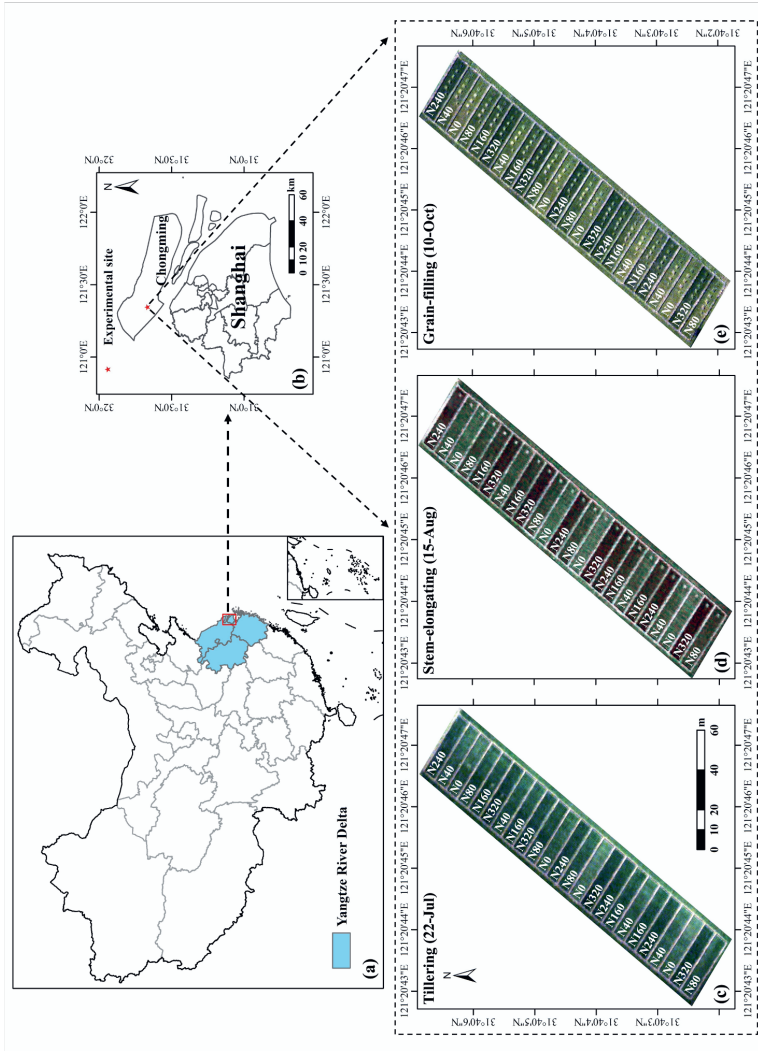


Fig. 4.1 Study area and the experiment site of rice (a, b). Illustrated true colour images (Red: 638 nm, Green: 550 nm, Blue: 470 nm) are drawn from time series hyperspectral images in 2020 (c-e). Different rates of nitrogen (N) application are denoted as N0, N40, N80, N160, N240 and N320 (see Table 4.2 for details).

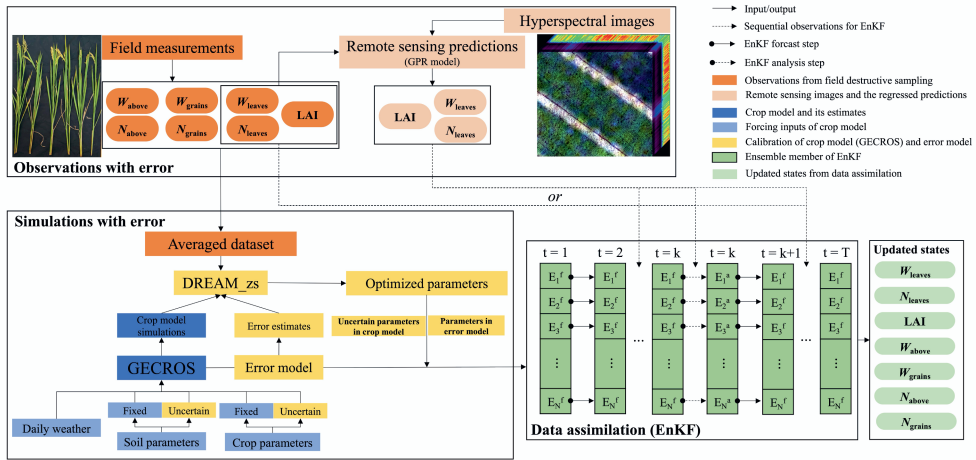


Fig. 4.2 Framework of the developed methodology in this study that applies systematically analysed errors of observations and simulations in the data assimilation procedure to enhance crop status forecast. Field experiments in two successive years were conducted for the acquisition of the necessary dataset to validate this method. Daily weather data served as the forcing input of the crop model GECROS and all field observations in the first year were averaged for model calibration. Before conducting the calibration procedure, the parameters in GECROS were fixed or treated as uncertain and the error model that describes the uncertainty of crop model simulation was assumed. The uncertain parameters in GECROS and parameters in the error model were determined simultaneously by an efficient Markov Chain Monte Carlo approach (DREAM_zs). To further improve the forecast ability of the crop model, the in-season observations of leaf traits in the second year were incorporated by the commonly used data assimilation procedure, Ensemble Kalman Filter (EnKF), which integrates the sequential observations into crop model simulations of crop growth processes. Two types of observations were collected. The first one was from field destructive measurements. The second type was from the remote sensing predictions, which were regressed from the machine learning method of Gaussian Process Regression (GPR). The uncertainties of field measurements were derived from the replications in the 2nd year's field experiment, while that of remote sensing predictions were estimated from the GPR model itself. With the systematically analysed uncertainties of crop model simulations and observations, in-season leaf traits at the canopy level, leaf weight (W_{leaves}), leaf nitrogen (N) content (N_{leaves}) and leaf area index (LAI) were updated directly by EnKF, and other crop carbon and N statuses, including aboveground biomass (W_{above}), grain weight (W_{grains}), aboveground N content (N_{above}) and grain N content (N_{grains}), were updated accordingly as well. The performance of the updated states in the second year was evaluated by their *in-situ* measurements, accordingly.

Table 4.1 Basic information of initial soil conditions at the experimental site.

	Soil type	Clay percentage (%)	Total organic carbon (g C m ⁻²)	Ammonium-N (g N m ⁻²)	Nitrate-N (g N m ⁻²)
2019	loam	18.1	7039	0.19	4.92
2020	loam	21.0	6376	0.17	4.43

4.2.1 Study site and field data

4.2.1.1 Experimental design

Field experiments for rice (*Oryza sativa* L., cv. Nanjing 46) were conducted at Chongming, Shanghai, China for two years (2019-2020). Basic soil information was collected before the experimentation (Table 4.1). The meteorological records of the closest weather station from our experimental site (ca. 5.7 km away) were downloaded from the website of the China meteorological data service centre (<http://data.cma.cn>). The weather variables include daily solar radiation, minimum and maximum temperatures, precipitation, vapour pressure, and wind speed. The mean air temperature and precipitation at the experimental site for the studied years are shown in Fig. S4.1, illustrating the typical weather conditions for rice cultivation in Yangtze River Delta.

Details about the field experiment were given previously (Wang et al., 2023). There were six treatments with varied N application amounts from 0 to 320 kg N ha⁻¹ per season in a complete randomised block design and three and four blocks in 2019 and 2020, respectively, in which each plot occupied an area of 180 m². Rice was sown with a row spacing of 20 cm and a within-row seed spacing of 2-3 cm on 14-Jun and 4-Jun in 2019 and 2020, respectively. The application time and amount of N were split according to the schedule in Table 4.2. Based on the local practice, sufficient phosphate (112.5 kg P₂O₅ ha⁻¹) and potash (112.5 kg K₂O ha⁻¹) fertilisers were applied per season to prevent phosphorus and potassium deficiencies. Other field managements were the same in each plot following local standard practices, including irrigation, and pest, weed and disease control.

Table 4.2 Split-applied nitrogen (N) fertiliser rates at different growth stages during the rice growing season in two experimental years.

Year	Application stage	N rate (kg N ha ⁻¹)					
		0	40	80	160	240	320
2019	Beginning of tillering	0	16	32	64	96	128
	Two weeks after tillering	0	8	16	32	48	64
	Panicle initiation	0	10	20	40	60	80
	Two weeks after flowering	0	6	12	24	36	48
2020	Seedling	0	12	24	48	72	96
	Beginning of tillering	0	8	16	32	48	64
	Two weeks after tillering	0	8	16	32	48	64
	Panicle initiation	0	12	24	48	72	96

Table 4.3 Destructive sampling dates at their corresponding rice growth stages in the experimental years 2019 and 2020.

	Tillering	Stem- elongating	Booting	Flowering	Grain-filling	Grain-filling	Maturity
2019	10-Jul	14-Aug	11-Sep	17-Sep	-	8-Oct	30-Oct
2020	22-Jul	15-Aug	-	8-Sep	21-Sep	10-Oct	30-Oct

4.2.1.2 Observations from field destructive sampling

Field destructive sampling was conducted in both years and its regimes in 2019 and 2020 are shown in Table 4.3. Seven types of observations at the canopy level were conducted: W_{above} , grain weight (W_{grains}), leaf weight (W_{leaves}), aboveground N content (N_{above}), grain N content (N_{grains}), leaf N content (N_{leaves}), and LAI. All observations in 2019 from different treatments were used for model calibration and those in 2020 were used for model validation.

Aboveground plants in an area of 1 and 2.25 m² in each plot were cut during each field measurement in 2019 and 2020, respectively. At harvest, two sampling sites were selected in each plot and their sampling areas were 2 and 4 m² in 2019 and 2020, respectively. The total fresh samples were bagged and weighed, while ca. 20% was dissected into component plant parts, including green leaves, yellow leaves, stems and leaf sheaths and panicle chaffs, and grains. The area of fresh green leaves and other green parts like stems was measured immediately using an LI-3100C Area Meter (Li-Cor, Lincoln, NE, USA). The dissected plant organs were weighed after oven-drying at 70 °C until constant weight and then ground to powder to assess N concentration using the Micro-Kjeldahl method. The N content in each plant organ was calculated as the product of their corresponding weight and N concentration. As all green parts are considered as functional leaves in GECROS (Yin and van Laar, 2005), the measured LAI, W_{leaves} and N_{leaves} were adjusted correspondingly following equations Eqns (S4.1-S4.3) (in Supplementary Appendix B) to make the crop model simulations and field measurements comparable.

4.2.2 Remote sensing observation and its uncertainty quantification

4.2.2.1 Collecting hyperspectral images

The field data collected in 2020 for validating GECROS were also utilised to validate the methodology proposed in this study. Thus, to obtain the remote sensing predictions, the

hyperspectral imaging data were collected in 2020 at the same date before each destructive sampling (Table 4.3). Details about canopy hyperspectral reflectance measurements were previously given (Wang et al., 2023). In brief, a DJI M600 PRO hexacopter (DJI, Shenzhen, China), equipped with a Cubert S185 hyperspectral snapshot camera (Cubert GmbH, Ulm, Baden-Württemberg, Germany), was flown over the experimental field between 10 a.m. and 2 p.m. We captured 125 spectral bands in the range of 450-950 nm with a sampling interval of 4 nm. Compared with using reflectance only, the feature set combining reflectance, vegetation indices and texture information worked better when predicting leaf traits (Wang et al., 2023). Thus, the complete dataset used for remote sensing prediction in this study was consistent with the combined features from hyperspectral image and targeted three leaf traits, including W_{leaves} , N_{leaves} and LAI.

4.2.2.2 Gaussian Process Regression (GPR) model calibration and validation

GPR establishes the relationship between the input features $x \in \mathbb{R}^B$ in the number of features B and the output variables (leaf traits) $y \in \mathbb{R}$ via the kernel function k , which defines the relationship between the pair of data points. The output variable values \mathbf{Y} and \mathbf{Y}_* of all training (\mathbf{x}) and testing (\mathbf{x}_*) data points are considered to be from a joint multivariate normal distribution (Rasmussen and Williams, 2006):

$$\begin{pmatrix} \mathbf{Y} \\ \mathbf{Y}_* \end{pmatrix} \sim \mathcal{N} \left(0, \begin{bmatrix} k(\mathbf{x}, \mathbf{x}) + \sigma_0^2 \mathbf{I} & k(\mathbf{x}, \mathbf{x}_*) \\ k(\mathbf{x}_*, \mathbf{x}) & k(\mathbf{x}_*, \mathbf{x}_*) \end{bmatrix} \right) \quad (4.1)$$

where $k(\mathbf{x}, \mathbf{x}_*)$ denotes the matrix of the covariances evaluated at all pairs of training and testing data points; the same applies to other entries of $k(\mathbf{x}, \mathbf{x})$, $k(\mathbf{x}_*, \mathbf{x})$ and $k(\mathbf{x}_*, \mathbf{x}_*)$. The observed output variables are assumed with i.i.d. Gaussian noise ($\mathcal{N}(0, \sigma_0^2)$) and \mathbf{I} represents the identity matrix.

The posterior distribution of \mathbf{Y}_* is estimated following (Rasmussen and Williams, 2006):

$$\mathbf{Y}_* | \mathbf{Y}, \mathbf{x}; \mathbf{x}_* \sim \mathcal{N}(y_{*,\mu}, y_{*,var}) \quad (4.2)$$

where the predicted posterior mean $y_{*,\mu}$ and variance $y_{*,var}$ are calculated as $k(\mathbf{x}_*, \mathbf{x})[k(\mathbf{x}, \mathbf{x}) + \sigma_0^2 \mathbf{I}]^{-1} \mathbf{Y}$ and $k(\mathbf{x}_*, \mathbf{x}_*) - k(\mathbf{x}_*, \mathbf{x})[k(\mathbf{x}, \mathbf{x}) + \sigma_0^2 \mathbf{I}]^{-1} k(\mathbf{x}, \mathbf{x}_*)$, respectively.

The commonly used anisotropic squared exponential kernel function is also adopted here (Verrelst et al., 2013a):

$$k(x_i, x_j) = \nu \exp \left(- \sum_{b=1}^B \frac{(x_i^{(b)} - x_j^{(b)})^2}{2\sigma_b^2} \right) + \sigma_0^2 \delta_{ij} \quad (4.3)$$

where ν is a scaling factor, σ_b is the length-scale per input feature b , controlling the spread of the relations for each input feature, and δ_{ij} is the Kronecker's symbol. The hyperparameters, which are denoted as $\theta_k = \{\nu, \sigma_b, \sigma_0\}$, are determined by maximising the log likelihood in the training set (Rasmussen and Williams, 2006):

$$\ell(\mathbf{Y}|\mathbf{x}, \theta_k) = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|k(\mathbf{x}, \mathbf{x}) + \sigma_0^2\mathbf{I}| - \frac{1}{2}\mathbf{Y}^T(k(\mathbf{x}, \mathbf{x}) + \sigma_0^2\mathbf{I})^{-1}\mathbf{Y} \quad (4.4)$$

where n represents the size of training dataset.

For calibrating and validating the GPR model, the acquired complete dataset was split into a training (75%) and a testing (25%) dataset. To avoid local maxima, the values of the hyperparameters in the GPR model were averaged from 100 iterations, and in each run, two-thirds of the training data were randomly selected from the whole training dataset (Verrelst et al., 2013a; Wang et al., 2019).

4.2.3 Crop model calibration and uncertainty estimation

4.2.3.1 Pre-set and uncertain parameters in GECROS

Before conducting the calibration procedure, key photosynthetic parameters in GECROS were pre-set with the values derived from our previous study (Wang et al., 2022). Meanwhile, phenological parameters, MTDV and MTD, and parameters like HTMX and STEMNC were determined by the field measurements (Table 4.4).

Uncertain parameters in GECROS influencing crop carbon and N assimilation, organ formation, and soil water and N movement were set free and allowed to vary within the referenced ranges (Table 4.4). Moreover, there was a newly added parameter, rASSA, accounting for the uncertainty of estimated photosynthetic parameters. This was considered necessary because photosynthetic parameter values estimated in our previous study based on gas exchange and chlorophyll fluorescence measurements (Wang et al., 2022), had uncertainties. Firstly, the cultivars differed between our previous study and this study. Secondly, the data in our previous study was only limited to certain growing stages rather than in the whole growing season like this study. Thirdly, plant growth conditions differed, in greenhouse vs in field conditions, and photosynthetic parameters are known to acclimate to growth conditions (e.g., Cai et al., 2020; Yin et al., 2019).

The remaining parameters in GECROS were set as default (Yin and van Laar, 2005).

Table 4.4 List of preset and uncertain parameters in the crop model GECROS with their definitions and references.

Parameter	Value or range	Units	Description	References
Preset				
SLNMIN	0.3	g N m^{-2}	Base leaf N for photosynthesis	Yin & Struik (2017)
MSR0	2.5	-	The ratio of mesophyll to stomatal resistance	Derived from data of Wang et al. (2022)
DD0C3	0.38	-	The slope of the decrease of $C_c:C_a$ ratio with increasing VPD	Derived from data of Wang et al. (2022)
XVN	210	$\mu\text{mol s}^{-1} \text{g}^{-1}$	Linear slope of maximum Rubisco activity at 25°C versus leaf N content	Derived from data of Wang et al. (2022)
XJN	255	$\mu\text{mol s}^{-1} \text{g}^{-1}$	Linear slope of maximum PSII e^- transport rate at 25 °C versus leaf N content	Derived from data of Wang et al. (2022)
MTDV	80.86	days	Minimum thermal days for vegetative growth phase	Field records
MTDR	27.74	days	Minimum thermal days for reproductive growth phase	Field records
HTMX	0.8	m	Maximum plant height	Field measurements
STEMNC	0.005	g N g^{-1}	Minimum N concentration in structural stem material	Field measurements
Free				
NUPTX	0.095-0.230	$\text{g N m}^{-2} \text{d}^{-1}$	Maximum daily crop nitrogen uptake	Derived from field measurements
rASSA	0.5-1.0	-	The factor for carbon assimilates, to collectively correct for the uncertainty of photosynthetic parameters	-
LNCI	0.02-0.05	g N g^{-1}	Initial nitrogen concentration in living leaves	Perturbation of default
SLA0	0.0150-0.0166	$\text{m}^2 \text{leaf g}^{-1}$	Specific leaf area constant	Field measurements
RDLV	0.01-0.10	-	Relative leaf death rate under conditions where seed fill is happened due to environmental stress, like drought	Perturbation of default
PMEH	0.6-0.9	-	Fraction of sigmoid curve inflexion in entire plant height growth period	Yin & van Laar (2005)

(To be continued)

Table 4.4 (Continued)

Parameter	Value or range	Units	Description	References
Free				
CDMHT	400-750	$\text{g m}^{-2} \text{m}^{-1}$	Stem dry weight per unit of plant height	Yin & van Laar (2005)
RVNC	0.001-0.025	g N g^{-1}	Nitrogen concentration in the reserve part of stem	Field measurements
SDF	15-35	-	Factor for biomass determined total seed number	Perturbation of default
PNPRE	0.6-1.0	-	Proportion of seed nitrogen that comes from none-structural nitrogen in vegetative organs accumulated before end of seed-number determining period	Yin & van Laar (2005)
SEEDW	0.0243-0.0278	g seed^{-1}	Seed weight	Field measurements
SEEDNC	0.0058-0.0130	g N g^{-1}	Standard seed (storage organ) nitrogen concentration	Field measurements
WCMIN	0.05-0.15	$\text{m}^3 \text{m}^{-3}$	Minimum soil water content	Yin & van Laar (2005)
WCFC	0.25-0.40	$\text{m}^3 \text{m}^{-3}$	Soil water content at holding capacity	Yin & van Laar (2005)
WCMAX	0.40-0.55	$\text{m}^3 \text{m}^{-3}$	Soil water content at maximum holding capacity	Yin & van Laar (2005)
BHC	2500-4000	g C m^{-2}	Initial value for microbial biomass and humified organic matter in the soil	Perturbation of default
HUMR	0.01-0.02	yr^{-1}	Decomposition rate constant for humified organic matter in the soil	Perturbation of default
BIOR	0.50-0.66	yr^{-1}	Decomposition rate constant for microbial biomass in the soil	Perturbation of default

4.2.3.2 The residual error assumption of the crop model

The relationship between the observations \mathbf{Y} and crop model simulations $\hat{\mathbf{Y}}(\cdot)$ can be presented as,

$$Y_{t,i} = \hat{Y}_t(\theta) + \varepsilon_t + \sigma_{t,i} \quad (4.5)$$

where $Y_{t,i}$ represents the i th replicated observation at the field destructive sampling date t , and $\hat{Y}_t(\theta)$ is the simulation given the crop model parameter set θ at t . ε_t represents the residual error ε of the simulation process at t and $\sigma_{t,i}$ is the observation error for $Y_{t,i}$. As $\sigma_{t,i}$ cannot be separated, especially in circumstances where there are replications for observations (Kennedy and O'Hagan, 2001), for simplicity, the averages of field destructive sampling data \bar{Y}_t were applied instead and the estimation of $\sigma_{t,i}$ was not included. As local information, including weather, soil and field managements, has been recorded in great detail and treated as *a priori* knowledge, the errors caused by those forcing inputs are reduced (Hansen and Jones, 2000). Thus, the residual error ε in this study represents mainly the structural error of the crop model. Accordingly, Eqn (4.5) is rewritten as,

$$\bar{Y}_t = \hat{Y}_t(\theta) + \varepsilon_t \quad (4.6)$$

To stabilise ε_t and reduce heteroscedasticity, the Box-Cox transformation of the measurements and simulations are introduced at first (Box and Cox, 1964),

$$\tau(Y, \lambda) = \begin{cases} (Y^\lambda - 1)/\lambda & \text{if } \lambda \neq 0 \\ \ln(Y) & \text{if } \lambda = 0 \end{cases} \quad (4.7)$$

where λ is the transformation parameter and varies in the range from 0 to 1.

To further deal with the heteroscedasticity and nonnormality, residual errors are proposed as according to Schoups and Vrugt (2010):

$$\Phi_p \varepsilon_t = \eta_t a_t \text{ with } a_t \sim \text{SEP}(0, 1, \xi, \beta) \quad (4.8)$$

where Φ_p represents the p th order autoregressive model, η_t and a_t account for the heteroscedasticity and non-Gaussian residual error at t , respectively. The heteroscedasticity is assumed to increase linearly with $\hat{\mathbf{Y}}$ ($\eta_t = \sigma_1 \hat{Y}_t$), where σ_1 ranges from 0 to 1. a_t denotes a random error with zero mean and unit standard deviation (STD), described by a standardised skew exponential power (SEP) density with parameter ξ and β to account for the nonnormality. The skewness parameter ξ ($\xi > 0$) affects the asymmetry of pdf and varies from 0.1 to 10, in which the density is symmetric when $\xi = 1$ and positively or negatively skewed when $\xi > 1$ or $\xi < 1$, respectively. The kurtosis parameter β determines the peakedness of the pdf and varies between -1 and $+1$. In the case of a symmetric density, the pdf is a uniform distribution

when $\beta = -1$, a Gaussian distribution when $\beta = 0$, and a Laplace or double-exponential distribution when $\beta = 1$. The marginal pdf for autoregressive models with Laplace innovations is typically quite complicated and the commonly used approximating approach is only valid for moderate to large sample sizes (Schoups and Vrugt, 2010), not suitable for the situations with limited field observations. Thus, in this study, the autoregressive models were excluded and β was assumed to be zero. Consequently, the simplified pdf of the SEP(0, 1, ξ , 0) in Eqn (4.8) is expressed as:

$$p(a_t|\xi) = \frac{1}{\sqrt{2\pi}} \frac{2\sigma_\xi}{\xi + \xi^{-1}} \exp\left\{-\frac{1}{2}|a_{\xi,t}|^2\right\} \quad (4.9)$$

where $a_{\xi,t} = \xi^{-\text{sign}(\mu_\xi + \sigma_\xi a_t)}(\mu_\xi + \sigma_\xi a_t)$, $\mu_\xi = M_1(\xi + \xi^{-1})$ and $\sigma_\xi = \sqrt{(M_2 - M_1^2)(\xi^2 + \xi^{-2}) + 2M_1^2 - M_2}$. For the standardised exponential power pdf, M_1 and M_2 can be obtained as $2/\sqrt{2\pi}$ and 1, respectively (Fernández and Steel, 1998; Schoups and Vrugt, 2010).

The resulting expression for the log-likelihood function of Eqns (4.6-4.9) is:

$$\ell(\theta, \lambda, \sigma_1, \xi|\bar{Y}) = -\frac{T}{2}\ln(2\pi) - \sum_{t=1}^T \ln \frac{(\tau(\bar{Y}_t, \lambda) - \tau(\bar{Y}_t(\theta), \lambda))}{\sigma_1 \bar{Y}_t(\theta)} + T \ln \frac{2\sigma_\xi}{\xi + \xi^{-1}} - \frac{1}{2} \sum_{t=1}^T |a_{\xi,t}|^2 \quad (4.10)$$

where T is the times of field sampling across the whole crop growing season.

4.2.3.3 Methods of combining likelihood values

To calibrate crop model parameters with multivariate observations, the Bayes' multiplication method was applied here (He et al., 2010). The combined log-likelihood $\ell_{combined}$ can be written as:

$$\ell_{combined}(\theta, \lambda, \sigma_1, \xi) = \sum_{m=1}^M \ell(\theta, \lambda_m, \sigma_{1_m}, \xi_m|\bar{Y}_m) \quad (4.11)$$

where λ , σ_1 and ξ denote the parameter sets of the residual error model for the multivariate observations. M represents the number of observation types. For the particular observation type m , λ_m , σ_{1_m} and ξ_m denote its residual error parameters, and \bar{Y}_m represents its averaged field destructive sampling data.

4.2.3.4 Calibration process and uncertainty estimation

After the specification of a prior parameter pdf, the uncertain parameters in crop model and residual error model can be calibrated by conditioning on the data through the combined log likelihood function of Eqn (4.11). In this study, the prior pdf of uncertain parameters was treated as a normal distribution. The prior means of crop model parameters were defined as midpoints

of their identified range, while their prior STDs were derived from the fixed coefficient of variation (CV) at 0.25. Regarding residual error parameters for multivariate observations, the prior means of λ , σ_1 and ξ were set as one, zero and one, respectively, and their prior STDs were 0.25. The DREAM-zs (DiffeREntial Evolution Adaptive Metropolis) algorithm was adopted to generate a representative sample from the posterior distribution. DREAM-zs is designed to accelerate convergence for high-dimensional problems, by sampling from an archive of past parameter candidates. Furthermore, DREAM-zs increases the diversity of candidate points by generating jumps beyond parallel direction updates (Schoups and Vrugt, 2010). Although the original DREAM-zs does not require outlier detection, the step of detecting an outlier chain (Vrugt et al., 2009a) was included in this study, as outlier chains can significantly deteriorate the performance of the MCMC sampler, especially when there is more than one type of observation.

The number of chains was set as two times the number of calibrated parameters following the suggestion of Vrugt et al. (2009b) and in total, 100 000 evaluations were conducted in the process of MCMC sampling. The outlier chain was detected until 60% evaluations were finished. The last 10% evaluations of each chain were compiled to calculate the mean and STD of the posterior pdf of each parameter.

As for the uncertainty estimation in the crop model, 2000 parameter sets were sampled from the posterior pdf to generate the corresponding simulations $\hat{\mathbf{Y}}$ and residual errors $\boldsymbol{\varepsilon}$. By calculating the 2.5% and 97.5% percentiles of simulations with or without $\boldsymbol{\varepsilon}$ for each type of observation, the uncertainties of the crop model caused by uncertain crop model parameters with or without model structural error were obtained. The estimation of $\boldsymbol{\varepsilon}$ involves generating independent samples from a SEP distribution. The followed sampling algorithm is described completely in Schoups and Vrugt (2010). With sampled residual error parameters, the corresponding residual errors $\boldsymbol{\varepsilon}$ were calculated using Eqn (4.8).

4.2.4 Data assimilation procedure

4.2.4.1 Ensemble Kalman Filter

In EnKF, the relation of the observation \mathbf{Y}_t to the model simulated state $\hat{\mathbf{Y}}_t$ can be described as:

$$\mathbf{Y}_t = \mathbf{H}\hat{\mathbf{Y}}_t + \boldsymbol{\varepsilon}_t \quad (4.12)$$

where \mathbf{H} is the measurement operator that maps the model state to the observation. The ensembles of observations and simulations of model state at t are stored in \mathbf{Y}_t and $\hat{\mathbf{Y}}_t$,

respectively. Both have a dimension of $N_y \times N_{ens}$, in which N_y and N_{ens} represent the dimension of observed states and the ensemble size, respectively.

The observations $\mathbf{Y}_{t,j}$ in ensemble \mathbf{Y}_t were drawn from a N_y -variate Gaussian distribution with mean equal to the observation, \mathbf{Y}_t^y , and covariance equal to \mathbf{R}_t ,

$$\mathbf{Y}_{t,j} = \mathbf{Y}_t^y + \boldsymbol{\sigma}_t^y \quad (4.13)$$

in which $\boldsymbol{\sigma}_t^y \sim \mathcal{N}(0, \mathbf{R}_t)$.

The model error covariance \mathbf{P}_t^f of $\hat{\mathbf{Y}}_t$ is calculated, using:

$$\mathbf{P}_t^f = (N_{ens} - 1)^{-1} \sum_{j=1}^{N_{ens}} (\hat{\mathbf{Y}}_{t,j} - \bar{\mathbf{Y}}_t^f)(\hat{\mathbf{Y}}_{t,j} - \bar{\mathbf{Y}}_t^f)^T \quad (4.14)$$

in which $\hat{\mathbf{Y}}_{t,j}$ and $\bar{\mathbf{Y}}_t^f$ represent the single simulation trajectory and the mean of the ensemble $\hat{\mathbf{Y}}_t$, respectively and the superscript T denotes the transpose of the matrix. $\bar{\mathbf{Y}}_t^f$ is calculated as $(\sum_{j=1}^{N_{ens}} \hat{\mathbf{Y}}_{t,j}) / N_{ens}$.

Under the linear assumptions, the updated analysis state \mathbf{Y}_t^a and its error covariance \mathbf{P}_t^a are calculated following:

$$\mathbf{Y}_t^a = \hat{\mathbf{Y}}_t + \mathbf{K}(\mathbf{Y}_t - \mathbf{H}\hat{\mathbf{Y}}_t) \quad (4.15)$$

$$\mathbf{P}_t^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}_t^f \quad (4.16)$$

in which \mathbf{I} is the identity matrix, and \mathbf{K} is the Kalman gain, defined as:

$$\mathbf{K} = \mathbf{P}_t^f \mathbf{H}^T (\mathbf{H} \mathbf{P}_t^f \mathbf{H}^T + \mathbf{R}_{t,cov})^{-1} \quad (4.17)$$

where $\mathbf{P}_t^f \mathbf{H}^T = (N_{ens} - 1)^{-1} \sum_{j=1}^{N_{ens}} (\hat{\mathbf{Y}}_{t,j} - \bar{\mathbf{Y}}_t^f)(\mathbf{H}\hat{\mathbf{Y}}_{t,j} - \mathbf{H}\bar{\mathbf{Y}}_t^f)^T$.

In summary, in the light of an ensemble of model trajectories, the EnKF approximates the probability density of the model states at each time step t . The updated mean of this ensemble represents the “best” state estimate, whereas the spread of the updated ensemble members provides a measure of the output uncertainty (Evensen, 1994).

4.2.4.2 Integration of the estimated uncertainties into Ensemble Kalman Filter

In this study, the simulated states were analysed and updated by the observations of W_{leaves} , N_{leaves} and LAI, which were from the field measurements or from the remote sensing observations. Thus, N_y was set at three. Meanwhile, N_{ens} was set at 100 in this study, due to the reasonable trade-off between efficiency and pdf representation (de Wit and van Diepen, 2007). The residual error $\boldsymbol{\varepsilon}_t$ of the crop model simulations was generated following the

abovementioned procedure (in Section 4.2.3.4). For the field observations, \mathbf{Y}_t^y was the measured data at each sampling date, and \mathbf{R}_t was derived from their replications, assuming that the CV for each single measurement was the same as the CV of the replications of each treatment at each sampling date. As for remotely sensed observations, both \mathbf{Y}_t^y and \mathbf{R}_t were predicted from the regressed GPR model. Moreover, the performances of crop model simulations without assimilating observations (*open-loop*) were also evaluated for comparison. To test the end-of-season forecast ability of the updated crop model, the assimilation of leaf traits ended at the grain-filling stage (20 days before maturity).

4.2.5 Evaluation metrics

With respect to the corresponding observations y_i , the performance of \hat{y}_i , simulations in model calibration and validation or updated simulations in DA, was evaluated by the coefficient of determination (R^2), the root mean square error ($RMSE$) and the normalised $RMSE$ ($NRMSE$):

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (4.18)$$

$$RMSE = \sqrt{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / (n - 1)} \quad (4.19)$$

$$NRMSE = RMSE / \bar{y} \quad (4.20)$$

where n represents the number of evaluated data points, and \bar{y} is the mean value of observations across the whole growing season.

Moreover, to assess the filter behaviour of EnKF, we introduced f_{rc} , which represents the relative change of the states after and before updating compared with the measured y_i ,

$$f_{rc} = \frac{\bar{y}_i^a - y_i}{\bar{y}_i^f - y_i} \quad (4.21)$$

in which \bar{y}_i^a is calculated as $(\sum_{j=1}^{N_{ens}} Y_{t,j}^a) / N_{ens}$. The closer f_{rc} is to one, the more likely does filter divergence occur.

4.3 Results

4.3.1 Bayesian calibration of GECROS

The varying range of residual error parameter σ_1 affected the simulation performance in calibration using the data from the 2019 field experiment (Fig. 4.3). When the upper bound of σ_1 was set at 1.0, the simulated W_{grains} , N_{leaves} and LAI had the highest $NRMSE$. After decreasing the upper bound of σ_1 , their $NRMSE$ tended to decrease, although the $NRMSE$ of simulated

N_{grains} tended to increase (Fig. 4.3a). Meanwhile, R^2 of simulations in calibration also varied with the upper bound of σ_1 . Especially, the R^2 of simulations while setting the upper bound of σ_1 at 0.2 tended to be higher than those while setting upper bounds of σ_1 at other levels. Due to the overestimation (results not shown), even though simulated N_{grains} achieved the lowest $NRMSE$ while setting the upper bound of σ_1 at 1.0, its R^2 values were rather low (Fig. 4.3b). The estimated means of posterior distribution of σ_1 for most of the observations were close to the fixed upper bound (Table 4.5, Fig. S4.2a-c). Although fixing the value of upper bound of σ_1 at 1.0 reduced this phenomenon, due to the unexpected residual error from the upper bound expansion of σ_1 , the estimated σ_1 and ξ did not converge well during the Bayesian calibration process (see Fig. S4.2 for the simulated W_{above} as an example). Consequently, based on the sensitivity analysis of the upper bound of σ_1 , the upper bound of σ_1 was set at 0.2 to conduct the following analysis.

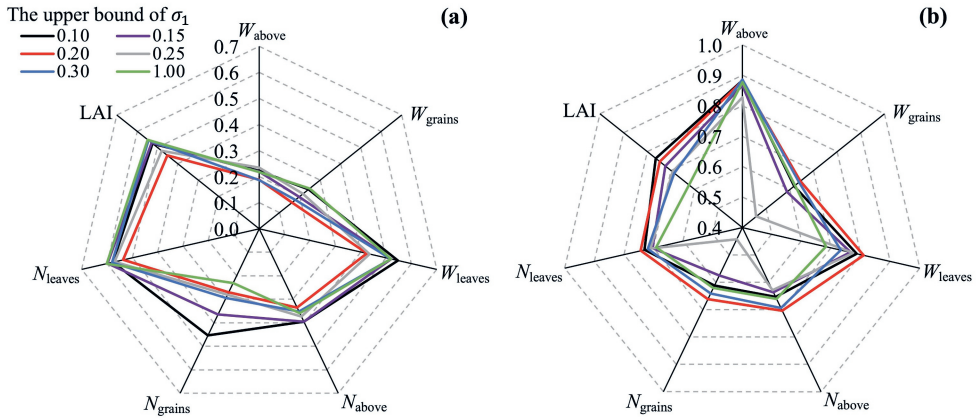


Fig. 4.3 Comparison of simulation performance in calibration, $NRMSE$ (a) and R^2 (b), in which the uncertain parameters in GECROS were calibrated with different upper bounds of the heteroscedasticity parameter σ_1 . W_{above} , aboveground biomass; W_{grains} , grain weight; W_{leaves} , leaf weight; N_{above} , aboveground N content; N_{grains} , grain N content; N_{leaves} , leaf N content in the canopy; and LAI, leaf area index.

