TOWARDS IMPROVED CROP GROWTH AND YIELD ESTIMATION: OBSERVATION CONSTRAINED WHEAT MODELLING

Yuxi Zhang

A Thesis
Submitted in Fulfilment of the Requirements for the
Degree of Doctor of Philosophy

Department of Civil Engineering
Monash University
Melbourne, Australia

October 2020

Copyright notice

Declaration of statement

This thesis describes my research carried out in the Department of Civil Engineering at Monash University in Australia during the candidature period from February 2016 to May 2020. The thesis is submitted to Monash University in total fulfilment of the requirements for the degree of Doctor of Philosophy.

This thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

Yuxi Zhang

October 2020

Acknowledgment

First and foremost, I would like to express my deepest appreciation to my main supervisor, Professor Jeffrey Walker for the continuous support and guidance of my PhD study. Without his invaluable encouragement, guidance, supervising and persistent help, this dissertation would not have been possible. I would also like to thank my co-supervisor, A/Professor Valentijn Pauwels, for his support in all the time of my PhD study and writing of this thesis.

Next, I wish to express my heartfelt gratitude to my families for their support, financially and mentally, that accompanied me during this PhD study. I would also like to express my special thanks to my partner, Dr Wei Wang, who supported, accompanied, and cared for me during the four-year study. I would also like to thank my friend's kitten, Hetao, who saved my sanity from the thesis writing and self-isolation during the COVID-19 pandemic.

I also appreciate the assistance and support from my colleagues, Dr Nan Ye, Dr Xiaoling Wu, Dr Mei Sun Yee, Dr Stefania Grimaldi, Dr Liujun Zhang, Dr Ying Gao, Dr Chaoxun Hang, Mr Yuval Sadeh, Ms Nithyapriya Boopathy, Mr Xiaoji Shen, and Mr Jaya Pudashine.

I would like to thank the hard-working Monash laboratory staff for their support and assistance, especially Ms Kiri Mason, Mr Long Goh, Mr Michael Leach, Mr Richard Williamson, and Ms Tracy Warner. I also appreciate the support from Ms Kristie Hunter, Ms Jenny Mason, Ms Fatin Hasan, and Ms Min Major, Ms Lilian Khaw, Ms Noi Souvandy, and Mr Sohail Khan from the Department of Civil Engineering.

Abstract

Predicting crop yield in response to the spatial and temporary variability within the field provides essential information for field management to improve productivity and reduce input.

Compared to solely using either the remotely sensed data or the crop models, which can only handle one aspect between the spatial and temporal variability, the data assimilation techniques provide an approach to assimilate remotely sensed data into the crop models. The assimilation of remotely sensed data and crop modelling for yield prediction was expected to integrate the advantages of the remote sensing that provides spatial variability and coverage, and the crop modelling that provides temporal variability and insights of the causes of such variability. Moreover, the data assimilation could reduce the uncertainties sourced from input and parameters to address an improved yield estimation.

Existing studies of crop model data assimilation are often limited to regional scale that cannot provide sufficient information for field management, based on simple assumptions on uncertainties, and with limited types of remote sensing observations been assimilated. The main contribution of this research was to comprehensively and systematically explore the assimilation of all potential state variables into the crop model APSIM-Wheat. The research aim was addressed through three data assimilation studies.

First, the APSIM-Wheat model was tested for its sensitivity to all input variables and prognostic state variables. Simultaneously, a data assimilation framework was developed for the APSIM-Wheat model, using the Ensemble Kalman filter algorithm and can be easily extended to other crop types and state-updating data assimilation method.

The first data assimilation study used a synthetically-created dataset to assimilate all

wheat and soil states into the APSIM-Wheat. This study confirmed the potential that the assimilation of wheat and soil state variables could provide an improved estimation in wheat states and yield when using the leaf area index, biomass, plant nitrogen, soil moisture and soil nitrogen.

The second data assimilation study used a dataset collected in a wheat season to assimilate all measurable state variables by ground measurements into the APSIM-Wheat. This study showed that the wheat states and yield were better estimated by assimilating wheat states (leaf area index, biomass) only, while the soil states (soil moisture and nitrogen) showed no improvement, given the soil water and nitrogen modules in the model well calibrated against these soil states. Moreover, the assimilation of wheat states improved the yield estimation even when the model parameters were uncalibrated.

The third data assimilation study used the same ground dataset as the second study, but with the assimilated observations obtained from remote sensing. This study confirmed the results of the previous study that the leaf area index was capable of improving yield estimation whenever the model was calibrated or not, while the soil moisture showed no improvement. Moreover, this study showed that the spatial information provided by the fine-resolution LAI has the potential to provide sub-field spatial variation to the model.

Contents

List of Figures

List of Tables

List of symbols

Symbols	Units	Descriptions
A	m^2	Catchment area in the lateral outflow calculation.
ALX	-	A variable to relate the current day of the year to the time of the hottest day of the year.
AMP	°C	Annual amplitude in mean monthly temperature.
AirDry	m^3/m^3	Soil moisture of air-dry soil.
BD	g/cm^3	Bulk density.
C_N	g/g	Nitrogen concentration of plant.
$C_{N,min}$, $C_{N,crit}$	g/g	Critical and minimum nitrogen concentration of plant, respectively.
ConA	-	A regression coefficient that specifies the change of cumulative evaporation in the second stage of soil evaporation calculation.
CNR	-	Carbon:nitrogen ratio.
d	mm	The depth of saturation in the soil layer.
DaS	-	Day after (simulation) start.
DepthLag	NA	A depth lag factor.
decay $rate_{pool}$	/day	The decomposition rate of substances in specific pool.
$Decomposition_{pool}$	kg/ha	The amount of decomposition substances in specific pool.
D_{head}, D_{grain}	kg/ha	Biomass demand of head and grain, respectively.

Symbols	Units	Descriptions
Diffusivity	mm^2/day	The diffusivity between two adjacent soil layers.
DiffusConst	mm^2/day	Parameter in the soil water diffusivity calculation.
DiffusSlope	/mm	Parameter in the soil water diffusivity calculation.
DoY	-	Day of year.
DUL	m^3/m^3	Drained upper limit.
e	-	The error vector between model estimated and the true state vector.
$E(\cdot)$	-	The expectation operator.
$factor(\cdot)$	-	Function of stress factors relating to soil water (W), soil temperature (T), CNR, etc.
F_{biom}	NA	The biom pool carbon as a fraction of the hum carbon that is subject to decomposition (0-1).
$f_{c,N}$	NA	
F_{inert}	NA	The fraction of the hum pool that is considered to be non-susceptible to decomposition (0-1).
$Flow_{sat}$, $Flow_{unsat}$, $Flow_{lat}$	mm	Saturated, unsaturated and lateral flow, respectively.
F_{leaf}	NA	The fraction for biomass to be partitioned to leaf.
$f_{N,pheno}$, $f_{N,photo}$, $f_{N,expan}$	NA	Nitrogen stress factor for phenology, photosynthesis and leaf expansion, respectively.
fom_c	NA	Fresh organic matter carbon in organic matter pools.
$f_{W,pheno}$, $f_{W,photo}$, $f_{W,expan}$	NA	Factors of water stress effect on phenology, photosynthesis and leaf expansion, respectively.
GrainWt, GrainN	g/m^2	Dry weight and nitrogen amount of grain, respectively.
H	-	Linearized observation operator, a matrix mapping state vector to observations.
$\mathcal{H}(\cdot)$		Observation operator.
$H(\cdot)$		Linear observation operator.

Symbols	Units	Descriptions
$h_{N,pheno}$, $h_{N,photo}$, $h_{N,expan}$	-	Multipliers of nitrogen deficit effect on phenology, photosynthesis and leaf expansion, respectively.
$h_{W,pheno}$, $h_{W,photo}$, $h_{W,expan}$	-	Functions relating soil water availability to the soil water stress of phenology, photosynthesis and leaf expansion, respectively.
I	MJ	Intercepted radiation.
$inert_c$	kg/ha	Carbon in the organic matter pools that is non-susceptible to decomposition.
K_{lat}	mm/day	Lateral hydraulic conductivity.
KS	mm/day	Amount of water that is allowed to drain from the layer when the the soil water is above saturation.
K		Kalman gain.
k, $k-1$		The k^{th} and the $k-1^{th}$ timestep.
LAI, LAI_c, LAI_s	m^2/m^2	Leaf area index, carbon-limited LAI, and stressed LAI.
LL15	mm^3/mm^3	15Bar lower limit of soil water content, equivalent to WheatLL or LL in APSIM.
WheatLL, LL	mm^3/mm^3	The lower limit to water extraction for each layer.
LeafWt, StemWt, PodWt, GrainWt	g/m^2	Dry weight of leaf, stem, pod and grain, respectively.
LeafN, StemN, PodN, GrainN	g/m^2	Nitrogen amount of leaf, stem, pod and grain, respectively.
L	m	Catchment discharge width in lateral outflow calculation.
LeafWt, LeafN	g/m^2	Dry weight and nitrogen amount of leaf, respectively.
M		Linearized predictive model, a matrix passing state vector in one timestep to the next.

Symbols	Units	Descriptions
$\mathcal{M}(.)$		A Markov chain-based state transfer function passes the state of the system at time step $k-1$ into the next step k.
$M(.)$		Linear Markov chain-based model estimates the state of the system at time step k given the estimated state in $k-1$.
N_g	/plant	Number of grains per plant.
NO3N1, NO3N2, etc	kg/ha	Soil nitrogen in the form of nitrate in soil layer 1, 2, etc.
NH4N1, NH4N2, etc	kg/ha	Soil nitrogen in the form of ammonium in soil layer 1, 2, etc.
oc	kg/ha	Organic carbon in organic matter pools.
P	-	Error covariance of the state vector.
PodWt, PodN	g/m^2,	Dry weight and nitrogen amount of pod, respectively
Q	kg/ha, or g/m^2	Dry biomass.
Q	-	Model error covariance.
R	-	Observational error covariance.
R^2 or R-squared	-	Coefficient of determination.
R_g	$gram^{-1}$	Number of grain per gram stem.
RUE	NA	Radiation use efficiency.
SLA, SLA_{max}	m^2/g	Specific leaf area, maximum specific leaf area.
s	m/m	Slope in the lateral outflow calculation.
Sat	mm^3/mm^3	Saturated soil moisture.
SM1, SM2, etc.		Volumetric soil moisture in soil layer 1, 2, etc
ST	°C	Soil temperature.
SW, SWC, or SM	mm^3/mm^3	Volumetric soil water content or soil moisture.