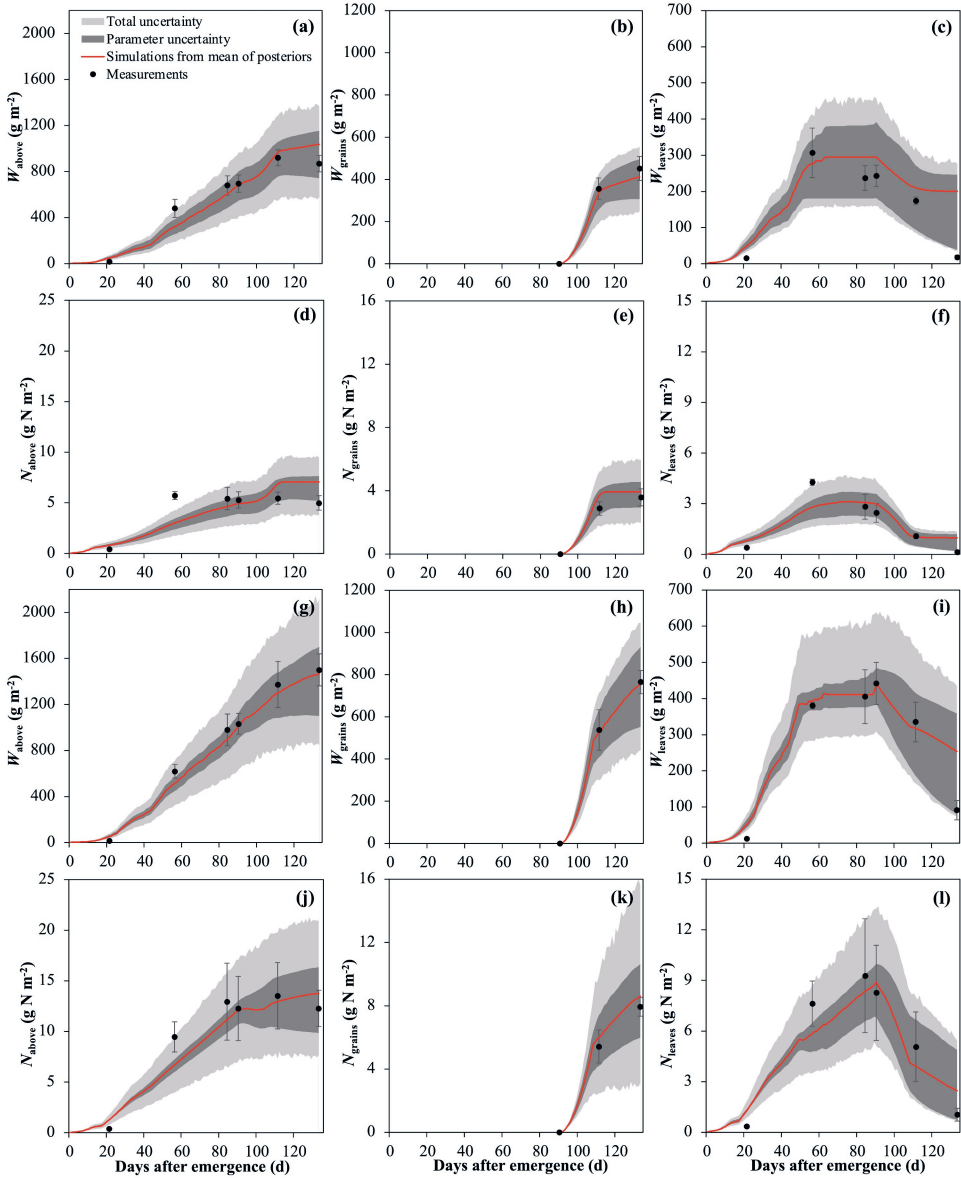


Fig. 4.4 Estimation of parameter (dark grey region) and total (light grey region) uncertainty of GECROS with the nitrogen input of 0 kg ha⁻¹ (a-f) and 240 kg ha⁻¹ (g-l) in 2019. Red lines represent the simulations from estimated mean values of uncertain parameters. The solid circles with vertical bars denote the means and standard deviations of measurements. W_{above} , aboveground biomass; W_{grains} , grain weight; W_{leaves} , leaf weight; N_{above} , aboveground N content; N_{grains} , grain N content; and N_{leaves} , leaf N content in the canopy.

Table 4.5 Estimated values of means and standard deviations (STDs) of posterior distributions of uncertain parameters in GECROS and parameters in error model while setting the upper bound of the heteroscedasticity parameter σ_1 at 0.2.

Parameters in the GECROS model ^a	Mean	STD	Parameters in the error model ^b	Mean	STD
NUPTX	0.206	0.018	$\sigma_{1,W_{above}}$	0.189	0.009
rASSA	0.705	0.024	$\xi_{W_{above}}$	0.976	0.171
LNCI	0.0203	0.0005	$\lambda_{W_{above}}$	1.000	0.000
SLA0	0.0164	0.0003	$\sigma_{1,W_{grains}}$	0.162	0.026
RDLV	0.017	0.013	$\xi_{W_{grains}}$	2.860	1.309
PMEH	0.873	0.028	$\lambda_{W_{grains}}$	0.999	0.001
CDMHT	627.3	89.0	$\sigma_{1,W_{leaves}}$	0.191	0.007
RVNC	0.008	0.006	$\xi_{W_{leaves}}$	0.427	0.180
SDF	29.28	3.99	$\lambda_{W_{leaves}}$	0.999	0.001
PNPRE	0.877	0.086	$\sigma_{1,N_{above}}$	0.196	0.005
SEEDW	0.026	0.001	$\xi_{N_{above}}$	0.952	0.122
SEEDNC	0.011	0.001	$\lambda_{N_{above}}$	0.977	0.024
WCMIN	0.099	0.026	$\sigma_{1,N_{grains}}$	0.149	0.034
WCFC	0.304	0.040	$\xi_{N_{grains}}$	1.746	1.077
WCMAX	0.459	0.039	$\lambda_{N_{grains}}$	0.718	0.229
BHC	3899	100	$\sigma_{1,N_{leaves}}$	0.190	0.009
HUMR	0.018	0.001	$\xi_{N_{leaves}}$	0.745	0.140
BIOR	0.629	0.026	$\lambda_{N_{leaves}}$	0.944	0.063
			$\sigma_{1,LAI}$	0.190	0.007
			ξ_{LAI}	0.646	0.174
			λ_{LAI}	0.943	0.069

^a GECROS-model parameters and their units are defined in Table 4.4;

^b σ_1 , ξ and λ denote the heteroscedasticity parameter, the skewness parameter and the Box-Cox transformation parameter in the error model, respectively (see the text). Definition of the traits used in the subscript of symbols: W_{above} , aboveground biomass; W_{grains} , grain weight; W_{leaves} , leaf weight; N_{above} , aboveground N content; N_{grains} , grain N content; N_{leaves} , leaf N content in the canopy; and LAI, leaf area index.

While fixing the upper bound of σ_1 at 0.2, the simultaneously calibrated uncertain parameters in GECROS and residual error parameters are shown in Table 4.5. Compared with W_{above} and N_{above} , the estimated error parameters ξ for W_{leaves} , N_{leaves} , LAI, W_{grains} and N_{grains} were farther away from one, indicating that the residual errors in those types of observations were supposed to be non-Gaussian (Table 4.5). Besides the residual error parameters, the uncertain parameters in GECROS were calibrated reasonably well (Table 4.5). Meanwhile, the

uncertainty problem caused by the pre-set photosynthetic parameters was solved by introducing the parameter *rASSA* (Table 4.4), and the overestimation in the simulations was reduced accordingly (Fig. S4.3).

With the estimated posterior distributions of the uncertain parameters in the crop model GECROS (Fig. S4.4), the parameter uncertainty was analysed for the field measurements in 2019, while the total uncertainty was calculated by further integrating the estimated residual error (Fig. 4.4). Taking no N input (0 kg N ha⁻¹) and locally common N input (240 kg N ha⁻¹) as examples, the simulations of different types of observations agreed well with the measured data and most of the points of the average measurements were within the range of simulated total uncertainty (Fig. 4.4).

4.3.2 Performance validating of the calibrated GECROS model

With the calibrated posterior distributions of uncertain parameters in GECROS, the performance of the simulations was validated against the data from the 2020 field experiment. The *NRMSE* and *R*² are listed in Table 4.6.

Compared with that in the calibration, the performance of the crop model GECROS in the validation tended to become worse, especially for *W*_{leaves}, *N*_{leaves} and LAI (Table 4.6). For simulated *W*_{grains} and *N*_{grains}, however, the performance improved in the validation. This was probably caused by the extra field destructive sampling in 2020 at the stage of early grain filling (Fig. S4.5). The poor performance in simulating *W*_{leaves}, *N*_{leaves} and LAI, probably stemmed from the underestimation when leaf growth approached the peak (Fig. S4.6).

Table 4.6 *NRMSE* and *R*² of simulations in the calibration (year 2019) and the validation (year 2020) using estimated mean values of uncertain parameters in the crop model GECROS ^a.

		<i>W</i> _{above}	<i>W</i> _{grains}	<i>W</i> _{leaves}	<i>N</i> _{above}	<i>N</i> _{grains}	<i>N</i> _{leaves}	LAI
Calibration	<i>NRMSE</i>	0.162	0.171	0.332	0.280	0.266	0.431	0.362
(2019)	<i>R</i> ²	0.882	0.645	0.810	0.704	0.662	0.743	0.747
Validation	<i>NRMSE</i>	0.282	0.153	0.486	0.386	0.180	0.590	0.388
(2020)	<i>R</i> ²	0.873	0.893	0.515	0.732	0.849	0.725	0.579

^a To make the performance in the calibration and the validation comparable, the mean values of two years' field measurements for each type of observation were used here for calculating *NRMSE*. *W*_{above}, aboveground biomass; *W*_{grains}, grain weight; *W*_{leaves}, leaf weight; *N*_{above}, aboveground N content; *N*_{grains}, grain N content; *N*_{leaves}, leaf N content in the canopy; and LAI, leaf area index.

4.3.3 Observations from remote sensing prediction

The leaf traits, like W_{leaves} , N_{leaves} and LAI, were predicted using the GPR modelling of hyperspectral image data, as leaf traits are more likely observed and thus able to be robustly predicted from collected remote sensing features, compared with other observations such as W_{above} , W_{grains} , N_{above} and N_{grains} . The performance of their site-specific predictions was evaluated in the training (Fig. 4.5a-c) and testing (Fig. 4.5d-f) datasets. We tried to reduce the overfitting by averaging the values of estimated hyperparameters from the repeated training subsets (Section 4.2.2.2). The predicted leaf traits in the training set performed better than those in the testing set. The R^2 values of the predicted leaf traits in the training set were higher than 0.95, while those in the testing set were lower than 0.88. The $NRMSE$ of predicted leaf traits in training set ranged from 0.110 to 0.172, while the $NRMSE$ in the testing test increased and varied from 0.175 to 0.336 (Fig. 4.5). Compared with N_{leaves} (Fig. 4.5b, e), W_{leaves} (Fig. 4.5a, d) and LAI (Fig. 4.5c, f) fitted with their measurements better in both the training set and the testing set.

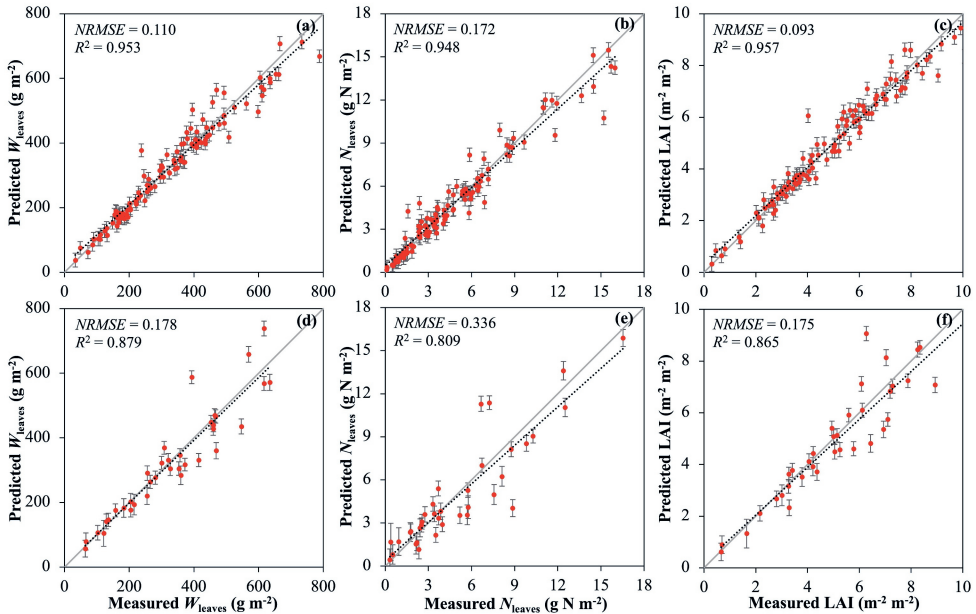


Fig. 4.5 Scatter plots of predicted leaf weight (W_{leaves}), leaf N content in the canopy (N_{leaves}) and leaf area index (LAI) set from hyperspectral image data with respect to their measurements in the training (a-c) and testing (d-f) set, using the Gaussian Process Regression (GPR) algorithm. The vertical bars denote the predicted standard deviations from the GPR. The grey lines indicate a 1:1 relationship.

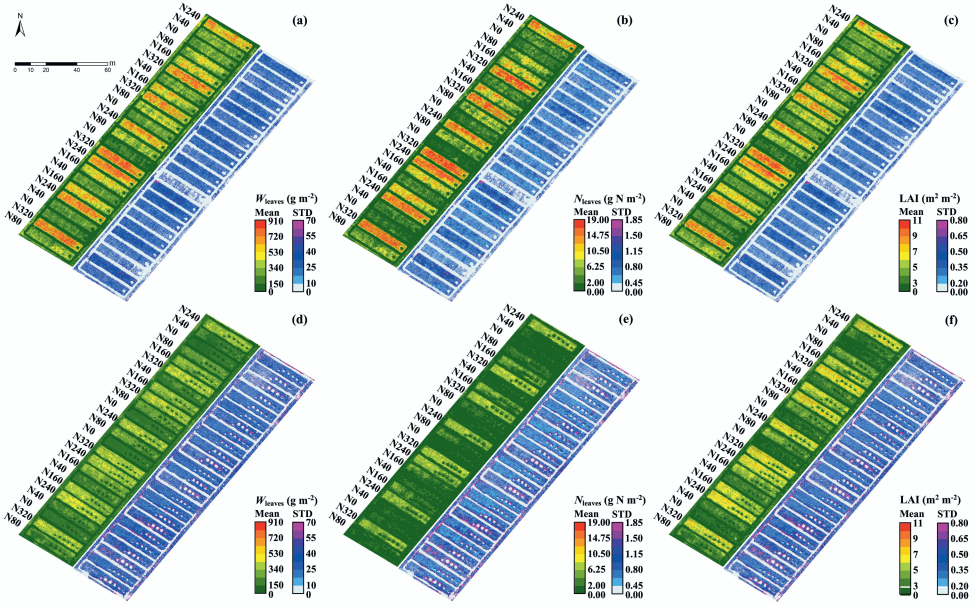


Fig. 4.6 Maps of the predicted leaf weight (W_{leaves}), leaf N content in the canopy (N_{leaves}) and leaf area index (LAI) at the stage of stem-elongating (15-Aug, 2020) (a-c) and grain-filling (10-Oct, 2020) (d-f) at the experimental site by regressed Gaussian Process Regression (GPR) models. The predicted value (mean) and its standard deviation (STD) were predicted from the GPR models and mapped in each panel. N0, N40, N80, N160, N240 and N320 denote varied nitrogen (N) application rates (see Table 4.2 for details).

Maps of the predicted leaf traits at the experimental site were generated, in which the temporal and spatial differences of crop growth at different growing stages were predicted (Fig. 4.6). In the treatments with low N input, the mean values of predicted leaf traits were low at the stem-elongating stage (Fig. 4.6a-c), and it later caused early senescence, which was reflected by the low values of the predicted leaf traits at the grain-filling stage (Fig. 4.6d-f). The mean and STD of the predicted leaf traits tended to be higher in experimental plots with higher N input at the stem-elongating stage (Fig. 4.6a-c). However, the predicted STD for leaf traits with lower predicted mean values at grain-filling stage tended to be even higher than those with higher predicted mean values at the stem-elongating stage (Fig. 4.6). The diverse uncertainties across the different growing seasons and the spatial variance between and within the experimental plots demonstrated the necessity of assimilating *in situ* observations into crop growth simulations for accurate forecasting of crop growth status.

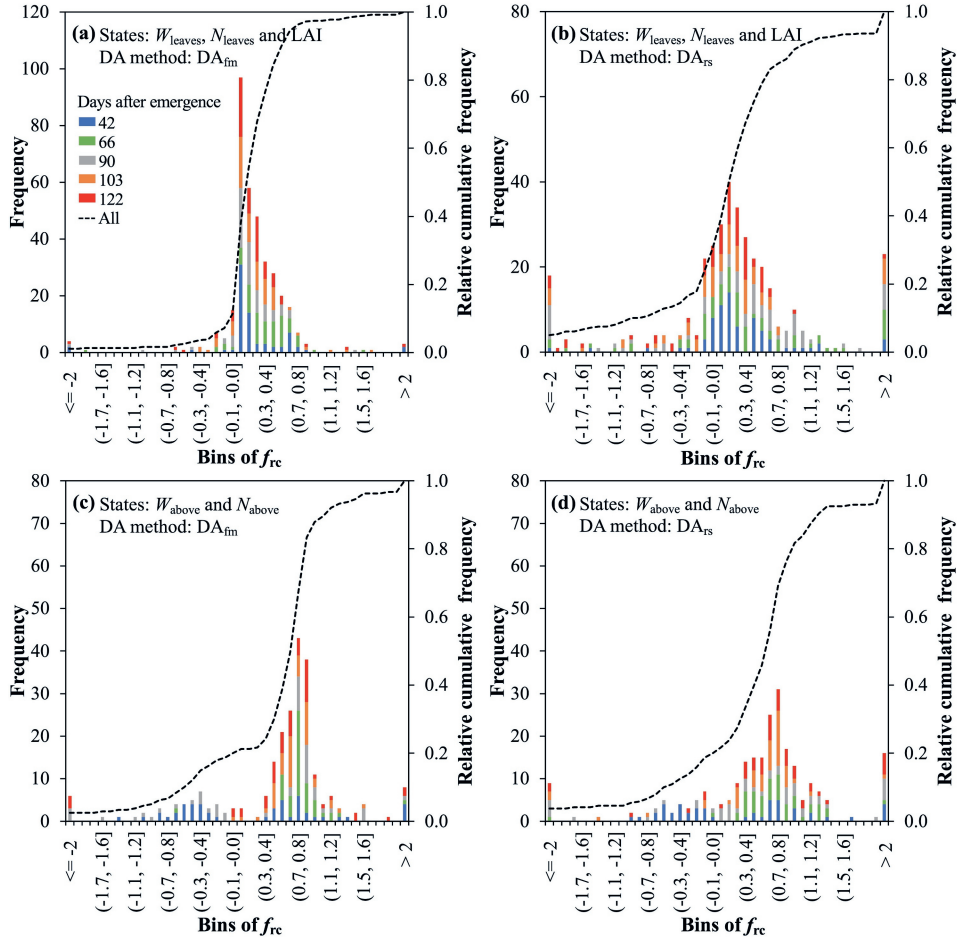


Fig. 4.7 The distribution of f_{rc} (see Eqn (4.21) for its definition) at each day conducting data assimilation (DA) by assimilating field measurements (DA_{fm}) (a, c) or remote sensing observations (DA_{rs}) (b, d) and its overall relative cumulative frequency. The f_{rc} of directly updated states, leaf weight (W_{leaves}), leaf N content in the canopy (N_{leaves}) and leaf area index (LAI), were shown in Panel a and b, while that of indirectly updated states, aboveground biomass (W_{above}) and aboveground N accumulation (N_{above}), are shown in Panel c and d.

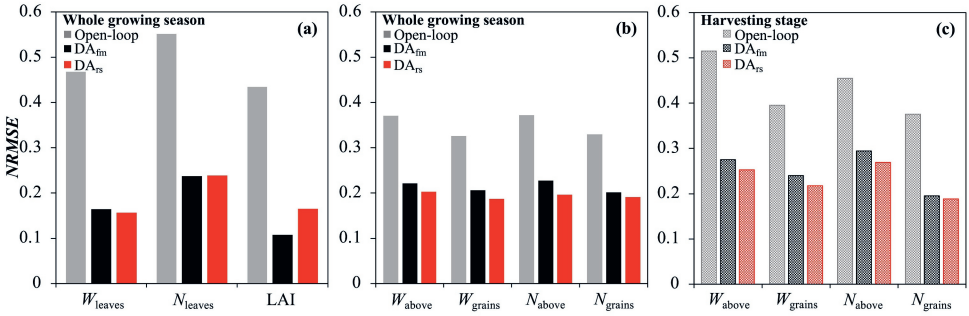


Fig. 4.8 Comparison of *NRMSE* of states simulated from the GECROS model without assimilating observations (open-loop) and of states analysed from assimilating observations of field measurements (DA_{fm}) and of remote sensing predictions (DA_{rs}) in year 2020. The overall *NRMSE* in the whole growing season of the directly updated states, leaf weight (W_{leaves}), leaf N content in the canopy (N_{leaves}) and leaf area index (LAI), are shown in Panel a. The overall *NRMSE* in the whole growing season and *NRMSE* at the harvesting stage of the indirectly updated states, aboveground biomass (W_{above}), grain weight (W_{grains}), aboveground N accumulation (N_{above}) and grain N content (N_{grains}), are shown in Panel b and c, respectively.

4.3.4 Performance of assimilating observations into crop growth simulations

While assimilating observations into crop simulations, the filter divergence did not show up and the updated simulations agreed better with the measurements. In line with the illustrated trajectories of time series states from open-loop without assimilating observations and updated states by DA_{fm} and DA_{rs} (Fig. S4.7), the differences in DA_{fm} or DA_{rs} between updated states and measured states tended to diminish with progress of the growing season (Fig. S4.8). Moreover, the performance of updated states hardly changed with the varied ensemble size from 50 to 500 in EnKF (Fig. S4.9). The f_{rc} for directly updated leaf traits by DA_{fm} and DA_{rs} centred around 0.1 and 0.2 (Fig. 4.7a-b), respectively, indicating improved filter performance after assimilating observations and no filter divergence occurrence. Although the f_{rc} for traits indirectly updated by DA_{fm} and DA_{rs} centred around 0.8, there was no pattern that the f_{rc} tended to be close to one with the progress of the growing season (Fig. 4.7c-d), which was another indication of the absence of filter divergence. Due to the method for simulating grain formation in the crop model GECROS, unlike with W_{above} and N_{above} , W_{grains} and N_{grains} could not be updated immediately once the leaf traits were updated, but only could be updated gradually in the following growth days (Fig. S4.7). The f_{rc} values of W_{grains} and N_{grains} were not included here.

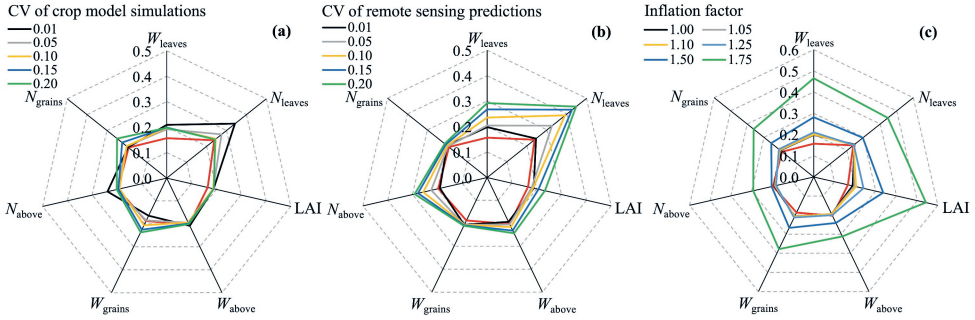


Fig. 4.9 Performance comparison of the data assimilation system with the assumed uncertainties alone, the assumed uncertainties combined with the introduced inflation factor in EnKF, and the quantified uncertainties. The *NRMSE* of the updated states in the growing season was shown, based on different values of coefficient of variation (CV) of crop model simulations with the CV of remote sensing predictions fixed at 0.01 (a), different values of CV of remote sensing predictions with the CV of crop model simulations fixed at 0.1 (b), and different values of inflation factor in EnKF with the CVs of crop model simulations and remote sensing predictions fixed at 0.1 and 0.01, respectively (c). The inflation factor was fixed at 1.0 for panels (a) and (b), assuring the assumed uncertainties functioned alone. The red lines in each panel indicate the *NRMSE* based on the estimated uncertainties of crop model simulations and remote sensing predictions from the proposed Bayesian methodology in this study. W_{leaves} , leaf weight; N_{leaves} , leaf N content in the canopy; LAI, leaf area index; W_{above} , aboveground biomass; W_{grains} , grain weight; N_{above} , aboveground N content; and N_{grains} , grain N content.

The *NRMSE* of the updated states by DA_{fm} and by DA_{rs} decreased, compared with those of open-loop (Fig. 4.8). Analysed states like W_{leaves} , N_{leaves} and LAI by DA_{fm} and DA_{rs} were directly updated and unsurprisingly performed better than those by open-loop across the whole growing season, in which their *NRMSE* on average decreased from 0.468, 0.551 and 0.434 to 0.161, 0.228 and 0.136, respectively (Fig. 4.8a). More importantly, the *NRMSE* of the indirectly updated states, W_{above} and N_{above} , by DA_{fm} across the whole growing season decreased to 0.222 and 0.227, respectively, while those by DA_{rs} further decreased to 0.203 and 0.196 (Fig. 4.8b). Especially at the harvesting stage, those indirectly updated states of DA_{fm} and DA_{rs} also agreed better with the field measurements than those of open-loop (Fig. 4.8c). Taking advantage of the *in-situ* prediction of crop growth status by the GPR modelling of remote sensing images, updated states of DA_{rs} tended to perform better than those of DA_{fm} (Fig. 4.8b-c). Particularly, compared with arbitrarily assumed uncertainties of crop model simulations and remote sensing predictions, analysed states of DA_{rs} based on the estimated uncertainties from the proposed Bayesian methodology showed better performance (Fig. 4.9a-b). Fixing the CVs of crop model simulations and remote sensing predictions at 0.1 and 0.01, respectively, resulted in a better performance than using other combinations of CVs. Thus, these CV values were selected for

further evaluating the performance of varied inflation factor (Fig. 4.9), in which the inflation factor was expected to enlarge \mathbf{K} in the DA process (Eqn (4.15)) for better filter performance. The updated states of N_{leaves} and N_{above} benefited from the further incorporated inflation factor, on which ranged from 1.05 to 1.25. However, the *NRMSE* of updated states based on the assumed uncertainties combined with the inflation factor was in general higher than that based on the assumed uncertainties only, which was also higher than that based on the proposed Bayesian methodology (Fig. 4.9c).

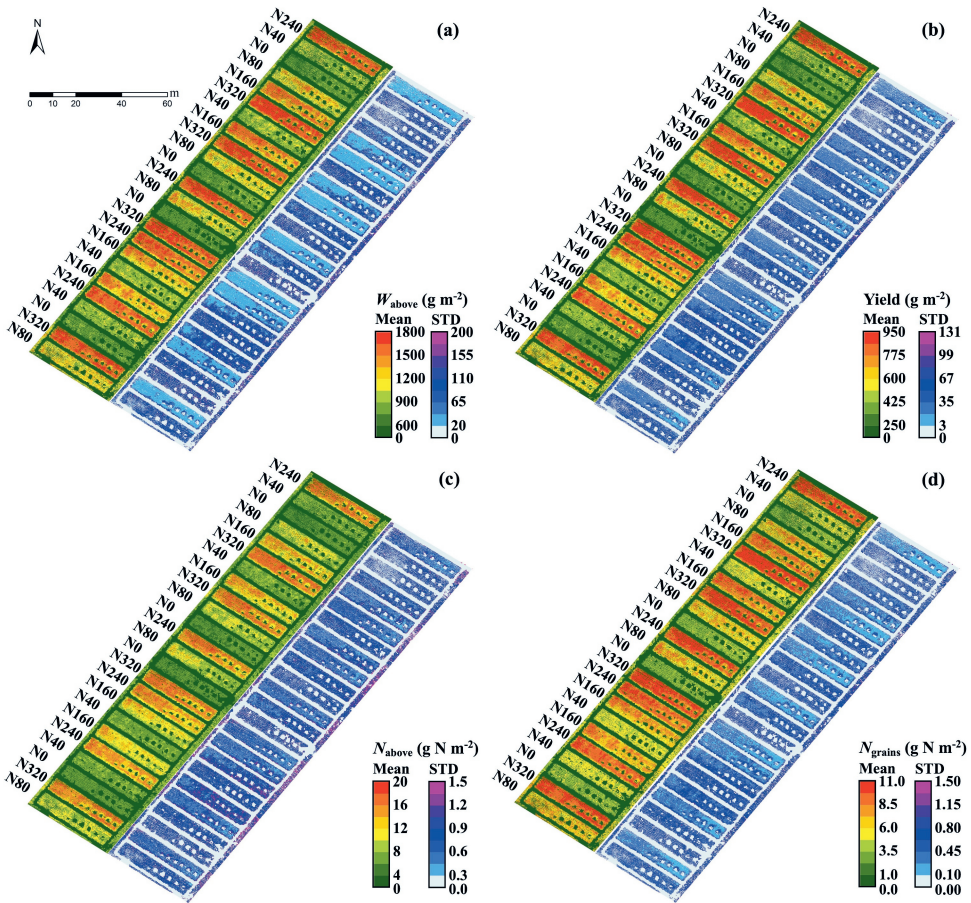


Fig. 4.10 Maps of the updated simulations of aboveground biomass (W_{above}) (a), yield (b), aboveground N content (N_{above}) (c) and grain N content (N_{grains}) (d) at maturity in year 2020 at the experimental site. Both mean and standard deviation (STD) of each state were derived from their updated ensemble values in EnKF. N0, N40, N80, N160, N240 and N320 denote varied nitrogen (N) application rates (see Table 4.2 for details).

4.3.5 Mapping carbon and nitrogen content of aboveground plant and grain at maturity

In view of the better performance of DA_{rs} , the maps of updated W_{above} , yield, N_{above} and N_{grains} at the experimental site at maturity were generated (Fig. 4.10) by assimilating the in-season crop growth simulation and leaf traits' prediction (Fig. 4.6). Differences in the indirectly updated biomass and N content of aboveground plant and grain caused by the different N input were observed between the experimental plots. The variances of those updated simulations within each experimental plot were also shown in their corresponding maps (Fig. 4.10), in line with the existing variances in the maps of predicted leaf traits (Fig. 4.6).

4.4 Discussion

4.4.1 Impacts of residual error assumption for crop model calibration and data assimilation

The residual error assumption heavily affects the performance of the formal Bayesian approach (Beven et al., 2008). In our study, σ_1 denoted the heteroscedastic residual error, assuming that the residual error linearly increased with the crop model simulations, as the STDs in field measurements tended to vary with the averages (Fig. 4.4). Similarly, due to the increase of both the averages and the STDs of the W_{above} throughout the seasons, the likelihood function was revised with the observational variances by Dumont et al. (2014) for the non-stationary residual errors. In line with their results, our study showed that σ_1 depicted the heteroscedasticity well for the crop model simulations (Fig. 4.4), although the calibration of the uncertain parameters tended to be sensitive to the range of σ_1 (Fig. 4.3). Instead of the heteroscedastic residual error, Zhang et al. (2021a) used time-series variance across the growing season to investigate the heteroscedasticity of the uncertain parameters in the crop model AquaCrop, and showed that it significantly improved the effectiveness of the particle filter as well, when assimilating remotely sensed canopy cover into the crop model. In addition to the hypothesised heteroscedasticity, non-Gaussian errors were also introduced in our study and it was shown that the residual errors were more likely to be negatively skewed ($\xi < 1$), except for those of simulated W_{grains} ($\xi_{W_{grain}} = 2.86$) and N_{grains} ($\xi_{N_{grain}} = 1.75$) (Table 4.5).

Compared with the complex residual error assumptions, the i.i.d. Gaussian errors worked well in the MCMC approach for the parameter calibration of the crop model (Dumont et al.,

2014; Kang and Özdoğan, 2019). However, while further assimilating crop model simulated and remotely sensed crop traits by EnKF, instead of the pre-assumed i.i.d. Gaussian errors, the inflation factor is introduced for accounting for model uncertainties (Kang and Özdoğan, 2019). Our results showed that the estimated uncertainties resulted in a better performance of the DA system, compared with the ones when the pre-assumed uncertainty was used alone or when it was combined with the inflation factor being introduced into EnKF (Fig. 4.9). Moreover, the inflation factor tends to not only be sensitive to the filter performance (Kang and Özdoğan, 2019; Whitaker and Hamill, 2012), but also differ between updated traits of crop and soil (Ines et al., 2013). With the rapid development of various satellites and unmanned aerial vehicles, a wide range of crop and soil information will be accessible to be incorporated into crop models (Jin et al., 2018). However, due to the uncertainties of crop and soil traits, rigorous determination of inflation factors in DA is difficult. Rather than the obscure adjustment of inflation factors, quantifying crop model uncertainties by the MCMC process together with the adapted residual error assumption in this study has great potential. With the Bayes' multiplication method, we calibrated the crop model parameters well by exploiting multiple crop traits (Table 4.6, Fig. 4.4), in line with the results of He et al. (2010). The uncertainties of multiple crop traits in crop model simulation were estimated simultaneously (Table 4.5). The estimated crop model uncertainties worked well in EnKF while assimilating remotely sensed leaf traits into the crop model GECROS, while the $NRMSE$ of updated W_{grains} and N_{grains} increased slightly, compared with simulations using mean values of uncertain parameters (Table 4.6, Fig. 4.8). Recently, next to inflation factors and residual error models, the parameter and model structural errors of the simulation of rice phenology were determined by a multi-model ensemble method with assumed i.i.d. Gaussian errors, in which the different simulations generated from the different models were treated as the samples from the distribution of the unknown true model (Gao et al., 2021). Thus, more research is needed regarding the exploration of quantified crop model uncertainty and of their performance comparison in DA.

4.4.2 Uncertainty estimation of remote sensing predictions and its application in data assimilation

Like with the regression performance of GPR for leaf-level traits like specific leaf weight (Wang et al., 2019) and chlorophyll content (Verrelst et al., 2013a), our results showed high predictive performance for the traits W_{leaves} , N_{leaves} and LAI, at canopy level, not only in the training dataset ($R^2 > 0.95$), but also in the testing dataset ($R^2 > 0.81$) (Fig. 4.5). Agreeing with

the threshold proposed by Global Climate Observing System for the ecological application of remote sensing observations (GCOS, 2011), the relative uncertainties of predicted leaf traits in our results were below 20% (Fig. 4.6). The predicted low uncertainty was in line with other related GPR research (Verrelst et al., 2016; Wang et al., 2019). Similar to the results of Wang et al. (2019), relatively high uncertainty always came with high N addition in the vegetative phase (Fig. 4.6a-c). We also found that the uncertainty tended to increase after entering the reproductive phase, especially for the treatments with low N input, caused by early leaf senescence (Fig. 4.6d-e). Meanwhile, as GPR captured temporal and spatial variations in crop growth well, DA_{rs} performed better than DA_{fm} (Fig. 4.8), which agrees with the results of yield forecasting by assimilating LAI into the APSIM-Wheat model (Zhang et al., 2021b). The estimated uncertainty of remote sensing predictions likely affected inflation factor estimation as well, as the inflation factor is determined from the updated posteriors based on the observation errors (Whitaker and Hamill, 2012). The contribution of the GPR model should be further evaluated given that the uncertainty of remote sensing observations is commonly overlooked in DA (Huang et al., 2019) and arbitrarily set based on a general guess (e.g., Kang and Özdoğan, 2019; Ines et al., 2013; Fig. 4.9b).