Symbols	Units	Descriptions
SWCON	/mm	A fraction of the water is allowed to drain into the next deeper layer.
StemWt, StemN	g/m^2	Dry weight and nitrogen amount of stem, respectively
t	day	Time.
TAV	°C	Annual average ambient temperature.
Tb_v, Tb_h	K	Brightness temperature in the vertical and the horizontal polarization directions.
T_c	°C	Crown temperature.
TempChange	°C	The change in soil temperature since the hottest day of the year.
total carbon	kg/ha	Total carbon in the organic matter pools.
MPDI	-	Multi-polarization difference index.
TT	day ·° C	Thermal time.
U	mm	Empirical parameter for soil evaporation estimation.
$\mathbf{v}$	-	Observational uncertainty.
$\mathbf{w}$	-	Model uncertainty.
W_s	gram	The dry weight of stem at anthesis.
W_u	mm	Total daily water update from the root.
W_d	mm	Soil water demand of leaf and head.
$\mathbf{x}$	-	True state vector in a dynamic system.
$\hat{\mathbf{x}}$	-	Estimated state vector.
$\mathbf{x}^b$, $\mathbf{x}^a$	-	Background and analysis state vector.
$\mathbf{X}$	-	State matrix, an ensemble of state vectors.
Yield	kg/ha	Wheat yield, the grain weight in kg/ha at harvest.
$\mathbf{y}$	-	Observation vector in a dynamic system.
θ_{av}	mm	Average of extractable soil water.

Symbols	Units	Descriptions
μ^b, μ^a	-	The ensemble mean of background and analysis state vectors.
$\triangle_{\bullet}$	-	Increment operator.

List of abbreviations

ACRM	A two-layer Canopy Reflectance Model
ALOS	Advanced Land Observing Satellite
AMSR2	Advanced Microwave Scanning Radiometer 2
APES	The Agricultural Production and Externalities Simulator
APSIM	Agricultural Production Systems sIMulator
ASTER	Advanced Spaceborne Thermal Emission and Reflection Radiometer
BLUE	Best Linear Unbiased Estimator
CERES	Crop Environment Resource Synthesis
COSMO	Constellation of Small Satellites for Mediterranean basin Observation
COUP	Coupled heat and mass transfer model for soil-plant-atmosphere systems
CropSyst	Cropping Systems Simulation Model
CSM	Cropping System Model
DUL	Drained Upper Limit
DVI	Difference Vegetation Index
DSSAT	Decision-Support System for Agro-technology Transfer
EKF	Extended Kalman filter
EnKF	Ensemble Kalman filter
EPIC	Environmental Policy Integrated Climate

ET	Evapotranspiration
ETM+	Enhanced Thematic Mapper Plus
EVI	Enhanced Vegetation Index
FASSET	Farm ASSEssment Tool
GPS	Global Positioning System
HDAS	Hydraprobe Data Acquisition System
IPVI	Infrared Percentage Vegetation Index
KF	Kalman Filter
KOMPSAT	Korea Multi-Purpose Satellite
LAI	Leaf Area Index
MAP	Mono-ammonium phosphate
MODIS	Moderate Resolution Imaging Spectroradiometer
MONICA	Model for NItrogen and CArbon dynamics in agroecosystems
MPDI	Multi-Polarization Difference Index
N	Nitrogen
NDVI	Normalized Difference Vegetation Index
NIR	Near-infrared
NRMSE, or RRMSE	Normalized (relative) RMSE
OM	Organic Matter
OSSE	Observing System Simulation Experiment
PALSAR	Phased Array type L-band Synthetic Aperture Radar
Pan	Panchromatic
PET	Potential evapotranspiration
PEST	Parameter ESTimation
PF	Particle filter
PROSAIL	A coupled PROSPECT-SAIL model
PROSPECT	
PSO	Particle Swarm Optimization

PVI	Perpendicular VI
RD	Relative difference
RMSE	Root mean square error
RTM	Radiative Transfer Model
RVI or SR	Ratio Vegetation Index, or Simple (Vegetation) Ratio
SAFY	Simple Algorithm for Yield model
SAIL	Scattering by Arbitrarily Inclined Leaves
SAR	Synthetic-aperture radar
SAVI	Soil Adjusted VI
SCE-UA	Shuffled Complex Evolution - University of Arizona
SIRF	Sequential Importance Resampling Filter
SMAP	Soil Moisture Active Passive
SMOS	Soil Moisture and Ocean Salinity
SoilWat	The soil water module in the APSIM
SoilN	The soil nitrogen module in the APSIM
SoilOM	The soil organic matter module in the APSIM
SoilTemp	The soil temperature module in the APSIM
SPOT	Satellite Pour l'Observation de la Terre, or Satellite for observation of Earth
ST	Soil temperature
STICS	Simulateur mulTIdisciplinaire pour les Cultures Standard, or multidisciplinary simulator for standard crops
SUCROS	Simple and Universal Crop growth Simulator
SWAP	Soil-Water-Atmosphere-Plant model
SWI	Soil Water Index
SWIR	short-wavelength infrared
TIR	Thermal infrared
VIS	Visible
VP	Vapor pressure

WDVI	Weighted Difference VI
WOFOST	WOrld FOod STudies

Chapter 1

Introduction

This thesis presents a crop model-data assimilation framework for improved wheat yield estimation. The principle of the proposed method is to integrate the observations of wheat and soil states into a wheat simulation model using an advanced data assimilation technique. The main novelty of this work is the comprehensive exploration of all prognostic wheat and soil states that are observable with either ground measurements or current remote sensing techniques for their impact on yield prediction when assimilated. This crop model-data assimilation framework was evaluated with field experiment data collected by the candidate during a whole wheat growing season specifically for this thesis. While the current framework is specific to wheat, it can be extended to other crop types available in the simulation model with only minor modification to the source code. Although the application of remote sensing data is limited to the point scale in this work, the assimilation of two-dimensional remote sensing images is a potential direction for future research that can be easily adapted from the work presented here.

1.1 Statement of the problem

With increased human population comes a demand for more food, but optimised crop productivity of farms requires management in response to the spatial and temporal variability of yield. The spatial variability is primarily caused by soil characteristics (Whelan, 2006; Maestrini & Basso, 2018), while temporal variability caused by season-to-season climate variations (Robertson, 2006; Basso *et al.* , 2007). Predicting within-farm crop yield

variation in response to the spatial and temporal variability is the key to improved field management (Batchelor *et al.* , 2002) for achieving increased farm productivity, reduced cost and a reduced impact on the environment (Panda *et al.* , 2010; Noori & Panda, 2016; Shaw *et al.* , 2016; Paustian & Theuvsen, 2017).

While crop simulation models provide an understanding of crop growth in a time-continuous manner, and can be run at fine temporal and spatial resolutions, the data required for accurate simulations across spatial scales are not usually available (Batchelor *et al.* , 2002; Mosleh *et al.* , 2015). Remote sensing is now providing observations for crucial wheat and soil states with high-resolution remote sensing (e.g., Sentinel-2) data that meets the requirements for agricultural application. Examples of such states include leaf area index (LAI) that strongly impacts crop interception of solar radiation, and is therefore considered a crucial indicator for photosynthesis productivity and grain yield (Huang *et al.* , 2016); and soil moisture (SM) that is directly related to plant water stress. In contrast to crop modelling, remote sensing data provide information on spatial variability, but the current methods to make use of them does not account for temporal variability (Basso *et al.* , 2007; Mosleh *et al.* , 2015) or provide insights on the interaction of the crop with the environment for management purposes (Mosleh *et al.* , 2015). Combination of crop simulation models and remote sensing data therefore provides a potential pathway for providing spatially variable information on crop status and expected yield.

The approach of combining crop simulation models and remotely sensed data has been summarized by many researchers (Wiegand *et al.* , 1986; Maas, 1988b; Bouman, 1995; Jin *et al.* , 2018). These strategies include using remotely sensed data: (1) as model input (e.g., Maas, 1988b); (2) for direct-insertion to substitute model simulated states (e.g., Maas, 1988b); (3) for calibration whereby model initial conditions and/or parameters are optimized by comparing modelled and remotely sensed states (e.g., Launay & Guerif, 2005; Thorp *et al.* , 2012; Jin *et al.* , 2016b; Novelli *et al.* , 2019); and, (4) state-updating data assimilation techniques that sequentially update the model states (e.g., Curnel *et al.* , 2011; Nearing *et al.* , 2012; Ines *et al.* , 2013; Li *et al.* , 2017a). Strategies (3) and (4) are the most commonly used in current practice.

Re-calibration approaches usually do not account for errors from remote sensing data (Delécolle *et al.* , 1992; Jin *et al.* , 2018; Kang & Özdoğan, 2019), and therefore the model

states and parameters can be inaccurate. Moreover, the re-calibration process requires a lot of optimization iterations, which is computationally expensive for complex model systems (Jin *et al.* , 2018). The optimization process is also unsuitable for real-time applications, where the iteration needs to be repeated whenever a new observation is obtained. State-updating algorithms allow the model states to be corrected in real-time without iteration at the instant when the observation is obtained (Jamet & Loisel, 2013). Moreover, advanced data assimilation techniques account for the uncertainties of both the model and the observations (Houser *et al.* , 2010), allowing yield to be estimated together with its uncertainty.

The existing practice of data assimilation has often been focused at a regional scale to monitor food security, with the results evaluated using official statistics to support policy-making (e.g., Ines *et al.* , 2013; Zhao *et al.* , 2013; Huang *et al.* , 2016). However, limited types of remote sensing observations have been assimilated. Moreover, the application of crop modelling and data assimilation to support agricultural decision-making requires studies to be undertaken at the field or sub-field scale with the integration of fine-resolution remote sensing data to provide spatial information at sub-paddock scale. A comprehensive and systematical exploration of assimilating all potential state variables into crop models is also lacking, due to the difficulties of measuring the state variables.

Based on the review of the literature (further specified in Chapter 2), three research questions were proposed in this thesis as:

1. Does the assimilation of all prognostic wheat and/or soil state variables in a synthetic study results in any improvement for yield estimation of APSIM-Wheat?

2. By making use of in-situ observations of wheat and soil states, does the assimilation of state variables in a case study results in any improvement for yield estimation of APSIM-Wheat?

3. By making use of available remote sensing optical and passive microwave data, does the assimilation of state variables in a case study results in any improvement for yield estimation of APSIM-Wheat?

1.2 Objective and scope

The Agricultural Production Systems sIMulator (APSIM, https://www.apsim.info/apsim-model), a highly advanced simulator of the agricultural system, has been selected for this study. A range of wheat and soil states in APSIM can be linked with external observations, including (1) remotely sensed LAI and surface soil moisture; (2) in-situ measured soil moisture for each soil layer, and ammonium and nitrate in the top two layers; and (3) destructively sampled weight of above-ground wheat organs (including leaf and stem and the total biomass). The primary purpose of this research was to assimilate observations of wheat and soil states into the APSIM-Wheat for predicting yield at farm scale. The Ensemble Kalman Filter (EnKF) data assimilation algorithm has been developed for non-linear models such as APSIM, and thus was the assimilation approach in this study.

1.3 Outline of the approach

The philosophy taken in developing the APSIM-EnKF framework was to start from a sensitivity analysis to understand the importance of all the model inputs, parameters and state variables in impacting yield estimation. The project was divided into three main stages to progressively develop and test the proposed yield forecasting framework through a series of data assimilation experiments as outlined in Figure 1.1. Importantly, this research involved developing the code that integrates APSIM with the EnKF, called APSIM-EnKF herein, and in collecting an extensive field dataset for testing purposes.

Based on the results of the sensitivity analysis, a synthetic study was undertaken to explore the yield forecast impact of assimilating the variables shown to be most sensitive in the synthetic study. Subsequently, a case study assimilated the ground-measured observations of wheat and soil states into the APSIM-Wheat model found most effective from the synthetic study. Finally, a second case study assimilated available remotely sensed observations of wheat and soil states into the APSIM-Wheat model based on the findings of the first case study. The remote sensing data included a fine spatio-temporal resolution LAI dataset and a tower-based soil moisture dataset.

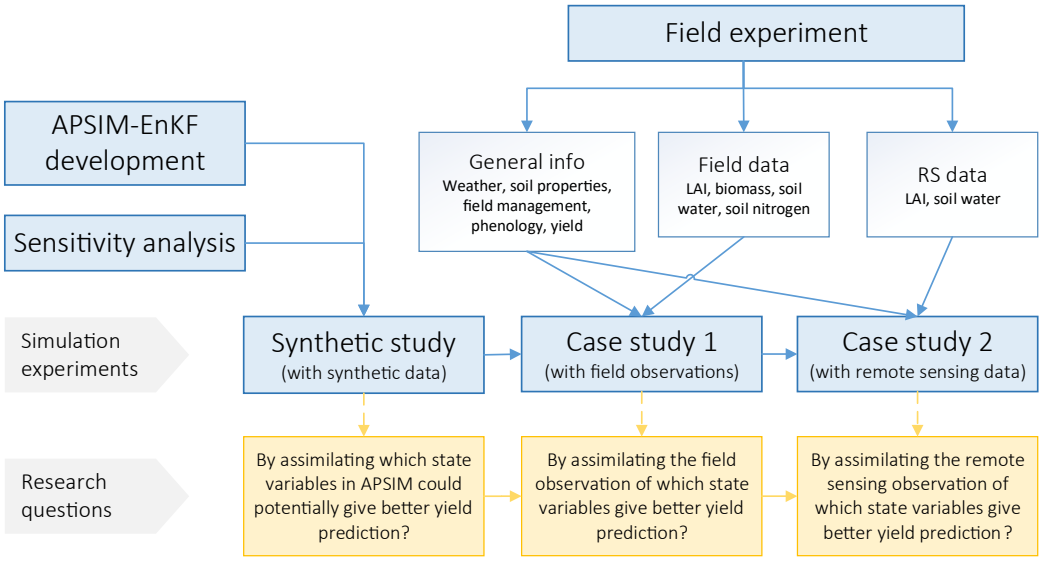

Figure 1.1: Research questions and experimental design of this project.

1.3.1 The sensitivity analysis and data assimilation studies

The sensitivity analysis was undertaken through the aspects below:

1. Which are the parameters most likely to have an impact on yield and which need to be carefully calibrated or measured?

2. Which are the model states most likely to have an impact on yield and which need to be well known?

3. Which are the weather forcing, states or variables that have the most impact on model prediction and thus need careful perturbation; and,

4. At which phenology stage, the perturbation of states have the most impact on model prediction?

The synthetic study was undertaken to answer the first research question through the aspects below:

1. The assimilation of which state variables whose observations are available from current remote sensing techniques (e.g., LAI, surface soil moisture, biomass) contribute to a better yield estimation?

2. Given observations of state variables that are easily obtained, is it possible that the

model gains any further improvement in yield estimation by assimilating state variables that are not currently available from remote sensing (e.g., phenology, soil nitrogen, root-zone soil moisture)?

3. Does the impact of particular state variable assimilation become important for particular phenology conditions?

4. What is the minimum requirement for observation data regarding repeat time, growing stage and accuracy to achieve better yield prediction?

The case studies were undertaken to answer the second and third research questions through the aspects below:

1. The assimilation of which state variables whose observations were collected in the field experiment contributes to a better yield estimation? Do they still provide an improved yield estimation by being replaced by the remotely sensed observations?

2. Does the impact of a particular state variable assimilation become important for particular phenology conditions?

3. Does the assimilation of these state variables further contribute to a better yield estimation by constraining the wheat phenology with the field observations?

4. Does the assimilation of these state variables into an uncalibrated APSIM model provide a comparative accuracy in yield estimation relative to a calibrated model, so that the observation of state variables can compensate the lack of some model parameters that are usually difficult to measure?

1.3.2 The APSIM model

The APSIM-Wheat model (version: APSIMX, known as the APSIM next generation) was selected for this study for several reasons:

- Localized to Australia: A wide range of crop simulation models are available, among which only several have been coupled with data assimilation algorithms. Although these models are widely applied in Europe, America and China, they require long-term and extensive experiments for calibration and validation before they can be used

in Australia. APSIM has been developed and used in Australia for more than 20 years (Probert *et al.* , 1998; Asseng *et al.* , 1998, 2000; Keating *et al.* , 2001, 2003; Holzworth *et al.* , 2014), making it among the most suitable crop simulation models in Australia. Moreover, it is preferable for Australian farmers, researchers and companies to use a model that is more familiar to them.

- Reliable: APSIM is an advanced crop simulation model. The accuracy and reliability of APSIM in crop simulation have been confirmed in several model comparison articles. More than 100 articles have been published with APSIM, simulating diverse crop types and under stressed conditions (e.g., Asseng *et al.* , 1998, 2000; Thorburn *et al.* , 2001; Asseng *et al.* , 2002; Farré *et al.* , 2002; Van Ittersum *et al.* , 2003; Zhang *et al.* , 2012; Zhao *et al.* , 2014; Ahmed *et al.* , 2016; Thorburn *et al.* , 2010; Robertson *et al.* , 2002).

- Flexible and user friendly: APSIM is a crop simulation system that integrates a range of crop and soil modules. Thus, skilled users can assemble model components based on their own need with maximum flexibility. For general users, a user interface and a detailed user manual are provided (Zheng *et al.* , 2014).

- Well-designed: APSIM was coded with parallel programming, which allows the program to run on advanced parallel research computers. The architecture of APSIM makes it easy to be coupled with the EnKF or any other state-updating algorithms. Moreover, the new version APSIMX is under re-design, with wheat module been completed at the start of this study. This new version was expected to be faster and can be used in multi-platform.

1.4 Organization of the thesis

The research embodied in this thesis is divided into nine chapters that describe a series of synthetic and realistic experiments. An outline of the thesis is presented below:

- Chapter 2 presents a review of the literature relevant to yield monitoring techniques, data assimilation algorithms and recent data assimilation experiments on crop modelling.

- Chapter 3 describes the APSIM-Wheat model, the EnKF data assimilation algorithm, and the coupled APSIM-EnKF framework.
- Chapter 4 presents a field experiment conducted in the Cora Lynn area to provide a dataset for evaluating this data assimilation method in real case studies.
- Chapter 5 presents the sensitivity analysis of the APSIM-Wheat model.
- Chapter 6 presents the results of a synthetic study to assimilate prognostic state variables of APSIM-Wheat with the EnKF.
- Chapter 7 presents the results of a case study to assimilate observations of APSIM-Wheat state variables collected from field observations.
- Chapter 8 presents the results of a case study to assimilate observations of APSIM-Wheat state variables obtained from remote sensing.
- Chapter 9 outlines the overall conclusions of this study and a perspective for further study.

Chapter 2

Literature review

The topic of applying remote sensing and data assimilation techniques to agricultural modelling for improved crop yield prediction has attracted attention in recent years. This chapter aims to present a clear understanding of the current status of remote sensing and model in agriculture for the purpose of improving crop monitoring and yield prediction. This chapter is divided into four sections. The first section presents the current status of food production and the requirement for understanding yield variability at field scale in the framework of precision agriculture. The second section describes current yield estimation methods and highlights the need for merging remote sensing information with crop models for the purpose of yield estimation and in understanding yield variability at field scale. The third section critically reviews the current practices of using data assimilation methods in crop modelling. Finally, based on the knowledge gaps identified, the research questions of this project are presented in the fourth section.

2.1 Background

Wheat is one of the largest harvest crops in the world. According to the statistics reported by the Food and Agriculture Organization (FAO) of the United Nations, wheat production contributes to more than 8% of the world crop production, taking account of over 12% of the harvest area. More than half of the world's wheat production originates from China, India, the Russian Federation and the United States of America, with Australia also being one of the top 10 wheat producers in the world. Wheat produces 70% of the total cereal production

in Western Australia and generates 2 to 3 billion dollars of the state revenue every year.

With increasing population comes a demand for more food, but the total food productivity of a farm is limited by the spatial and temporal variability of yield. Numerous factors contribute to yield variability. Temporally, crop yield varies as a result of season-to-season climate variability in terms of rainfall, temperature and solar radiation (Robertson, 2006). Variability and extreme events in climate strongly impact the crop growth and the timing of management events (e.g. sowing, fertilizer use, etc.) (Robertson, 2006). Spatially, yield variability is mostly attributed to variations in the soil chemical and physical characteristics, including soil water availability, soil texture, soil structure, soil depth, organic matter, nutrients and pH (Batchelor *et al.* , 2002; Whelan & Palmer, 2006). Biotic factors (e.g., weeds, pest and diseases) and past management also influence the spatial variability (Whelan, 2006). Therefore, understanding within-farm crop yield variation in response to these factors is the key to improved field management, for increasing farm productivity, and for reducing the impact on the environment (Batchelor *et al.* , 2002).

Driven by increased crop demand, precision agriculture, also called site-specific management (Plant, 2001), has attracted great attention in recent decades. One of the main purposes of precision agriculture is to maximize production efficiency (Panda *et al.* , 2010; Noori & Panda, 2016) by optimizing field management "at a spatial scale smaller than the field" (Plant, 2001) using information techniques. With an understanding of crop yield variability, growers can make better decisions on when and where to sow, irrigate and fertilize, and by how much. Growers can thus manage farms based on the actual need of the crop to achieve the best profit from minimum input, rather than simply applying the same water and fertilizer amount for the entire paddock. Therefore, predicting crop yield with explicit accounting of variability and understanding of the yield variability factors is of great significance in field management for precision agriculture.

2.2 Yield estimation techniques

Techniques for crop monitoring and yield estimation are divided into three categories: (1) modelling crop growth through agrometeorological crop models; (2) relating historical yield with specific remote sensing indices through statistical methods; and (3) joining remote

sensing and crop models with data assimilation methods. These yield estimation approaches are introduced in this section.

2.2.1 Crop modelling

Agrometeorological crop models (referred to as crop models hereafter) have been developed as physics-based decision support tools to describe the development of plants and their interaction with weather, atmosphere, soil and field management. A definition introduced by Wiegand *et al.* (1986) stated that agrometeorological crop models are designed to describe the crop behaviours from planting to maturity at field scale, by simulating various plant processes (i.e., phenology, photosynthesis, respiration, evapotranspiration, dry matter accumulation) driven by daily weather data and soil properties as input data.