4.4.3 Performance of the analysed system uncertainties in data assimilation

With the estimated uncertainties in crop model simulations and remote sensing observations, the inflation factor was no longer needed in this study. Filter divergence did not occur during the DA process with the analysed system uncertainties (Fig. 4.7). 96% of f_{rc} of leaf traits in DA_{fm} and 76% in DA_{rs} ranged from -0.9 to 0.9, indicating at least 10% uncertainty reduction of those leaf traits after conducting DA (Fig. 4.7a-b), and $NRMSE$ of updated leaf traits on average decreased by 66% and 62% by DA_{fm} and DA_{rs} , respectively (Fig. 4.8a). This large improvement of directly updated states was in line with the results of Kivi et al. (2022) who estimated system uncertainties for EnKF while assimilating in situ observed daily soil moisture into the crop model APSIM. Even though the forecast accuracy of soil moisture was improved, the performance of updated LAI and yield tended to be worse than that of simulations without assimilating soil moisture (Kivi et al., 2022), which conflicts with previous studies (de Wit and van Diepen, 2007; Ines et al., 2013). As with the similar concerns of Schoups and Vrugt (2010) about the approach of simultaneous optimisation and data assimilation proposed by Vrugt et al. (2005), crop growth simulation seems to be affected by measured soil moisture during joint estimation of system uncertainty, causing poorer performance in updated LAI and yield (Kivi

et al., 2022). Without such tangled system uncertainties, our results showed that the performance of in-season updated and end-of-season forecasted crop carbon and N status improved (Fig. 4.8). However, there is still scope for further improvement. For instance, the distributions of f_{rc} of updated aboveground status tended to be more diverged and closer to one (Fig. 4.7c-d), while those of updated leaf traits were more converged and closer to zero (Fig. 4.7a-b), implying a limitation in updating leaf traits for an accurate forecast of the W_{above} and N_{above} . Meanwhile, the end-of-season forecast of W_{above} and N_{above} might also be impacted by their insufficiently updated in-season status (Figs. 4.8b-c, S4.8a, c).

4.4.4 Perspective of applying sophisticated crop models in data assimilation

Compared with applying simpler models like the crop model SAFY over large scales (Kang and Özdoğan, 2019), it is a challenge to use sophisticated crop models like GECROS for practical applications. For instance, GECROS is distinguished by a highly detailed photosynthetic process (Yin and Struik, 2017) and the related photosynthetic parameters were determined based on our previous study (Wang et al., 2022) and pre-set in this study, although the uncertainty still exists (Fig. S4.3). Also, the states updated in DA should be more carefully chosen, due to the complex feedback of the updated states in the sophisticated crop models that incorporate complex feedback mechanisms among biological processes. For instance, updating LAI in the crop model APSIM directly contributes to the improved performance of crop growth and yield formation (e.g., Zhang et al., 2022), due to its physiological mechanism that LAI influences the biomass accumulation by directly controlling the intercepted solar radiation in the canopy (Monsi and Saeki, 2005). However, as LAI is determined by both carbon and N status in the crop in the crop model GECROS (Yin et al., 2000), updating LAI alone hardly generated feedback for crop growth and yield forecast (results not shown) and thus updating states of W_{leaves} and N_{leaves} were incorporated together in this study as well (Fig. 4.8a). Similarly, Ines et al. (2013) indicated that updating states like W_{leaves} and specific leaf area, which feedback to LAI, might reduce the sensitivity of EnKF to model bias of the crop model DSSAT.

On the other hand, simple models have their own weaknesses. Due to the simplified physiological process, the key parameters in simple crop models like SAFY tend to be not only site-specific, but also year-specific (Claverie et al., 2012; Kang and Özdoğan, 2019). As the year-specific parameters should be calibrated with the actual in-season meteorological data, forecasting of crop growth and end-of-season yield is very uncertain (Kang and Özdoğan, 2019). In contrast to this, without any year-specific parameter calibration, GECROS performed

reasonably well in the validation year (Table 4.6), especially for W_{above} ($R^2 > 0.87$, $NRMSE = 0.28$). By only updating remotely sensed leaf traits, the performance of the simulated W_{above} improved and its $NRMSE$ at maturity further decreased to 0.25 (Fig. 4.8b-c). Relying on the integrated simulation of physiological processes relating to N dynamics, N_{above} and N_{grains} were simulated by GECROS and further updated more accurately in EnKF than that of open-loop (Fig. 4.8b-c), which probably forms a better basis for predicting traits like grain quality (Ma et al., 2022).

4.4.5 Prospect of updated crop status forecast for field-level in-season management

Since most remote sensing algorithms have been adapted to large areas, making them applicable to the field-level is a priority now, especially in countries like China (Weiss et al., 2020) where the agricultural system is dominated by millions of smallholders and N is always overdosed (Cui et al., 2018). Spatial and temporal predictions of N_{leaves} (Fig. 4.6) are crucial for determining the management of fertilisation timing to be performed at specific growth stages (Weiss et al., 2020). However, the use of remote sensing data solely is not sufficient to quantify top-dressing requirements (Weiss et al., 2020), due to the manifold interactions in the soil-crop-atmosphere continuum. Assimilating supplementary information from a crop model with remote sensing data has been identified as one of the most promising methods to make field management decisions (Jin et al., 2018; Weiss et al., 2020). For instance, Baret et al. (2007) demonstrated that an optimised in-season N application map can be generated by assimilating remotely sensed LAI and N_{above} into the crop model STICS, in which the N application rate in each map pixel ($20 \times 20 \text{ m}^2$) was optimised by maximising the productivity, using the historical meteorological data between the time of decision and harvest as the unknown future weather conditions. With optimised real-time fertiliser management, severe environmental issues caused by overfertilisation can be reduced, thereby enabling smart farming and sustainable agricultural production (Berger et al., 2020b).

4.5 Conclusions

In this study, we developed an approach that combines several disparate quantitative methods into one framework for explicitly quantifying the uncertainties of crop model simulations and remotely sensed observations, contributing to an accurate crop growth forecast while avoiding filter divergence in DA. Our results showed that calibration and uncertainty estimation of a

complex crop model benefited from an MCMC approach using the adapted residual error model. The calibrated uncertain parameters in crop model performed reasonably well in validation. The GPR models for analysing remote sensing data provided temporal and spatial predictions and corresponding uncertainties of leaf traits. Applying the quantified uncertainties into EnKF to update the leaf traits W_{leaves} , N_{leaves} and LAI enabled the crop model simulation to agree better with the measurements, without filter divergence. More importantly, updated simulations of in-season and end-of-season W_{above} , W_{grains} , N_{above} and N_{grains} also performed better than those of simulations without assimilating observations. The developed method is geared toward cases where multiple crop traits are observed and *in-situ* crop and soil information becomes increasingly available with the rapid development of remote sensing technologies. Armed with the precise forecast of in-season crop carbon and N status, field management can be better optimised to support sustainable smart farming.

Acknowledgement

We thank the financial support from the High-tech Industry and Scientific and Technological Innovation Project of Lin-gang Special Area, Shanghai (grant number: SH-LG-GK-2020-02-19). We greatly appreciate the expert input from Dr. Marcel van Oijen, whose critical comments, enlightening suggestions and detailed edits helped improve earlier drafts of this manuscript considerably.

References

- Acevedo, M., Pixley, K., Zinyengere, N., Meng, S., Tufan, H., Cichy, K., Bizikova, L., Isaacs, K., Ghezzi-Kopel, K., & Porciello, J. (2020). A scoping review of adoption of climate-resilient crops by small-scale producers in low- and middle-income countries. *Nature Plants*, 6, 1231-1241. doi:10.1038/s41477-020-00783-z
- Anderson, J. L., & Anderson, S. L. (1999). A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Review*, 127, 2741-2758.
- Baret, F., Houles, V., & Guerif, M. (2007). Quantification of plant stress using remote sensing observations and crop models: the case of nitrogen management. *Journal of Experimental Botany*, 58, 869-880. doi:10.1093/jxb/erl231
- Berger, K., Verrelst, J., Féret, J.-B., Hank, T., Woche, M., Mauser, W., & Camps-Valls, G. (2020a). Retrieval of aboveground crop nitrogen content with a hybrid machine learning method. *International Journal of Applied Earth Observation and Geoinformation*, 92, 102174. doi:10.1016/j.jag.2020.102174
- Berger, K., Verrelst, J., Féret, J.-B., Wang, Z., Woche, M., Strathmann, M., Danner, M., Mauser, W., & Hank, T. (2020b). Crop nitrogen monitoring: Recent progress and principal developments in the context of imaging spectroscopy missions. *Remote Sensing of Environment*, 242, 111758. doi:10.1016/j.rse.2020.111758
- Beven, K., & Freer, J. (2001). Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *Journal of Hydrology*, 249, 11-29.
- Beven, K. J., Smith, P. J., & Freer, J. E. (2008). So just why would a modeller choose to be incoherent? *Journal of Hydrology*, 354, 15-32. doi:10.1016/j.jhydrol.2008.02.007
- Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 26, 211-243.
- Box, G. E. P., & Tiao, G. C. X. (1973). *Bayesian inference in statistical analysis*. Reading, Pennsylvania, The United States: Addison-Wesley.
- Cai, C., Li, G., Di, L., Ding, Y., Fu, L., Guo, X., Struik, P. C., Pan, G., Li, H., Chen, W., Luo, W., & Yin, X. (2020). The acclimation of leaf photosynthesis of wheat and rice to seasonal temperature changes in T-FACE environments. *Globe Change Biology*, 26, 539-556. doi:10.1111/gcb.14830
- Carrassi, A., Bocquet, M., Bertino, L., & Evensen, G. (2018). Data assimilation in the geosciences: An overview of methods, issues, and perspectives. *WIREs Climate Change*, 9, 535. doi:10.1002/wcc.535
- Claverie, M., Demarez, V., Duchemin, B., Hagolle, O., Ducrot, D., Marais-Sicre, C., Dejoux, J.-F., Huc, M., Keravec, P., Béziat, P., Fieuzal, R., Ceschia, E., & Dedieu, G. (2012). Maize and sunflower biomass estimation in southwest France using high spatial and temporal resolution remote sensing data. *Remote Sensing of Environment*, 124, 844-857. doi:10.1016/j.rse.2012.04.005
- Cui, Z., Zhang, H., Chen, X., Zhang, C., Ma, W., Huang, C., Zhang, W., Mi, G., Miao, Y., Li, X., Gao, Q., Yang, J., Wang, Z., Ye, Y., Guo, S., Lu, J., Huang, J., Lv, S., Sun, Y., Liu, Y., Peng, X., Ren,

- J., Li, S., Deng, X., Shi, X., Zhang, Q., Yang, Z., Tang, L., Wei, C., Jia, L., Zhang, J., He, M., Tong, Y., Tang, Q., Zhong, X., Liu, Z., Cao, N., Kou, C., Ying, H., Yin, Y., Jiao, X., Zhang, Q., Fan, M., Jiang, R., Zhang, F., & Dou, Z. (2018). Pursuing sustainable productivity with millions of smallholder farmers. *Nature*, 555, 363-366. doi:10.1038/nature25785
- de Wit, A. J. W., & van Diepen, C. A. (2007). Crop model data assimilation with the Ensemble Kalman filter for improving regional crop yield forecasts. *Agricultural and Forest Meteorology*, 146, 38-56. doi:10.1016/j.agrformet.2007.05.004
- de Wit, C. T. (1965). *Photosynthesis of leaf canopies*. In Agricultural Research Reports no. 663. Wageningen, The Netherland: Pudoc.
- de Wit, C. T., & Penning de Vries, F. W. T. (1985). Predictive models in agricultural production. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 310, 309-315.
- Dumont, B., Leemans, V., Mansouri, M., Bodson, B., Destain, J. P., & Destain, M. F. (2014). Parameter identification of the STICS crop model, using an accelerated formal MCMC approach. *Environmental Modelling & Software*, 52, 121-135. doi:10.1016/j.envsoft.2013.10.022
- Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, 99, 10143-10162.
- Evensen, G. (2003). The Ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dynamics*, 53, 343-367. doi:10.1007/s10236-003-0036-9
- Fernández, C., & Steel, M. F. J. (1998). On Bayesian modeling of fat tails and skewness. *Journal of the american statistical association*, 93, 359-371.
- Gao, F., Anderson, M. C., Zhang, X., Yang, Z., Alfieri, J. G., Kustas, W. P., Mueller, R., Johnson, D. M., & Prueger, J. H. (2017). Toward mapping crop progress at field scales through fusion of Landsat and MODIS imagery. *Remote Sensing of Environment*, 188, 9-25. doi:10.1016/j.rse.2016.11.004
- Gao, Y., Wallach, D., Hasegawa, T., Tang, L., Zhang, R., Asseng, S., Kahveci, T., Liu, L., He, J., & Hoogenboom, G. (2021). Evaluation of crop model prediction and uncertainty using Bayesian parameter estimation and Bayesian model averaging. *Agricultural and Forest Meteorology*, 311. doi:10.1016/j.agrformet.2021.108686
- GCOS. (2011). Systematic observation requirements for satellite-based data products for climate, 2011 update, supplemental details to the satellite-based component of the Implementation plan for the global observing system for climate in support of the UNFCCC (2010 update, GCOS-154) (pp. 138).
- Gettelman, A., Geer, A. J., Forbes, R. M., Carmichael, G. R., Feingold, G., Posselt, D. J., Stephens, G. L., van den Heever, S. C., Varble, A. C., & Zuidema, P. (2022). The future of Earth system prediction: Advances in model-data fusion. *Science Advances*, 8, eabn3488. doi:10.1126/sciadv.abn3488
- Han, X., & Li, X. (2008). An evaluation of the nonlinear/non-Gaussian filters for the sequential data assimilation. *Remote Sensing of Environment*, 112, 1434-1449. doi:10.1016/j.rse.2007.07.008

- Hansen, J. W., & Jones, J. W. (2000). Scaling-up crop models for climate variability applications. *Agricultural Systems*, 65, 43-72.
- Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57, 97-109.
- He, J., Jones, J. W., Graham, W. D., & Dukes, M. D. (2010). Influence of likelihood function choice for estimating crop model parameters using the generalized likelihood uncertainty estimation method. *Agricultural Systems*, 103, 256-264. doi:10.1016/j.agsy.2010.01.006
- Houser, P. R., De Lannoy, G. J. M., & Walker, J. P. (2012). Hydrologic data assimilation. In J. Tiefenbacher (Eds.), *Approaches to Managing Disaster—Assessing Hazards, Emergencies and Disaster Impacts* (pp. 41-64). Rijeka, Croatia: InTech.
- Hu, S., Shi, L., Zha, Y., Williams, M., & Lin, L. (2017). Simultaneous state-parameter estimation supports the evaluation of data assimilation performance and measurement design for soil-water-atmosphere-plant system. *Journal of Hydrology*, 555, 812-831. doi:10.1016/j.jhydrol.2017.10.061
- Huang, J., Gómez-Dans, J. L., Huang, H., Ma, H., Wu, Q., Lewis, P. E., Liang, S., Chen, Z., Xue, J.-H., Wu, Y., Zhao, F., Wang, J., & Xie, X. (2019). Assimilation of remote sensing into crop growth models: Current status and perspectives. *Agricultural and Forest Meteorology*, 276-277. doi:10.1016/j.agrformet.2019.06.008
- Idso, S. B., Jackson, R. D., & Reginato, R. J. (1977). Remote-sensing of crop yields: Canopy temperature and albedo measurements have been quantitatively correlated with final harvests of wheat. *Science*, 196, 19-25.
- Ines, A. V. M., Das, N. N., Hansen, J. W., & Njoku, E. G. (2013). Assimilation of remotely sensed soil moisture and vegetation with a crop simulation model for maize yield prediction. *Remote Sensing of Environment*, 138, 149-164. doi:10.1016/j.rse.2013.07.018
- Jazwinski, A. H. (1970). *Stochastic processes and filtering theory*. New York, The United States: Academic Press.
- Jin, X., Kumar, L., Li, Z., Feng, H., Xu, X., Yang, G., & Wang, J. (2018). A review of data assimilation of remote sensing and crop models. *European Journal of Agronomy*, 92, 141-152. doi:10.1016/j.eja.2017.11.002
- Johnson, D. M. (2014). An assessment of pre- and within-season remotely sensed variables for forecasting corn and soybean yields in the United States. *Remote Sensing of Environment*, 141, 116-128. doi:10.1016/j.rse.2013.10.027
- Jones, J. W., Hoogenboom, G., Porter, C. H., Boote, K. J., Batchelor, W. D., Hunt, L. A., Wilkens, P. W., Singh, U., Gijsman, A. J., & Ritchie, J. T. (2003). The DSSAT cropping system model. *European Journal of Agronomy*, 18, 235-265.
- Kalnay, E., Li, H., Miyoshi, T., Yang, S.-C., & Ballabrera-Poy, J. (2007). 4-D-Var or ensemble Kalman filter? *Tellus A: Dynamic Meteorology and Oceanography*, 59, 758-773. doi:10.1111/j.1600-0870.2007.00261.x
- Kang, Y., & Özdoğan, M. (2019). Field-level crop yield mapping with Landsat using a hierarchical data assimilation approach. *Remote Sensing of Environment*, 228, 144-163. doi:10.1016/j.rse.2019.04.005

- Katzfuss, M., Stroud, J. R., & Wikle, C. K. (2016). Understanding the Ensemble Kalman Filter. *The American Statistician*, 70, 350-357. doi:10.1080/00031305.2016.1141709
- Keating, B. A., Carberry, P. S., Hammer, G. L., Probert, M. E., Robertson, M. J., Holzworth, D., Huth, N. I., Hargreaves, J. N. G., Meinke, H., & Hochman, Z. (2003). An overview of APSIM, a model designed for farming systems simulation. *European Journal of Agronomy*, 18, 267-288.
- Kennedy, M. C., & O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63, 425-464.
- Kivi, M. S., Blakely, B., Masters, M., Bernacchi, C. J., Miguez, F. E., & Dokoohaki, H. (2022). Development of a data-assimilation system to forecast agricultural systems: A case study of constraining soil water and soil nitrogen dynamics in the APSIM model. *Science of the Total Environment*, 820, 153192. doi:10.1016/j.scitotenv.2022.153192
- Lobell, D. B., & Azzari, G. (2017). Satellite detection of rising maize yield heterogeneity in the U.S. Midwest. *Environmental Research Letters*, 12, 014014. doi:10.1088/1748-9326/aa5371
- Ma, J., Zheng, B., & He, Y. (2022). Applications of a hyperspectral imaging system used to estimate wheat grain protein: A review. *Frontiers in Plant Science*, 13, 837200. doi:10.3389/fpls.2022.837200
- Martínez-Ferrer, L., Moreno-Martínez, Á., Campos-Taberner, M., García-Haro, F. J., Muñoz-Marí, J., Running, S. W., Kimball, J., Clinton, N., & Camps-Valls, G. (2022). Quantifying uncertainty in high resolution biophysical variable retrieval with machine learning. *Remote Sensing of Environment*, 280. doi:10.1016/j.rse.2022.113199
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21, 1087-1092.
- Monsi, M., & Saeki, T. (2005). On the factor light in plant communities and its importance for matter production. *Annals of Botany*, 95, 549-567. doi:10.1093/aob/mci052
- Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian processes for machine learning* (Vol. 2). Cambridge, Massachusetts, The United States: MIT press.
- Schoups, G., & Vrugt, J. A. (2010). A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors. *Water Resources Research*, 46, W10531. doi:10.1029/2009wr008933
- Verrelst, J., Alonso, L., Rivera Caicedo, J. P., Moreno, J., & Camps-Valls, G. (2013a). Gaussian process retrieval of chlorophyll content from imaging spectroscopy data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 6, 867-874. doi:10.1109/jstars.2012.2222356
- Verrelst, J., Malenovsky, Z., Van der Tol, C., Camps-Valls, G., Gastellu-Etchegorry, J.-P., Lewis, P., North, P., & Moreno, J. (2019). Quantifying vegetation biophysical variables from imaging spectroscopy data: A Review on Retrieval Methods. *Surveys in Geophysics*, 40, 589-629. doi:10.1007/s10712-018-9478-y
- Verrelst, J., Muñoz, J., Alonso, L., Delegido, J., Rivera, J. P., Camps-Valls, G., & Moreno, J. (2012). Machine learning regression algorithms for biophysical parameter retrieval: Opportunities for Sentinel-2 and -3. *Remote Sensing of Environment*, 118, 127-139. doi:10.1016/j.rse.2011.11.002

- Verrelst, J., Rivera, J. P., Gitelson, A., Delegido, J., Moreno, J., & Camps-Valls, G. (2016). Spectral band selection for vegetation properties retrieval using Gaussian processes regression. *International Journal of Applied Earth Observation and Geoinformation*, 52, 554-567. doi:10.1016/j.jag.2016.07.016
- Verrelst, J., Rivera, J. P., Moreno, J., & Camps-Valls, G. (2013b). Gaussian processes uncertainty estimates in experimental Sentinel-2 LAI and leaf chlorophyll content retrieval. *ISPRS Journal of Photogrammetry and Remote Sensing*, 86, 157-167. doi:10.1016/j.isprsjprs.2013.09.012
- Vrugt, J. A., Diks, C. G. H., Gupta, H. V., Bouten, W., & Verstraten, J. M. (2005). Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation. *Water Resources Research*, 41, W01017. doi:10.1029/2004wr003059
- Vrugt, J. A., Ter Braak, C. J. F., Diks, C. G. H., Robinson, B. A., Hyman, J. M., & Higdon, D. (2009a). Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. *International journal of nonlinear sciences and numerical simulation*, 10, 273-290.
- Vrugt, J. A., ter Braak, C. J. F., Gupta, H. V., & Robinson, B. A. (2009b). Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling? *Stochastic Environmental Research and Risk Assessment*, 23, 1011-1026. doi:10.1007/s00477-008-0274-y
- Wang, D., Rianti, W., Gálvez, F., van der Putten, P. E. L., Struik, P. C., & Yin, X. (2022). Estimating photosynthetic parameter values of rice, wheat, maize and sorghum to enable smart crop cultivation. *Crop and Environment*, 1, 119-132. doi:10.1016/j.crope.2022.05.004
- Wang, D., Struik, P. C., Liang, L., & Yin, X. (2023). Estimating leaf and canopy nitrogen contents in major field crops across the growing season from hyperspectral images using nonparametric regression. *Authorea*. doi:10.22541/au.170111047.73824045/v1.
- Wang, Z., Townsend, P. A., Schweiger, A. K., Couture, J. J., Singh, A., Hobbie, S. E., & Cavender-Bares, J. (2019). Mapping foliar functional traits and their uncertainties across three years in a grassland experiment. *Remote Sensing of Environment*, 221, 405-416. doi:10.1016/j.rse.2018.11.016
- Weiss, M., Jacob, F., & Duveiller, G. (2020). Remote sensing for agricultural applications: A meta-review. *Remote Sensing of Environment*, 236, 111402. doi:10.1016/j.rse.2019.111402
- Whitaker, J. S., & Hamill, T. M. (2012). Evaluating methods to account for system errors in ensemble data assimilation. *Monthly Weather Review*, 140, 3078-3089. doi:10.1175/mwr-d-11-00276.1
- Yin, X., Schapendonk, A., & Struik, P. C. (2019). Exploring the optimum nitrogen partitioning to predict the acclimation of C₃ leaf photosynthesis to varying growth conditions. *Journal of Experimental Botany*, 70, 2435-2447. doi:10.1093/jxb/ery277
- Yin, X., Schapendonk, A. H. C. M., Kropff, M. J., van Oijen, M., & Bindraban, P. S. (2000). A generic equation for nitrogen-limited leaf area index and its application in crop growth models for predicting leaf senescence. *Annals of Botany*, 85, 579-585.
- Yin, X., & Struik, P. C. (2017). Can increased leaf photosynthesis be converted into higher crop mass production? A simulation study for rice using the crop model GECROS. *Journal of Experimental Botany*, 68, 2345-2360. doi:10.1093/jxb/erx085

- Yin, X., & van Laar, H. H. (2005). *Crop systems dynamics: an ecophysiological simulation model for genotype-by-environment interactions*. Wageningen, The Netherlands: Wageningen Academic Publishers.
- Ying, Y., & Zhang, F. (2015). An adaptive covariance relaxation method for ensemble data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 141, 2898-2906. doi:10.1002/qj.2576
- Zhang, T., Su, J., Liu, C., & Chen, W.-H. (2021a). State and parameter estimation of the AquaCrop model for winter wheat using sensitivity informed particle filter. *Computers and Electronics in Agriculture*, 180, 105909. doi:10.1016/j.compag.2020.105909
- Zhang, Y., Walker, J. P., Pauwels, V. R. N., & Sadeh, Y. (2021b). Assimilation of wheat and soil states into the APSIM-Wheat crop model: A case study. *Remote Sensing*, 14, 65. doi:10.3390/rs14010065
- Zhang, Y., Walker, J. P., & Pauwels, V. R. N. (2022). Assimilation of wheat and soil states for improved yield prediction: The APSIM-EnKF framework. *Agricultural Systems*, 201, 103456. doi:10.1016/j.agry.2022.103456

Supplementary Appendix A in Chapter 4

Brief description of the GECROS model

The crop model Genotype-by-Environment interaction on CROp growth Simulator (GECROS) used in this study is a generic model that can simulate a broad range of crops (Yin and van Laar (2005); Yin and Struik (2017)). The water, carbon and N dynamics in the plant and soil are simulated on a daily basis. Its adaptability has been validated in many field crops, such as potato (Khan et al., 2014), rice (Gu et al., 2014), wheat and maize (Lenz-Wiedemann et al., 2010). It is characterised by the detailed algorithms for photosynthesis of C_3 and C_4 crops. Specifically, the instantaneous leaf photosynthesis is calculated by the biochemical model of Farquhar et al. (1980) for C_3 crops and the equivalent model for simulating C_4 photosynthesis (von Caemmerer and Furbank, 1999; Yin and Struik, 2009, 2012). Leaf N content linearly affects photosynthetic capacity parameters including the maximum rate of Rubisco activity-limited carboxylation (V_{cmax}) and the maximum rate of linear electron transport under saturated light (J_{max}) (Evans, 1983; Wang et al., 2022). Photosynthesis at the leaf level is upscaled to the canopy level by the sun/shade model of De Pury and Farquhar (1997). Temporal upscaling of the instantaneous rates to daily integrals is performed using the five-point Gaussian integration (Goudriaan, 1986). N uptake is considered as the minimum of crop N demand and soil N supply. Partitioning of the newly formed carbon (C) and N assimilates between root and shoot is based on the equations of Yin and Schapendonk (2004). The development of LAI is modelled as the minimum of C- and N-determined LAI. The C-determined LAI is the same as LAI calculated in older Wageningen crop models from biomass accumulated in leaves, while the N-determined LAI is simulated based on leaf N content in the canopy (Yin et al., 2000) [their Eqn (5)].

Supplementary Appendix B in Chapter 4

Equations for adjusting field measured leaf traits, LAI, leaf weight (W_{leaves}), leaf N content (N_{leaves})

As green stems are also considered as part of the functional leaves in GECROS, LAI values were adjusted following,

$$LAI = LAI_{leaves} + LAI_{stems} = \frac{S_{leaves}}{S_{ground}} + \frac{S_{stems}}{S_{ground}} * \frac{\pi}{2} \quad (S4.1)$$

where LAI_{leaves} and LAI_{stems} were calculated LAI accounting for green leaves and stems, respectively. S_{leaves} and S_{stems} were the scanned leaf and stem area, respectively, and S_{ground} denoted their corresponding ground area. $\frac{\pi}{2}$ was the projection factor for the cylindrical stem.

Similarly, W_{leaves} and N_{leaves} were adjusted accordingly,

$$W_{leaves} = W_{leaves,g} + W_{stems,g} \quad (S4.2)$$

$$N_{leaves} = N_{leaves,g} + N_{stems,g} \quad (S4.3)$$

where $W_{leaves,g}$ and $W_{stems,g}$, and $N_{leaves,g}$ and $N_{stems,g}$ were the W_{leaves} and N_{leaves} accounting for green leaves and stems, respectively. $W_{stems,g}$ were calculated as LAI_{stems}/SLA , assuming that the green stem area had the same specific leaf area (SLA) with measured green leaves. $N_{stems,g}$ was calculated as $W_{stems,m} * (NC_{stems,m} - STEMNC)$, in which $W_{stems,m}$ and $NC_{stems,m}$ represented the measured stem weight and stem N concentration, and STEMNC is the minimum N concentration in structural stem material (Table 4.4).

Supplementary Figures in Chapter 4

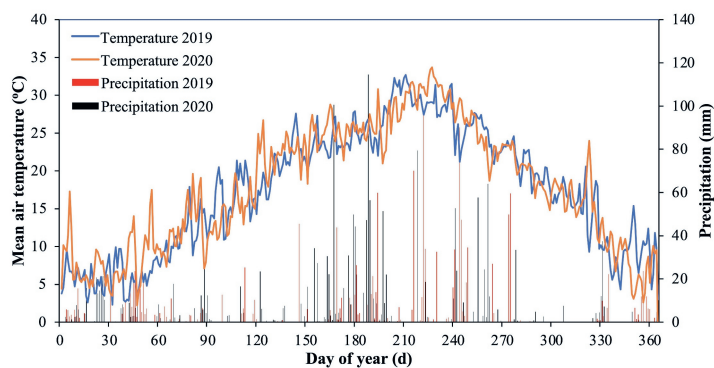


Fig. S4.1 Daily mean air temperature and precipitation in year 2019 and 2020 at experimental site.

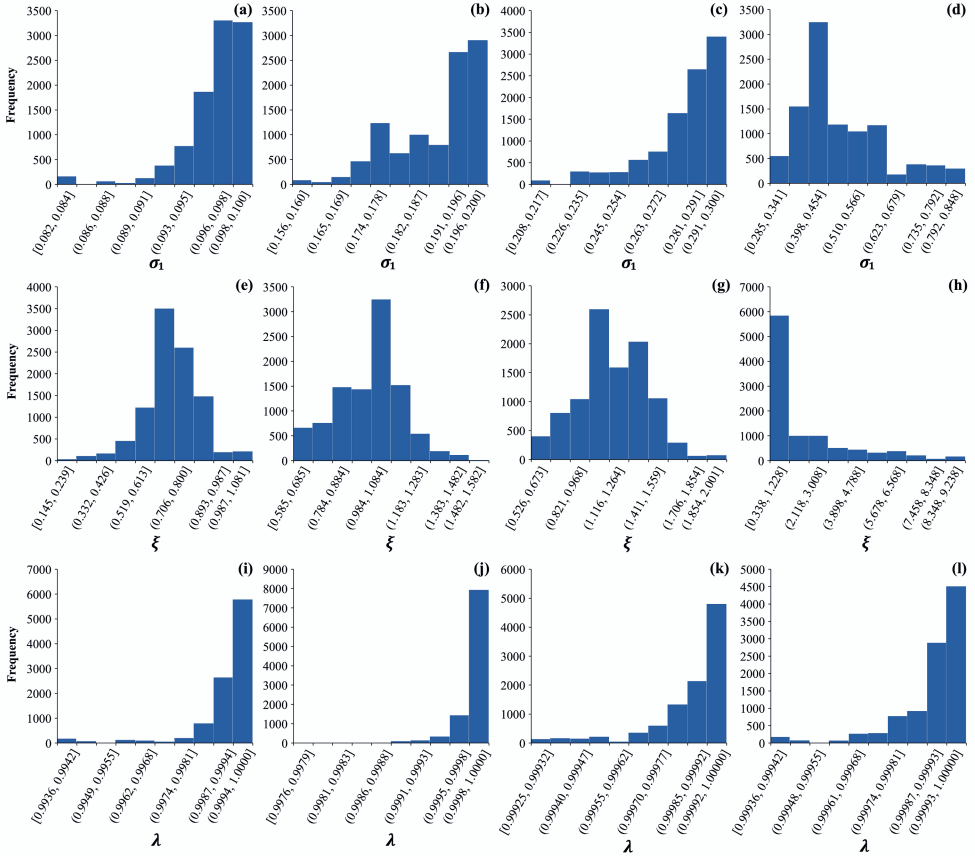


Fig. S4.2 Histogram of estimated residual error parameters, the heteroscedasticity parameter σ_1 , the skewness parameter ξ and the Box-Cox transformation parameter λ , for simulated aboveground biomass with different upper bounds of σ_1 , which were 0.1 (a, e, i), 0.2 (b, f, j), 0.3 (c, g, k) and 1.0 (d, h, l). The values of estimated parameters are from last 10000 evaluations of the Bayesian calibration process, which were used for generating the posterior distributions. The frequency of estimated parameter value represents the number of its occurrences in the last 10000 evaluations for each specified data range.

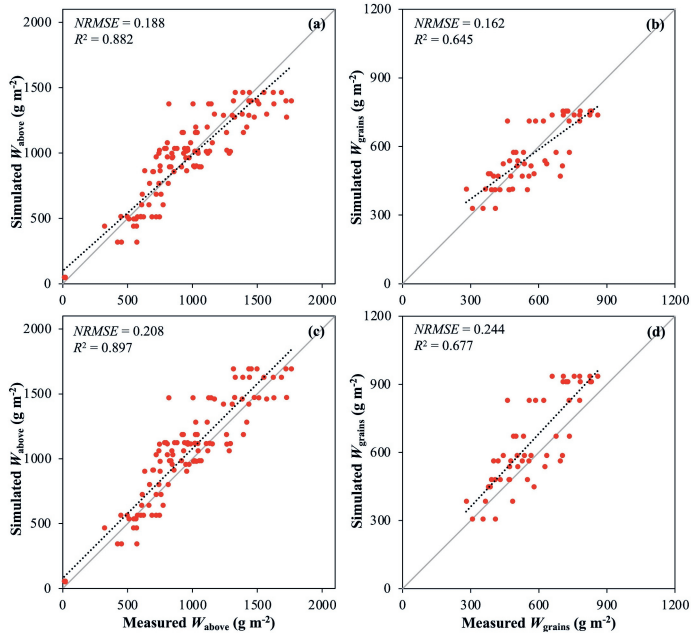


Fig. S4.3 The scatter plots of simulated aboveground biomass (W_{above}) and grain weight (W_{grains}) with (a, b) and without (c, d) the adjusting factor rASSA for the simulated carbon assimilation in GECROS using the estimated photosynthetic parameters. The grey lines indicate a 1:1 relationship. The dotted lines indicate the linear regression.