Table 2.1 reproduced from Palosuo *et al.* (2011) was supplemented to include more models, to list the main features of several popular crop simulation models/systems. The classification of the main features is also based on the work of Palosuo *et al.* (2011). Most models are generic, applied to several crop types, while a few are crop-specified (APSIM and DSSAT). Crop models consider the light utilization process either based on detailed photosynthesis-respiration processes, or a simple radiation use efficiency approach. The leaf area development and light interception approaches could be detailed, accounting for the leaf area dynamics under the impact of phenology stages, temperature or light; or simple, estimating with the specific leaf area and biomass partitioning, or with a maximum leaf area index (Palosuo *et al.* , 2011). Crop phenology is based on a temperature accumulation concept, known as the growth degree-days or the thermal time, defined as the mean daily temperature between certain lower/upper threshold) accumulated over a period of time. For winter wheat, most of the models consider the impact of vernalization, a key process for some winter crops that requires a period of low temperature to change from the vegetative stage to the reproductive stage, in phenology development. Yield is determined by a fixed harvest index multiplied with the above-ground biomass, or biomass partitioning during the reproductive stages, and some consider the grain number per plant as a factor in yield development. Soil water and nitrogen were usually considered the main stress factor for crop modelling. Most models use a capacity model to estimate soil water dynamics, while a few use a simplification or numerical solution of Richard's equations. Evapotranspiration

Table 2.1: Main features of several popular crop simulation models. Reproduced and updated from Palosuo *et al.* (2011).

Main feature	Model											
	APSIM-Wheat	AquaCrop	APES	COUP	CropSyst	Daisy	CERES-Wheat	FASSET	MONICA	Sirius	STICS	WOFOST
Model type [a]	S	G	G	G	G	G	S	G	G	G	G	G
Light utilization[b]	RUE	None	RUE	P-R	RUE	P-R	RUE	RUE	P-R	RUE	RUE	P-R
Leaf area development and light interception[c]	S	None	D	D	S	D	S	D	D	D	D	D
Yield formation[d]	Y(Gn, Prt)	Y(HI, B)	Y(Prt)	Y(Prt)	Y(HI, B)	Y(Prt)	Y(Gn, Prt)	Y(HI, B)	Y(HI, B)	Y(HI(Gn), B)	Y(HI(Gn), B)	Y(Prt, B)
Crop phenology[e]	f(T, DL, V)	f(T, DL)	f(T, DL, V)	f(T, DL)	f(T, DL, V)	f(T, DL)	f(T, DL, V)	f(T, DL)	f(T, DL, V)	f(T, DL, V)	f(T, DL, V)	f(T, DL)
Stresses involved[f]	W, N	W	W, N	W, N, T, S	W, N	W, N	W, N, H, C, L	W, N	W, N, O, H	W, N	W, N	W, N, K, P
Soil water dynamics[g]	C	C	C	R	C or R	R	C	C	C	C	C	C
Evapotranspiration[h]	PT	PM	P	PM	PM or PT	PM	PT, P-F or PM	Makkink	PM	P	P	P
Model description	Keating *et al.* (2003)	Raes *et al.* (2009)	Donatelli *et al.* (2007, 2010)	Jansson & Karl-berg (2010)	Stöckle *et al.* (2003)	Abrahamsen & Hansen (2000)	Jones *et al.* (2003)	Lægdsmand (2011)	Nendel *et al.* (2011)	Jamieson *et al.* (1998b)	Brisson *et al.* (2003)	Van Diepen *et al.* (1989)

[a] G=generic, S=crop-specific. [b] D=detailed approaches consider photosynthesis and respiration processes, RUE=simple radiation use efficiency approach. [c] D=detailed approaches that consider leaf development varies by phenology states, S=simple approach. [d] Y(.) is the yield formation equation. HI=harvest index, Gn=grain number, B=total biomass, Pat=biomass partitioned to grain during the productive stages. [e] f(.) is the phenology development function. T=temperature, DL=day length (photoperiod), V=vernalization. [f] Stress factors, W=water, N=nitrogen, K=potassium, P=phosphorus, O=oxygen, C=carbon dioxide, L=light, H=heat, S=salinity. [g] C=capacity model, R=Richard's equation. [h] Evapotranspiration approaches, P=Penman, PT=Priestly-Talyer, TW=Turc-Wendling, P-F=Penman-FAO, PM=Penman-Moteich, Makk=Makkink.

were estimated by several approaches, including Penman (Penman, 1956), Penman-FAO (Doorenbos, 1975), Penman-Moteich (Monteith, 1986), Priestly-Talyer (Priestley & Taylor, 1972), Turc-Wendling (Wendling *et al.* , 1991), etc.

In the review of Jin *et al.* (2018), a timeline describing the evolution of several widely applied crop models over the world was presented. An inter-comparison of crop models can be found in several articles, including under drought (Jamieson *et al.* , 1998a), non-limiting conditions (Porter *et al.* , 1993), and different conditions of soil (Eitzinger *et al.* , 2004), climate (Palosuo *et al.* , 2011; Rötter *et al.* , 2012), climate change (Wolf *et al.* , 1996; Semenov *et al.* , 1996) and nitrogen (Olesen *et al.* , 2002; Salo *et al.* , 2016) scenarios.

Crop model simulation methods focus on the monitoring of crop growth in a time-continuous manner, providing the simulation of crop growth in temporal variability. These models provide an understanding of the causes of variability by accounting for the interaction of the crop with environmental factors, and thus widely applied in estimating yield in response to stress from water (Chapman, 2008), nitrogen (Lisson & Cotching, 2011), diseases (Rouse, 1988), and climate change (Luo & Kathuria, 2013).

The accuracy of crop models has been validated in a range of experiments. For example, Otter & Ritchie (1985) reported a coefficient of determination (R^2) of 0.81 of CERES-Wheat, using a dataset including 245 measured yields from 25 sites. A review by Timsina & Humphreys (2006) has reported the grain yield estimation with CERES-Wheat showed the RMSE between 0.11 and 0.53 t/ha and NRMSE between 2% to 16% in seven experiments with a total of 137 data pairs. More validation experiments were presented in Table 2.2. Overall, these process-based models were reported a high accuracy in yield estimation with sufficient input data provided.

Despite the many benefits of crop models above, there are several drawbacks of crop models that limit their application in crop growth and yield estimation. These models are driven by a range of weather data and require soil properties and field management information to run. Although weather data are usually available from weather stations, field management information needs to be provided by the grower, and soil properties require long-term calibration over several growing seasons or in-situ measurements that are labour-intensive and time-consuming to collect (Batchelor *et al.* , 2002; Mosleh *et al.* , 2015). The requirement for multiple types of data hinders the extensive application of crop models in new fields

Table 2.2: Yield estimation accuracy reported by validation experiments for crop simulation models.

Experiment(s) in literature	Model	Statistic criteria[a]			
		RMSE (t/ha)	NRMSE (%)	d-index	R^2
Ahmed *et al.* (2016)	APSIM-Wheat	0.15 to 0.40	1.30 to 2.98	NA	0.92 to 0.97
Asseng *et al.* (1998)	APSIM-Wheat	0.4	NA	NA	NA
Zhang *et al.* (2012)	APSIM-Wheat	0.53 to 0.91	7 to 22	NA	NA
Ahmed *et al.* (2016)	CERES-Wheat	0.08 to 0.12	0.94 to 2.01	NA	0.93 to 0.98
Xiong *et al.* (2008)	CERES-Wheat	NA	22.8	0.85	NA
Timsina & Humphreys (2006) (review)	CERES-Wheat	0.11 to 0.53	2 to 16	0.86 to 1 .00	NA
Patel *et al.* (2010)	CERES-Wheat	0.05 to 0.36	NA	-0.03 to 1.00	NA
Otter & Ritchie (1985)	CERES-Wheat	NA	NA	NA	0.81
Zhang *et al.* (2013)	AquaCrop	0.50 to 1.44	NA	0.95 to 0.98	NA
Pannkuk *et al.* (1998)	CropSyst	0.23 to 0.51	NA	NA	NA
Singh *et al.* (2013)	CropSyst	0.33	NA	NA	0.84

[a] RMSE: root mean square error. NRMSE: normalized (relative) RMSE. d-index: Wilmott index of agreement. R^2: coefficient of determination. NA: not available.

without adequate field information available.

Due to the difficulties of collecting input data, crop models are unsuitable for the growth and yield estimation across a variety of fields or locations, or at scales smaller than a field to address the spatial variability. Crop models usually aim to simulate crop growth on a point basis, and thus assume that all the inputs are uniform over the paddock (Batchelor *et al.* , 2002). To handle the yield spatial variability, simulations can be implemented with a group of point-based models, using the dense measurements of soil properties specific to the site (Batchelor *et al.* , 2002). Thus, accurate simulation at high spatial resolution is usually not available due to the limitation of spatial inputs.

2.2.2 Remote sensing observations

With the development of earth-observing satellites, remote sensing techniques now provide a large range of observations from airborne and space-borne sensors, with broad spatial coverage, long-term observation, and high spatial and temporal resolution. Remote sensing data provide information on the soil and vegetation canopy from three main sources: solar reflectance from the vegetation for optical sensors (Table 2.3), microwave emission from the soil for radiometers, and microwave scattering from the soil for radar (Table 2.4).

Table 2.3: Current multispectral remote sensing missions for agriculture purpose.

Mission	Bands included[a]					Launched year	Resolution (multispectral)	
	Pan	VIS	NIR	SWIR	TIR		Spatial (m)	Temporal (days)
ASTER		×	×	×	×	1999	15	16
IKONOS	×	×	×			1999	3.28	3
Landsat 7 ETM+	×	×	×	×	×	1999	30	16
GeoEye-1	×	×	×			2008	1.84	2.1 to 8.3
RapidEye		×	×			2008	5	1 to 5.5
Quickbird	×	×	×			2001	2.62	Sub-daily to 3.5
SPOT-5	×	×	×	×		2011	10	2 to 3
Pleiades-1A/B	×	×	×			2011 to 2012	2	1
SPOT-6/7	×	×	×			2012	6	1
Landsat 8 OLI	×	×	×	×	×	2013	30	16
SkySat	×	×	×			2013	1	sub-daily
WorldView-3	×	×	×	×		2014	1.24	1 to 4.5
KOMPSAT-3A	×	×	×		×	2015	2.2	1.4
Gaofen-2	×	×	×			2014	3.2	5
TripleSat	×	×	×			2015	3.2	4
Sentinel-2A/B		×	×	×		2015 to 2016	10	5
SuperView-1	×	×	×			2018	2	2

[a] Pan: Panchromatic. VIS: visible. NIR: near-infrared. SWIR: short-wavelength infrared. TIR: thermal infrared.

Table 2.4: Current microwave remote sensing missions.

Mission	Type[a]	Band(s)	Launched year	Spatial resolution	Revisit (days)
TerraSAR-X	A	X	2007	1 to 18.5 m	11
ALOS-2/PALSAR-2	A	L	2014	1 to 100 m	14
Radarsat-1	A	C	1995	8 to 100 m	24
Radarsat-2	A	C	2007	3 to 100 m	24
ALOS/PALSAR	A	L	2005	7 to 100 m	46
COSMO-SkyMed	A	X	2017 - 2010	3 to 15 m	5
Sentinel-1A/B	A	C	2014, 2016	5 to 40 m	12
AMSR2	P	C, X	2012	24 to 62 km	16
SMOS	P	L	2009	35 to 50 km	2 to 3
SMAP	A/P	L	2014	36 (P), 3(A), 9 (A-P)	3

[a] Sensor type: A=active, P=passive.

Optical data are observed by detecting the energy reflected and transmitted by the object in the channels of visible and infrared, usually with multi-spectral instruments. Optical data have been used as a variety of indices calculated as a combination of reflectance in multiple bands, as indicators of vegetation health and crop production potentials (Kayad *et al.* , 2016), or as input of process-based radiative transfer models (RTM). Spectral reflectance in the visible to infrared bands, particularly the near-infrared and the red bands, are widely used to estimate important vegetation canopies such as LAI, photosynthetically active radiation, canopy cover, biomass, and canopy water content because plant leaves have a high reflectance in the near-infrared band and high absorption in the red band (Guyot, 1990).

Radiometer and radar systems measure vegetation canopy and soil properties in the microwave bands. The differences between the two systems are that radiometers measure the brightness temperature of the object by passively receiving the microwave emission from the ground surface, in contrast of the radar systems that measure by actively sending and receiving microwave energy in different wavelengths and polarization. Microwave data are sensitive to water and thus commonly used in detecting soil and vegetation water conditions. Owing to the ability of microwave that can penetrate objects, observation in with radiometers and radars can be used to retrieve near-surface soil moisture of bare soil of depth 1 to 5 cm (usually in the L-\C-\X- bands), and is suitable for cloudy weather conditions, where spectral bands are often blocked by the cloud cover.

The initial attempts of using remote sensing information to estimate crop yield used a statistical basis (Delécolle *et al.* , 1992). The VIs were related to ground-observed yield with approaches ranging from multiple linear correlations to advanced non-linear techniques (Kayad *et al.* , 2016). For example, models that are based on linear and quadratic functions were employed by Son *et al.* (2013), searching for the best-fit relationship of rice yield with the composite of MODIS EVI and LAI. In recent years, machine learning techniques have been applied in yield estimation for precision agriculture, with successful techniques including Artificial Neural Networks, Support Vector Regression, M5-Prime Regression Trees, and k-nearest neighbour (Chlingaryan *et al.* , 2018).

Studies using statistical methods have demonstrated the potential of remote sensing data for providing spatial variability of soil and vegetation canopy information related to yield with the information provided by remote sensing data. However, statistical methods, such as

empirically-based methods have several disadvantages in yield estimation. The relationship between VIs and yield is usually established at certain growing stages. Thus the estimation does not allow for real-time yield monitoring. These methods lack an understanding of the causes underlying variability, especially when more than one factor is affecting the yield (Plant, 2001). Therefore, they cannot account for variability caused by extreme weather events and temporal variability without long-term observations (Mosleh *et al.* , 2015). Moreover, the statistical relationship is only valid to specific crop and location. When estimating the yield of a different crop or at a different location, either calibration of parameters or a new empirical relationship is needed.

2.2.3 Crop modelling and remote sensing observation

Compared to solely using remote sensing data or crop models in crop monitoring and yield estimation, numerous benefits can be found in integrating remote sensing data with crop simulation models. While crop simulation models can be run at fine temporal and spatial resolutions, the data required for accurate simulation at a spatial scale are usually unavailable. In contrast, remote sensing data provide information on spatial variability, but the way statistical methods make use of them does not account for temporal variability or provide insights on the interaction of the crop with the environmental factors for forecasting purposes. The combination of crop simulation models and remote sensing data therefore provides a potential pathway for providing spatially variable information on current crop status together with expected yield.

Early studies on integrating remote sensing data with crop simulation models arose in the 1980s, with three principal approaches: the "driving" method, the "direct insertion" and the "re-calibration" (Wiegand *et al.* , 1986; Bouman, 1995; Bouman *et al.* , 1997). The driving method uses remote sensing data as an alternative source of input weather data such as precipitation that drive the model. The direct insertion method replaces and resets the state variables such as LAI estimated by the model with the estimates from remote sensing data. With the re-calibration approaches, sometimes known as "simulation steering" (Bouman, 1995; Bouman *et al.* , 1997), an objective function representing the mismatch between model states and the relevant remote sensing observations are optimized by adjusting model initial conditions and parameters (known as re-initialization and re-parametrization), to "steer" the

model to give a best estimation of state variables that close to the remote sensing data.

Pioneering attempts to apply these strategies have been reported Maas (1988a,b); Bouman (1995); Kergoat *et al.* (1995); Clevers & Van Leeuwen (1996); Guérif & Duke (1998). For instance, Maas (1988b) re-initialized green leaf area index (GLAI) and above-ground dry biomass at emergence by fitting model-derived GLAI to remote sensing with an interactive method, showing remote sensing data is a labour-saving alternative of in-situ observations. In their following work (Maas, 1988a), remotely-sensed LAI was used in four methods: (1) as input data of a simple model for simulating the growth and yield of grain sorghum, and as an external data source of (2) direct insertion, (3) re-initialization and (4) re-parameterization methods. It was reported that re-calibration methods were more computationally expensive but more attractive than direct insertion in cancelling errors of data. Early studies have been mentioned in several good reviews (Wiegand *et al.* , 1986; Delécolle *et al.* , 1992; Fischer *et al.* , 1997; Moulin *et al.* , 1998; Plummer, 2000; Dorigo *et al.* , 2007). These studies have demonstrated the potential to predict crop yield by integrating crop simulation models with discrete remote sensing data in order to reduce the demand for expensive in-situ observations. The re-initialized timing of sowing or emergence can also be used as complementary information where those data are not recorded. Whilst the strategies of integrating remote sensing data been explored by early studies, the accuracy of these methods are often limited by the simple structure of the empirical models, the simple assumption that remote sensing observations are accurate, and the coarse resolution of early remote sensing data.

The development of crop models and fine spatial and temporal resolution land surface remote sensing observations in the recent two decades has provided the possibility of daily, time-continuous crop monitoring using process-based modelling and within-field spatial information for crop and soil states, and eventually more accurate and site-specific yield prediction. Techniques of merging remote sensing data with crop simulation models in the recent ten years were divided into two groups. The first group follows the early studies on re-calibration approaches, and explore various cost functions and optimization algorithms. The other group put efforts on using remote sensing information in crop modelling with advanced data assimilation methods that allow the model states to be updated according to the error characteristics of both the model and the observation (Reichle, 2008). These methods are further elaborated in the next section.

2.3 Techniques of integrating RS data with crop models

As a broad concept, data assimilation is the process of updating dynamic models with the observation data from an external source (Reichle, 2008; Stuart & Zygalakis, 2015). Data assimilation techniques have been extensively applied in improving the predictive performance of meteorology (e.g., Kanamitsu, 1989), land surface (e.g., Rodell *et al.* , 2004) and hydrologic (e.g., Reichle *et al.* , 2002) models using remote sensing data with great success.

Although the term data assimilation can be used for any processes of model updating, including input, state and parameter updating, and error correction (Houser *et al.* , 2010), the state updating, in a narrow sense, is usually what has been meant for data assimilation in meteorology, land surface and hydrology data assimilation research. In the remainder of this review, the discussion to data assimilation is limited to state updating. Therefore, although many researchers working on crop model data assimilation accept the broad concept of data assimilation to include the approaches that optimize cost functions through the re-estimation of initial conditions or parameters, this technique is a part of parameter updating and is termed "re-calibration" hereafter.

2.3.1 Re-calibration techniques

The re-calibration approaches are important and widely-used in integrating remote sensing data with crop models. These approaches are based on the assumption that external observations of model state variables are valid and the model can be steered to approach the observations by adjusting initial conditions or model parameters. Under this assumption, observations are completely trusted without querying any observational uncertainties possibly caused by instruments, observing techniques, or spatial resolution. The primary process of this technique is to construct and minimize a cost function representing the mismatch of model state variables between model estimates and remote sensing techniques.

Numerous re-calibration experiments have been conducted for wheat, maize, rice, sugar beet and cotton crops, using models and remote sensing information from a variety of sources. These experiments are summarized in Table 2.5, with cost functions usually constructed by the root mean squared deviation between the state variables of model estimates and remote sensing observations, and minimized by adjusting management parameters, crop establish-

ment parameters or initial state variables. Optimization algorithms or software involved in these studies to minimize the cost function includes Price search algorithm (Price, 1977), Parameter ESTimation (PEST, Doherty, 1994; Doherty *et al.* , 1994), Shuffled Complex Evolution method developed at The University of Arizona (SCE-UA, Duan *et al.* , 1993) and Particle Swarm Optimization (PSO, Eberhart & Kennedy, 1995; Kennedy & Eberhart, 1995). These experiments have confirmed the potential of using remote sensing data to calibrate crop models to provide the estimation of crop yield with an accuracy usually less than 1 t/ha (Table 2.5).

LAI is one of the predominant state variables that links the crop model with remote sensing information. The usage of remote sensing LAI products can be found in Fang *et al.* (2008); Thorp *et al.* (2012); Ma *et al.* (2013a); Huang *et al.* (2015b,a); Bai *et al.* (2019); Ban *et al.* (2019); Novelli *et al.* (2019). Other experiments have established links between crop model states and remote sensing data using a radiative transfer or microwave model, to allow predicted and observed spectral reflectance (usually red and near-infrared bands) or radar backscatter information to be directly compared. For instance, Launay & Guerif (2005) minimized the differences between estimated and remote sensing observations of TSAVI by re-calibrating parameters related to crop establishment and water use in a coupled SUCROS sugar beet model. In their experiment, SUCROS was coupled with a SAIL reflectance model to computed TSAVI. Thorp *et al.* (2012) linked DSSAT-CSM crop model with two radiative transfer models, PROSAIL and Choudhury, through NDVI and LAI, respectively. They adopted PEST estimation algorithm to optimize the objective function by adjusting the leaf area growth parameters. The strategies of integrating remote sensing data with crop models are further discussed in Section 2.4.2.

There is a variety of cost functions applied in these experiments. Although the mainstream technique is to use a cost function constructed by the sum or mean of the squared differences, or the sum of absolute differences between model estimates and remote sensing observations, attempts of using more complex cost functions appeared in the work of Huang *et al.* (2015a,b); Li *et al.* (2015b); Jin *et al.* (2016a). Huang *et al.* (2015b) compared the yield estimation of a traditional sum of squared difference cost function and a new cost function constructed by the sum of the generalized vector angles from the LAI and ET time series, and concluded that the new cost function gave a better yield estimation. However,

Table 2.5: Re-calibration experiments for yield improvement.