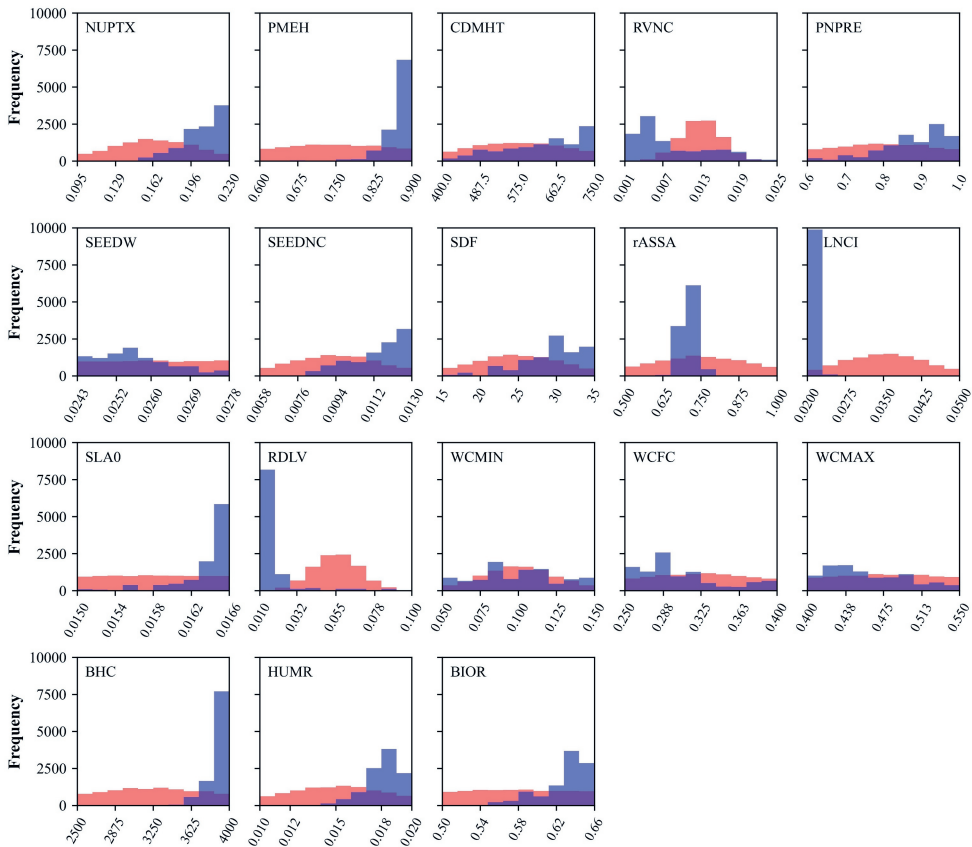


Fig. S4.4 Histograms of prior (red) and posterior (blue) distributions of crop model parameters of GECROS, which are defined in Table 4.4. The prior distributions were drawn from the 10,000 random samples of normal distributions, while the values of estimated parameters from the last 10,000 evaluations of the Bayesian calibration process were used for generating the posterior distributions. The frequency of the parameter values from the prior and posterior distribution represents the number of its occurrences in the 10,000 samples or evaluations for each specified data range.

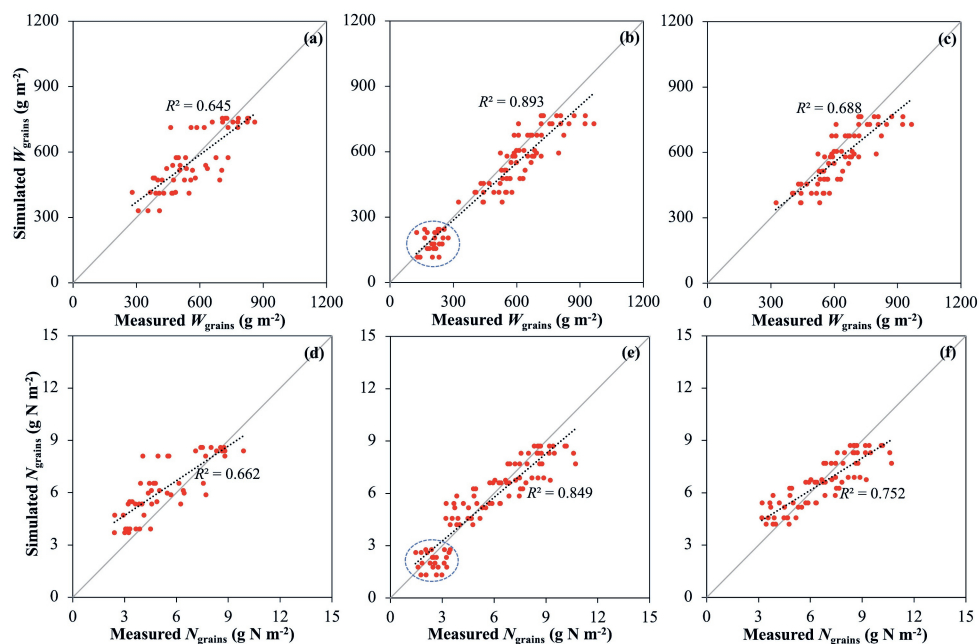


Fig. S4.5 The scatter plots of simulated and measured grain weight (W_{grains}) and grain N content (N_{grains}) in the calibration (a, d) and the validation (b-c, e-f). The grey lines indicate a 1:1 relationship. The dotted lines indicate the linear regression. The blue dashed-line ovals represent the field measurements conducted at the stage of early grain-filling (21-Sep) in the validation (year 2020). The simulated and measured W_{grains} and N_{grains} without those data points are shown in Panel e and f, respectively.

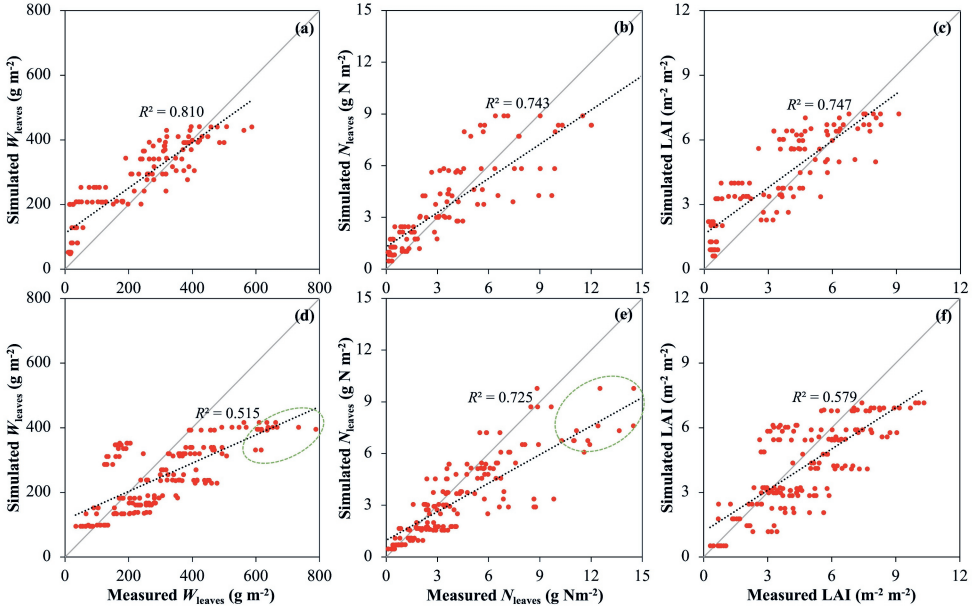


Fig. S4.6 The scatter plots of simulated and measured leaf weight (W_{leaves}), leaf N content in the canopy (N_{leaves}) and LAI in the calibration (a-c) and the validation (d-f). The grey lines indicate a 1:1 relationship. The dotted lines indicate the linear regression. The green dashed-line ovals represent the underestimated simulations when the values of measured W_{leaves} and N_{leaves} were high.

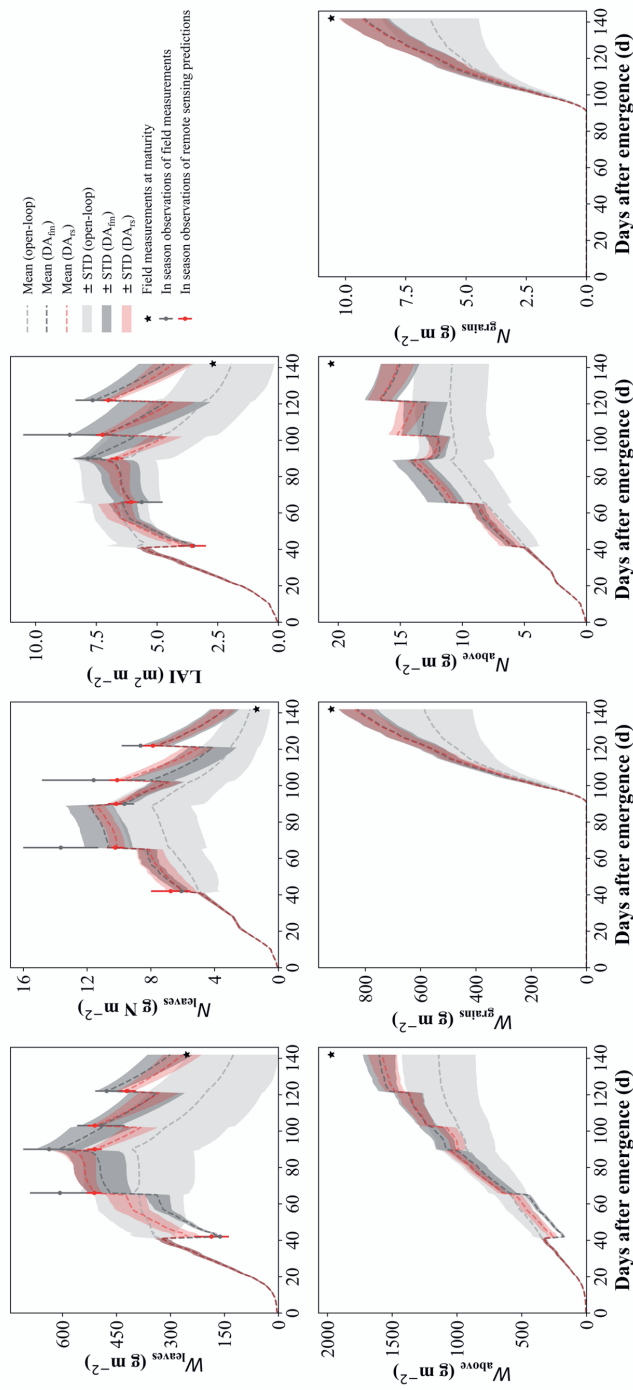


Fig. S4.7 Trajectory of simulations from the GECROS model only (open-loop) and of states analysed from assimilating observations of field measurement (DA_{fm}) and of remote sensing predictions (DA_{rs}) in year 2020, taking the sampling site of maturity in plot 1 with the nitrogen input of 240 kg N ha^{-1} as an example. In-season observations of leaf weight (W_{leaves}), leaf N content in the canopy (N_{leaves}) and leaf area index (LAI) were assimilated to update state variables of aboveground biomass (W_{above}), grain weight (W_{grains}), aboveground N accumulation (N_{above}) and grain N content (N_{grains}). Mean and standard deviation (STD) were derived from ensembles of EnKF.

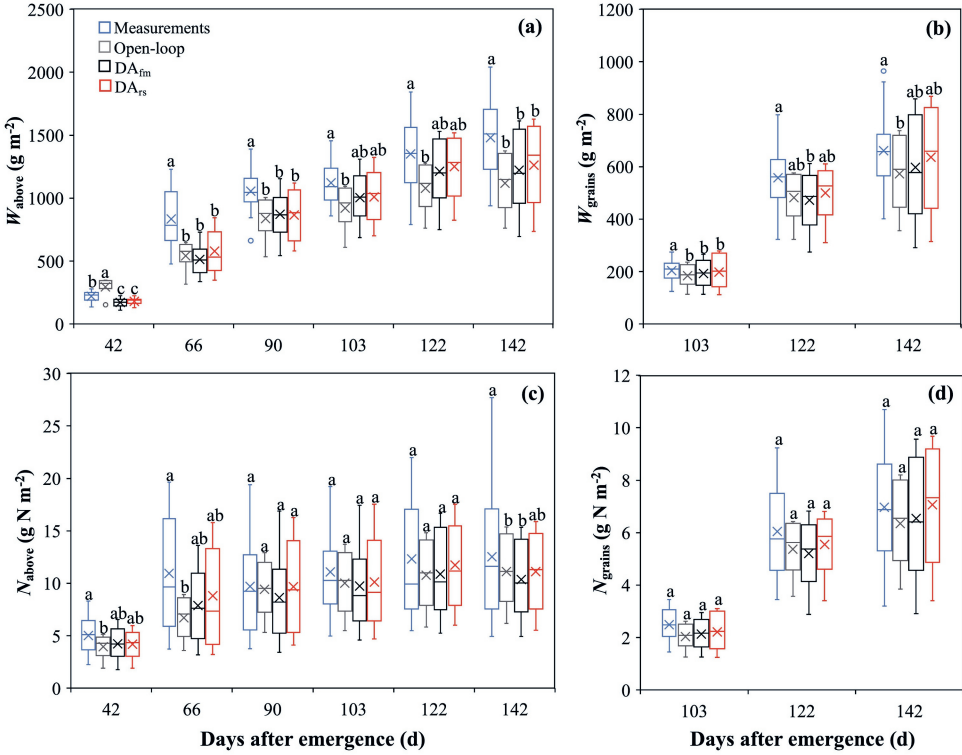


Fig. S4.8 Boxplot for measured states (measurements), simulated states from the GECROS model without assimilating observations (open-loop), and analysed states from data assimilation based on observations of field measurements (DA_{fm}) or of remote sensing (DA_{rs}) for aboveground biomass (W_{above}) (a), grain weight (W_{grains}) (b), aboveground N content (N_{above}) (c) and grain N content (N_{grains}) (d) across various N treatments at each sampling date in year 2020. The cross marks represent the mean value of the state across various N treatments at each sampling date. The different letters (a, b, c) above the bars indicate significant differences according to the HSD Tukey's test at the HSD Tukey's test ($\alpha = 0.05$).

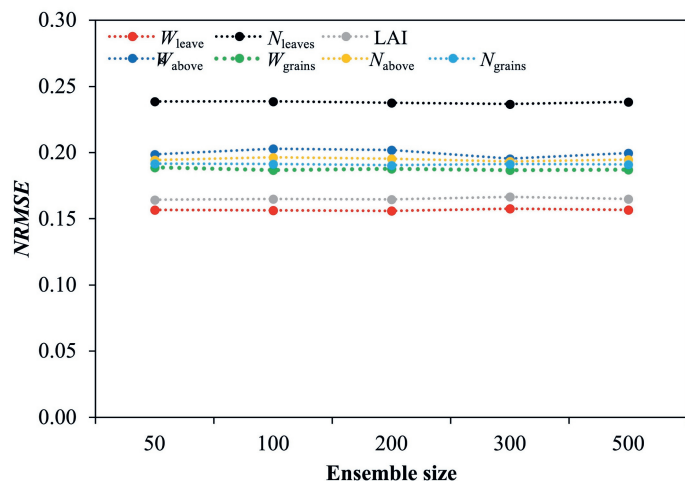


Fig. S4.9 Comparison of *NRMSE* of updated states with different ensemble size in EnKF. W_{leave} , leaf weight; N_{leave} , leaf N content in the canopy; LAI, leaf area index; W_{above} , aboveground biomass; W_{grains} , grain weight; N_{above} , aboveground N content; and N_{grains} , grain N content.

References

- De Pury, D. G. G., & Farquhar, G. D. (1997). Simple scaling of photosynthesis from leaves to canopies without the errors of big-leaf models. *Plant, Cell and Environment*, 20, 537-557. doi:10.1111/j.1365-3040.1997.00094.x
- Evans, J. R. (1983). Nitrogen and photosynthesis in the flag leaf of wheat (*Triticum aestivum* L.). *Plant Physiology*, 72, 297-302.
- Farquhar, G. D., von Caemmerer, S., & Berry, J. A. (1980). A biochemical model of photosynthetic CO₂ assimilation in leaves of C₃ species. *Planta*, 149, 78-90.
- Goudriaan, J. (1986). A simple and fast numerical method for the computation of daily totals of crop photosynthesis. *Agricultural and Forest Meteorology*, 38, 249-254.
- Gu, J., Yin, X., Zhang, C., Wang, H., & Struik, P. C. (2014). Linking ecophysiological modelling with quantitative genetics to support marker-assisted crop design for improved yields of rice (*Oryza sativa*) under drought stress. *Annals of Botany*, 114, 499-511. doi:10.1093/aob/mcu127
- Khan, M. S., Yin, X., van der Putten, P. E. L., & Struik, P. C. (2014). An ecophysiological model analysis of yield differences within a set of contrasting cultivars and an F1 segregating population of potato (*Solanum tuberosum* L.) grown under diverse environments. *Ecological Modelling*, 290, 146-154. doi:10.1016/j.ecolmodel.2013.11.015
- Lenz-Wiedemann, V. I. S., Klar, C. W., & Schneider, K. (2010). Development and test of a crop growth model for application within a Global Change decision support system. *Ecological Modelling*, 221, 314-329. doi:10.1016/j.ecolmodel.2009.10.014
- von Caemmerer, S., & Furbank, R. T. (1999). Modelling C₄ photosynthesis. In R. F. Sage, & R. K. Monson (Eds.), *C₄ plant biology* (pp. 173-211). Toronto, Canada: Academic Press.
- Wang, D., Rianti, W., Gálvez, F., van der Putten, P. E. L., Struik, P. C., & Yin, X. (2022). Estimating photosynthetic parameter values of rice, wheat, maize and sorghum to enable smart crop cultivation. *Crop and Environment*, 1, 119-132. doi:10.1016/j.crope.2022.05.004
- Yin, X., & Schapendonk, A. H. C. M. (2004). Simulating the partitioning of biomass and nitrogen between roots and shoot in crop and grass plants. *NJAS: Wageningen Journal of Life Sciences*, 51, 407-426. doi:10.1016/s1573-5214(04)80005-8
- Yin, X., Schapendonk, A. H. C. M., Kropff, M. J., van Oijen, M., & Bindraban, P. S. (2000). A generic equation for nitrogen-limited leaf area index and its application in crop growth models for predicting leaf senescence. *Annals of Botany*, 85, 579-585.
- Yin, X., & Struik, P. C. (2009). C₃ and C₄ photosynthesis models: An overview from the perspective of crop modelling. *NJAS - Wageningen Journal of Life Sciences*, 57, 27-38. doi:10.1016/j.njas.2009.07.001
- Yin, X., & Struik, P. C. (2012). Mathematical review of the energy transduction stoichiometries of C₄ leaf photosynthesis under limiting light. *Plant, Cell and Environment*, 35, 1299-1312. doi:10.1111/j.1365-3040.2012.02490.x

- Yin, X., & Struik, P. C. (2017). Can increased leaf photosynthesis be converted into higher crop mass production? A simulation study for rice using the crop model GECROS. *Journal of Experimental Botany*, 68, 2345-2360. doi:10.1093/jxb/erx085
- Yin, X., & van Laar, H. H. (2005). *Crop systems dynamics: an ecophysiological simulation model for genotype-by-environment interactions*. Wageningen, The Netherlands: Wageningen Academic Publishers.

Chapter 5

Developing remote sensing- and crop model-based methods to optimise nitrogen management in rice fields

Dong Wang^{1,2}, Paul C. Struik¹, Lei Liang² and Xinyou Yin¹

¹ Centre for Crop Systems Analysis, Department of Plant Sciences, Wageningen University & Research, 6700 AK Wageningen, The Netherlands

² Shanghai Lankuaikei Technology Development Co. Ltd., No. 888 Huanhu West 2nd Road, Pudong New District, Shanghai, China

Abstract

Physiological principles-based crop modelling and *in situ* sensor technology provide opportunities for smart nitrogen (N) management for sustainable agricultural production. We propose two optimisation methods – the crop model-based method (CM), and an integrated remote sensing-crop model method by data assimilation (RSCM) – to optimise N management. Data collected from a field experiment of rice with six N treatments (each with four times of topdressing) were used to illustrate the methods, where the first two N topdressings (N_{top}) were applied as in the experiment while the last two N_{top} were optimised. The two methods were compared with three reference methods: farmer practice optimised by the yield response curve (FP_{opt}), and the Sufficiency Index- or Response Index-based remote sensing (RS) methods, based on simulated crop growth with actual weather data. Compared with FP_{opt} , the sum of the optimised N_{top} of the CM method on average decreased by 37.9%, while that of the RSCM method decreased by 61.2%. The methods of CM, RSCM and RS decreased the simulated yield by 0.9%, 1.2%, and 4.4%, respectively, while they increased the profit by 2.8%, 4.4%, and -0.4%, respectively, compared with FP_{opt} . The CM method relying on crop physiological principles tended to perform better than the methods of FP_{opt} and RS in optimising in-season N application, while the RSCM method further benefited from assimilating data from *in situ* remote sensing information into the CM framework, thereby potentially best suiting to guide smart fertiliser management.

Keywords: Crop model; remote sensing; nitrogen management; smart farming; sustainable agricultural production

5.1 Introduction

Productivity of cereal crops in China has doubled, while chemical fertiliser application increased by threefold, over the past forty years (National Bureau of Statistics of China, <http://data.stats.gov.cn>). China's agriculture is dominated by smallholder farms, and nitrogen (N) fertilisers tend to be over-used in farmer practice (FP) (Cui et al., 2018; Moebius-Clune et al., 2013). While about 45% of current grain yield productivity can be attributed to this high N input (Yu et al., 2019), the FP also caused severe environmental issues, including leaching, eutrophication, greenhouse gas emission, and potential human health hazards (e.g., Cui et al., 2018; Zhang et al., 2013). The Chinese government has invested 10.7 billion dollars from 2007 to 2022 to control the eutrophication of Taihu (Taihu Basin Authority of Ministry of Water Resources, <http://www.tba.gov.cn>). For both economic and environmental reasons, there is a need for the precise field N management that synchronises soil N supply and crop demand (Cui et al., 2010). However, considering the spatial and temporal variability in soil properties and crop needs (Pierce and Nowak, 1999), the optimum N of whole-season total N application (N_{tot}) and in-season N topdressing (N_{top}) varies (Dumont et al., 2016; Mamo et al., 2003). This poses a great challenge for the sustainable agricultural production (Moebius-Clune et al., 2013).

Production functions have been commonly adopted for fitting the yield response curve showing the relationship of yield versus N_{tot} to determine the optimal N_{tot} (e.g., Cui et al., 2010; Fageria and Baligar, 2005). However, nitrogen fertiliser management, particular the appropriate rate and timing of application, is an important factor for the change of soil N in crop root zones (Dinnes et al., 2002). The so-called *yield goal method* can help to determine the whole-season N_{tot} and in-season N_{top} based on the soil N cycle and plant N uptake (Stanford, 1973). The N_{tot} is divided between two or three applications within the growing season, with the optimal N rate for the basal or N_{top} being determined from soil nitrate test in the root zone and a target N value of the corresponding crop growth period given the yield goal (Chen et al., 2006). Field experiments have shown that applying this strategy for the wheat-maize system in the North China Plain can reduce N application compared with FP without sacrificing crop yield (Chen et al., 2006; Cui et al., 2010). However, due to the unpredictability of the grain formation caused by the uncertain crop growth conditions, the yield goal might not give an optimum N rate (Lory and Scharf, 2003). Additionally, to account for the between- and within-field variability, considerable effort might be required for the destructive sampling and laboratory testing, especially when this method is applied at large scale. Moreover, due to the lack of the standard

soil sampling designs, including for example sampling intensity or interpolating methods, there are obstacles for creating soil nutrient maps for *in situ* management in large areas (Pierce and Nowak, 1999). Within the comparison of 31 N recommendation tools that mainly relied on a soil nitrate test, all tools failed to work as a universally reliable tool over diverse environmental conditions across eight corn belt states of the US (Ransom et al., 2020). There is a need for developing N-recommendation methods that are more responsive to soil and weather conditions (Ransom et al., 2020).

The proximal or remote sensing technologies have shown the potential for monitoring the spatial and temporal variability of crop growth for determining the *in situ* field N management (Hansen and Schjoerring, 2003; Pierce and Nowak, 1999). With an attainable yield target and the agronomic N efficiency (AE_N), an in-season adjustment of N_{top} for rice at tillering and panicle initiation can be made by estimating leaf N status from a chlorophyll meter or a leaf colour chart (Peng et al., 2010). Huang et al. (2015) guided the N_{top} for rice at the panicle initiation using satellite multispectral images, relying on the estimated N nutrition index, defined as the ratio of the actual to the critical crop aboveground N concentration. Due to the frequent cloudy weather during the growing season of rice and coarse spatial resolution of satellite remote sensing, Unmanned Aerial Vehicle-based images, especially the hyperspectral images with higher spectral resolution, are recommended (Huang et al., 2015). For instance, by applying the Unmanned Aerial Vehicle acquired multispectral images based Sufficiency Index (SI) to wheat at stem-elongating, the N_{tot} was reduced by 15.4% and AE_N increased from 9.07 to 10.36 kg kg⁻¹, compared with the local FP (Zhang et al., 2021). Similarly, Yao et al. (2012) demonstrated that the N_{tot} of rice was reduced by 33.3% without yield loss over FP by optimising the N_{top} at stem-elongating using the portable sensor based Response Index (RI), following the Nitrogen Fertiliser Optimisation Algorithm developed by Raun et al. (2002).

Although the variance of crop growth status can be revealed by remote sensing for *in situ* N optimisation, the temporally dynamic interactions of stress like N availability on crop growth and yield cannot be totally accounted for. Compared with the spatial variability in crop N requirements, the temporal variability of weather conditions is more important, and should be better managed (Miao et al., 2011). Relying on the ability of simulating the comprehensive interaction between the soil and crop processes under a given atmospheric condition (e.g., Yin and van Laar, 2005), process-based crop models have been considered as a powerful tool (Paz et al., 1999). As future weather is always uncertain, the historical weather data over a long time have been commonly utilised for crop growth simulation to optimise N_{top} (e.g., Paz et al., 1999;

Wang et al., 2021a). However, due to the gathering cost of soil data, the between- and within-field variabilities of crop growth cannot be simulated sufficiently by crop model only (Basso et al., 2007). To simulate the space-time continuum of crop production by a crop model, Paz et al. (1999) divided the field into a grid for optimising N_{tot} , in which the crop model was separately run with calibrated soil parameters by measured yield for each grid cell. Coupling a crop model and remote sensing technologies has recently been shown to better optimise field N management. For instance, based on the remotely sensed vegetation index or yield map, a field was delineated into a few stable zones of similar crop response and the crop model DSSAT was executed with specific inputs in each of these zones (Basso et al., 2007). Instead, Baret et al. (2007) generated a map of N_{top} for a targeted field at pixel level (20 m \times 20 m), based on the updated crop growth status by assimilating in-season remote sensing observations into the crop model STICS. With the improved forecasting accuracy of crop growth status, the optimised N_{top} is supposed to benefit from the integrated system as well (Baret et al., 2007). Thus, more work is needed about assimilating remote sensing information into crop models for optimising in-season *in situ* N management (Morari et al., 2020).

In this study, methods for optimising field N management based on remote sensing image and crop model were developed, and evaluated in comparison with earlier remote sensing-based and yield response curve-based methods. Due to the great N loss potential caused by frequent irrigation, timely N management is especially important for rice, compared with other major field crops (Fageria and Baligar, 2005). Our field experiments of rice conducted in years 2019 and 2020, reported earlier (Wang et al., 2023a, 2023b), were adopted here as a case study. The crop model GECROS was used, because of its generality and physiological robustness (Yin and van Laar, 2005) (see a brief description of the crop model GECROS in Supplement Appendix A).

5.2 Materials and methods

5.2.1 Field experiment design

Details of the experimental design were provided previously (Wang et al., 2023a, 2023b), and here, only summary information is given. The field study was conducted on a paddy field, located in Chongming, Shanghai, China (31.67N, 121.35E), a typical site for rice cultivation in Yangtze River Delta (Fig. 5.1). The soil type of the experimental site was silt loam. The meteorological records of the local weather station around our experimental site were collected

from the website of the China meteorological data service centre (<http://data.cma.cn>). Monthly average temperature and rainfall from 1989 to 2020 are shown in Fig. S5.1.

The N_{tot} of the local FP is 240 kg N ha⁻¹ per season. To assess the optimal N_{tot} , six treatments with varied N application amounts from 0 to 320 kg N ha⁻¹ were arranged, in a complete randomised block design, with three (2019) or four (2020) blocks, in which each single plot occupied an area of 180 m². Rice was sown with a row spacing of 20 cm and a within-row seed spacing of 2-3 cm on 14-Jun (2019) or 4-Jun (2020). The application time and amount of N were split as scheduled (Table 5.1). Sufficient phosphate (112.5 kg P₂O₅ ha⁻¹) and potash (112.5 kg K₂O ha⁻¹) fertilisers were applied per season, based on the local practice, to prevent phosphorus and potassium deficiencies. Other soil and crop management practices (including irrigation, and pest, weed and disease control) were the same in each plot following local standard practices.

The optimisation of N_{top} was established in the field experiment in 2020. In this study, the decision date was set at two weeks after tillering. Before the decision date, half of the N_{tot} for each treatment was applied at the stages of seedling and beginning of tillering, following the field experimental design (Table 5.1). Thus, six treatments with the applied N varied from 0 to 160 kg N ha⁻¹ were generated for evaluating N optimisation methods regarding the third and fourth N_{top} . The treatments were separated into two categories of N conditions accordingly, the deficient (for the treatments with applied N before the decision date of 0, 20, 40 and 80 kg N ha⁻¹) and the sufficient (for those of 120 and 160 kg N ha⁻¹).

5.2.2 Field measurements and the collection of hyperspectral images

At harvest, two sampling sites were selected in each plot and their sampling areas were 2 m² (2019) or 4 m² (2020). The total fresh samples were weighed and ca. 20% was dissected into component plant parts, including grains. The sampled grains were weighed after oven-drying at 70 °C until constant weight. Finally, grain yield of rice was adjusted to a moisture content of 14%.

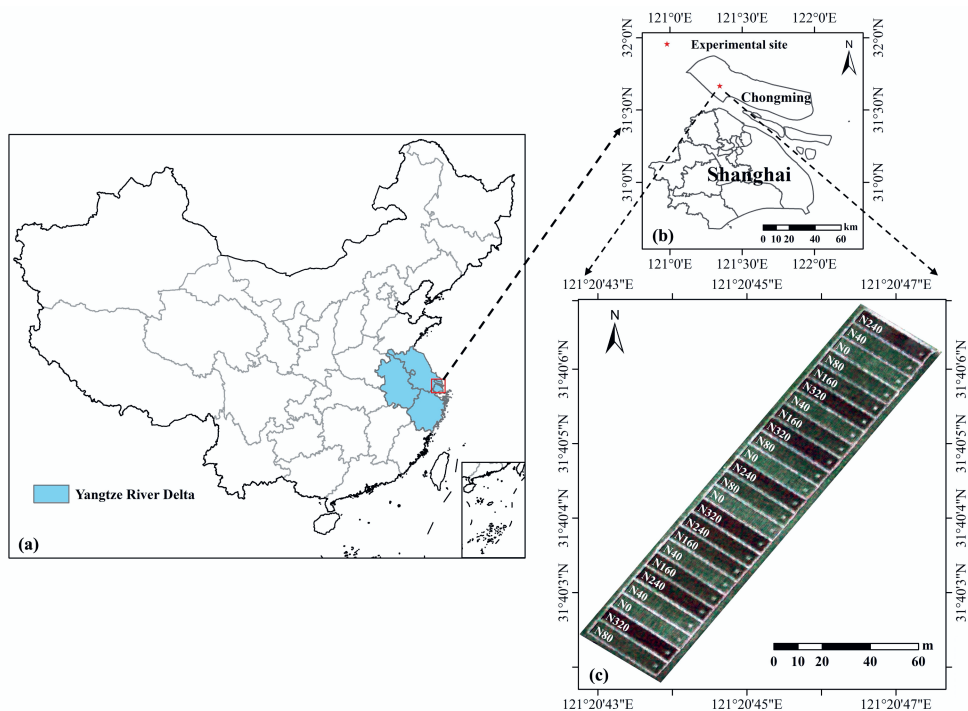


Fig. 5.1 Study site of the field experiment of rice (a, b). The true colour image (Red: 638 nm, Green: 550 nm, Blue: 470 nm) at the stem-elongating stage in 2020 is illustrated here (c). N0, N40, N80, N160, N240 and N320 denote different rates of nitrogen (N) application (see Table 5.1 for details).

Table 5.1 Split-applied nitrogen (N) fertiliser rates at different growth stages during the rice growing season in two experimental years following farmer practice, based on Wang et al. (2023a, 2023b).

Year	Application stage	N rate (kg N ha ⁻¹)					
		0	40	80	160	240	320
2019	Beginning of tillering	0	16	32	64	96	128
	Two weeks after tillering	0	8	16	32	48	64
	Panicle initiation	0	10	20	40	60	80
	Two weeks after flowering	0	6	12	24	36	48
2020	Seedling	0	12	24	48	72	96
	Beginning of tillering	0	8	16	32	48	64
	Two weeks after tillering	0	8	16	32	48	64
	Panicle initiation	0	12	24	48	72	96

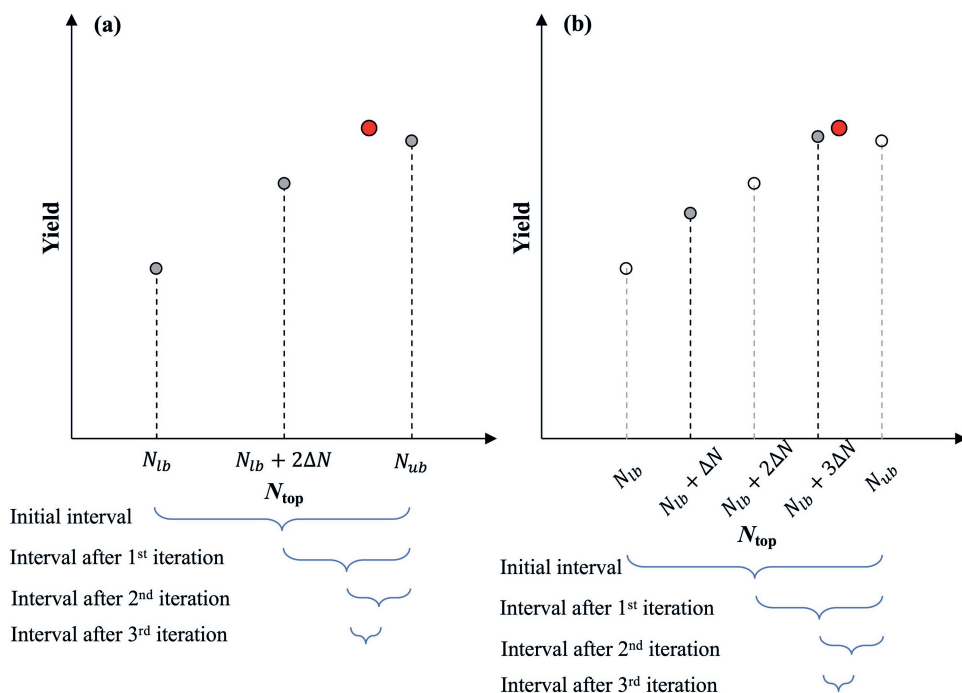


Fig. 5.2 Illustration of the adaptation of the bisection algorithm to optimise N topdressing (N_{top}) in our study. The red dot is the root of the optimal N_{top} searching from N_{lb} to N_{ub} . ($N_{lb} + 2\Delta N$) is the midpoint of the initial range of N_{top} , with ΔN being equal to $(N_{ub} - N_{lb})/4$.

At the decision date of two weeks after tillering, the canopy reflectance data of the hyperspectral images were acquired. A DJI M600 PRO hexacopter (DJI, Shenzhen, China), equipped with a Cubert S185 hyperspectral snapshot camera (Cubert GmbH, Ulm, Baden-Württemberg, Germany), was flown over the experimental paddy field between 10 a.m. and 2 p.m. 125 spectral bands were captured in the range of 450-950 nm with a sampling interval of 4 nm. The spatial resolution of remote sensing images and fertilisation maps was set at 1 m² in this study, due to the distinguished spatial variability recommended from previous studies (e.g., Solie et al., 1996).

5.2.3 Nitrogen optimisation strategies

In this study, we proposed two methods, a crop model-based (CM) method and an integrated remote sensing-crop model (RSCM) method, to optimise N fertilisation. These two methods

were compared with three reference methods (two remote sensing-based (RS) methods, and yield response curve-based method), in terms of either yield or profit. Profit, P , was defined as,

$$P = yp_y - xp_x \quad (5.1)$$

where y and x represent the yield and N application amount, respectively, and p_y and p_x represent their prices, respectively. Other costs, like labor, fuel and machinery, are not included here. The values of p_y and p_x were from local government (<https://www.shcm.gov.cn>) and defined as 0.407 and 0.875 \$ kg⁻¹, respectively.

Moreover, the varied N_{tot} following the fixed ratio of N partitioning from FP was also incorporated for comparison. As there was no optimisation of partitioning ratios of N_{top} in FP, the amount of total N_{top} , the sum of the third and fourth N_{top} , was as same as the amount of N applied before the decision date (Table 5.1). Within the RS methods, as only the third N_{top} could be optimised after acquiring hyperspectral images at the decision date of two weeks after tillering, the fourth N_{top} was kept as same as that of yield response curve-based method for consistency with the limitation that the N_{tot} should not exceed the optimised N_{tot} from the production function (see later).

5.2.3.1 The crop model-based optimisation method

In the CM method, the N_{top} optimisation was based on the crop model with a bisection algorithm (Conte and De Boor, 1965; Supplement Appendix B). While applying the method for optimising N_{top} in the interval of the lower bound (N_{lb}) to the upper bound (N_{ub}) to maximise yield or profit, the midpoint ($N_{lb} + 2\Delta N$) of the interval is replaced with either N_{lb} or N_{ub} that comes with lower simulated yield or profit, where ΔN is calculated as $(N_{ub} - N_{lb})/4$ (Fig. 5.2a). As yield always increases with an increase in N apply until the optimal N_{top} and then may slightly decrease, to reduce the required times of simulation and improve searching efficiency, the new interval can be selected by comparing the simulated yield or profit of the midpoints of candidate intervals of $[N_{lb}, N_{lb} + 2\Delta N]$ and $[N_{lb} + 2\Delta N, N_{ub}]$ (Fig. 5.2b).

To optimise the third and fourth N_{top} simultaneously, a double-bisection algorithm was designed as follows:

1. The amount of leftover N after the first two N_{top} was obtained according to the whole-season N_{tot} optimised from the production function for yield response curve (see later). Both the third and fourth N_{top} were varied in the initial interval of zero (lower bound, N_{lb}) to the amount of leftover N (upper bound, N_{ub}).

2. The ΔN of the third and fourth N_{top} was set to $(N_{ub,3rd} - N_{lb,3rd})/4$ and $(N_{ub,4th} - N_{lb,4th})/4$, respectively. The candidate intervals of the third N_{top} were bisected as $[N_{lb,3rd}, N_{lb,3rd} + 2\Delta N_{3rd}]$ and $[N_{lb,3rd} + 2\Delta N_{3rd}, N_{ub,3rd}]$, while that of the fourth N_{top} were $[N_{lb,4th}, N_{lb,4th} + 2\Delta N_{4th}]$ and $[N_{lb,4th} + 2\Delta N_{4th}, N_{ub,4th}]$.
3. The midpoints of candidate intervals of the third N_{top} , $(N_{lb,3rd} + \Delta N_{3rd})$ and $(N_{lb,3rd} + 3\Delta N_{3rd})$, were paired with that of the fourth N_{top} , $(N_{lb,4th} + \Delta N_{4th})$ and $(N_{lb,4th} + 3\Delta N_{4th})$. The pair that total amount of the third and fourth N_{top} was higher than the amount of leftover N was recognised as illegal and excluded.
4. The yields or profits were derived from the simulated crop growth by the crop model with weather data and the candidate N_{top} pairs.
5. To maximise yield or profit, the targeted N_{top} pair with maximal yield or profit was selected among the candidates. If the relative increment of yield or profit was less than 0.1%, the N_{top} pair with the minimal N_{tot} was selected to avoid excessive use of N.
6. The new intervals of the third and fourth N_{top} were updated to that of the selected N_{top} pair. ΔN_{3rd} and ΔN_{4th} were updated accordingly.
7. The ranges of the third and fourth N_{top} were narrowed by iterating the steps from 2 to 6. When both the differences of the upper and lower bound of the optimised ranges of third and fourth N_{top} were less than 1 kg N ha^{-1} , the iteration stopped and the N_{top} pair with the maximal yield or profit was considered as the optimal.

For the CM method, the weather data fusion approach (Chen et al., 2020) was incorporated. Take optimising N_{top} in 2020 for example, thirty single-year historical (1990-2019) weather files were generated and the historical daily weather data until the decision date in each weather file were replaced by the daily records in 2020. The third and fourth N_{top} of rice were optimised based on the proposed CM method with fused weather data series. Finally, the averaged N_{top} was served as the final optimised decision.

5.2.3.2 The integrated remote sensing-crop model method

A method of RSCM was proposed for N_{top} optimisation, in which the CM method was further incorporated into the integrated system for crop status monitoring and data assimilation developed in our previous study (Wang et al., 2023a). Within the integrated system, Monte Carlo-based Ensemble Kalman Filter (Evensen, 1994), one of the most popular methods for data assimilation (Carrassi et al., 2018), is applied to assimilate the crop model simulations and

remotely sensing observations based on their quantified uncertainties. A Markov Chain Monte Carlo approach (Vrugt et al., 2009; Schoups and Vrugt, 2010) was adopted to calibrate the uncertain parameters in crop model GECROS and estimate their posterior probability distribution functions. The uncertainties of crop growth simulations were estimated simultaneously by incorporating an assumed residual model. Remotely sensed canopy-level leaf traits, leaf weight, leaf N content and leaf area index, were predicted by a Gaussian Process Regression model (Rasmussen and Williams, 2006), with the uncertainties of remote sensing observations estimated from the model itself. The feature set combining reflectance, vegetation indices and texture information was extracted from the collected hyperspectral images for the better prediction of leaf traits (Wang et al., 2023b). While applying Ensemble Kalman filter, the procedure iteratively alternates between model forecasting and state updating. Each forecasting step produces an ensemble of different crop model simulations with uncertainties from an ensemble of parameter sets, which were sampled from the estimated posterior probability distribution functions. Each updating step uses the predicted remotely sensed observations, weighted by the estimated uncertainties from the Gaussian Process Regression model, to correct the ensemble forecast. Consequently, the crop model simulations of canopy-level leaf traits were updated directly in the integrated system and those of aboveground biomass, grain weight, aboveground N content and grain N content were updated indirectly for their better forecasting (Wang et al., 2023a).

To this end, based on the acquired hyperspectral images in 2020, the crop status at two weeks after tillering in the experimental area was updated pixel by pixel by the integrated system accordingly. The third and fourth N_{top} were optimised in a pixelated manner in the RSCM method. Like with the CM method, historical weather records after the decision date were used in the RSCM method to optimise N_{top} under different weather conditions and the averaged N_{top} map was generated as the final decision.

5.2.3.3 Remote sensing methods

Two remote-sensing methods are evaluated. The two methods are based on Sufficiency Index (SI) and Response Index (RI), respectively. The SI is expressed as:

$$SI = \frac{V_{sen}}{V_{ref}} \quad (5.2)$$

where V_{sen} and V_{ref} represent the sensed crop property and that from the referenced crop of N-rich treatments under the same measurement condition, respectively. SI ranges from zero to

one; the closer SI is to one, the more sufficient the N supply is. V_{ref} in this study was from the treatment 240 kg N ha⁻¹, which is also the local FP. Following Holland and Schepers (2013), V_{sen} or V_{ref} was calculated as $(reflectance_{nir}/reflectance_{re} - 1)$, i.e. based on the reflectance of the waveband of near infrared (nir, 850 nm) and red-edge (re, 730 nm) determined from our previous study (Wang et al., 2023b), respectively.

The RS method with SI (RS_{SI}) developed by Holland and Schepers (2010) was adopted for comparison, in which the N_{top} is formed as,

$$N_{top} = (N_{tot,opt} - N_{PreDeci} - N_{OM}) \sqrt{\frac{1-SI}{\Delta SI(1+0.1e^{m(SI_{Threshold}-SI)}}} \quad (5.3)$$

where $N_{tot,opt}$ is the optimal N_{tot} for maximising yield or profit, $N_{PreDeci}$ is the amount of N applied before the decision date, and N_{OM} is the N credit for the organic matter content within the field. ΔSI is defined as the sufficiency index difference parameter and calculated as $1 - SI_0$, in which SI_0 denotes the SI when there is no N application, representing the difference between the SI of healthy plants (SI = 1.0) and those hardly able to recover by topdressing N. m and $SI_{Threshold}$ are the back-off parameters controlling the reducing rate of N_{top} for situations with reduced yield potential (i.e., plant density), which are described as the back-off rate variable and the back-off cut-on point, respectively. In this study, the $N_{tot,opt}$ was determined from the production function (see later). $N_{PreDeci}$ values for different treatments were derived from field records following Table 5.1. The values of ΔSI were set at 0.41, as the measured SI_0 in our field experiment was 0.59. m was set at 20, according to the results of Holland and Schepers (2010) and Zhang et al. (2021). $SI_{Threshold}$ was set at 0.59 in this study, as it is supposed to coincide with the SI_0 point (Holland and Schepers, 2010). N_{OM} was estimated to be 20-30 kg N ha⁻¹ per 1% soil organic matter (Holland and Schepers, 2013), and set to 60 kg N ha⁻¹ for this study, based on the tested soil organic matter of 3% at the experimental site.

The in-season RI, proposed by Raun et al. (2002), is equivalent to the reciprocal of SI in our context, i.e.: $RI = V_{ref}/V_{sen}$. N_{top} optimised by the RS method with RI (RS_{RI}), known as the Nitrogen Fertiliser Optimisation Algorithm, is described as (Raun et al., 2002; Raun et al., 2005):

$$N_{top} = \frac{N_{grain}}{NUE} YP_0 (RI - 1) \quad (5.4)$$

where N_{grain} represents the average N uptake (2.0 kg) per 100 kg grains for Japonica rice cultivars in Yangtze River Delta (Ling et al., 2005). Nitrogen use efficiency (NUE, harvested grain N (kg) per 1 kg of applied N fertiliser) is set to 0.6, as the theoretical NUE of an in-season

N application is ranged from 0.5 to 0.7 (Raun et al., 2005). The prediction of yield potential with no added fertilisation (YP_0 , kg ha⁻¹) is followed the form of $YP_0 = aINSEY + b$, in which In-Season Estimated Yield (INSEY, day⁻¹) is determined by dividing V_{sen} by the number of growing degree days (GDD) > 0 from emergence to senescence. The GDD is calculated as $GDD = (T_{min} + T_{max})/2 - T_b$. T_{min} and T_{max} represent the daily minimum and maximum air temperature, respectively, while T_b is the base temperature for phenological development and set at 8.0 °C for rice (Yin and van Laar, 2005). The potential yield with additional N fertiliser, the product of YP_0 and RI, should not exceed the maximum of field measurements (Raun et al., 2002), which is equal to 11.2 t ha⁻¹ in this study. The RI is capped at 2.0 as the in-season N_{tot} would unlikely lead to the potential yield two times greater than the baseline YP_0 (Raun et al., 2002).

5.2.3.4 Production function and yield response curve

The optimal N_{tot} in this study was defined as the N_{tot} at which the yield or profit was maximised. Here, yield response curve to N observed in our experiments was fitted to a cubic model, a simplified application of the growth functions proposed by Yin et al. (2003):

$$y = y_0 + \frac{(y_{max} - y_0)(3x_e - 2x)x^2}{x_e^3} \quad (5.5)$$

where y_0 denotes the value of yield when there is no N input, y_{max} denotes the maximum value of yield and x_e represents the amount of N when y_{max} is achieved. In the form of this equation, all the three parameters have a straightforward biological meaning.

The economically optimum N_{tot} is defined as the N_{tot} when the first derivative of the production function equals to the price ratio of fertiliser N to yield (e.g., Cerrato and Blackmer, 1990). Based on Eqn (5.1), the derivative of profit with respect to N is expressed as:

$$\frac{dP}{dx} = \frac{dy}{dx} p_y - p_x \quad (5.6)$$

in which $\frac{dy}{dx}$ denotes the derivative of y with respect to x (or called the marginal yield); it can be solved from Eqn (5.5) as $\frac{dy}{dx} = \frac{6(y_{max} - y_0)}{x_e^3} x(x_e - x)$. The highest profit achieves when $\frac{dP}{dx} = 0$.

Thus, the optimal N for maximising profit, x_p , is solved as:

$$x_p = \left(kx_e + \sqrt{(kx_e)^2 - 4k \frac{p_x}{p_y}} \right) / (2k) \quad (5.7)$$

where k is $\frac{6(y_{max} - y_0)}{x_e^3}$. x_p is available when $\frac{p_x}{p_y} \leq \frac{kx_e^2}{4}$.

The measured yield across varied N_{tot} in both experimental years was used to fit the yield response curve. The determined optimal N_{tot} for maximising yield or profit was able to be used directly to optimise the FP for the corresponding target, named FP_{opt} thereafter, in which the leftover N was split-applied for the last two N_{top} with fixed partitioning ratio of FP.

5.2.4 Evaluation of the nitrogen optimisation methods

The above five methods were applied for optimising the last two N_{top} based on our field experiment in 2020. The crop growth with the applied first two N_{top} was simulated and updated at two weeks after tillering in 2020 based on the integrated system (described in Section 5.2.3.2). The last two N_{top} optimised from different methods was fertilised in the forecast of their corresponding crop growth. Based on the simulated crop growth with actual weather data, the performance of the obtained N_{top} from different strategies was evaluated by yield, profit and AE_N . AE_N was calculated as:

$$AE_N = (y - y_0)/N_{\text{tot}} \quad (5.8)$$

5.3 Results

5.3.1 Field nitrogen optimisation based on yield response curve

The yield response to N_{tot} is shown in Fig. 5.3. Measured yield in both experimental years increased significantly ($P < 0.05$) from no N input until N_{tot} reached 240 kg N ha⁻¹ (Fig. 5.3a). Yield with N_{tot} at 320 kg N ha⁻¹ decreased by 0.16% (2019) and 1.89% (2020), compared with that at 240 kg N ha⁻¹. As fitted curves for individual years of 2019 and 2020 did not differ significantly ($P > 0.05$), a general yield response curve was generated using the pooled data (Fig. 5.3b). Derived from the cubic yield response model, the optimal N_{tot} for maximising yield was 282.5 kg N ha⁻¹, located in the indicated range of optimal N_{tot} (Fig. 5.3). Based on Eqn (5.7), the optimal N_{tot} for maximising profit was calculated as 273.3 kg N ha⁻¹, which was slightly lower than that for yield (Fig. 5.3b).

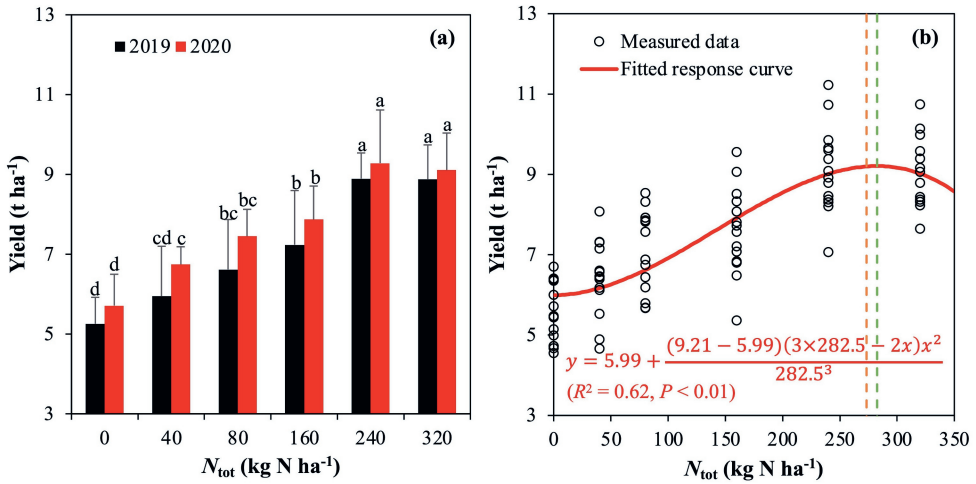


Fig. 5.3 Bar plots of measured yield (a) and yield response curve (b) to the varied N application amount (N_{tot}). The error bars in panel a denote the standard deviations derived from replications. The response equation for yield in panel b was fitted by two years' data in 2019 and 2020, and the green and orange vertical lines represent the optimised N_{tot} for maximising yield and profit, respectively.

5.3.2 Topdressing nitrogen based on the remote sensing methods

The maps of N_{top} optimised by the RS methods were generated based on the SI or RI map (Fig. 5.4). As described in the RS_{SI} method, the optimised N_{top} targeting at maximising yield and profit changed not only with the levels of SI, but also with the levels of applied N before the decision date (Fig. S5.2). Due to the slight differences between the optimised N_{tot} for yield and profit, the optimised N_{top} for yield and profit changed accordingly (Figs. 5.4b-c, S5.2). Regarding the RS_{RI} method, as it recommends N just based on the potential yield (Fig. S5.3), only the objective of maximising yield was included here (Fig. 5.4e). In N deficient treatments, the pixel-level third N_{top} optimised by the RS_{SI} method tended to be higher than that of the uniform N_{top} of FP_{opt}, while that by the RS_{RI} method tended to be lower. For instance, the third N_{top} of the RS_{SI} method for maximising yield for N treatments with applied N in levels of 0, 20 and 40 kg N ha⁻¹ increased by 66.9%, 57.0% and 19.0%, respectively, and that of the RS_{RI} method on average decreased by 8.7% (Table S5.1). The plot-level total amount of the last two N_{top} (hereafter referred to as total N_{top}) of the RS_{SI} method in all treatment for maximising yield and profit decreased, compared with that of FP_{opt}, by 39.4-51.8% and by 39.6-51.1%, respectively (Table 5.2). Similarly, the total N_{top} of the RS_{RI} method for maximising yield decreased by 42.9-61.5% (Table 5.2).

Table 5.2 The plot-level averaged topdressing N (N_{top} , kg N ha⁻¹) in different treatments of levels of N applied before the decision date (N_{FeDeci})^a. The N_{top} for maximising yield or profit was optimised from different strategies, the farmer practice (FP), the optimised farmer practice (FP_{opt}), the method of remote sensing with Sufficiency Index (RS_{SI}), the method of remote sensing with Response Index (RS_{RI}), the method of crop model (CM), and the integrated method of remote sensing-crop model (RSCM).

Target	N_{PreDeci}	Third N_{top}				Fourth N_{top}				RSCM	CM	RSCM	CM
		FP	FP _{opt}	RS _{SI}	RS _{RI}	CM	RSCM	FP	FP _{opt}				
Yield	0.0	0.0	113.0	103.6	69.4	110.0	57.8	0.0	169.5	43.8	46.5	19.1	9.3
	20.0	8.0	105.0	106.6	64.1	125.1	75.6	12.0	157.5	52.4	54.8	27.0	14.5
	40.0	16.0	97.0	73.5	39.7	119.3	72.1	24.0	145.5	52.6	53.6	22.1	13.3
	80.0	32.0	81.0	46.4	29.8	106.3	70.0	48.0	121.5	59.9	59.7	24.0	18.3
	120.0	48.0	65.0	17.1	13.4	96.8	62.9	72.0	97.5	61.3	60.3	23.2	14.9
	160.0	64.0	49.0	7.7	9.8	69.6	44.5	96.0	73.5	62.8	60.1	25.4	17.1
Profit	0.0	-	109.3	99.4	-	103.7	53.5	-	164.0	42.9	-	17.7	8.8
	20.0	-	101.3	101.8	-	110.8	69.9	-	152.0	51.1	-	27.0	14.5
	40.0	-	93.3	69.8	-	110.5	67.0	-	140.0	51.0	-	25.2	13.9
	80.0	-	77.3	43.5	-	98.6	65.9	-	116.0	58.2	-	28.2	17.7
	120.0	-	61.3	15.6	-	88.8	55.6	-	92.0	59.4	-	21.7	15.8
	160.0	-	45.3	6.6	-	58.8	38.8	-	68.0	60.8	-	23.3	17.2

(To be continued)

Table 5.2 (Continued)

Target	N_{PreDeci}	Total N_{top}	FP	FP _{opt}	RS _{St}	RS _{RI}	CM	RSCM
Yield	0.0	0.0	0.0	282.5	147.4	115.8	129.1	67.1
	20.0	20.0	20.0	262.5	159.0	118.9	152.1	90.1
	40.0	40.0	40.0	242.5	126.0	93.3	141.4	85.4
	80.0	80.0	80.0	202.5	106.4	89.5	130.3	88.3
	120.0	120.0	120.0	162.5	78.4	73.7	120.0	77.8
Profit	160.0	160.0	160.0	122.5	70.6	70.0	95.0	61.6
	0.0	-	-	273.3	142.3	-	121.4	62.3
	20.0	-	-	253.3	152.9	-	137.9	84.4
	40.0	-	-	233.3	120.9	-	135.7	80.9
	80.0	-	-	193.3	101.6	-	126.7	83.7
	120.0	-	-	153.3	75.0	-	110.5	71.4
	160.0	-	-	113.3	67.4	-	82.0	56.0

^a The decision date was set at two weeks after tillering. Two categories of N conditions were separated, the deficient (for the treatments with the N_{PreDeci} of 0, 20, 40 and 80 kg N ha⁻¹) and the sufficient (for those of 120 and 160 kg N ha⁻¹).

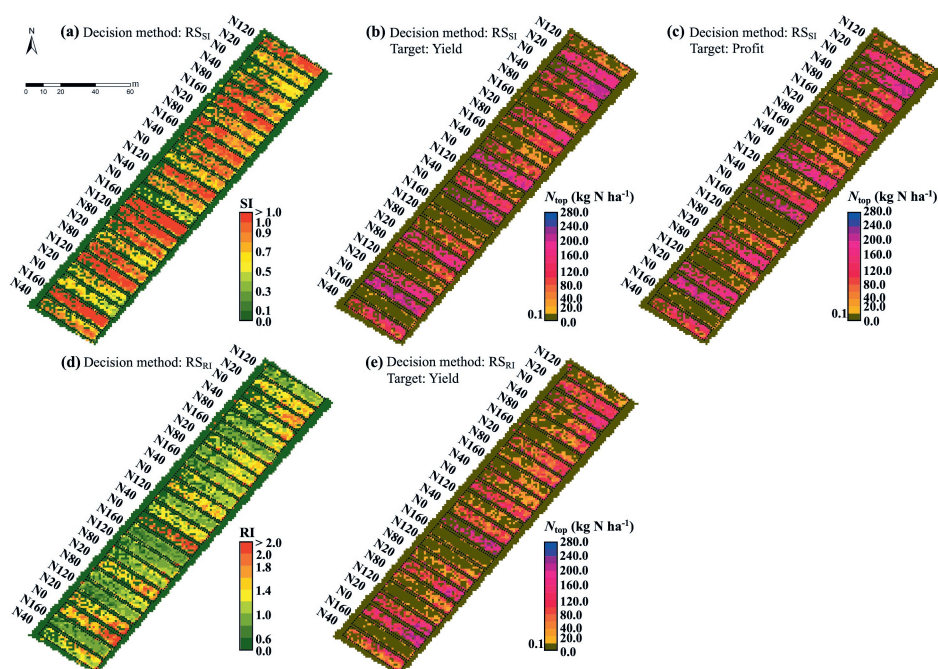


Fig. 5.4 Maps of Sufficiency Index (SI) (a), Response Index (RI) (d) and optimised topdressing N (N_{top}) by the method of remote sensing with SI (RS_{Si}) or RI (RS_{Ri}). Within the method of RS_{Si} , N_{top} was optimised for maximising yield (b) and profit (c), respectively. As the RS_{Ri} method recommends N just based on the potential yield, the objective of maximising yield was only included (e). N0, N20, N40, N80, N120 and N160 denote varied N rates applied before the decision (the decision date was set at two weeks after tillering, see the text). The dotted lines within the maps depict the boundaries of experimental plots.

5.3.3 Determining the optimal topdressing nitrogen with the crop model method

As the optimisation of N_{top} from the CM method used historical weather records, its applicability was validated at first. Compared with using the actual weather data, simulated yield by the historical records agreed well with the field measurements in both experimental years of 2019 and 2020 (Table S5.2). The response curve from the simulated yield by the model ($R^2 = 0.60$, $P < 0.01$, Fig. S5.4) was similar to that from the measured data shown in Fig. 5.3b. The amount of total N_{top} of the CM method was consistently decreased in different N treatments and the decrease for maximising yield varied from 22.4% to 54.3% and that for maximising profit was from 27.6% to 55.6%, compared with FP_{opt} (Table 5.2). Similar to FP, the partitioning ratio of the third N_{top} in FP_{opt} was fixed at 40% of the total N_{top} , while that of the CM method increased to ca. 80% and tended to be increased with the decrease of applied N

before (Table 5.2). Compared with FP_{opt} , the third N_{top} of different treatments optimised by the CM method increased by -5.1% to 48.9%, while the fourth N_{top} decreased by 65.4% to 89.2% (Table 5.2).

5.3.4 Optimisation of topdressing nitrogen by the method of remote sensing-crop model

The maps of N_{top} optimised by the integrated method of RSCM targeting at maximising yield and profit were generated, based on updated crop growth status using the integrated system for crop status monitor and forecast (Fig. 5.5). Due to the difference between the optimised N_{top} for yield and profit, like with the method of CM, the optimised third and fourth N_{top} of RSCM for yield tended to be slightly higher than that for profit in both pixel- and plot-level (Tables 5.2, S5.1). Taking the N sufficient treatment with the 120 kg N ha⁻¹ applied as an example, ca. 70% pixels of the optimised third N_{top} for maximising yield were in the range from 85 to 100 kg N ha⁻¹, while that for maximising profit ranged from 75 to 90 kg N ha⁻¹ (Fig. 5.6). The total N_{top} of RSCM in both pixel- and plot-level was lower than that of RS, and like with the CM method, more N tended to be allocated to the third N_{top} (Tables 5.2, S5.1). Compared with the CM method, benefiting from the *in situ* crop growth status, the averaged pixel-level total N_{top} of different treatments optimised by the RSCM method decreased by 4.6-13.9% (Table S5.1). Relying on the *in situ* optimisation, the plot-level total N_{top} of RSCM further decreased, by 49.7-77.2%, compared with FP_{opt} (Table 5.2).

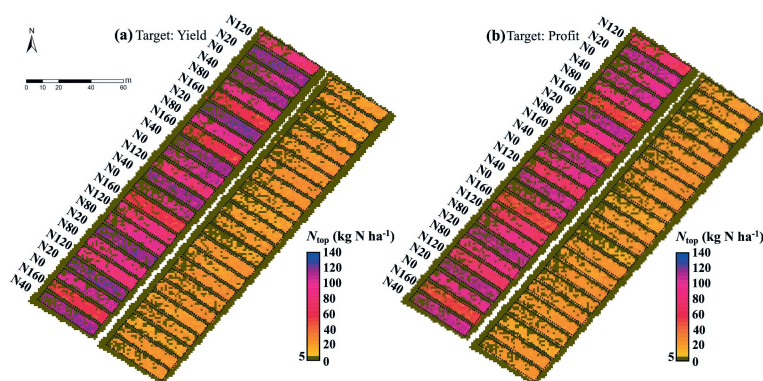


Fig. 5.5 Maps of optimised topdressing N (N_{top}) by the integrated method of remote sensing-crop model targeting at maximising yield (a) and profit (b). The third and fourth N_{top} maps were shown in upper left and lower right, respectively, in each panel. N0, N20, N40, N80, N120 and N160 denote varied N rates applied before the decision (the decision date was set at two weeks after tillering, see the text). The dotted lines within the maps depict the boundaries of experimental plots.

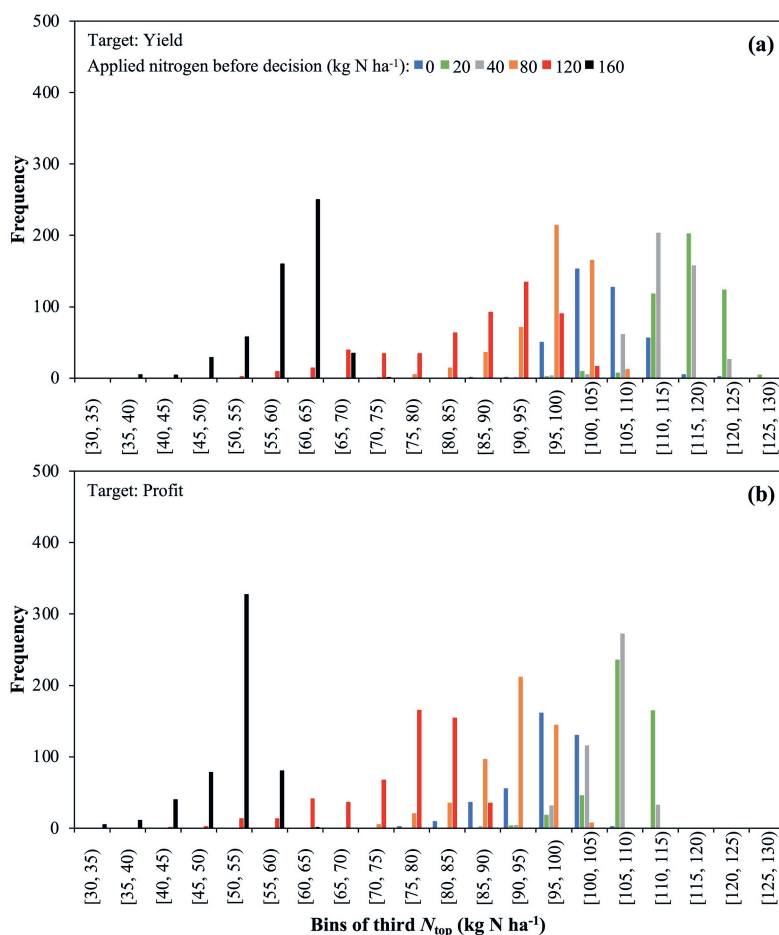


Fig. 5.6 The distribution of pixel-level optimised third topdressing N (N_{top}) by the method of remote sensing-crop model targeting at maximising yield (a) and profit (b). The decision date was set at two weeks after tillering (see the text).

5.3.5 Performance evaluation of optimised field nitrogen

Yield maps of the different N_{top} strategies were simulated using the actual weather data (Fig. 5.7), although the simulated yield tended to be higher than the measurements (Fig. S5.4). Even though the AE_N increased while decreasing N_{tot} to 40 kg N ha^{-1} following the strategy of FP, the yield and profit decreased greatly (Table 5.3). The method of FP_{opt} achieved the highest yield among the N optimisation strategies, with the cost of the highest N_{top} and lowest AE_N (Tables 5.2, 5.3). Compared with FP_{opt} , in N deficient treatments for maximising yield or profit, the simulated yield of the methods of CM and RSCM decreased by 0.1% to 1.1% and 0.2 to

1.4%, respectively (Table 5.3), while their profit increased by 1.1% to 9.4% and 2.3% to 13.2%, respectively, due to the saved N (Table 5.3). As for the sufficient N treatment with 160 kg N ha⁻¹ applied, both the simulated yield and profit of the methods of CM and RSCM decreased, but the AE_N of them for maximising yield and profit on average increased by 9.2% and 23.0%, respectively (Table 5.3). As there was hardly N_{top} for N treatments with sufficient input, not only the simulated yield and profit but also the AE_N of RS were lower than those of RSCM (Fig. 5.7, Table 5.3). Moreover, compared with the local FP, the simulated yield and profit of FP_{opt} increased by 140 kg ha⁻¹ and 20 \$ ha⁻¹, respectively, but its AE_N decreased from 19.2 to 16.8 kg kg⁻¹, while the yield and profit of the methods of CM and RSCM increased by 54 and 15 kg ha⁻¹ and 22 and 43 \$ ha⁻¹, respectively, and their AE_N increased to 19.9 and 22.6 kg kg⁻¹, respectively (Table 5.3).

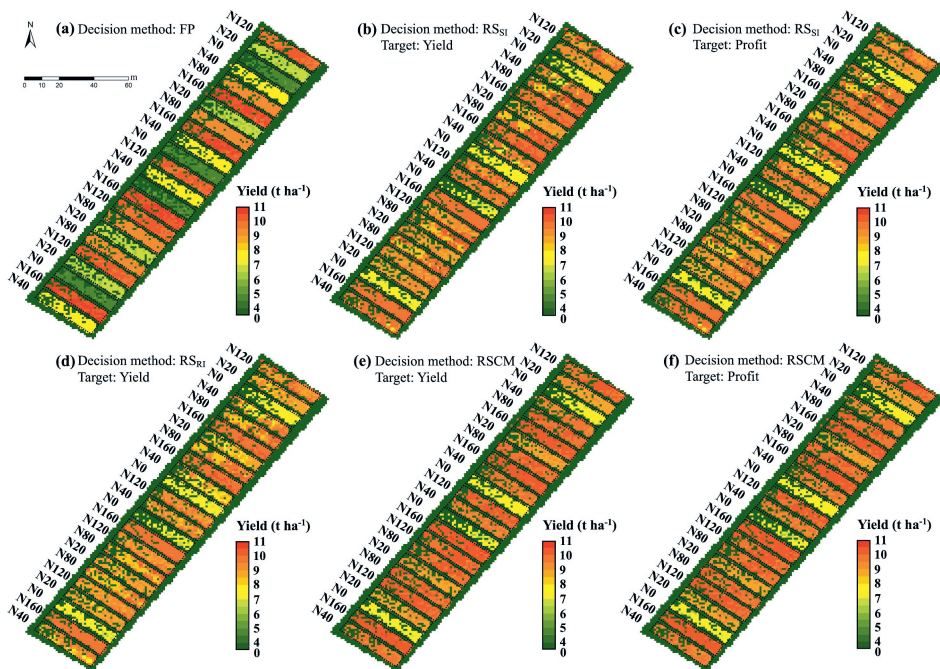


Fig. 5.7 Yield maps from different strategies, the farmer practice (FP) (a), the method of remote sensing with Sufficiency Index (RS_{Si}) (b, c) or Response Index (RS_{Ri}) (d), and the integrated method of remote sensing-crop model (RSCM) (e, f), for optimising topdressing N targeting at maximising yield and profit. As the RS_{Ri} method recommends N just based on the potential yield, the objective of maximising yield was only included (d). The yield for different strategies was simulated from the crop model GECROS using actual weather data. N0, N20, N40, N80, N120 and N160 denote varied N rates applied before the decision (the decision date was set at two weeks after tillering, see the text). The dotted lines within the maps depict the boundaries of experimental plots.

Table 5.3 The plot-level averaged yield, profit and agronomic efficiency of N (AE_N) in different treatments of levels of N applied before the decision date ($N_{PreDeci}$)^a. The strategies for maximising yield or profit included the farmer practice (FP), the optimised farmer practice (FP_{opt}), the method of remote sensing with Sufficiency Index (RS_{SI}), the method of remote sensing with Response Index (RS_{RI}), the method of crop model (CM), and the integrated method of remote sensing-crop model (RSCM). The yield for different strategies was simulated from the crop model GECROS using actual weather data.

Target	$N_{PreDeci}$	Yield (t ha ⁻¹)			Profit (\$ ha ⁻¹)			RS_{SI}	RS_{RI}	CM	RSCM	FP	FP_{opt}	RS_{SI}	RS_{RI}	CM	RSCM
		FP	FP_{opt}		FP	FP											
Yield	0.0	2.70	4.07		4.06	1099	1408	1525	1545	1540	1594						
	20.0	4.25	5.79		5.77	1696	2110	2198	2169	2195	2243						
	40.0	4.94	6.09		6.03	1939	2232	2273	2152	2296	2340						
	80.0	6.56	7.20		7.13	2530	2685	2604	2514	2716	2747						
	120.0	7.31	7.45		7.37	2766	2786	2626	2592	2788	2809						
	160.0	7.86	7.87		7.76	2920	2954	2826	2809	2936	2948						
Profit	0.0	-	4.07		4.06	-	1416	1529	-	1546	1596						
	20.0	-	5.79		5.75	-	2118	2203	-	2203	2243						
	40.0	-	6.09		6.03	-	2239	2274	-	2301	2339						
	80.0	-	7.17		7.13	-	2680	2602	-	2721	2741						
	120.0	-	7.43		7.34	-	2786	2624	-	2786	2806						
	160.0	-	7.86		7.72	-	2961	2825	-	2930	2942						

(To be continued)

Table 5.3 (Continued)

Target	N_{PreDeci}	AE_N (kg kg ⁻¹)	FP	FP _{opt}	RS _{SI}	RS _{RI}	CM	RSCM
Yield	0.0	-	-	4.8	8.4	11.0	10.5	18.4
	20.0	38.8	38.8	10.9	17.2	21.1	17.8	27.6
	40.0	27.9	27.9	12.0	19.6	21.6	18.4	26.5
	80.0	24.1	24.1	15.9	22.0	22.6	21.0	26.2
	120.0	19.2	19.2	16.8	21.1	21.1	19.4	23.4
	160.0	16.1	16.1	18.3	20.6	20.4	19.9	22.6
Profit	0.0	-	-	5.0	8.7	-	11.2	19.7
	20.0	-	-	11.3	17.8	-	19.3	29.0
	40.0	-	-	12.4	20.1	-	19.0	27.3
	80.0	-	-	16.4	22.5	-	21.4	26.8
	120.0	-	-	17.3	21.4	-	20.1	24.1
	160.0	-	-	18.9	20.8	-	20.7	23.1

^a The decision date was set at two weeks after tillering. Two categories of N conditions were separated, the deficient (for the treatments with the N_{PreDeci} of 0, 20, 40 and 80 kg N ha⁻¹) and the sufficient (for those of 120 and 160 kg N ha⁻¹).