Ref.	Model	Re-calibrated paras or inits[b]	Remotely sensed data	Crop type	Accuracy[c]
Launay & Guerif (2005)	SUCROS [a]	Crop and water use paras	TSAVI	Sugar beet	Yield RMSE reduced in one of the two scenarios.
Fang *et al.* (2008)	CERES-Maize[a]	Management paras	LAI	Maize	Yield RD was 3.1-17.9% in the six scenarios.
Ma *et al.* (2008)	WOFOST[a]	Management paras	SAVI	Wheat	Estimates of the aggregate weights of storage organ closer to the official data.
Thorp *et al.* (2012)	DSSAT-CSM[a]	Crop paras	LAI, spectral reflectance	Wheat	Yield NRMSE was 12.8% and 10.0% in the two scenarios.
Ma *et al.* (2013a)	WOFOST	Emergence date, initial states	LAI	Wheat	Yield RMSE reduced from 0.98 to 0.47 and 0.669 in the two scenarios.
Huang *et al.* (2015b)	SWAP	Management paras	LAI, ET	Wheat	Yield RMSE was 0.6-1.6 in the three scenarios.
Huang *et al.* (2015a)	WOFOST	Management, crop paras and inits	LAI	Wheat	Yield RMSE reduced from 0.37 to 0.15.
Li *et al.* (2015b)	CERES-Wheat	Cultivar and management para	Canopy N accumula-tion	Wheat	Yield RMSE was 0.63.
Dong *et al.* (2016)	SAFY	Phenology paras	GreenLAI	Wheat	Yield RMSE was 0.23 and 0.19 in the two scenarios.
Jin *et al.* (2016a)	CERES-Maize[a]	Management, phenology and RTM paras	Spectral reflectance	Maize	LAI accuracy was enhanced.
Jin *et al.* (2016b)	AquaCrop	Crop paras	Biomass	Wheat	Biomass RMSE was 1.80 and 0.55 in the two scenarios.
Silvestro *et al.* (2017)	AquaCrop	Management and crop paras	Canopy cover	Wheat	Yield RMSE was 1.09.
Bai *et al.* (2019)	WOFOST	Crop para	LAI (single observation)	Jujube tree	Yield RMSE was 0.74 and 0.87 in the two years.
Ban *et al.* (2019)	CERES-Maize	Management and crop paras	LAI	Maize	Yield RMSE was between 0.75 and 3.00 in the twelve scenarios.
Huang *et al.* (2019)	WOFOST[a]	Crop paras	Spectral reflectance	Wheat	Yield RMSE was 0.60, 1.29, and 0.60 in the three years.
Novelli *et al.* (2019)	EPIC	Crop paras	LAI	Wheat	Yield RMSE was reduced from 0.57 to 0.32.

[a] Crop model was coupled with a radiative transfer model or microwave model in these experiments.
[b] paras: parameters, inits: initial conditions.
[c] The unit of yield and biomass RMSE is t/ha.

the efficiency of using various cost functions needs to be validated with more extensive experiments.

Re-calibration techniques have the drawback that they usually do not account for errors in the remote sensing data (Delécolle *et al.* , 1992; Jin *et al.* , 2018; Kang & Özdoğan, 2019). Model states and parameters can be inaccurately driven with poor-quality data (Delécolle *et al.* , 1992). With re-calibration methods, the application of the approaches to crop prediction for any other area is limited. Moreover, in real-time simulation, the process of re-parameterization will need to be repeated when a new observation is obtained, which is computationally expensive for complex model systems with increasingly more data observed over time. To overcome the drawbacks of the re-calibration methods, advanced data assimilation methods have been introduced.

2.3.2 Data assimilation techniques

Data assimilation aims to seek the best estimation of states of a dynamic system, given a model and external observations both with uncertainties (Houser *et al.* , 2010). There are two basic types of data assimilation approaches: smoothing and filtering. In practice, smoothing seeks to determine the best estimation of system states by minimizing an objective function constructed by the initial state vector, observations and their uncertainties over a period (known as the "data assimilation window") (Stuart & Zygalakis, 2015). Typical smoother algorithms include variational method (Var), Kalman smoother, etc.

Filtering sequentially updates the system states, given the probability distributions of the prior state and external observation. Filter algorithms update model states and the error covariance based on the background states from the previous timestep (Reichle, 2008), and allows real-time and successive updating during the evolution of the system whenever observations are available (Jamet & Loisel, 2013). Such a sequential-updating process makes filter algorithms popular in crop simulation processes, as crop modelling requires in-season prediction of yield to guide field management.

The filtering method was established to account for random noise, but errors in the model and data are often systematic (Dee, 2005; Li *et al.* , 2009). Systematic errors are unavoidable and difficult to attribute to the model or the observations, making a proper treatment of biases critical for the success of data assimilation systems (Dee & Da Silva, 1998; Kumar

et al. , 2012). For example, off-line bias correction schemes (e.g., a priori scaling approach, Dee, 2005) estimate biases from model predictions and observations prior to data assimilation (De Lannoy *et al.* , 2007). "Bias aware" (as opposed to "bias blind" methods where systematic errors are ignored, Dee, 2005) data assimilation methods to dynamically update parameters simultaneously with the model states to correct model errors (De Lannoy *et al.* , 2007, e.g., augmented-state approach, Evensen, 2003). Different from the off-line scheme, bias aware data assimilation is an on-line scheme designed to estimate errors in the model parameters and based on the assumption that observations are unbiased.

The Kalman filter (KF) is a typical filtering algorithm based on the hypothesis that the state in one timestep is transferred to the next timestep with a linear operator, and the probability distribution of uncertainties are Gaussian. The Kalman filter is only applicable to linear predictive models, but the crop growth processes are complex and are usually modelled by non-linear models. The Extended Kalman filter (EKF), Ensemble Kalman filter (EnKF) and Particle filter (PF) are two approaches designed to solve filtering problems in non-linear systems, popularly applied in the field of crop model data assimilation. Compared to the KF that was initially developed for linear models, the EnKF is a popular DA algorithm designed for non-linear models, being easy to code and implement (Li *et al.* , 2009), and therefore selected for this study. The practices of data assimilation (in the narrow sense) in crop yield estimation is further discussed in the next section.

2.4 Practices of data assimilation in yield estimation

Involvement of data assimilation techniques for the purpose of yield prediction has been initiated in the simulation of the land surface process, as crop yield is usually one of the model outputs; however, the main focus of land surface modelling is typically the water and energy exchange between the land surface and the atmosphere. For instance, Pellenq & Boulet (2004) presented synthetic experiments that apply EnKF to a coupled crop-SVAT model. Numerous studies of assimilating remote sensing data into coupled crop models with land surface models or hydrology models have been conducted (Olioso *et al.* , 2005; Pauwels *et al.* , 2007). However, instead of crop outputs, they concentrated on improving the estimation of water conditions such as evaporation and soil moisture.

The application of data assimilation in crop monitoring and yield prediction was initiated when De Wit & Van Diepen (2007) developed an EnKF data assimilation framework for the WOFOST crop simulation model, and assimilated the soil water index that derived from coarse resolution surface soil moisture. The results showed that the relationship between simulated yield and actual crop yield could be improved by the assimilation of a soil water index only 33 out of the 88 regions studied. The poor assimilation results were possibly caused by the irrigation in the study area, which was not considered by the model.

Curnel *et al.* (2011) compared the assimilation efficiency of two assimilation strategies, EnKF and the re-calibration method, in assimilating LAI into the WOFOST model in an Observing System Simulation Experiment (OSSE) (Arnold Jr & Dey, 1986). Results demonstrated a worse performance for the EnKF, which was attributed to the discrepancy in phenology stage of different ensemble members when the state variable LAIs of each ensemble was updated.

Another OSSE conducted by Nearing *et al.* (2012) explored the performance of two sequential data assimilation algorithms, EnKF and sequential importance resampling filter (SIRF), in assimilating synthetic observation of LAI and SM respectively, using a CERES-Wheat model. They concluded that state-updating algorithms were unable to improve yield estimation under realistic uncertainty conditions because the link between estimated yield and LAI was weak due to the model structure, while yield was strongly impacted by cultivar parameters. Besides, the surface soil moisture did not fully demonstrate water availability conditions.

Although the EnKF failed to improve end-of-season yield prediction in some OSSEs, the poor assimilation performance caused by model structure can possibly be cancelled by using more appropriate crop prediction models. It should be noted that synthetic studies were based on a series of assumptions and limitations, as a simplification of reality. In contrast to synthetic studies, realistic experiments have been conducted with some success (Table 2.6). The EnKF has been employed at the regional scale (Ines *et al.* , 2013; Zhao *et al.* , 2013; Huang *et al.* , 2016) and the results were validated with official yield statistics. In these experiments, improvements in simulated yield were observed, demonstrating the potential of EnKF in yield prediction in realistic conditions at the aggregate scale.

Table 2.6: Sequential data assimilation experiments for yield improvement.

Ref.	Model	DA method	Assimilated state(s)	Crop type	Scale	Accuracy[b, c]
De Wit & Van Diepen (2007)	WOFOST	EnKF	SM	Wheat	Regional	DA outperformed OL in only one of the four regions.
Vazifedoust *et al.* (2009)	SWAP	BLUE	LAI, relative ET	Wheat	Field	Yield RD reduced in all the three fields.
Curnel *et al.* (2011)	WOFOST	EnKF	LAI, SM	Wheat	Synthetic	Poor result due to the phenology shift in the simulated and observed LAI.
Nearing *et al.* (2012)	CERES-Wheat	EnKF, SIRF	LAI	Wheat	Synthetic	State-updating DA did not improve yield estimation due to lack of RZSM information, error LAI and a lack of correlation between leaf and grain growth.
Ines *et al.* (2013)	CSM-Maize	EnKF	LAI, SM	Maize	Regional	Yield RMSE reduced from 4.7 to 4.2, 3.6 and 2.9 in the three scenarios.
Ma *et al.* (2013b)	WOFOST[a]	EnKF	NDVI	Wheat	Regional	Yield RMSE reduced from 2.17 to 0.78.
Zhao *et al.* (2013)	WOFOST	EnKF	LAI	Wheat	Regional	Yield error reduced from 14.04% to 10.44-12.71% in the four scenarios.
Huang *et al.* (2016)	SWAP	EnKF	LAI	Wheat	Regional and field	Yield RMSE reduced from 0.65 to 0.27 in one of the two scenarios.
Li *et al.* (2017b)	CERES-Wheat	PF, POD4DVar	LAI	Wheat	Regional	Yield RMSE reduced from 0.91 to 0.44 and 0.31 by using two different DA algorithm.
Li *et al.* (2017a)	CERES-Wheat	EnKF	LAI	Wheat	Regional	Yield RMSE reduced from 0.49 to 0.44 and 0.32 in the two scenarios.
Silvestro *et al.* (2017)	SAFY	EnKF	LAI	Wheat	Regional	Yield RMSE was 1.09. No OL was given.
Guo *et al.* (2019)	WheatGrow[a]	EnKF	LAI	Wheat	Regional and field	Yield RMSE was between 0.92 and 1.57 in the four scenarios and two fields. No OL was given.
Kang & Özdoğan (2019)	SAFY	EnKF	LAI	Wheat	Regional	Yield RMSE reduced in all the five scenarios.

[a] Crop model was coupled with a radiative transfer model or microwave model in these experiments.

[b] The unit of yield RMSE is t/ha.

[c] OL: open-loop. DA: data assimilation.OL: open-loop. DA: data assimilation.

2.4.1 Existing crop model data assimilation frameworks

Data assimilation frameworks have been established for several widely-used crop models. A list of the combination of crop model and data assimilation algorithms is presented in Table 2.7.