5.4 Discussion

5.4.1 Characteristics of in-season nitrogen topdressing from different nitrogen optimisation methods

Guerrero et al. (2021) indicated that more N is suggested to be fed to the poor fertility fields and less to the rich fields. In line with the N_{top} at stem-elongating of winter wheat optimised by the RS_{SI} method (Zhang et al., 2021), our results also showed that higher N_{top} at the decision date of two weeks after tillering for rice tended to be recommended for the N deficient treatments (Table 5.2). Regarding the N sufficient treatments, less N was recommended by the RS methods, as SI tended to be higher than 1.0 or RI tended to be lower than 1.0 (Fig. 5.4). However, Huang et al. (2015) suggested that as the optimal or sufficient crop N status in the vegetative phase only indicates the N status at that particular stage, a certain amount of N fertiliser should still be recommended for the required N needs of the crop until the maturity. Our results showed that for the N sufficient treatments, compared with FP_{opt} , the third N_{top} optimised from the methods of CM and RSCM varied by 29.7% to 48.9% and -14.5% to -3.2% while that of the RS method decreased by 73.7% to 85.5% (Table 5.2).

While applying the strategy of FP_{opt} , more yield or profit can be achieved only through directly optimising N_{tot} , in which N_{top} was adjusted with a fixed ratio of N partitioning. Like with FP and FP_{opt} in this study, the ratio of the third (mid-tillering) and fourth (panicle initiation) N_{top} is recommended as two to three (Peng et al., 2006). However, in the RS methods, for N deficient treatments, 33.3% to 70.3% of total N_{top} was allocated to the third N_{top} , while only 9.8% to 21.8% was allocated to the third N_{top} in N sufficient treatments (Table 5.2). Especially, the N partitioning of the methods of CM and RSCM performed consistently among different treatments and on average, 80.4% of total N_{top} was applied in the third N_{top} (Table 5.2), indicating that more N should be allocated for tillering. Thus, the difference in early growth vigour of rice might be caused by the adjusted in-season N partitioning, which is always overlooked in N optimisation methods with the fixed one (Peng et al., 2010).

5.4.2 Optimisation of whole-season nitrogen application amount

By means of the production function in the form of the quadratic model, the economically optimal N_{tot} of rice along the Yangtze River Basin was determined within the range from 180 to 285 kg N ha⁻¹ (Chen et al., 2011) and higher optimal N_{tot} was found in regions with better

environmental conditions, like the Yangtze River Delta (Ren et al., 2022). However, it is noticeable that the identified optimal N_{tot} tends to differ between forms of production functions (Cerrato and Blackmer, 1990). For instance, in this study, the optimised N_{tot} for yield from the commonly used linear-plateau and quadratic model was 153.6 and 373.3 kg N ha⁻¹, respectively (Eqns (S5.1-S5.3) in Supplement Appendix C, Fig. S5.5), while that from our cubic model was 282.5 kg N ha⁻¹ (Fig. 5.3b). The measured data indicated that the linear plateau model overestimated yield in the section of the response curve close to the optimised N_{tot} and thus resulted in a lower optimum N_{tot} , while the quadratic model indicated an even higher optimum N_{tot} (Figs. 5.3a, S5.5), in line with the statements of Cerrato and Blackmer (1990). Consequently, the cubic model performed better in our results (Fig. 5.3b).

The *in situ* N managements based on the production function are able to decrease the required N inputs, as both spatial and temporal variability are considered appropriately (Mamo et al., 2003). For instance, compared with the optimised N_{tot} (250 kg N ha⁻¹) from the production function, the optimised N_{tot} for rice from the *in situ* approaches was lower and ranged from 166 to 233 kg N ha⁻¹ (Xue and Yang, 2008). Likewise, our results also showed that the N_{tot} optimised for maximising yield for different treatments varied in the range from 115.8 to 230.6 kg N ha⁻¹ (from the RS methods), from 129.1 to 255.0 kg N ha⁻¹ (from the CM method) and from 67.1 to 221.6 kg N ha⁻¹ (from the RSCM method), all lower than that from the production function (282.5 kg N ha⁻¹) (Table 5.2). Moreover, Xue and Yang (2008) recommended that not only N_{top} should be fine-tuned, but also the proportion of the basal N applied could be properly reduced. The optimisation of the basal fertiliser will be incorporated together in the further study for the more efficient whole-season N arrangement.

5.4.3 Responses of optimised field nitrogen management to yield, profit and nitrogen use efficiency

In line with the optimised N_{top} for wheat using the method of N nutrition index (Jiang et al., 2023), our simulated results showed that the performance of the optimised N_{top} from our CM and RSCM methods was reasonable among varied N treatments of rice. Compared with FP, the yield increment of these optimisation methods in N deficient treatments was similar and on average ranged from 27.5% to 29.9%, while the decrement of the optimised N_{tot} for the N sufficient treatment with 160 kg N ha⁻¹ applied varied from 20.3% to 32.5% (Tables 5.2, 5.3). Compared with the CM method, the RS methods tended to save more N by the sacrifice of grain yield (Table 5.3). The N_{tot} of the RSCM method was able to be further decreased to a lower

level than that of the RS methods, but a similar grain yield with that of CM was maintained (Tables 5.2, 5.3). The in-season N_{top} optimisation considering both soil N supply and crop N demand contributes to a higher N use efficiency (Flowers et al., 2004). AE_N of different treatments with optimised N_{top} by the RSCM method increased to the range from 18.4 to 27.6 kg kg⁻¹, generally agreeing with values for the developed countries (20-25 kg kg⁻¹) (Zhang et al., 2008), while that of FP_{opt} decreased to the range from 4.8 to 18.3 kg kg⁻¹ (Table 5.3).

In reviews of N optimisation, the *in situ* optimised N seems more profitable than the uniform one (e.g., Pedersen et al., 2020). For instance, after delineating the management zones, the profit from the *in situ* N_{top} for maize simulated by the crop model SALUS increased by 12 € ha⁻¹ (13 \$ ha⁻¹), compared with that from the uniform FP (Basso et al., 2016). Pedersen et al. (2021) integrated the simulated remote sensing proxies into the DAISY crop model and found that there was an expected profit increase from 6 to 29 € ha⁻¹ (7 to 32 \$ ha⁻¹). Our results further demonstrated the utility of the RSCM method with actual remote sensing images. Especially, compared with local FP, the profit of the methods of FP_{opt} , CM and RSCM for maximising yield increased by 20, 22 and 43 \$ ha⁻¹, respectively, although that of the RS methods decreased by 140-174 \$ ha⁻¹ (Table 5.3). Even though the profit of the method of FP_{opt} achieved the highest in N sufficient treatments, higher increment of yield and profit was achieved under N deficient conditions by the methods of CM and RSCM, in which the profit of them on average increased by 24.0% and 26.6%, respectively, compared with that of FP (Table 5.3). The results were consistent with the conclusions from Guerrero et al. (2021). However, field N under N sufficient conditions still should be optimised carefully to prevent the yield and profit loss, taking the RS methods for example (Table 5.3).

5.4.4 The applicability of crop model in field nitrogen management

Like with previous studies (Paz et al., 1999; Wang et al., 2021a), our results also showed that the crop model GECROS could be used for guiding field N management (Tables 5.2, 5.3). However, it has been claimed that the use of the crop model is limited due to the required inputs and parameters (e.g., Guerrero et al., 2021). Thus, machine learning models have been regressed by combining soil, weather and management data with remotely sensed crop data (Wang et al., 2021b). However, the optimisation of N_{top} is affected by the selection of machine learning algorithms and might fail due to the inaccurately simulated yield response (Wang et al., 2021b; Zhang et al., 2023). Meanwhile, due to the black-box nature of the machine learning algorithms, the lack of the process-based response of the crop growth hinders the in-season optimisation of

N_{top} , especially when considering it over a long period of time. Instead, the strength of a crop model for determining the optimal N_{tot} or N_{top} lays in interactions between genotypes, soil, weather and management and the long-term simulations under different weather conditions (Basso et al., 2016; Jiang et al., 2019).

Nonetheless, until now, most crop model-based approaches are still used for optimising the averaged N rates in years or sites, rather than in-season or *in situ* (Bai and Gao, 2021). In line with previous reports for N optimisation of maize at planting and V8 stage (Wang et al., 2021a), our results also showed that the weather data fusion approach enables the crop model to forecast the rice growth of entire growing season for optimising the in-season N_{top} (Tables 5.2, S5.1). Moreover, within the integrated system of RSCM, the remote sensing observation provides the missing heterogeneously spatial information required by the crop model for the improved yield forecasting (Morari et al., 2020). Like with the results of Baret et al. (2007), relying on the RSCM method, the third and fourth N_{top} was optimised *in situ* without the cost of yield or profit loss (Table 5.3). Especially, compared with previous studies that optimises N_{top} by various N levels (e.g., Baret et al., 2007; Wang et al., 2021a), the N_{top} in this study was optimised continuously.

5.4.5 Prospect of the long-term application of the proposed nitrogen optimisation methods

Although the N optimisation methods were evaluated comprehensively by the simulations from the calibrated crop model with the actual weather data, their actual performance in real farmers' fields is of importance, especially when considering the effect of the long-term N optimisation for the sustainable agricultural production. When applying these methods to real farmers' fields, the costs from information acquisition, information processing to decision implementation, need to be further included. Pedersen et al. (2020) estimated the all included costs and pointed out that it is ranged from 5 to 80 € ha⁻¹ (5 to 88 \$ ha⁻¹) depending on farm size. Based on the cost estimation, the varied N application from the method of RSCM in this study could be relevant on farms with 100 ha and above. By further coupling with soil information, it could be relevant on even smaller farms (Pedersen et al., 2021). Moreover, after a long-term application of the *in situ* management, the between- and within-field variability shall be diminished (Pierce and Nowak, 1999) and thus the field management can be simplified, contributing to the increase of profitability. While there are obstacles like acquiring remote sensing images and implementing *in situ* decisions, the CM method is recommended based on our results (Table

5.3). With the optimised long-term fertiliser management plan, severe environmental issues from over or under fertilisation can be prevented for the sustainable agricultural production in the era of smarting farming (Berger et al., 2020).

5.5 Conclusions

In this study, two new methods, CM and RSCM, for N optimisation were developed and evaluated by the simulated crop growth from the crop model GECROS in varied field N conditions. The optimised in-season N_{top} from the CM and RSCM methods was compared with that of FP_{opt} and the RS methods. Compared with FP_{opt} , although the simulated yield of the methods of CM and RSCM slightly decreased, their profit on average increased by 2.8% and 4.4%, respectively, and their AE_N on average increased from 13.35 kg kg⁻¹ to 18.24 and 24.54 kg kg⁻¹, respectively. Compared to the local FP (240 kg N ha⁻¹), after optimising N_{top} , the simulated yield of the methods of CM and RSCM increased by 0.6% and 0.1%, respectively, whereas the RS methods decreased the simulated yield by 6.0-7.2%. In general, the developed methods of CM and RSCM benefiting from the crop physiological principles and *in situ* remote sensing information provide promising opportunities to improve the productivity, the profit and resource use efficiency for the sustainable agricultural production. However, this potential needs field experimentation or farmer's demonstration trials to verify.

Acknowledgement

This study was partly supported by the High-tech Industry and Scientific and Technological Innovation Project of Lin-gang Special Area, Shanghai (grant number: SH-LG-GK-2020-02-19).

References

- Bai, Y., & Gao, J. (2021). Optimization of the nitrogen fertilizer schedule of maize under drip irrigation in Jilin, China, based on DSSAT and GA. *Agricultural Water Management*, 244, 106555. doi:10.1016/j.agwat.2020.106555
- Baret, F., Houles, V., & Guerif, M. (2007). Quantification of plant stress using remote sensing observations and crop models: the case of nitrogen management. *Journal of Experimental Botany*, 58, 869-880. doi:10.1093/jxb/erl231
- Basso, B., Bertocco, M., Sartori, L., & Martin, E. C. (2007). Analyzing the effects of climate variability on spatial pattern of yield in a maize–wheat–soybean rotation. *European Journal of Agronomy*, 26, 82-91. doi:10.1016/j.eja.2006.08.008
- Basso, B., Dumont, B., Cammarano, D., Pezzuolo, A., Marinello, F., & Sartori, L. (2016). Environmental and economic benefits of variable rate nitrogen fertilization in a nitrate vulnerable zone. *Science of the Total Environment*, 545-546, 227-235. doi:10.1016/j.scitotenv.2015.12.104
- Berger, K., Verrelst, J., Féret, J.-B., Wang, Z., Woche, M., Strathmann, M., Danner, M., Mauser, W., & Hank, T. (2020). Crop nitrogen monitoring: Recent progress and principal developments in the context of imaging spectroscopy missions. *Remote Sensing of Environment*, 242, 111758. doi:10.1016/j.rse.2020.111758
- Carrassi, A., Bocquet, M., Bertino, L., & Evensen, G. (2018). Data assimilation in the geosciences: An overview of methods, issues, and perspectives. *WIREs Climate Change*, 9, 535. doi:10.1002/wcc.535
- Cerrato, M. E., & Blackmer, A. M. (1990). Comparison of models for describing; corn yield response to nitrogen fertilizer. *Agronomy Journal*, 82, 138-143.
- Chen, J., Huang, Y., & Tang, Y. (2011). Quantifying economically and ecologically optimum nitrogen rates for rice production in south-eastern China. *Agriculture, Ecosystems & Environment*, 142, 195-204. doi:10.1016/j.agee.2011.05.005
- Chen, S., Jiang, T., Ma, H., He, C., Xu, F., Malone, R. W., Feng, H., Yu, Q., Siddique, K. H. M., Dong, Q. g., & He, J. (2020). Dynamic within-season irrigation scheduling for maize production in Northwest China: A method based on weather data fusion and yield prediction by DSSAT. *Agricultural and Forest Meteorology*, 285-286. doi:10.1016/j.agrformet.2020.107928
- Chen, X., Zhang, F., Römhild, V., Horlacher, D., Schulz, R., Böning-Zilkens, M., Wang, P., & Claupein, W. (2006). Synchronizing N supply from soil and fertilizer and N demand of winter wheat by an improved N_{min} method. *Nutrient Cycling in Agroecosystems*, 74, 91-98. doi:10.1007/s10705-005-1701-9
- Conte, S. D., & De Boor, C. (1965). *Elementary Numerical Analysis: An Algorithmic Approach* (2nd ed.). New York, The United States: McGraw-Hill.
- Cui, Z., Zhang, F., Chen, X., Dou, Z., & Li, J. (2010). In-season nitrogen management strategy for winter wheat: Maximizing yields, minimizing environmental impact in an over-fertilization context. *Field Crops Research*, 116, 140-146. doi:10.1016/j.fcr.2009.12.004
- Cui, Z., Zhang, H., Chen, X., Zhang, C., Ma, W., Huang, C., Zhang, W., Mi, G., Miao, Y., Li, X., Gao, Q., Yang, J., Wang, Z., Ye, Y., Guo, S., Lu, J., Huang, J., Lv, S., Sun, Y., Liu, Y., Peng, X., Ren,

- J., Li, S., Deng, X., Shi, X., Zhang, Q., Yang, Z., Tang, L., Wei, C., Jia, L., Zhang, J., He, M., Tong, Y., Tang, Q., Zhong, X., Liu, Z., Cao, N., Kou, C., Ying, H., Yin, Y., Jiao, X., Zhang, Q., Fan, M., Jiang, R., Zhang, F., & Dou, Z. (2018). Pursuing sustainable productivity with millions of smallholder farmers. *Nature*, 555, 363-366. doi:10.1038/nature25785
- Dinnes, D. L., Karlen, D. L., Jaynes, D. B., Kaspar, T. C., Hatfield, J. L., Colvin, T. S., & Cambardella, C. A. (2002). Nitrogen management strategies to reduce nitrate leaching in tile-drained Midwestern soils. *Agronomy Journal*, 94, 153-171.
- Dumont, B., Basso, B., Bodson, B., Destain, J. P., & Destain, M. F. (2016). Assessing and modeling economic and environmental impact of wheat nitrogen management in Belgium. *Environmental Modelling & Software*, 79, 184-196. doi:10.1016/j.envsoft.2016.02.015
- Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, 99, 10143-10162.
- Fageria, N. K., & Baligar, V. C. (2005). Enhancing nitrogen use efficiency in crop plants. *Advances in Agronomy*, 88, 97-185. doi:10.1016/s0065-2113(05)88004-6
- Flowers, M., Weisz, R., Heiniger, R., Osmond, D., & Crozier, C. (2004). In-season optimization and site-specific nitrogen management for soft red winter wheat. *Agronomy Journal*, 96, 124-134.
- Guerrero, A., De Neve, S., & Mouazen, A. M. (2021). Data fusion approach for map-based variable-rate nitrogen fertilization in barley and wheat. *Soil and Tillage Research*, 205, 104789. doi:10.1016/j.still.2020.104789
- Hansen, P. M., & Schjoerring, J. K. (2003). Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. *Remote Sensing of Environment*, 86, 542-553. doi:10.1016/s0034-4257(03)00131-7
- Holland, K. H., & Schepers, J. S. (2010). Derivation of a variable rate nitrogen application model for in-season fertilization of corn. *Agronomy Journal*, 102, 1415-1424. doi:10.2134/agronj2010.0015
- Holland, K. H., & Schepers, J. S. (2013). Use of a virtual-reference concept to interpret active crop canopy sensor data. *Precision Agriculture*, 14, 71-85. doi:10.1007/s11119-012-9301-6
- Huang, S., Miao, Y., Zhao, G., Yuan, F., Ma, X., Tan, C., Yu, W., Gnyp, M., Lenz-Wiedemann, V., Rascher, U., & Bareth, G. (2015). Satellite remote sensing-based in-season diagnosis of rice nitrogen status in Northeast China. *Remote Sensing*, 7, 10646-10667. doi:10.3390/rs70810646
- Jiang, J., Wu, Y., Liu, Q., Liu, Y., Cao, Q., Tian, Y., Zhu, Y., Cao, W., & Liu, X. (2023). Developing an efficiency and energy-saving nitrogen management strategy for winter wheat based on the UAV multispectral imagery and machine learning algorithm. *Precision Agriculture*, 24, 2019-2043. doi:10.1007/s11119-023-10028-6
- Jiang, R., He, W., Zhou, W., Hou, Y., Yang, J. Y., & He, P. (2019). Exploring management strategies to improve maize yield and nitrogen use efficiency in northeast China using the DNDC and DSSAT models. *Computers and Electronics in Agriculture*, 166, 104988. doi:10.1016/j.compag.2019.104988

- Ling, Q., Zhang, H., Dai, Q., Ding, Y., ling, L., Su, Z., Xu, M., Que, J., & Wang, S. (2005). Study on precise and quantitative N application in rice (in Chinese with English abstract). *Scientia Agricultura Sinica*, 38, 2457-2467. doi:10.3321/j.issn:0578-1752.2005.12.014
- Lory, J. A., & Scharf, P. C. (2003). Yield goal versus delta yield for predicting fertilizer nitrogen need in corn. *Agronomy Journal*, 95, 994-999.
- Mamo, M., Malzer, G. L., Mulla, D. J., Huggins, D. R., & Strock, J. (2003). Spatial and temporal variation in economically optimum nitrogen rate for corn. *Agronomy Journal*, 95, 958-964.
- Miao, Y., Stewart, B. A., & Zhang, F. (2011). Long-term experiments for sustainable nutrient management in China. A review. *Agronomy for Sustainable Development*, 31, 397-414. doi:10.1051/agro/2010034
- Moebius-Clune, B., Van Es, H., & Melkonian, J. (2013). Adapt-N uses models and weather data to improve nitrogen management for corn. *Better Crops*, 97, 7-9.
- Morari, F., Zanella, V., Gobbo, S., Bindi, M., Sartori, L., Pasqui, M., Mosca, G., & Ferrise, R. (2020). Coupling proximal sensing, seasonal forecasts and crop modelling to optimize nitrogen variable rate application in durum wheat. *Precision Agriculture*, 22, 75-98. doi:10.1007/s11119-020-09730-6
- Paz, J. O., Batchelor, W. D., Babcock, B. A., Colvin, T. S., Logsdon, S. D., Kaspar, T. C., & Karlen, D. L. (1999). Model-based technique to determine variable rate nitrogen for corn. *Agricultural Systems*, 61, 69-75.
- Pedersen, M. F., Gyldegren, J. G., Pedersen, S. M., Diamantopoulos, E., Gislum, R., & Styczen, M. E. (2021). A simulation of variable rate nitrogen application in winter wheat with soil and sensor information - An economic feasibility study. *Agricultural Systems*, 192, 103147. doi:10.1016/j.agsy.2021.103147
- Pedersen, S. M., Pedersen, M. F., Ørum, J. E., Fountas, S., Balafoutis, A. T., van Evert, F. K., van Egmond, F., Knierim, A., Kernecker, M., & Mouazen, A. M. (2020). Economic, environmental and social impacts. In A. Castrignano, G. Buttafuoco, R. Khosla, A. Mouazen, D. Moshou, & O. Naud (Eds.), *Agricultural Internet of Things and Decision Support for Precision Smart Farming* (pp. 279-330). Elsevier.
- Peng, S., Buresh, R. J., Huang, J., Yang, J., Zou, Y., Zhong, X., Wang, G., & Zhang, F. (2006). Strategies for overcoming low agronomic nitrogen use efficiency in irrigated rice systems in China. *Field Crops Research*, 96, 37-47. doi:10.1016/j.fcr.2005.05.004
- Peng, S., Buresh, R. J., Huang, J., Zhong, X., Zou, Y., Yang, J., Wang, G., Liu, Y., Hu, R., Tang, Q., Cui, K., Zhang, F., & Dobermann, A. (2010). Improving nitrogen fertilization in rice by sitespecific N management. A review. *Agronomy for Sustainable Development*, 30, 649-656. doi:10.1051/agro/2010002
- Pierce, F. J., & Nowak, P. (1999). Aspects of precision agriculture. *Advances in agronomy*, 67, 1-85.
- Ransom, C. J., Kitchen, N. R., Camberato, J. J., Carter, P. R., Ferguson, R. B., Fernández, F. G., Franzen, D. W., Laboski, C. A. M., Nafziger, E. D., Sawyer, J. E., Scharf, P. C., & Shanahan, J. F. (2020). Corn nitrogen rate recommendation tools' performance across eight US midwest corn belt states. *Agronomy Journal*, 112, 470-492. doi:10.1002/agi2.20035

- Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian processes for machine learning* (Vol. 2). Cambridge, Massachusetts, The United States: MIT press.
- Raun, W. R., Solie, J. B., Johnson, G. V., Stone, M. L., Mullen, R. W., Freeman, K. W., Thomason, W. E., & Lukina, E. V. (2002). Improving nitrogen use efficiency in cereal grain production with optical sensing and variable rate application. *Agronomy Journal*, *94*, 815-820.
- Raun, W. R., Solie, J. B., Stone, M. L., Martin, K. L., Freeman, K. W., Mullen, R. W., Zhang, H., Schepers, J. S., & Johnson, G. V. (2005). Optical sensor-based algorithm for crop nitrogen fertilization. *Communications in Soil Science and Plant Analysis*, *36*, 2759-2781. doi:10.1080/00103620500303988
- Ren, K., Xu, M., Li, R., Zheng, L., Liu, S., Reis, S., Wang, H., Lu, C., Zhang, W., Gao, H., Duan, Y., & Gu, B. (2022). Optimizing nitrogen fertilizer use for more grain and less pollution. *Journal of Cleaner Production*, *360*, 132180. doi:10.1016/j.jclepro.2022.132180
- Schoups, G., & Vrugt, J. A. (2010). A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors. *Water Resources Research*, *46*, W10531. doi:10.1029/2009wr008933
- Solie, J. B., Raun, W. R., Whitney, R. W., Stone, M. L., & Ringer, J. D. (1996). Optical sensor based field element size and sensing strategy for nitrogen application. *Transactions of the ASAE*, *39*, 1983-1992.
- Stanford, G. (1973). Rationale for optimum nitrogen fertilization in corn production. *Journal of Environmental Quality*, *2*, 159-166.
- Vrugt, J. A., Ter Braak, C. J. F., Diks, C. G. H., Robinson, B. A., Hyman, J. M., & Higdon, D. (2009). Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. *International Journal of Nonlinear Sciences and Numerical Simulation*, *10*, 273-290.
- Wang, D., Struik, P. C., Liang, L., & Yin, X. (2023a). Enhancing field-level forecasting of crop growth status by incorporating the analytically estimated system uncertainties into a data assimilation procedure. *Authorea*. doi:10.22541/au.170111043.37713869/v1
- Wang, D., Struik, P. C., Liang, L., & Yin, X. (2023b). Estimating leaf and canopy nitrogen contents in major field crops across the growing season from hyperspectral images using nonparametric regression. *Authorea*. doi:10.22541/au.170111047.73824045/v1
- Wang, X., Miao, Y., Batchelor, W. D., Dong, R., & Kusnierek, K. (2021a). Evaluating model-based strategies for in-season nitrogen management of maize using weather data fusion. *Agricultural and Forest Meteorology*, *308-309*, 108564. doi:10.1016/j.agrformet.2021.108564
- Wang, X., Miao, Y., Dong, R., Zha, H., Xia, T., Chen, Z., Kusnierek, K., Mi, G., Sun, H., & Li, M. (2021b). Machine learning-based in-season nitrogen status diagnosis and side-dress nitrogen recommendation for corn. *European Journal of Agronomy*, *123*, 126193. doi:10.1016/j.eja.2020.126193
- Xue, L., & Yang, L. (2008). Recommendations for nitrogen fertiliser topdressing rates in rice using canopy reflectance spectra. *Biosystems Engineering*, *100*, 524-534. doi:10.1016/j.biosystemseng.2008.05.005

- Yao, Y., Miao, Y., Huang, S., Gao, L., Ma, X., Zhao, G., Jiang, R., Chen, X., Zhang, F., Yu, K., Gnyp, M. L., Bareth, G., Liu, C., Zhao, L., Yang, W., & Zhu, H. (2012). Active canopy sensor-based precision N management strategy for rice. *Agronomy for Sustainable Development*, 32, 925-933. doi:10.1007/s13593-012-0094-9
- Yin, X., Goudriaan, J., Lantinga, E. A., Vos, J., & Spiertz, H. J. (2003). A flexible sigmoid function of determinate growth. *Annals of Botany*, 91, 361-371. Doi:10.1093/aob/mcg029
- Yin, X., & van Laar, H. H. (2005). *Crop systems dynamics: an ecophysiological simulation model for genotype-by-environment interactions*. Wageningen, The Netherlands: Wageningen Academic Publishers.
- Yu, C., Huang, X., Chen, H., Godfray, H. C. J., Wright, J. S., Hall, J. W., Gong, P., Ni, S., Qiao, S., Huang, G., Xiao, Y., Zhang, J., Feng, Z., Ju, X., Ciais, P., Stenseth, N. C., Hessen, D. O., Sun, Z., Yu, L., Cai, W., Fu, H., Huang, X., Zhang, C., Liu, H., & Taylor, J. (2019). Managing nitrogen to restore water quality in China. *Nature*, 567, 516-520. Doi:10.1038/s41586-019-1001-1
- Zhang, F., Chen, X., & Vitousek, P. (2013). Chinese agriculture: An experiment for the world. *Nature*, 497, 33-35.
- Zhang, F., Wang, J., Zhang, W., Cui, Z., Ma, W., Chen, X., & Jiang, R. (2008). Nutrient use efficiencies of major cereal crops in China and measures for improvement. *Acta Pedologica Sinica*, 45, 915-924.
- Zhang, J., Fu, Z., Zhang, K., Li, J., Cao, Q., Tian, Y., Zhu, Y., Cao, W., & Liu, X. (2023). Optimizing rice in-season nitrogen topdressing by coupling experimental and modeling data with machine learning algorithms. *Computers and Electronics in Agriculture*, 209, 107858. doi:10.1016/j.compag.2023.107858
- Zhang, J., Wang, W., Krienke, B., Cao, Q., Zhu, Y., Cao, W., & Liu, X. (2021). In-season variable rate nitrogen recommendation for wheat precision production supported by fixed-wing UAV imagery. *Precision Agriculture*, 23, 830-853. doi:10.1007/s11119-021-09863-2

Supplementary Appendix A in Chapter 5

Brief description of the crop model GECROS

The crop model Genotype-by-Environment interaction on CROp growth Simulator (GECROS) is a generic model and able to simulate a broad range of crops, in which the water, carbon and N dynamics of plant and soil are simulated on a daily basis (Yin and van Laar (2005); Yin and Struik (2017)). Its adaptability has been validated in many field crops, such as potato (Khan et al., 2014), rice (Gu et al., 2014), wheat and maize (Lenz-Wiedemann et al., 2010). One of the characteristics of the crop model GECROS is the detailed algorithms describing photosynthesis for C_3 and C_4 crops. In detail, the instantaneous leaf photosynthesis is calculated by the biochemical model of Farquhar et al. (1980) for C_3 crops and the equivalent model for simulating C_4 photosynthesis (von Caemmerer and Furbank, 1999; Yin and Struik, 2009, 2012). The photosynthetic capacity parameters including the maximum rate of Rubisco activity-limited carboxylation (V_{cmax}) and the maximum rate of linear electron transport under saturated light (J_{max}) is proved to be linearly affected by leaf N content (Evans, 1983; Wang et al., 2022). Relying on the sun/shade model of De Pury and Farquhar (1997), photosynthesis at the leaf level is upscaled to the canopy level. Temporal upscaling of the instantaneous rates to daily integrals is performed using the five-point Gaussian integration (Goudriaan, 1986). Crop N uptake is limited by both crop N demand and soil N supply. Partitioning of the newly formed carbon (C) and N assimilates between root and shoot is based on the equations of Yin and Schapendonk (2004). The development of LAI is modelled as the minimum of C- and N-determined LAI. The C-determined LAI is calculated as same as older Wageningen crop models from biomass accumulated in leaves, while the N-determined LAI is simulated based on leaf N content (Yin et al., 2000).

Supplementary Appendix B in Chapter 5

Description of bisection algorithm and its application to optimise N topdressing

The bisection algorithm is a root-searching method for a continuous function (F) defined for an interval $[a_1, b_1]$ with the opposite sign of function values for $F(a_1)$ and $F(b_1)$. The midpoint (b_2) of the interval and its function value ($F(b_2)$) are calculated and replaced with either $(a_1, F(a_1))$ or $(b_1, F(b_1))$ so that there is a zero crossing within the new interval. The interval is reduced by 50% at each iteration until it is sufficiently small (Fig. S.B5.1).

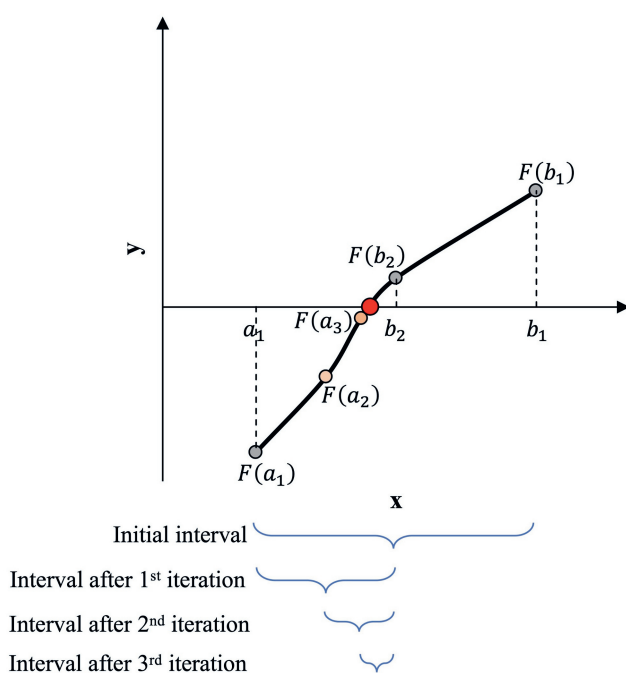


Fig. S.B5.1 Illustration of the bisection algorithm for searching the root of function F . The general principle of the bisection method, in which the red dot is the root of the F searching in the range from a_1 to b_1 ; and b_2 , a_2 and a_3 are the midpoints of interval $[a_1, b_1]$, $[a_1, b_2]$ and $[a_2, b_2]$, respectively.

Supplementary Appendix C in Chapter 5

Description of response models of the linear plateau model and the quadratic model

The linear plateau model is defined as,

$$y = \begin{cases} y_0 + AE_{N,e}x & \text{if } x < x_e \\ y_{max} & \text{if } x \geq x_e \end{cases} \quad (S5.1)$$

where y and x represent the yield and N application amount, respectively. y_0 denotes the value of yield when there is no N input. y_{max} denotes the maximum value of yield and x_e represents the amount of N when y_{max} is achieved. The linear coefficient is interpreted as the agronomic efficiency of N ($AE_{N,e}$), which is defined as $(y_{max} - y_0)/x_e$.

The quadratic model is normally defined as,

$$y = a + bx + cx^2 \quad (S5.2)$$

As the parameters, a , b and c , lack of specifically biological meaning, the quadratic model was reformed, following the indication from beta growth function proposed by Yin et al. (2003), as follows,

$$y = y_0 + \frac{(y_{max}-y_0)(2x_e-x)x}{x_e^2} \quad (S5.3)$$

Supplementary Tables in Chapter 5

Table S5.1 The pixel-level averaged topdressing N (N_{top} , kg N ha⁻¹) in different treatments of levels of N applied before the decision date ($N_{PreDeci}$)^a. The N_{top} for maximising yield and profit was optimised from strategies of the methods of remote sensing with Sufficiency Index (RS_{SI}) or Response Index (RS_{RI}) and the integrated method of remote sensing-crop model (RSCM).

Target	$N_{PreDeci}$	Total N_{top}			Third N_{top}			Fourth N_{top}		
		RS _{SI}	RS _{RI}	RSCM	RS _{SI}	RS _{RI}	RSCM	RS _{SI}	RS _{RI}	RSCM
Yield	0.0	268.4	214.4	122.0	188.6	129.7	105.1	79.8	84.8	16.9
	20.0	245.7	184.2	139.5	164.8	99.6	117.1	80.9	84.6	22.4
	40.0	198.2	146.7	134.8	115.4	62.2	113.8	82.8	84.4	21.0
	80.0	147.5	124.1	122.7	64.3	41.3	97.3	83.2	82.9	25.4
	120.0	107.3	100.8	106.2	23.6	18.5	85.9	83.7	82.3	20.3
Profit	160.0	93.9	93.1	81.9	10.5	13.4	59.1	83.4	79.7	22.7
	0.0	259.0	-	113.1	180.9	-	97.1	78.2	-	16.0
	20.0	236.4	-	130.7	157.4	-	108.3	79.0	-	22.4
	40.0	190.1	-	127.7	109.7	-	105.7	80.4	-	22.0
	80.0	140.9	-	116.3	60.2	-	91.6	80.7	-	24.7
	120.0	102.6	-	97.5	21.5	-	75.9	81.1	-	21.6
	160.0	89.7	-	74.4	9.0	-	51.5	80.8	-	22.9

^a The decision date was set at two weeks after tillering.

Table S5.2 The Normalised Root Mean Square Error (*NRMSE*) and coefficient of determination (R^2) of simulated yield in 2019 and 2020 from crop model simulations using actual weather data (CM_{act}) and 30 years' historical weather data (CM_{his})^a.

	2019		2020	
	<i>NRMSE</i>	R^2	<i>NRMSE</i>	R^2
CM_{act}	0.17	0.91	0.15	0.96
CM_{his}	0.10	0.80	0.12	0.92

^a The mean values of two years' field measurements of yield used here for calculating *NRMSE*. Historical weather data from 1989-2018 and 1990-2019 were used for generating CM_{his} of 2019 and 2020, respectively.

Supplementary Figures in Chapter 5

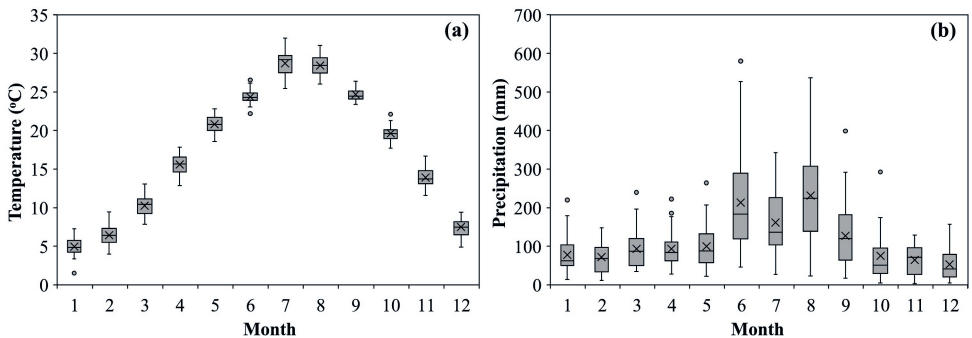


Fig. S5.1 Box plots of monthly mean air temperature (a) and total precipitation (b) from 1989 to 2020.

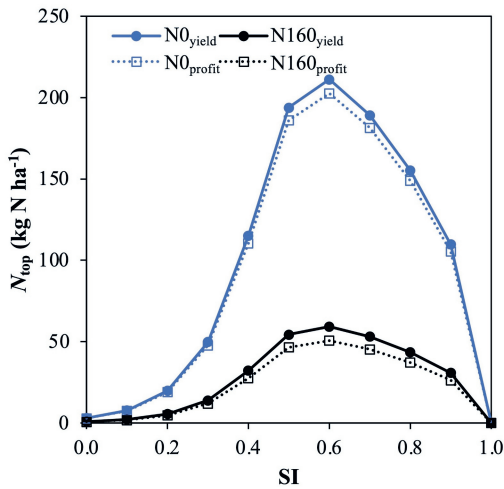


Fig. S5.2 The optimised topdressing N (N_{top}) in response to the Sufficiency Index (SI) under different levels of applied N for maximising yield (solid lines) or profit (dash lines). N0 and N160 represent the amount of N applied before the decision date of two weeks after tillering of 0 and 160 kg N ha⁻¹, respectively.

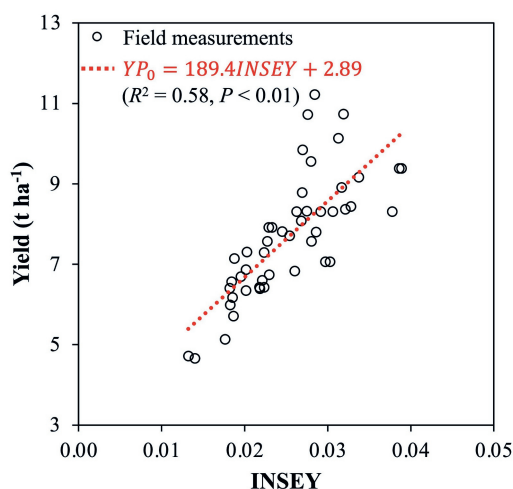


Fig. S5.3 Relationship between measured yield and the In Season Estimated Yield (INSEY) for predicting the yield potential with no added fertilisation (YP_0).

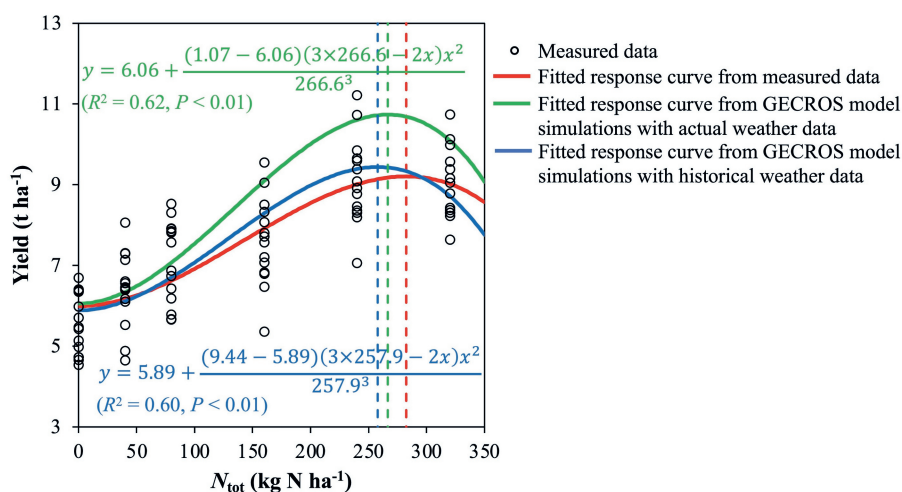


Fig. S5.4 Yield response curves to the varied N application amounts (N_{tot}) fitted from GECROS model simulations with actual (solid green curve) and historical (solid blue curve) weather data, compared with that from measured data (solid red curve). The green, blue and red vertical lines represent their optimised N_{tot} for maximising yield. The fitted equations with actual and historical weather data were texted with the colours of green and blue, respectively.

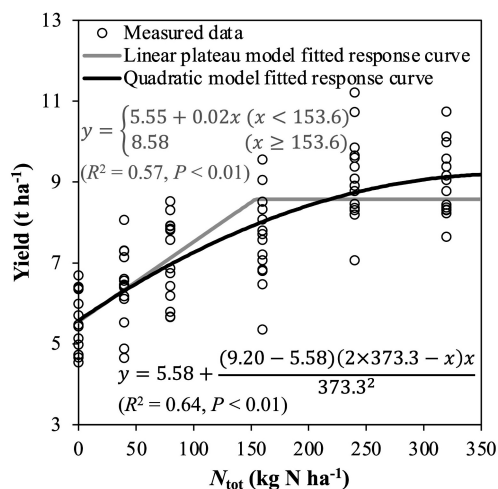


Fig. S5.5 Yield response curves to the varied N application amounts (N_{tot}) fitted by the linear plateau model and quadratic model. The models were fitted by two years' data in 2019 and 2020. Fitted equations of linear plateau model and quadratic model were texted with the colours of grey and black, respectively.

References

- De Pury, D. G. G., & Farquhar, G. D. (1997). Simple scaling of photosynthesis from leaves to canopies without the errors of big-leaf models. *Plant, Cell and Environment*, 20, 537-557. doi:10.1111/j.1365-3040.1997.00094.x
- Evans, J. R. (1983). Nitrogen and photosynthesis in the flag leaf of wheat (*Triticum aestivum* L.). *Plant Physiology*, 72, 297-302.
- Farquhar, G. D., von Caemmerer, S., & Berry, J. A. (1980). A biochemical model of photosynthetic CO₂ assimilation in leaves of C₃ species. *Planta*, 149, 78-90.
- Goudriaan, J. (1986). A simple and fast numerical method for the computation of daily totals of crop photosynthesis. *Agricultural and Forest Meteorology*, 38, 249-254.
- Gu, J., Yin, X., Zhang, C., Wang, H., & Struik, P. C. (2014). Linking ecophysiological modelling with quantitative genetics to support marker-assisted crop design for improved yields of rice (*Oryza sativa*) under drought stress. *Annals of Botany*, 114, 499-511. doi:10.1093/aob/mcu127
- Khan, M. S., Yin, X., van der Putten, P. E. L., & Struik, P. C. (2014). An ecophysiological model analysis of yield differences within a set of contrasting cultivars and an F1 segregating population of potato (*Solanum tuberosum* L.) grown under diverse environments. *Ecological Modelling*, 290, 146-154. doi:10.1016/j.ecolmodel.2013.11.015
- Lenz-Wiedemann, V. I. S., Klar, C. W., & Schneider, K. (2010). Development and test of a crop growth model for application within a Global Change decision support system. *Ecological Modelling*, 221, 314-329. doi:10.1016/j.ecolmodel.2009.10.014
- von Caemmerer, S., & Furbank, R. T. (1999). Modelling C₄ photosynthesis. In R. F. Sage & R. K. Monson (Eds.), *C₄ plant biology* (pp. 173-211). Toronto, Canada: Academic Press.
- Wang, D., Rianti, W., Gálvez, F., van der Putten, P. E. L., Struik, P. C., & Yin, X. (2022). Estimating photosynthetic parameter values of rice, wheat, maize and sorghum to enable smart crop cultivation. *Crop and Environment*, 1, 119-132. doi:10.1016/j.crope.2022.05.004
- Yin, X., Goudriaan, J., Lantinga, E. A., Vos, J., & Spiertz, H. J. (2003). A flexible sigmoid function of determinate growth. *Annals of Botany*, 91, 361-371. doi:10.1093/aob/mcg029
- Yin, X., & Schapendonk, A. H. C. M. (2004). Simulating the partitioning of biomass and nitrogen between roots and shoot in crop and grass plants. *NJAS: Wageningen Journal of Life Sciences*, 51, 407-426. doi:10.1016/s1573-5214(04)80005-8
- Yin, X., Schapendonk, A. H. C. M., Kropff, M. J., van Oijen, M., & Bindraban, P. S. (2000). A generic equation for nitrogen-limited leaf area index and its application in crop growth models for predicting leaf senescence. *Annals of Botany*, 85, 579-585.
- Yin, X., & Struik, P. C. (2009). C₃ and C₄ photosynthesis models: An overview from the perspective of crop modelling. *NJAS - Wageningen Journal of Life Sciences*, 57, 27-38. doi:10.1016/j.njas.2009.07.001
- Yin, X., & Struik, P. C. (2012). Mathematical review of the energy transduction stoichiometries of C₄ leaf photosynthesis under limiting light. *Plant, Cell and Environment*, 35, 1299-1312. doi:10.1111/j.1365-3040.2012.02490.x

- Yin, X., & Struik, P. C. (2017). Can increased leaf photosynthesis be converted into higher crop mass production? A simulation study for rice using the crop model GECROS. *Journal of Experimental Botany*, 68, 2345-2360. doi:10.1093/jxb/erx085
- Yin, X., & van Laar, H. H. (2005). *Crop systems dynamics: an ecophysiological simulation model for genotype-by-environment interactions*. Wageningen, the Netherlands: Wageningen Academic Publishers.

Chapter 6

General discussion

The demand for crop production is expected to be roughly doubled by 2050, compared to 2005 (Tilman et al., 2011). Driven by the aim to achieve maximum productivity, farmers in China tend to apply excess N in the hope of preventing yield loss. This has already caused negative impacts on the environment. Agricultural production must change towards simultaneously maintaining or even increasing crop productivity, improving N use efficiency and minimising the negative impact on environment (Basso and Antle, 2020; Shanahan et al., 2008). A so-called smart N management system supported by means of crop modelling and remote sensing has been recognised as a promising strategy for achieving the sustainable agricultural production (Chapter 1). Such a system holds the potential to optimise the N application *in situ* for meeting the temporally and spatially variable crop needs.

Crop N demand is related to crop carbon and N dynamics for meeting photosynthetic and non-photosynthetic physiological requirements. In general, less N is required to produce a given level of biomass of C₄ crops (e.g., maize and sorghum) than that of C₃ crops (e.g., rice and wheat) (Gastal and Lemaire, 2002). Within fields, N supply varies due to interactions with changing weather conditions, soil status and management practices in the field regarding nitrogen husbandry. Major obstacles in N management in the field are therefore poor synchrony between N fertilisation and crop demand, mismatch between uniform N applications and spatial crop variations, and failure to account for temporal dynamics in crop N needs (Shanahan et al., 2008). Greater reliance has, thus, been paid recently to the variable N application based on in-season estimation of crop N needs (Chapter 1). To achieve this, a reliable N management system should be developed.

This thesis integrates the utilisation of the crop model GECROS and remote sensing images in the field for major field crops in China, aiming to develop a smart N management system with the ability to account for soil and crop N dynamics effected on both temporal and spatial variations. To achieve this aim, I first introduced the potential of incorporating crop modelling with remote sensing in field N optimisation (Chapter 1). Then, I estimated primary photosynthetic parameters and quantified their relationship with specific leaf N content (SLN) of major field crops based on the measured data in a greenhouse, enabling crop growth simulation from GECROS (Chapter 2). Unmanned Aerial Vehicle (UAV)-based hyperspectral images were acquired in field experiments for assessing the prediction of canopy and leaf N status by different machine learning algorithms and pathways in Chapter 3. I also developed a method for integrating crop model simulations and remote sensing predictions with better forecasting performance (Chapter 4). Finally, I investigated different remote sensing- and crop

model-based N optimisation methods in Chapter 5 to complete the developed smart N management system.

In this Chapter, main findings of each research chapter will be summarised in Section 6.1, and significant merits of the designed smart N management system will be highlighted (Section 6.2). Thereafter, the future avenues of improving and utilising the developed system will be discussed in Section 6.3. Finally, some concluding remarks will be made in Section 6.4.

6.1 Research findings of this thesis

6.1.1 Foundation of leaf nitrogen in crop photosynthesis for smart farming

Smart farming adapts in-season field management of nitrogen based on crop growth forecasting given the current crop status. Sensing technologies contribute to monitoring crop growth, leaving the gap between monitored crop status to end-of-season yield to be fulfilled by the simulation of crop model. Due to the primacy of photosynthesis in crop production (Lawlor, 1995), an accurate and robust modelling of photosynthesis or related parameters is a prerequisite for many crop models. The biochemical model of Farquhar et al. (1980) describing photosynthetic processes has been widely used. Its key photosynthetic parameters are known to correlate with SLN, reflecting the fundamental role of N in various photosynthetic machineries (e.g., Evans and Clarke, 2019). Yet, for the major crops, studies that estimate and compare their photosynthetic parameters together are rare, especially in view of smart farming. Thus, four major crops, two C_3 species (rice and wheat) and two C_4 species (maize and sorghum), were incorporated together in Chapter 2 to quantify the primary photosynthetic parameters and their relationship with SLN.

Photosynthetic capacity linearly increased with SLN in these four major crops and the regressed equations for C_4 species differed significantly from those for C_3 species in terms of both slope and intercept. The SLN differed significantly between C_3 and C_4 species. At the same SLN, values of estimated Rubisco carboxylation capacity (V_{cmax}) of rice and wheat tended to be higher than those of maize and sorghum, possibly indicating that for achieving the same rate of CO_2 assimilation, a lower amount of Rubisco is required in C_4 species than in C_3 species. Within the two C_3 species, wheat showed a lower regression slope of the maximum electron transport rate (J_{max}) versus SLN than rice, but this was compensated by wheat's higher SLN. Between the two C_4 crops, the regressed linear equations of the maximum ATP production rate versus SLN did not differ significantly. The estimate of triose phosphate utilisation for C_3 crops

and that of phosphoenolpyruvate carboxylation capacity for C_4 crops increased with SLN as well. However, there was no significant difference in their corresponding regression equations between rice and wheat, and between maize and sorghum, respectively. The estimated photosynthetic parameters and their quantified relationships with SLN were utilised in the crop model GECROS in Chapters 4 and 5.

6.1.2 Physiological basis for data-driven prediction of leaf nitrogen from remote sensing

Given the crucial role of leaf N in affecting photosynthetic parameters (Chapter 2) and other aspects of crop growth, a diagnosis of canopy and leaf N status is conducted based on the UAV-acquired hyperspectral images in Chapter 3. Three types of features, canopy reflectance, vegetation indices, and texture information, were extracted from the hyperspectral image. Together with crop development stage, these features were applied to predict crop N status of rice, wheat and maize, using five machine learning algorithms (or referred to nonparametric regression algorithms). Furthermore, relying on the physiological basis that both area- and mass-based leaf N can be scaled up to the canopy level, I investigated the direct and indirect pathways for obtaining more accurate crop N information.

The model performance of Partial Least Squares Regression and Support Vector Regression was more robust than that of other machine learning algorithms for the major field crops. The combined three types of hyperspectral features contributed to the performance improvement, although there was no improvement when further incorporating the feature of development stage. Compared to the area-based leaf N trait (SLN in g N m^{-2} , which was used in Chapter 2), the prediction of the mass-based trait (leaf N concentration in g N (g DM)^{-1}) was better. Values of SLN were better predicted via an indirect method (predicted via the specific leaf area and leaf N concentration) than the direct method. When further scaled up to canopy, the indirectly predicted canopy N content using leaf area index (LAI) and directly predicted SLN agreed better with measured canopy N content than the directly predicted canopy N content in rice and maize. The identified indirect pathways contributed to the further improvement of the prediction of leaf and canopy N traits, although they might be more uncertain due to the introduced more regression processes.

6.1.3 Strength of uncertainty quantification in the integration of remote sensing and crop model

Due to uncertainties in acquired remotely sensed observations (Chapter 3) and crop model simulations, the data assimilation procedure attracts ever-increasing interests in smart farming. This procedure combines remote sensing observations and crop model simulations and can provide a more accurate estimate of crop growth than any of the two approaches on its own (Houser et al., 2012; Jin et al., 2018). However, the uncertainties have rarely been quantified, but have arbitrarily been assigned or simply overlooked for decades (Kang and Özdoğan, 2019; Nearing et al., 2012). The unspecified uncertainties in remote sensing and crop modelling hinder their implementation in data assimilation. To leverage the advantage of assimilating remote sensing observations and crop model simulations, I explored the analytical system uncertainty and investigated its performance in a commonly used data assimilation procedure, the Ensemble Kalman filter. Furthermore, the developed procedure was validated in our field experiment using rice as a case (Chapter 4).

Leaf traits at the canopy level, leaf weight, canopy leaf-N content and LAI, were well predicted from the machine learning algorithm of the Gaussian Process Regression model, based on the collected hyperspectral features (Chapter 3). Estimated photosynthetic parameters in Chapter 2 and parameters from field records were treated as the fixed parameters in crop model GECROS. Using collected field observations, the uncertain parameters in the crop model GECROS were calibrated by a Markov Chain Monte Carlo approach; the parameters in the error model that describes the uncertainties of the crop model simulations were estimated simultaneously. Applying the quantified uncertainties into the Ensemble Kalman filter, the directly updated traits in the crop model GECROS agreed better with the measurements, without filter divergence. Indirectly updated in-season and end-of-season aboveground biomass, grain weight, aboveground N content and grain N content also agreed better with measurements than those without assimilating observations. More importantly, due to the additionally revealed spatial and temporal variations of traits, updated states from assimilating crop model simulation with remote sensing prediction had even better performance than those with field measurements.

6.1.4 Potential of using the integrated crop model- and remote sensing-based methods for nitrogen management in the field

Precise monitoring and forecasting of in-season crop growth are expected to support the description of temporal and spatial crop N uptake and demand (Chapters 3, 4). Addressing varied crop N needs via in-season N optimisation has been identified as key to optimising the trade-off among yield, profit and N use efficiency for attaining sustainable smart farming (Shanahan et al., 2008). Given that the effects of changing weather on field N optimisation are always ignored (Chapter 1), the crop model-based N optimisation method (CM) was developed in Chapter 5, aiming to temporally determine the optimal N management from the soil-crop-atmosphere continuum. Further taking spatial variations of crop growth into consideration, I also investigated an integrated remote sensing-crop model method by data assimilation (RSCM) to optimise N management *in situ*. The reference methods, including the farmer practice optimised by the yield response curve, and the remote sensing (RS) methods based on the Sufficiency Index or Response Index, were incorporated into the comparison.

Based on the simulated crop growth with actual weather data, the optimised in-season topdressing N (N_{top}) of the CM method tended to perform better in improving yield, profit and N use efficiency than that of reference methods. Benefiting from assimilating *in situ* remote sensing information into the CM method, the RSCM method further improved the performance. Therefore, the RSCM method potentially suited best to guide smart fertiliser management for the sustainable agricultural production. While there are obstacles like acquiring remote sensing images or implementing *in situ* decisions, the CM method that relies on crop physiological principles was recommended.

6.2 Signature merits of the designed smart nitrogen management system

6.2.1 Connecting leaf-level photosynthesis to field-scale nitrogen optimisation

To advance crop yield for the growing food demand globally (Ray et al., 2013), enhancing photosynthesis is widely regarded as critical (Long et al., 2015; Walter and Kromdijk, 2022) and potential manipulation targets for both C_3 and C_4 crop species have been proposed (von Caemmerer and Evans, 2010; von Caemmerer and Furbank, 2016). Cross-scale models connecting leaf photosynthesis and crop production have also been developed to guide photosynthetic manipulation efforts targeting crop biomass and yield improvement (Gu et al.,

2014; Wu et al., 2019). For instance, to increase rice biomass, the greater importance of improving J_{\max} , relative to improving V_{\max} , was highlighted by using the crop model GECROS (Yin et al., 2022).

However, Sinclair et al. (2019) stated that instead of focusing on increasing photosynthesis, more attention should be paid on the availability and uptake of N for increasing yield. Due to the conflict between the primary N limitation in soil supply and the large N requirement for photosynthesis and crop growth, the N fertiliser application for productive crops is of importance (Evans and Clarke, 2019). For instance, Gu et al. (2014) pointed out that any shortage of the required N supply hampers the yield potential of improving photosynthesis. However, increasing N fertilisation blindly to avoid N limitation has reached a critical stage, due to the severe impacts caused by low N recovery (Schlesinger, 2009). To this end, the connection between leaf photosynthesis and field production needs be further extended to N management in the field for sustainable agricultural production. Thus, the methods of CM and RSCM that are able to connect leaf photosynthesis capacity to field N supply were developed in Chapter 5. Both of them showed the potential of optimising N_{top} for higher yield, profit and N use efficiency, by means of the simulation from the crop model GECROS with estimated photosynthetic parameters and their relationship with SLN in Chapter 2.

The highlighted importance of N uptake for improving photosynthesis is in line with the past breeding strategy that selects more N input-responsive cultivars (Sinclair et al., 2019). Greater root uptake ability, such as increased root surface area and conductivity, may lead to higher SLN, which enables higher photosynthetic rates (Adachi et al., 2011; Hikosaka, 2010). The simulated results from Yin et al. (2022) showed that increasing season-long rice crop N uptake by 20% increased radiation use efficiency and biomass by 10.7% and 14.6%, respectively. While improving traits of photosynthetic parameters and N uptake together, radiation use efficiency and biomass increased by 29% and 31.5%, respectively (Yin et al., 2022), close to the projected effect of engineering the complete C₄ mechanism into rice (Yin and Struik, 2017).

6.2.2 Improving nitrogen management in the field via enhanced real-time monitoring and forecasting of in situ crop growth

Efforts to provide fertilisation advice based on spatial and temporal variations have been made for decades. In view of spatial variations within farmland, the N optimisation methods were

developed initially based on soil nutrient maps and the expected yield (Sarkadi and Várallyay, 1989; Várallyay et al., 1992). Relying on the crop model Hybrid-Maize, an integrated soil-crop system management approach was developed by Chen et al. (2011) to manage the N supply to match the requirement of the model-designed crop system. Cui et al. (2018) made the recommendations to smallholder farmers in a national campaign following the integrated soil-crop system management approach with the necessary adjustments from local experts and participating farmers. A site-specific nutrient management decision support tool, *Nutrient Expert*®, has been developed using computer-based decision support technology and a questionnaire (He et al., 2022). Although there are tremendous outcomes from those methods, the critical in-season N optimisation based on a real-time crop growth monitor is still missing.

Remote sensing technology has been incorporated as one of important components of smart farming (Chapter 1). As physiological and biochemical characteristics change during crop growth under varied management practices and micro-environmental conditions, canopy spectral reflectance differs thereafter (Strachan et al., 2002). In virtue of this, real-time crop N status was predicted from collected hyperspectral images by the machine learning algorithms through direct or indirect pathways (Chapter 3). The mapped differences of crop growth within and between experimental plots across successive growth stages demonstrate the necessity of *in situ* N optimisation (Chapter 3). Sensor-based N optimisation methods have also been developed for spatially varied N application (Colaço and Bramley, 2018). However, compared with local farmer's practice, the yield decrease tended to be caused by the optimised N_{top} from the RS methods, implying the N deficiency of the ensuing crop growth caused by the temporal variation (Chapter 5). Oppositely, with the diverse weather inputs derived from the historical daily weather data, higher yield and profit were achieved by the optimised N_{top} from the CM method, in which the model-proposed N management in the field temporally interacted with the forecasted final yield or profit (Chapter 5).

Due to the missing temporal interactions in remote sensing predictions and the lacking spatial variations in crop model simulations, remote sensing and crop modelling have long been believed to complement each other for more accurate temporal and spatial forecasting of grain yield (Huang et al., 2019; Jin et al., 2018). However, there are still obstacles to fully exploit the remote sensing and crop modelling for better crop growth forecasting (Chapter 1). Via explicitly quantified system uncertainties of crop model simulations and remote sensing predictions, more accurate crop growth forecasting was achieved through a data assimilation framework, compared to that of the assumed uncertainties in a traditional way (Chapter 4). By means of the

enhanced monitoring and forecasting of crop growth, the RSCM method maintained the high yield while maximising the profit and N use efficiency (Chapter 5).

6.2.3 Achieving multi-objectives simultaneously in real-time decision-making on field nitrogen

The major focus of optimising N management in the field is to improve the agricultural productivity and profitability for farmers while maintaining environmental sustainability. As the overuse of N commonly occurs in smallholder farmland in China (e.g., Cui et al., 2018), the main benefit of the various N optimisation methods is reduced N rates. For instance, compared with the local farmer's practice, the whole-season total N application of the developed CM and RSCM methods and the referenced RS methods decreased by 2.0-19.8%, so as to reduce the N loss and environmental cost (Chapter 5). Regarding the RS methods, the yield and profit, however, decreased by 6.0-7.2% and 5.1-6.3%, respectively (Chapter 5). In line with this, Colaço and Bramley (2018) concluded, based on reviewed reports of sensor-based N optimisation, that increasing yield is more effective than N saving for achieving an increased profitability.

As spatial variations occur across the field and within the referenced N-rich area, the inadequate representative of the selected reference area and the missing link with *in situ* soil conditions hinder the performance of the RS methods (Colaço and Bramley, 2018). In accordance with local soil and climatic conditions, the yield and profit of the CM and RSCM methods increased by 0.1-0.6% and 0.8-1.5%, respectively, compared with the local farmer's practice (Chapter 5). The *in situ* N optimisation relies on the accurately simulated impacts of surrounding environments on crop growth (Paz et al., 1999). By virtue of the enhanced forecast of final yield and grain quality (e.g., grain N content) (Chapter 4), the *in situ* N demand was better estimated by the RSCM method than by other methods, enabling the better match of the N application to production targets. Consequently, benefiting from assimilating data from *in situ* RS information into the CM method, the RSCM method has the best performance in achieving multi-objectives for the sustainable agricultural production, thereby potentially best suiting for guiding smart fertiliser management.

6.3 Avenues for future research

6.3.1 Exploitation of physiological and machine learning methods for better crop modelling

Uncertainty in simulating crop growth with crop models stems from the inadequate model structure, poor parameterisation, or biased forcing input (Chapter 4). In recent decades, various crop models with differentiated model structure and parameterisation have been developed. By comparing the simulation results of participating models, the sources of uncertainty in existing crop models are expected to be identified for model improvement via the Agricultural Model Intercomparison and improvement Program (e.g., Martre et al., 2015; Wallach et al., 2018). For instance, ensembles of current models identify the uncertainty in modelling LAI. Yet, such finding has already concluded in 1990s (Chapter 1), implying the limited algorithm improvements of the recently developed crop models (Yin et al., 2021). Moreover, as the attention in crop modelling has largely shifted to phenomenological or empirical models during last decades, some classical modelling philosophies even have largely been ignored (Connor, 2019). It is found that no single model fits all sites or environments and the averaged simulations of ensemble models tend to perform worse than simulations of the best model (Wallach et al., 2018). Thus, Yin et al. (2021) argued that the aim should be to achieve real improvement in biological mechanisms captured in crop models, in which the complex crop phenotypes can be predicted on the basis of underlying physiological principles (e.g., Zhou et al., 2023).

With the advent of the big ecological data, a hybrid modelling approach, coupling physical process models with data-driven machine learning, is also anticipated to improve crop model algorithms (Reichstein et al., 2019). Regarding the improvement of model structure, the processes that are not explicitly described in crop models can be replaced by well-trained machine learning models if there are sufficient auxiliary observations. The interpretable extrapolation beyond observed conditions can, therefore, be offered relying on the physiological principles in the hybrid model, while the data-driven processes may be amenable to finding unexpected patterns (Reichstein et al., 2019). As for parameter estimation, the concept that parameter values must be constant has become a classic doctrine since the birth of crop modelling (Forrester, 1961). However, crop growth constantly acclimates and adapts to the changing environments (e.g., Berry and Bjorkman, 1980; Fang et al., 2023). Thus, Luo and

Schuur (2020) stated that parameter values may have to vary over space and time, in response to the changing properties and unresolved processes in crop modelling. Alternatively, model structure should be improved to account for crop acclimation processes (Yin and Struik, 2010). Nevertheless, given the complications of crop growth processes and their interaction in response to environmental variables, machine learning models are inevitably required by learning from the appropriate sets of statistical covariates to obtain more dynamic, interdependent and contextual parameter values (Luo and Schuur, 2020; Tao et al., 2023).

6.3.2 Transfer learning for the preferable application of remote sensing

To predict crop growth from remotely sensed reflectance, machine learning models have attained outstanding results at local or global scales (e.g., Berger et al., 2020; Chapter 3). Yet training machine learning models often requires a substantial amount of ground-labelled data from labour-intensive field campaigns, which are costly to collect. Due to the changed collection conditions or setups, a change of data distribution, known as the “domain shift”, always causes poor performance of machine learning models when directly applied from one domain to other domains (Michau and Fink, 2021). Although the temporal and spatial transferability of the Gaussian Process Regression model adopted in Chapter 4 has been demonstrated before (Verrelst et al., 2013), transfer learning, a machine learning technique, has emerged as a promising strategy to improve the transferability of machine learning models in general (Ma et al., 2024).

By leveraging the feature or knowledge learned from the data-abundant source domain, transfer learning helps address the issue of domain shift by alleviating the need for local labelled data in the target domain (Pan and Yang, 2010). For instance, by coupling Transfer Component Analysis with Support Vector Regression, the transferability of a developed unsupervised domain adaptation model greatly improved and only 5% of the target domain samples was needed to achieve an effective prediction of leaf N concentration across different plant species (Wan et al., 2022). Instead of acquiring abundant labelled field data of the source domain, Zhang et al. (2021) trained a deep learning model from a simulated dataset generated by a radiative transfer model, and then transferred and fine-tuned the model with a few *in situ* field samples for the prediction of leaf chlorophyll content of winter wheat. Moreover, there are also attempts to build remote sensing foundation models (or referred to large pre-trained models), which are trained on large-scale data in a task-agnostic way and then transferred to particular downstream task by transfer learning (Ma et al., 2024). For instance, in cooperation with IBM

Research, NASA has released its first open-source foundation models for applications like landcover classification and yield prediction (Jakubik et al., 2023).

6.3.3 Upscaling the smart N management from site to region

Although the performance of the designed smart N management system has been evaluated at field level (Chapter 5), scaling its application up to the regional level should be further investigated. However, while applying crop models over a scale larger than individual fields, poor performance tends to be caused by a lack of knowledge about spatial variations of model parameters, soil properties, and field managements (Hansen and Jones, 2000; Nearing et al., 2012). To estimate spatially heterogeneous parameters, a deep learning model was developed to upscale the site-level estimates to the region- or globe-level by connecting site-level parameters with their associated environmental variables (Tao and Luo, 2022). Without punctual measurements, the soil status and field-specific agricultural practices, like sowing dates, fertilisation dates and amounts, have been identified through the surveys with the farmers for the forecast of regional rice growth and yield, while assimilating remotely sensed LAI into crop model STICS (Courault et al., 2021). Moreover, the remote sensing information might be functional as well. For instance, Courault et al. (2020) extracted the sowing dates from the method based on the spectral indices. With a coupled forward-inverse framework, the irrigation schedule was estimated by assimilating the simulations using the crop model SWAP and remote sensing observations (Wang and Cai, 2007).

The utilisation of multispectral remote sensing has become integral to agricultural practice for large area's cultivation in developed countries. However, while upscaling the smart N management in the currently dominant smallholder farmland in China (Cui et al., 2018), the application of satellite remote sensing is always impaired by the inadequate spatial resolution, along with insufficient revisiting frequencies (Marshall and Thenkabail, 2015). The fusion of multispectral and hyperspectral satellite images is expected, as it can reconstruct super-resolution remote sensing images with high spatial, temporal and spectral resolution (Li et al., 2022). Not only the reconstructed optical satellite images, but also satellite data of synthetic aperture radar and solar-induced fluorescence have emerged in recent years for better crop growth monitoring, such as LAI (Pipia et al., 2019) and photosynthetic activity (Mohammed et al., 2019). Consequently, the performance of assimilating multi-source heterogeneous inputs into a crop model for the upscaling forecast needs to be further evaluated in future work.

6.4 Concluding remarks

In an effort to design a smart N management system for major field crops in China to achieve the sustainable agricultural production, in this thesis, comprehensive studies have been conducted on crop modelling and remote sensing, which have been applied to forecasting and monitoring crop growth, respectively. I investigated the impacts of varied N application rates on leaf photosynthesis and crop growth for providing crucial links from leaf photosynthesis prediction to field N optimisation. The results revealed the significant roles of crop modelling and remote sensing in the designed smart N management system for real-time *in situ* N optimisation, which can be concluded as:

- Estimated photosynthetic parameters and their strong relationship with SLN contribute to the crop growth simulation via the cross-scale crop model GECROS.
- Better in-season crop growth monitoring benefits from the UAV-based remote sensing predictions by means of the proper selection of hyperspectral features, regression algorithms and prediction pathways.
- The crop growth forecast can be enhanced by assimilating crop model simulation with remote sensing predictions using a developed Bayesian methodology that systematically quantifies uncertainties.
- The optimised in-season N_{top} from the CM method achieves multi-objectives simultaneously by meeting the forecasted temporal crop needs, and that from the RSCM method further benefits from the monitored spatial variations of crop growth.

As discussed above, crop modelling and remote sensing are crucial modules in smart N management system. To further improve the designed system, the follow-up research should be exploring unresolved issues with regard to physiological or machine learning methods in crop modelling, and extending the applicability of remote sensing by applying the transfer learning and data fusion technique. The integration of improved crop modelling and remote sensing can result in a mighty tool that upscales the smart N management system from site to region. Ultimately, the regional or global regulation of field N can be better optimised, contributing to making the agricultural production and human life on the planet sustainable.