The EnKF is the predominant data assimilation algorithm coupled with crop models, being designed for non-linear models and so could be easily coupled with the structure of crop models. The success of EnKF data assimilation greatly depends on the estimation of background and observation uncertainties (Pellenq & Boulet, 2004; Ines *et al.* , 2013). Background uncertainties are those generated during model prediction and caused by several resources: weather forcing, parameters, initial state variables and model physics (Li *et al.* , 2014). Observational uncertainties are caused by the errors of instruments, sensing techniques and retrieval algorithms.

Existing studies of EnKF data assimilation are usually based on the background and observations error on simplistic assumptions. Table 2.8 presents the assumptions of these EnKF data assimilation experiments. Most experiments assumed that the model uncertainties were only caused by model parameters, management parameters and initial state variables (e.g., biomass and leaf area), and thus initialized the model ensembles by perturbing crop parameters and initial leaf area or biomass. For instance, Ines *et al.* (2013) made the assumption that uncertainties were from several model parameters, and error values of $\pm 10\%$ were assumed. Zhao *et al.* (2013) simply assumed model uncertainties are from model parameters that control target accumulated thermal time, and mentioned that observational errors were difficult to assess because of the lack of validation. While the weather forcing such as precipitation being one of the most sensitive input data affecting the simulation result of crop growth of crop models (further specified in the sensitivity analysis in Chapter 5), only a few studies (De Wit & Van Diepen, 2007; Nagarajan *et al.* , 2012; Nearing *et al.* , 2012; Liu *et al.* , 2016) considered the uncertainties of weather forcing.

The state vectors of the EnKF are constructed of the prognostic state variables of the crop model that are to be updated at each timestep when external observations are available. The state vectors involved in existing EnKF data assimilation experiments were usually constructed of the specific crop state variable(s) being assimilated, with only a few state variables

Table 2.7: Existing data assimilation frameworks for crop models.

Model	DA algorithm
WOFOST	EnKF (De Wit & Van Diepen, 2007; Curnel *et al.* , 2011; Zhao *et al.* , 2013; Huang *et al.* , 2015b), EnKF coupled with an RTM (Ma *et al.* , 2013b), KF (Huang *et al.* , 2016)
DSSAT-CERES-Wheat	EnKF (Nearing *et al.* , 2012), SIRF (Nearing *et al.* , 2012), Var (Dente *et al.* , 2008), PF (Xie *et al.* , 2017)
DSSAT-CERES-Maize	EnKF (Ines *et al.* , 2013; Liu *et al.* , 2019)
DSSAT-CROPGRO-Soybean	EnKF coupled with a microwave model (Liu *et al.* , 2016)
SVAT	EnKF coupled with a microwave model (Nagarajan *et al.* , 2012)
WheatGrow	EnKF coupled with an RTM (Guo *et al.* , 2019)
SAFY	EnKF (Silvestro *et al.* , 2017)
SWAP	EnKF (Huang *et al.* , 2016)
AquaCrop	ExKF (Linker & Ioslovich, 2017)
MCWLA-Wheat	Constant gain KF (Chen *et al.* , 2018)

updated at timesteps when external observations are available. The only exception is the synthetic study of Nearing *et al.* (2012), where 33 prognostic state variables of WOFOST were updated to account for the correlations among all state variables by assimilating the observations of several state variables.

Several data assimilation experiments explored the use of the variational method (Dente *et al.* , 2008), KF (Vazifedoust *et al.* , 2009; Huang *et al.* , 2016), EKF (Linker & Ioslovich, 2017), and PF (Li *et al.* , 2015a; Xie *et al.* , 2017; Li *et al.* , 2017b). Dente *et al.* (2008) employed a variational method to minimize the differences between prior LAI state estimated by CERES-Wheat model and remotely sensed LAI, by tuning three parameter values rather than solving the initial state of the assimilation window, while seeking values of the latter is a common practice of the variational method. Vazifedoust *et al.* (2009) updated LAI and relative evapotranspiration states in the SWAP model, using a Best Linear Unbiased Estimate (BLUE) approach to estimate the Kalman gain as a constant calculated from the background and observational error covariance. Linker & Ioslovich (2017) tested two different data assimilation strategies with the AquaCrop model for potato and cotton, by assimilating canopy cover and biomass individually and together: in the first data assimilation strategy, the state variables were updated with an EKF algorithm, while the second strategy performed a recalibration process to updated three model parameters after states were updated. A practice of using the particle filter was found in Xie *et al.* (2017), in which LAI and soil moisture at the first 20 cm derived from Landsat observations were assimilated into the CERES-Wheat

Table 2.8: Assumptions of the experiments of EnKF data assimilation.

Ref.	Model	Ensemble size	Quantities[a] perturbed?				State vector [b]	Inflation applied?	Crop type
			W	Init	Para	States			
De Wit & Van Diepen (2007)	WOFOST	50	×	×			SM		Wheat
Curnel *et al.* (2011)	WOFOST	50			×		LAI, SM		Wheat
Nagarajan *et al.* (2012)	SVAT (MB)	100	×		×		LSP, RZSM		Maize
Nearing *et al.* (2012)	CSRES-Wheat	100	×	×	×		33 states		Wheat
Ines *et al.* (2013)	CERES-Maize	40			×		LAI, SM	×	Maize
Ma *et al.* (2013b)	WOFOST (ACRM)	70		×			NDVI		Wheat
Zhao *et al.* (2013)	WOFOST	50			×		LAI		Wheat
Huang *et al.* (2016)	SWAP	100			×		LAI		Wheat
Liu *et al.* (2016)	DSSAT-CROPGRO (MB)	50	×		×		SSM, RZSM		Soybean
Li *et al.* (2017a)	CERES-Wheat	100		×	×		LAI		Wheat
Silvestro *et al.* (2017)	SAFY	100			×	×	LAI		Wheat
Guo *et al.* (2019)	WheatGrow	70				×	LAI		Wheat
Kang & Özdoğan (2019)	SAFY	40		×	×	×	LAI	×	Wheat
Liu *et al.* (2019)	CERES-Maize	30		×	×	×	SSM	×	Maize

[a] Quantities perturbed: W=Weather, Init=Initial states, Para=Parameters.

[b] SM: soil moisture. SSM: surface SM. RZSM: root -zone SM. LSP: land surface.

to produce the time series of LAI and soil moisture at a daily timestep with better accuracy. However, rather than running the crop model whose states have been updated with the PF until the harvesting season to obtain the yield, the wheat yield was estimated using a linear regression model trained with the LAI and soil moisture produced by data assimilation and field-measured yield. To further understand the performance of applying data assimilation in crop simulation, more investigations using various models and assimilation strategies are needed. There is also a lack of inter-comparison between re-calibration approaches and data assimilation techniques, and among different data assimilation techniques.

2.4.2 Strategies of integrating RS data with predictive models

Remote sensing data have been linked to certain states of crop models according to two strategies (Delécolle *et al.* , 1992): (1) converting radiation information to vegetation properties that are direct states of the crop models with independent inversion of the radiative transfer models (also known as retrieval models), and then assimilating those state variables; and (2) coupling crop models with radiation transfer models to directly assimilate radiation information from the satellite.

2.4.2.1 Using retrieved state variable products

Various variables can be retrieved from remote sensing data, including leaf area index (LAI), surface soil moisture (SSM), land surface temperature, vegetation water content (VWC), biomass, canopy nitrogen status and phenology stages, etc. Among all the satellite remote sensing products, the most prevailing observations integrated with crop simulation models is LAI due to its strong impact on crop interception of solar radiation, making it a crucial indicator for photosynthesis productivity and grain yield. Several assimilation studies that focused on LAI are found in Dente *et al.* (2008); Curnel *et al.* (2011); Zhao *et al.* (2013); Huang *et al.* (2016); Li *et al.* (2017b).

Apart from the LAI, water condition is another key factor impacting crop yield, and is likely to be more promising to be assimilated in crop modelling for yield estimation than LAI. In order to compensate for the uncertainty caused by rainfall data, soil water index (SWI) was assimilated into a distributed WOFOST model (De Wit & Van Diepen, 2007). Curnel *et al.* (2011) suggested that the assimilation of soil moisture is likely to be less

sensitive to phenology stages compared to the LAI assimilation. Moreover, Nearing *et al.* (2012) compared the assimilation performance of LAI and SM in an OSSE, and reported a better yield estimation performance of soil moisture than LAI. Joint assimilation of both LAI and SM was performed by Ines *et al.* (2013); Xie *et al.* (2017), with better improvement in simulated yield was observed when soil moisture and LAI were assimilated together (Ines *et al.* , 2013).

2.4.2.2 Using radiative transfer or microwave model

Apart from using state variables retrieved from remote sensing, another way to link remote sensing information to the models is through radiative transfer models or microwave models, to allow the direct use of optical reflectance and microwave data. Such coupled crop-radiative transfer or -microwave models were initially employed in re-calibration processes. Bouman (1995) re-calibrated the SUCROS crop model using optical reflectance and X-band radar backscatter, showing that the errors in canopy biomass estimation were decreased with optical information. In the study of Prévot *et al.* (2003), the STICS crop model was coupled with a first-order radiative transfer model to assimilate radar information, and a "multi-layer and multi-element" version of the turbid model SAIL was coupled with PROSPECT to assimilate optical information. Thorp *et al.* (2012) coupled the DSSAT-CSM with two radiative transfer models, Choudhury (Wiegand *et al.* , 1986) and a PROSAIL radiative transfer model (Jacquemoud & Baret, 1990), and employed a model inversion approach to improve wheat yield estimation.

Coupling models were also used in advanced data assimilation attempts. Vazifedoust *et al.* (2009) coupled the SUCROS crop model with the SAIL radiation transfer model, and assimilated LAI and relative evapotranspiration with the Kalman Filter. Ma *et al.* (2013b) assimilated fine resolution NDVI data from the Chinese HJ-1A/B into the WOFOST crop model, which was coupled with a two-layer canopy reflectance model (ACRM) to capture the key spatial and temporal variability in the growing season. Active and passive microwave observations were assimilated into the DSSAT model by Liu *et al.* (2016). In this study, the microwave backscatters were assimilated to update the total dry biomass, and the brightness temperature was assimilated to update soil moisture.

2.4.3 Practices at the field and regional scales

While the majority of data assimilation studies illustrated the effectiveness of crop model data assimilation in predicting yield at the regional scale, the performance of yield prediction was only evaluated with the official total yield over counties and not for the individual value of each pixel. The pixels in coarse-resolution remote sensing images usually contain a mixture of different vegetation types, or vegetation and bare soil. When it comes to the field scale, information with a higher spatial resolution is required for application in individual paddocks. However, assimilation studies using fine-resolution satellite data are rarely found. The requirements of precision agriculture for estimating the spatial variability of yield at field scale and fine resolution is not addressed. Thus it is necessary to make use of fine-resolution remote sensing data and simulated crop growth to estimate yield at the field scale.

2.5 Knowledge gaps and research questions

In reviewing the literature, knowledge gaps were identified as:

1. Various remote sensing observations of crop state variables (such as LAI, SM, VWC and phenological stages) can be assimilated into a crop simulation model with a sequential data assimilation method, while only LAI and SM have been explored in existing studies. A complete understanding of the assimilation performance of various crop states is lacking.