References

- Adachi, S., Tsuru, Y., Nito, N., Murata, K., Yamamoto, T., Ebitani, T., Ookawa, T., & Hirasawa, T. (2011). Identification and characterization of genomic regions on chromosomes 4 and 8 that control the rate of photosynthesis in rice leaves. *Journal of Experimental Botany*, 62, 1927-1938. doi:10.1093/jxb/erq387
- Basso, B., & Antle, J. (2020). Digital agriculture to design sustainable agricultural systems. *Nature Sustainability*, 3, 254-256. doi:10.1038/s41893-020-0510-0
- Berger, K., Verrelst, J., Féret, J.-B., Wang, Z., Woche, M., Strathmann, M., Danner, M., Mauser, W., & Hank, T. (2020). Crop nitrogen monitoring: Recent progress and principal developments in the context of imaging spectroscopy missions. *Remote Sensing of Environment*, 242, 111758. doi:10.1016/j.rse.2020.111758
- Berry, J., & Bjorkman, O. (1980). Photosynthetic response and adaptation to temperature in higher plants. *Annual Review of Plant Physiology*, 31, 491-543.
- Chen, X. P., Cui, Z. L., Vitousek, P. M., Cassman, K. G., Matson, P. A., Bai, J. S., Meng, Q. F., Hou, P., Yue, S. C., Romheld, V., & Zhang, F. S. (2011). Integrated soil-crop system management for food security. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 6399-6404. doi:10.1073/pnas.1101419108
- Colaço, A. F., & Bramley, R. G. V. (2018). Do crop sensors promote improved nitrogen management in grain crops? *Field Crops Research*, 218, 126-140. doi:10.1016/j.fcr.2018.01.007
- Connor, D. J. (2019). An experimentally-calibrated model of photosynthesis, assimilate partitioning and tuber yield in cassava in response to water supply to assist crop management and improvement. *Field Crops Research*, 242, 107606. doi:10.1016/j.fcr.2019.107606
- Courault, D., Hossard, L., Demarez, V., Dechatre, H., Irfan, K., Baghdadi, N., Flamain, F., & Ruget, F. (2021). STICS crop model and Sentinel-2 images for monitoring rice growth and yield in the Camargue region. *Agronomy for Sustainable Development*, 41, 1-17. doi:10.1007/s13593-021-00697-w
- Courault, D., Hossard, L., Flamain, F., Baghdadi, N., & Irfan, K. (2020). Assessment of agricultural practices from Sentinel 1 and 2 images applied on rice fields to develop a farm typology in the Camargue region. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 5027-5035. doi:10.1109/jstars.2020.3018881
- Cui, Z., Zhang, H., Chen, X., Zhang, C., Ma, W., Huang, C., Zhang, W., Mi, G., Miao, Y., Li, X., Gao, Q., Yang, J., Wang, Z., Ye, Y., Guo, S., Lu, J., Huang, J., Lv, S., Sun, Y., Liu, Y., Peng, X., Ren, J., Li, S., Deng, X., Shi, X., Zhang, Q., Yang, Z., Tang, L., Wei, C., Jia, L., Zhang, J., He, M., Tong, Y., Tang, Q., Zhong, X., Liu, Z., Cao, N., Kou, C., Ying, H., Yin, Y., Jiao, X., Zhang, Q., Fan, M., Jiang, R., Zhang, F., & Dou, Z. (2018). Pursuing sustainable productivity with millions of smallholder farmers. *Nature*, 555, 363-366. doi:10.1038/nature25785
- Evans, J. R., & Clarke, V. C. (2019). The nitrogen cost of photosynthesis. *Journal of Experimental Botany*, 70, 7-15. doi:10.1093/jxb/ery366

- Fang, L., Martre, P., Jin, K., Du, X., van der Putten, P. E. L., Yin, X., & Struik, P. C. (2023). Neglecting acclimation of photosynthesis under drought can cause significant errors in predicting leaf photosynthesis in wheat. *Global Change Biology*, 29, 505-521. doi:10.1111/gcb.16488
- Farquhar, G. D., von Caemmerer, S., & Berry, J. A. (1980). A biochemical model of photosynthetic CO₂ assimilation in leaves of C₃ species. *Planta*, 149, 78-90.
- Forrester, J. W. (1961). *Industrial dynamics*. Cambridge, Massachusetts, The United States: MIT Press.
- Gastal, F., & Lemaire, G. (2002). N uptake and distribution in crops: an agronomical and ecophysiological perspective. *Journal of Experimental Botany*, 53, 789-799.
- Gu, J., Yin, X., Stomph, T. J., & Struik, P. C. (2014). Can exploiting natural genetic variation in leaf photosynthesis contribute to increasing rice productivity? A simulation analysis. *Plant Cell and Environment*, 37, 22-34. doi:10.1111/pce.12173
- Hansen, J. W., & Jones, J. W. (2000). Scaling-up crop models for climate variability applications. *Agricultural Systems*, 65, 43-72.
- He, P., Xu, X., Zhou, W., Smith, W., He, W., Grant, B., Ding, W., Qiu, S., & Zhao, S. (2022). Ensuring future agricultural sustainability in China utilizing an observationally validated nutrient recommendation approach. *European Journal of Agronomy*, 132, 126409. doi:10.1016/j.eja.2021.126409
- Hikosaka, K. (2010). Mechanisms underlying interspecific variation in photosynthetic capacity across wild plant species. *Plant Biotechnology*, 27, 223-229.
- Houser, P. R., De Lannoy, G. J. M., & Walker, J. P. (2012). Hydrologic data assimilation. In J. Tiefenbacher (Eds.), *Approaches to Managing Disaster—Assessing Hazards, Emergencies and Disaster Impacts* (pp. 41-64). Rijeka, Croatia: InTech.
- Huang, J., Gómez-Dans, J. L., Huang, H., Ma, H., Wu, Q., Lewis, P. E., Liang, S., Chen, Z., Xue, J.-H., Wu, Y., Zhao, F., Wang, J., & Xie, X. (2019). Assimilation of remote sensing into crop growth models: Current status and perspectives. *Agricultural and Forest Meteorology*, 276-277, 107609. doi:10.1016/j.agrformet.2019.06.008
- Jakubik, J., Chu, L., Fraccaro, P., Gomes, C., Nyirjesy, G., Bangalore, R., Lambhate, D., Das, K., Oliveira Borges, D., Kimura, D., Simumba, N., Szwarcman, D., & Muszynski, E. (2023). Prithvi-100M. doi:10.57967/hf/0952
- Jin, X., Kumar, L., Li, Z., Feng, H., Xu, X., Yang, G., & Wang, J. (2018). A review of data assimilation of remote sensing and crop models. *European Journal of Agronomy*, 92, 141-152. doi:10.1016/j.eja.2017.11.002
- Kang, Y., & Özdoğan, M. (2019). Field-level crop yield mapping with Landsat using a hierarchical data assimilation approach. *Remote Sensing of Environment*, 228, 144-163. doi:10.1016/j.rse.2019.04.005
- Lawlor, D. W. (1995). Photosynthesis, productivity and environment. *Journal of Experimental Botany*, 46, 1449-1461.
- Li, T., Liu, T., Wang, Y., Li, X., & Gu, Y. (2022). Spectral reconstruction network from multispectral images to hyperspectral images: A multitemporal case. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1-16. doi:10.1109/tgrs.2022.3195748

- Long, S. P., Marshall-Colon, A., & Zhu, X. (2015). Meeting the global food demand of the future by engineering crop photosynthesis and yield potential. *Cell*, 161, 56-66. doi:10.1016/j.cell.2015.03.019
- Luo, Y., & Schuur, E. A. G. (2020). Model parameterization to represent processes at unresolved scales and changing properties of evolving systems. *Global Change Biology*, 26, 1109-1117. doi:10.1111/gcb.14939
- Ma, Y., Chen, S., Ermon, S., & Lobell, D. B. (2024). Transfer learning in environmental remote sensing. *Remote Sensing of Environment*, 301, 113924. doi:10.1016/j.rse.2023.113924
- Marshall, M., & Thenkabail, P. (2015). Advantage of hyperspectral EO-1 Hyperion over multispectral IKONOS, GeoEye-1, WorldView-2, Landsat ETM+, and MODIS vegetation indices in crop biomass estimation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 108, 205-218. doi:10.1016/j.isprsjprs.2015.08.001
- Martre, P., Wallach, D., Asseng, S., Ewert, F., Jones, J. W., Rotter, R. P., Boote, K. J., Ruane, A. C., Thorburn, P. J., Cammarano, D., Hatfield, J. L., Rosenzweig, C., Aggarwal, P. K., Angulo, C., Basso, B., Bertuzzi, P., Biernath, C., Brisson, N., Challinor, A. J., Doltra, J., Gayler, S., Goldberg, R., Grant, R. F., Heng, L., Hooker, J., Hunt, L. A., Ingwersen, J., Izaurrealde, R. C., Kersebaum, K. C., Muller, C., Kumar, S. N., Nendel, C., O'Leary, G., Olesen, J. E., Osborne, T. M., Palosuo, T., Priesack, E., Ripoche, D., Semenov, M. A., Shcherbak, I., Steduto, P., Stockle, C. O., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Travasso, M., Waha, K., White, J. W., & Wolf, J. (2015). Multimodel ensembles of wheat growth: Many models are better than one. *Global Change Biology*, 21, 911-925. doi:10.1111/gcb.12768
- Michau, G., & Fink, O. (2021). Unsupervised transfer learning for anomaly detection: Application to complementary operating condition transfer. *Knowledge-Based Systems*, 216, 106816. doi:10.1016/j.knsys.2021.106816
- Mohammed, G. H., Colombo, R., Middleton, E. M., Rascher, U., van der Tol, C., Nedbal, L., Goulas, Y., Perez-Priego, O., Damm, A., Meroni, M., Joiner, J., Cogliati, S., Verhoef, W., Malenovsky, Z., Gastellu-Etchegorry, J. P., Miller, J. R., Guanter, L., Moreno, J., Moya, I., Berry, J. A., Frankenberg, C., & Zarco-Tejada, P. J. (2019). Remote sensing of solar-induced chlorophyll fluorescence (SIF) in vegetation: 50 years of progress. *Remote Sensing of Environment*, 231, 111177. doi:10.1016/j.rse.2019.04.030
- Nearing, G. S., Crow, W. T., Thorp, K. R., Moran, M. S., Reichle, R. H., & Gupta, H. V. (2012). Assimilating remote sensing observations of leaf area index and soil moisture for wheat yield estimates: An observing system simulation experiment. *Water Resources Research*, 48, W05525. doi:10.1029/2011wr011420
- Pan, S. J., & Yang, Q. (2010). A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, 22, 1345-1359. doi:10.1109/tkde.2009.191
- Paz, J. O., Batchelor, W. D., Babcock, B. A., Colvin, T. S., Logsdon, S. D., Kaspar, T. C., & Karlen, D. L. (1999). Model-based technique to determine variable rate nitrogen for corn. *Agricultural Systems*, 61, 69-75.

- Pipia, L., Munoz-Mari, J., Amin, E., Belda, S., Camps-Valls, G., & Verrelst, J. (2019). Fusing optical and SAR time series for LAI gap filling with multioutput Gaussian processes. *Remote Sensing of Environment*, 235, 111452. doi:10.1016/j.rse.2019.111452
- Ray, D. K., Mueller, N. D., West, P. C., & Foley, J. A. (2013). Yield trends are insufficient to double global crop production by 2050. *PLoS One*, 8, e66428. doi:10.1371/journal.pone.0066428
- Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, 566, 195-204. doi:10.1038/s41586-019-0912-1
- Sarkadi, J., & Várallyay, G. (1989). Advisory system for mineral fertilization based on large-scale land-site maps. *Agrokémia és Talajtan*, 38, 775-789.
- Schlesinger, W. H. (2009). On the fate of anthropogenic nitrogen. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 203-208. doi:10.1073/pnas.0810193105
- Shanahan, J. F., Kitchen, N. R., Raun, W. R., & Schepers, J. S. (2008). Responsive in-season nitrogen management for cereals. *Computers and Electronics in Agriculture*, 61, 51-62. doi:10.1016/j.compag.2007.06.006
- Sinclair, T. R., Rufty, T. W., & Lewis, R. S. (2019). Increasing photosynthesis: Unlikely solution for world food problem. *Trends in Plant Science*, 24, 1032-1039. doi:10.1016/j.tplants.2019.07.008
- Strachan, I. B., Pattey, E., & Boisvert, J. B. (2002). Impact of nitrogen and environmental conditions on corn as detected by hyperspectral reflectance. *Remote Sensing of Environment*, 80, 213-224.
- Tao, F., Huang, Y., Hungate, B. A., Manzoni, S., Frey, S. D., Schmidt, M. W. I., Reichstein, M., Carvalhais, N., Ciais, P., Jiang, L., Lehmann, J., Wang, Y. P., Houlton, B. Z., Ahrens, B., Mishra, U., Hugelius, G., Hocking, T. D., Lu, X., Shi, Z., Viatkin, K., Vargas, R., Yigini, Y., Omuto, C., Malik, A. A., Peralta, G., Cuevas-Corona, R., Di Paolo, L. E., Luotto, I., Liao, C., Liang, Y. S., Saynes, V. S., Huang, X., & Luo, Y. (2023). Microbial carbon use efficiency promotes global soil carbon storage. *Nature*, 618, 981-985. doi:10.1038/s41586-023-06042-3
- Tao, F., & Luo, Y. (2022). PROcess-Guided Deep Learning and DATA-Driven Modelling (PRODA). *Land Carbon Cycle Modeling* (pp. 319-328).
- Tilman, D., Balzer, C., Hill, J., & Befort, B. L. (2011). Global food demand and the sustainable intensification of agriculture. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 20260-20264. doi:10.1073/pnas.1116437108
- Várallyay, G., Buzás, I., Kádár, I., & Németh, T. (1992). New plant nutrition advisory system in Hungary. *Communications in Soil Science and Plant Analysis*, 23, 2053-2073.
- Verrelst, J., Rivera, J. P., Moreno, J., & Camps-Valls, G. (2013). Gaussian processes uncertainty estimates in experimental Sentinel-2 LAI and leaf chlorophyll content retrieval. *ISPRS Journal of Photogrammetry and Remote Sensing*, 86, 157-167. doi:10.1016/j.isprsjprs.2013.09.012
- von Caemmerer, S., & Evans, J. R. (2010). Enhancing C₃ photosynthesis. *Plant Physiology*, 154, 589-592. doi:10.1104/pp.110.160952
- von Caemmerer, S., & Furbank, R. T. (2016). Strategies for improving C₄ photosynthesis. *Current opinion in plant biology*, 31, 125-134. doi:10.1016/j.pbi.2016.04.003

- Wallach, D., Martre, P., Liu, B., Asseng, S., Ewert, F., Thorburn, P. J., van Ittersum, M., Aggarwal, P. K., Ahmed, M., Basso, B., Biernath, C., Cammarano, D., Challinor, A. J., De Sanctis, G., Dumont, B., Eyshi Rezaei, E., Fereres, E., Fitzgerald, G. J., Gao, Y., Garcia-Vila, M., Gayler, S., Girusse, C., Hoogenboom, G., Horan, H., Izaurralde, R. C., Jones, C. D., Kassie, B. T., Kersebaum, K. C., Klein, C., Koehler, A. K., Maiorano, A., Minoli, S., Muller, C., Naresh Kumar, S., Nendel, C., O'Leary, G. J., Palosuo, T., Priesack, E., Ripoche, D., Rotter, R. P., Semenov, M. A., Stockle, C., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Wolf, J., & Zhang, Z. (2018). Multimodel ensembles improve predictions of crop-environment-management interactions. *Global Change Biology*, *24*, 5072-5083. doi:10.1111/gcb.14411
- Walter, J., & Kromdijk, J. (2022). Here comes the sun: How optimization of photosynthetic light reactions can boost crop yields. *Journal of Integrative Plant Biology*, *64*, 564-591. doi:10.1111/jipb.13206
- Wan, L., Zhou, W., He, Y., Wanger, T. C., & Cen, H. (2022). Combining transfer learning and hyperspectral reflectance analysis to assess leaf nitrogen concentration across different plant species datasets. *Remote Sensing of Environment*, *269*, 112826. doi:10.1016/j.rse.2021.112826
- Wang, D., & Cai, X. (2007). Optimal estimation of irrigation schedule – An example of quantifying human interferences to hydrologic processes. *Advances in Water Resources*, *30*, 1844-1857. doi:10.1016/j.advwatres.2007.02.006
- Wu, A., Hammer, G. L., Doherty, A., von Caemmerer, S., & Farquhar, G. D. (2019). Quantifying impacts of enhancing photosynthesis on crop yield. *Nature Plants*, *5*, 380-388. doi:10.1038/s41477-019-0398-8
- Yin, X., Gu, J., Dingkuhn, M., & Struik, P. C. (2022). A model-guided holistic review of exploiting natural variation of photosynthesis traits in crop improvement. *Journal of Experimental Botany*, *73*, 3173-3188. doi:10.1093/jxb/erac109
- Yin, X., & Struik, P. C. (2010). Modelling the crop: from system dynamics to systems biology. *Journal of Experimental Botany*, *61*, 2171-2183. doi:10.1093/jxb/erp375
- Yin, X., & Struik, P. C. (2017). Can increased leaf photosynthesis be converted into higher crop mass production? A simulation study for rice using the crop model GECROS. *Journal of Experimental Botany*, *68*, 2345-2360. doi:10.1093/jxb/erx085
- Yin, X., Struik, P. C., & Goudriaan, J. (2021). On the needs for combining physiological principles and mathematics to improve crop models. *Field Crops Research*, *271*, 108254. doi:10.1016/j.fcr.2021.108254
- Zhang, X., Mauzerall, D. L., Davidson, E. A., Kanter, D. R., & Cai, R. (2015). The economic and environmental consequences of implementing nitrogen-efficient technologies and management practices in agriculture. *Journal of Environmental Quality*, *44*, 312-324. doi:10.2134/jeq2014.03.0129
- Zhang, Y., Hui, J., Qin, Q., Sun, Y., Zhang, T., Sun, H., & Li, M. (2021). Transfer-learning-based approach for leaf chlorophyll content estimation of winter wheat from hyperspectral data. *Remote Sensing of Environment*, *267*, 112724. doi:10.1016/j.rse.2021.112724

Zhou, Z., Zhang, Z., van der Putten, P. E. L., Fabre, D., Dingkuhn, M., Struik, P. C., & Yin, X. (2023). Triose phosphate utilization in leaves is modulated by whole-plant sink-source ratios and nitrogen budgets in rice. *Journal of Experimental Botany*, 74, 6692-6707. doi:10.1093/jxb/erad329

Summary

Ensuring global food security requires increased crop production. Although nitrogen (N) fertilisation has significantly contributed to increasing global crop production, its efficiency is rather low. Severe environmental issues are aggravated as excessive N use leads to increased N loss. Blindly increasing N fertilisation for more crop production has reached a critical stage. To achieve a food-secure future, sustainable crop production is required. It is necessary to increase crop productivity, improve N use efficiency and minimise negative environmental impacts simultaneously. Decision tools for precise fertilisation have been developed, to match N application and crop demand. However, compared with *in situ* crop N needs, N rates can be easily under- or over-recommended, due to the lack of consideration of temporal and spatial variations in crop growth. To address this issue, a smart N management system that effectively acquires and uses “agricultural big data” should be explored to optimise N rate during the growing season based on *in situ* soil and weather conditions and crop growth status. Developing such a system was the main purpose of this thesis.

Chapter 1 outlines the existing use of crop modelling and remote sensing to support N management in the field. Making use of the complementarities of crop modelling and remote sensing, the integrated approach is considered a promising strategy to temporally simulate and spatially monitor crop N needs. This approach is expected to better optimise in-season N management to achieve the multi-objectives simultaneously in terms of productivity, economic profit, and sustainability. Using the mechanistic crop model GECROS, this thesis connects leaf-level photosynthesis to N optimisation in the field, with a focus on improved ability of predicting crop growth to determine the optimal N topdressing.

The primary process of crop growth and production is photosynthesis. A mechanistic prediction of crop growth requires the robust modelling of photosynthesis or related parameters in crop models. Due to the fundamental role of N in various photosynthetic machineries, the key photosynthetic parameters in GECROS and their relationship with specific leaf N content (SLN) were quantified in **Chapter 2**. Results showed that SLN varied under three N application levels and differed significantly between C₃ (rice and wheat) and C₄ (maize and sorghum) species. The linear relationships between photosynthetic capacity and SLN differed significantly between C₃ and C₄ species. The linear slope of Rubisco carboxylation capacity versus SLN for C₃ crops was significantly steeper than that for C₄ crops. The electron transport

parameters, the maximum electron transport rate for C_3 crops and the maximum ATP production rate for C_4 crops, significantly increased with SLN. The increases of triose phosphate utilisation for C_3 crops and phosphoenolpyruvate carboxylation capacity for C_4 crops with an increase in SLN were significant as well. Our results highlighted the importance of leaf N status in the estimation of photosynthesis capacity and crop productivity.

To obtain leaf N status, hyperspectral images were collected across the growing season of major field crops, rice, wheat, and maize, in **Chapter 3**. For the better prediction of leaf and canopy nitrogen contents, I investigated the performance of five machine learning algorithms (or referred to as nonparametric regression algorithms) with different feature sets and prediction pathways. The feature sets were combined from extracted hyperspectral features, including canopy reflectance, vegetation indices and texture information, and simulated crop development stage. The Partial Least Squares Regression and Support Vector Regression models performed better than the others, and their performance could be improved by the combined use of three types of hyperspectral features, but not by a further incorporation of development stage. Our results also showed that the prediction performance of the mass-based leaf N trait (leaf N concentration in g N (g DM)^{-1}) was better than that of the area-based trait (SLN in g N m^{-2}). Compared with the directly predicted SLN, the values of indirectly predicted SLN via the specific leaf area and leaf N concentration agreed better with the measurements. For the canopy scale, the values of canopy N content were better predicted by the indirect method using leaf area index and directly predicted SLN than those by the direct method in rice and maize. The results suggest that the prediction of leaf and canopy N traits can be improved by identifying indirect predicting pathways, although they might become much uncertain due to the introduced more regression processes.

Forecasting of crop status benefits from remote sensing by incorporating image information into crop model simulations. However, the uncertainties exist both in remote sensing predictions and in crop model simulations, and propagate across growing season, leading to a biased forecasting of in-season crop growth and end-of-season crop yield. In **Chapter 4**, I developed a Bayesian methodology that systematically quantified uncertainties of crop model simulations and remote sensing predictions to improve the forecasting ability of crop status. Ensemble Kalman Filter, one of the most popular methods for data assimilation, was applied based on the quantified uncertainties. For the parameter calibration and uncertainty estimation of the crop model GECROS, a Markov Chain Monte Carlo approach was used, while a Gaussian Process Regression model was adopted to obtain remote sensing predictions and

estimate their uncertainties. Therefore, several disparate methods were combined in one framework, and the framework was validated by a two years' field experiment. Seven types of measured canopy-level observations from the first year were included into the process, including aboveground biomass, grain weight, aboveground N content, grain N content and leaf traits like leaf dry weight, leaf N content and leaf area index. The calibrated parameters of crop model GECROS performed well in the validation year (the second year), except for the simulated leaf traits. Compared with the simulations from the crop model GECROS, remotely sensed leaf traits predicted by a Gaussian Process Regression model were more accurate, with simultaneously estimated uncertainties from the Gaussian Process Regression model itself. To integrate crop model simulations with remote sensing predictions, the simulated leaf traits from crop model GECROS were directly updated by the remotely sensed ones to agree better with the measurements. Those of aboveground biomass, grain weight, aboveground N content and grain N content were updated indirectly with the improved performance of forecasting. The estimated uncertainties resulted in a better performance of the data assimilation system, compared with the ones when the pre-assumed uncertainty was used alone or when it was combined with the inflation factor being introduced into Ensemble Kalman Filter. The results showed that the developed method can contribute to the better understanding of uncertainties analysis in data assimilation and more accurate forecasting of in-season and end-of-season crop carbon and N status.

With the precise monitoring and forecasting of crop growth (**Chapter 3** and **Chapter 4**), the description of temporal and spatial crop N uptake and demand was expected to be used to optimise in-season N management. The research aim of **Chapter 5** was to develop the in-season N optimisation methods to simultaneously maximise yield, profit, and N use efficiency for sustainable smarting farming. In this Chapter, I proposed two new methods for N management in the field. The performance of a crop model-based N optimisation method was investigated, which aimed to temporally determine the optimal N management from the soil-crop-atmosphere continuum. An integrated remote sensing-crop model method by data assimilation was also proposed to optimise N management *in situ*. Three reference methods were used for comparison, including the farmer practice optimised by the yield response curve, the remote sensing method based on Sufficiency Index, and the remote sensing method based on Response Index. The optimised in-season topdressing N of the newly proposed methods tended to perform better in improving yield, profit, and N use efficiency than that of reference methods, based on the simulated crop growth with actual weather data. Benefiting from the *in situ* remote sensing

information, the integrated crop model-remote sensing method performed better than that of the crop model-based method. Our results suggested that the integrated crop model-remote sensing method suits best to guide smart fertiliser management for the sustainable agricultural production. The crop model-based method that relies on crop physiological principles is recommended for situations when facing obstacles in obtaining remote sensing images or implementing *in situ* decisions.

In Chapter 6, main findings regarding the integrated utilisation of crop modelling and remote sensing in field N optimisation were summarised. Then, the significant merits of the designed smart N management system were discussed, including connecting leaf photosynthesis, enhancing monitoring, and forecasting of crop growth, and achieving multi-objectives simultaneously. Finally, routes of exploiting crop modelling and remote sensing were further discussed, especially in terms of upscaling the smart N management from site to region. Overall, I conclude that crop modelling- and remote sensing-based N management system can be considered as a promising strategy to achieve smart farming for sustainable agricultural production.

Acknowledgements

I deeply appreciate all the institutions and individuals for their support during my PhD journey. I could not have achieved so much in academia without all your presence.

First of all, I am grateful to Shanghai Lankuaikē Technology Development Company Limited (LTD, China) and Wageningen University & Research (WUR, the Netherlands). The collaborative project of these two institutions enlightened my academic path at the very beginning and shaped my thinking about the future of agriculture. I feel privileged to have the opportunity to be part of this project.

Secondly, I would like to express my sincere gratitude to my supervision team, which consisted of Prof. Paul C. Struik [Centre for Crop Systems Analysis (CSA), WUR], Dr. Xinyou Yin (CSA, WUR) and Dr. Lei Liang (LTD).

This team provided comprehensive support to me in completing my challenging work. Dear Paul, thank you so much for your considerate guidance and encouragement during my PhD life. Due to the interdisciplinary task, I always needed to stop and learn new things. Your kind praise for every small achievement of mine encouraged me to go further and further, until the final thesis writing. I also admire your high work efficiency and excellent interpersonal skills, which I believe have contributed to your remarkable life. Dear Xinyou, thank you so much for your meticulous supervision and profound influence on my academic career. I still remembered that on that thunderstorm day, you thoughtfully arranged to pick me up from Schiphol Airport, which has always moved me as a first-time overseas traveller. Your kind patience and guidance helped me get out of the swamp of bluntness and disorderliness. I deeply appreciate your carefulness and innovativeness throughout my PhD study. Your invaluable guidance in scientific thinking and article crafting has greatly ameliorated my academic standards. I would also like to extend my appreciation to your wife, Yarong, who encouraged and inspired me every time I visited. Dear Lei, thank you so much for your great trust and patience. You accompanied me to every workplace of LTD in Shanghai and spared no effort to fulfil my academic requests. Your enthusiasm and dedication to the field of smart farming always invigorated me. The supervision team supported me in every aspect of my PhD career. I am sincerely grateful to have you all.

I would like to extend my specific acknowledgement to Dr. Didi Qian and Mr. Blake Dong. Dear Didi, your kind encouragement and considerable supports during my PhD study lightened

my burden of trivial affairs especially and helped me get through some tough times. Dear Blake, your thoughtful advice and keen insights motivated my career pursuits. I also wish to express my gratitude to Dr. Haolu Shang. Your initial efforts in the project greatly laid the foundation of my entire PhD study.

I am thankful to the previous and current colleagues in LTD: Dawei Zhao, Heng Zhang, Huaitan Wang, Jiale Chai, Mingrui Gao, Shuangxi Xie, Shunkai Zhang, Xingchuan Zhang, Yufei Wang and Zhijie Zhang. All your dedicated efforts were crucial for supporting me in conducting my field experiments and data collection. I would also like to express my thank to Dr. Binzhen Du, Dr. Jiejiang Zhao, Dr. Xuwen Feng, Xiaohan Zhou and Xinyuan Min, for all the engaging conversations we have had. My gratitude also goes to other colleagues in LTD for their kind helps.

For my studying in WUR, I want to thank our secretaires in the CSA group at first: Nicole Wolffensperger and Petra Rozema, for assisting me in all kinds of administrative formalities and group affairs. I would like to give special thanks to Dr. Allard J. W. de Wit, Dr. Francisco Pinto, and Dr. Steven M. Driever for all the inspiring conversations we had. I would also like to express my sincere gratitude to Prof. Gerard B. M. Heuvelink and Dr. Sytze de Bruin for leading me into the realm of geo-information science. I deeply thank my great friends, Zhenxiang Zhou, Liang Fang and Liping Shao. Your thoughtful support and cheerful communications coloured my academic as well as social life. I would also like to thank my other PhD fellows and friends: Ambra Tosto, Bei Dong, David Kottelenberg, Hao Liu, Haozheng Li, Heng Wan, Hongyu Sun, Jiahui Gu, Jiali Cheng, Jian Liu, Jie Lu, Jin Wang, Ke Shan, Kostas Kipros, Mengxiao Sun, Panpan Fan, Phuong Thi Nguyen, Qian Li, Qian Liu, Quanxin Wan, Robbert van den Dool, Shuangwei Li, Shuangshuang Jia, Tao Song, Wenjing Ouyang, Xiaorui Guo, Yalin Liu, Yifeng Lu, Zhuo Xu, Zishen Wang and many others. Thank you a lot to all these people who were part of my memorable PhD life.

Last but by not least, I would express my deepest respect and love to my parents and sister. Your unconditional love healed me during moments of overwhelming doubt, pressure and despair and consistently supported me in pursuing my PhD abroad. I would like to give my final thank to my girlfriend Ziyi Chen. I always owe you more for being in a long-distance relationship for so long. I dedicate an ancient Chinese poem to commemorate our past: if love between both sides can last for aye, why need we stay together night and day? I propose my modified version to promise our future: if love between both sides can last for aye, why do not need we stay together night and day?

Best wishes to all my supervisors, friends, colleagues and family members.

Dong Wang

Wageningen, The Netherlands,

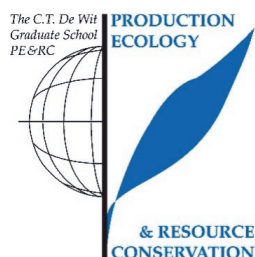
January, 2024

List of Publications

- Wang, D.,** Struik, P. C., Liang, L., & Yin, X. (2024). Developing remote sensing- and crop model-based methods to optimise nitrogen management in rice fields. *Computers and Electronics in Agriculture*: accepted.
- Wang, D.,** Struik, P. C., Liang, L., & Yin, X. (2023). Enhancing crop growth forecasting by incorporating estimated uncertainties for time series hyperspectral image data and crop model GECROS into Ensemble Kalman Filter. Submitted.
- Wang, D.,** Struik, P. C., Liang, L., & Yin, X. (2023). Estimating leaf and canopy nitrogen contents in major field crops across the growing season from hyperspectral images using nonparametric regression. Submitted.
- Wang, D.,** Rianti, W., Gálvez, F., van der Putten, P. E. L., Struik, P. C., & Yin, X. (2022). Estimating photosynthetic parameter values of rice, wheat, maize and sorghum to enable smart crop cultivation. *Crop and Environment*, 1, 119-132. doi:10.1016/j.crope.2022.05.004

PE&RC Training and Education Statement

With the training and education activities listed below the PhD candidate has complied with the requirements set by the C.T. de Wit Graduate School for Production Ecology and Resource Conservation (PE&RC) which comprises of a minimum total of 32 ECTS (= 22 weeks of activities)



Review/project proposal (6 ECTS)

- Integrating simulation model GECROS and sensor technology in the era of big data to design real-time smart crop management system in China

Post-graduate courses (6.5 ECTS)

- Fundamentals of crop physiology; PE&RC and University of Florida (2019)
- Root ecology; PE&RC and University of Copenhagen (2023)
- Advanced training and application of agrosystems modelling; Shanxi Agricultural University (2023)
- Geostatistics; Wageningen University & Research (2023)

Deficiency, refresh, brush-up courses (3 ECTS)

- Data management and machine learning in Python; Jiangsu Chuanzhiboke Education Technology Co. LTD. (2020)

Laboratory training and working visits (3 ECTS)

- Utilisation of sensor technologies in agriculture; National Engineering Research Centre for Information Technology in Agriculture (2021)
- Green agricultural production and field experiment visits; Science and Technology Backyard in Gusheng Village, China Agricultural University (2023)

Competence, skills and career-oriented activities (3.45 ECTS)

- Searching and organising literature; WGS (2019)
- Research data management; WGS (2019)
- Essentials scientific writing and presenting; WGS (2019)
- Personal growth and career planning in a developing company; Lankuaikei Agriculture Development Shanghai Co., Ltd. (2020)

Scientific integrity/ethics in science activities (0.3 ECTS)

- Ethics in plant and environmental sciences; WGS (2019)

PE&RC Retreat, PE&RC day, and other PE&RC events (1.5 ECTS)

- PE&RC First year weekend (2019)
- PE&RC Last year's retreat (2023)

Discussion groups/local seminars or scientific meetings (9.6 ECTS)

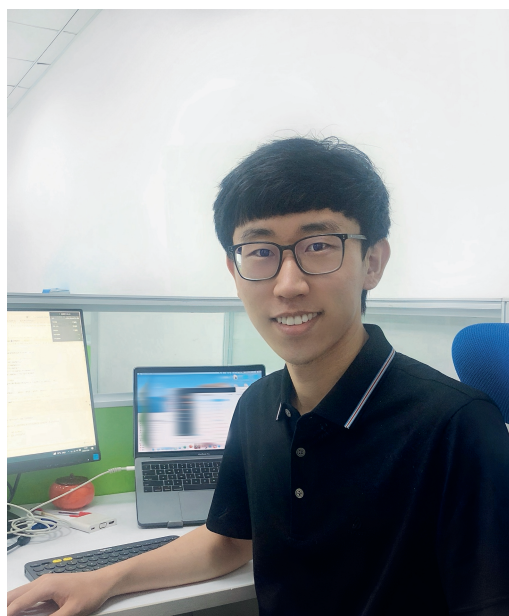
- Weekly seminar of centre for crop systems analysis (2019)
- Lankuaikēi agriculture development company meeting in China (2019-2022)
- 7th International symposium on plant growth modelling, simulation, visualisation, and applications (2022)
- The 8th International symposium on agricultural model development and application (2022)
- Workshop on agrosystems modelling and agrometeorology (2023)

International symposia, workshops and conferences (5.7 ECTS)

- World artificial intelligence conference; Shanghai, China (2022)
- The 11th international conference on agro-geoinformatics; Wuhan, China (2023)
- 13rd Workshop on hyperspectral image and signal processing: evolution in remote sensing; Athens, Greece (2023)

Curriculum Vitae

Dong Wang was born in Shanxi Province, China, on the 19th of August, 1995. In 2012, he started his BSc study at the College of Agronomy and Biotechnology at the China Agricultural University (CAU), Beijing, China. After completing his four-year BSc study programme, he pursued his MSc degree in the Research Centre of Regional Agriculture of CAU from 2016 to 2018. During his MSc studies, he majored in farming ecology and worked on a thesis entitled “Ecosystem services evaluation of maize-based multi-cropping systems in North China Plain”. In 2018, he acquired a



PhD position within the project “Big data quantification and modelling for modern agriculture in China”, a collaborative project involving Shanghai Lankuaikei Technology Development Company Limited (LTD) and Wageningen University & Research (WUR). In 2019, he started his PhD journey, in which he explored the potential of integrating crop modelling and remote sensing to design a smart nitrogen management system for sustainable crop production in China. This PhD research work was under the supervision of Prof. Paul C. Struik (WUR), Dr. Xinyou Yin (WUR), and Dr. Lei Liang (LTD).

Funding

The research described in this thesis was financially supported by Shanghai Lankuaikei Technology Development Co. Ltd.

Financial support from the Centre for Crop Systems Analysis, Wageningen University, for printing this thesis is gratefully acknowledged.

Cover designed by Yuxi Jing and Dong Wang

Printed by ProefschriftMaken