2. Most assimilation studies have focused on predicting total crop yield in regional scale over the area. To meet the requirement of precision agriculture, it is necessary to make use of fine-resolution remote sensing data and simulate crop growth and estimate yield at the field scale.

Based on the knowledge gaps presented above, three research questions were proposed□

1. Does the assimilation of all prognostic wheat and/or soil state variables in a synthetic study results in any improvement for yield estimation of APSIM-Wheat?

2. By making use of in-situ observations of wheat and soil states, does the assimilation

of state variables in a case study results in any improvement for yield estimation of APSIM-Wheat?

3. By making use of available remote sensing optical and passive microwave data, does the assimilation of state variables in a case study results in any improvement for yield estimation of APSIM-Wheat?

The hypothesis behind the research questions are listed below:

1. The data assimilation of observations of wheat and/or soil state variables in the synthetic study can help understand the constraint to which model state variable potentially benefits APSIM-Wheat yield prediction.

2. The uncertainty of APSIM-Wheat yield prediction is primarily caused by weather input and model parameters, and the uncertainties from model and field observations could be well estimated.

3. By replacing field observation of wheat and soil state variables with remote sensing data, it brings additional challenges and opportunities to the implementation of data assimilation.

2.6 Chapter summary

This chapter introduced the background and current progress in the literature related to this thesis. Against a background of growing world food demand, reliable crop monitoring, and yield prediction is necessary as it plays a vital role in guiding field management. Accordingly, crop modelling and remote sensing data must be merged with data assimilation techniques so as to provide a spatial and temporal understanding of crop growth for farmers in order to meet the increased demand. Based on the review of existing literature, three research questions were proposed to be answered at the completion of this research.

Chapter 3

APSIM model and the EnKF algorithm

This chapter describes the methodology of this thesis. A wheat modelling-data assimilation framework was built to allow external field and remote sensing observations of wheat and soil sates to be assimilated into the APSIM wheat model system. APSIM is a well-performing and widely-used and validated Australian model that is ideal for modelling experiments in Australia. The model is under active development, and the latest version involves more than 20 types of crop and considers the weather, soil water and chemical dynamics, and management factors to help growers to monitor and predict the in-season crop growth with maximum flexibility. The Ensemble Kalman filter is a widely-used data assimilation algorithm designed for nonlinear systems. Consequently, this chapter gives an introduction to the structure and key features of the APSIM-Wheat model, the EnKF data assimilation algorithm, and the development of the coupled APSIM-EnKF framework.

3.1 APSIM wheat modelling

APSIM is a software framework that integrates a range of crop modules and allows coupling with other modules Keating *et al.* (2003). This study uses a version equivlent to APSIM 7.5, but was newly transplanted to the APSIMX (APSIM Next Generation) platform without any changes to the model physics when this study began. The architecture of the new platform is suitable to be coupled with any state-updating data assimilation algorithm.

The APSIM-Wheat model simulates crop growth at a daily time step. Weather data are input to the model, including precipitation, solar radiation, daily maximum and minimum air temperature, vapour pressure and wind speed. Several fundamental modules are required to assemble a crop simulation model, including:

1. one or more crop modules to simulate the crop development on a daily time step;
2. soil modules to simulate the soil water and nitrogen balance and exchange with the plant and the atmosphere;
3. weather and management modules;
4. a set of infrastructure modules controlling the sequence of simulations and the reading/writing of input/output files.

This study uses a default assembly for wheat modelling, with main modules of Wheat, SoilWat (the default soil water module), SoilN and other supporting modules. A schematic diagram of the APSIM structure is shown in Figure 3.1.

3.1.1 Wheat module

The description of the Wheat module in this section was adapted according to the APSIM-Wheat documentation (Zheng *et al.* , 2014). In the Wheat module, plant organs are simulated individually through the partitioning of the daily income biomass. The concept of wheat organs is defined by:

1. the root (the underground part of the plant);
2. the leaf (accounts for leaf blades only);
3. the stem (stem and leaf sheaths);
4. the pod (wheat spike without the grain); and,
5. the grain (meal part of wheat spike only).

In addition to the individual names of each organ, the wheat spike is the combination of the pod and grain, called the head in the model. The above-ground part of the plant is the

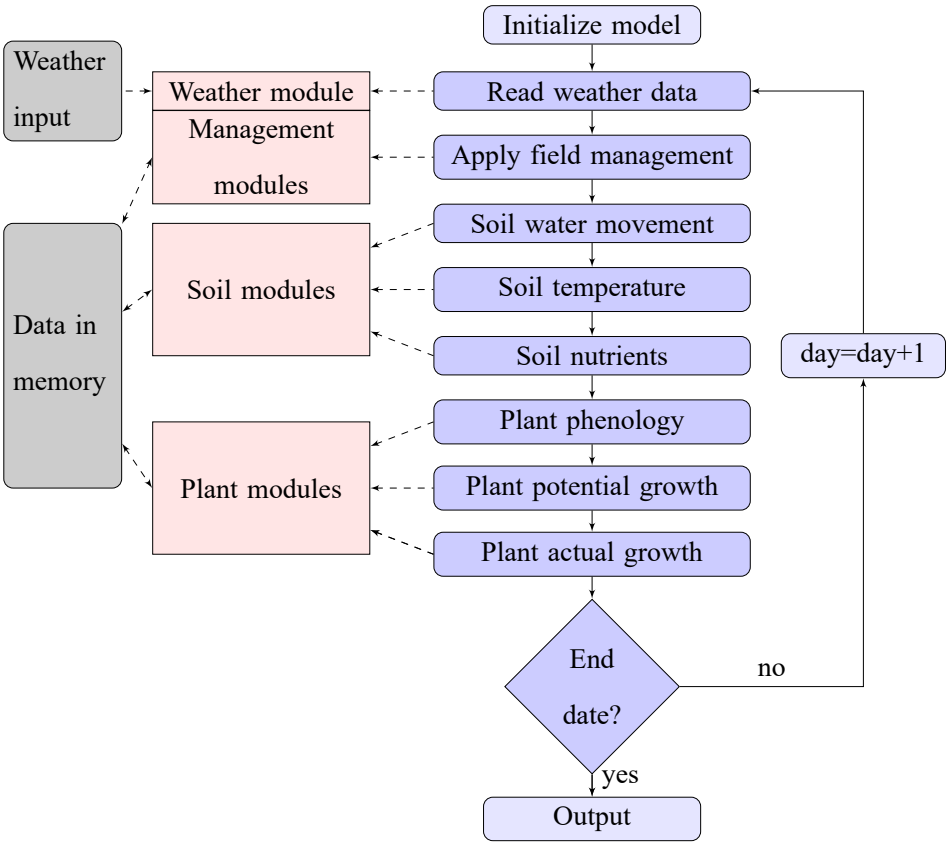

Figure 3.1: The APSIM structure.

combination of the leaf, stem and head, while the below-ground part is the root only. The Wheat module simulates the development of phenology, the accumulation and partitioning of biomass, leaf development, grain development processes under the effect of stress factors and the substance and energy exchanges between the plant and other modules.

3.1.1.1 Phenology

The development of wheat phenology is driven by thermal time (TT), similar to a more widely-used concept growing degree-day, and defined as the accumulation of temperature over time. A simple prognostic equation is:

$$TT_{t+1} = TT_t + \triangle TT, \tag{3.1}$$

with subscripts representing the states at timestep (in days) t and $t + 1$, respectively. The daily increment of thermal time ($\triangle$TT) is calculated as a function of the crown temperature

Table 3.1: Wheat phenology phases and stages in APSIM and Zadoks scale. The phenology stage represents the period between two phenology phases.

APSIM phase	APSIM stage	Phenology phase and description in Zadoks scale	Typical target TT (°C · day)
1 Sowing	1 Sowing to germination	00 Dry seed.	-
2 Germination	2 Germination to emergence	05 Radicle emerged from caryopsis.	-
3 Emergence	3 Emergence to end of juvenile	10 First leaf trough coleoptile.	-
4 End of juvenile	4 End of juvenile to floral initiation	10 First leaf trough coleoptile.	380-410
5 Floral initiation	5 Floral initiation to flowing	39 Flag leaf ligule visible to 40 booting.	520-630
6 Flowering	6 Flowering to start of grain filling	65 Anthesis half-way.	120-125
7 Start of grain filling	7 Grain filling (or start of grain filling to end of grain filling)	71 Caryopsis water ripe.	580-610
8 End of grain filling	End of grain filling to maturity	87 Hard dough.	-
9 Maturity	9 Maturity to harvest	90 Ripening.	-
10 Harvest	-	-	-

T_c (calculated as the average of the daily maximum and minimum air temperature) as:

$$\triangle \mathrm{TT_p} = \begin{cases} T_c, & 0 < T_c < 26 \\ 26/8 \times (34 - T_c), & 26 < T_c < 34 \\ 0, & T_c < 0, \ or \ T_c > 34 \end{cases}, \tag{3.2}$$

and modified by photoperiod, vernalization and environmental factors as:

$$\triangle \mathrm{TT} = \triangle \mathrm{TT_p} \times \mathrm{factor}(\mathrm{photoperiod},\ \mathrm{vernalization},\ \mathrm{environment}). \tag{3.3}$$

In the wheat module, 10 phenology phases (9 stages) are defined from sowing to harvest. The wheat phenology stages in APSIM can be linearly mapped to the more widely-used Zadoks phenology scale (Zadoks *et al.* , 1974) according to Table 3.1. In each phenology stage, a certain threshold (target thermal time) of in-phase accumulated thermal time is required to be reached, to allow the plant in one phase to transfer to the next phase. The target thermal time required for each stage varies among wheat cultivars with typical values shown in Table 3.1.

3.1.1.2 Light utilization and biomass partition

While some models use a physically-based photosynthesis-respiration approach (i.e., APES, DAISY, MONICA, WOFOST, etc.), APSIM-Wheat uses a simple radiation use efficiency approach to calculate incoming biomass accumulation (Zheng *et al.* , 2014). The above-ground dry biomass (hereafter referred to as "biomass" Q) is accumulated on a daily basis through photosynthesis. The daily biomass increment is calculated from the intercepted radiation (I) and the radiation use efficiency (RUE) under stress factors according to:

$$\triangle \mathrm{Q} = \mathrm{I} \times \mathrm{RUE} \times \mathrm{factor}(\mathrm{T},\ \mathrm{N},\ \mathrm{CO_2}), \tag{3.4}$$

where the actual biomass increment is reduced by stress factors of temperature (T), nitrogen (N), carbon dioxide (CO_2) concentration and water, and the intercepted radiation is a function of the LAI.

The daily incoming biomass is partitioned into the above-ground part to meet the demand of wheat organs in the sequence of grain, pod, leaf and stem. Specifically, the biomass is firstly partitioned to meet the demand of the head as:

$$\triangle \mathrm{Q_{head}} = \min(\triangle \mathrm{Q_{head}},\ \mathrm{D_{head}})\ , \tag{3.5}$$

and the biomass partitioned to the head is further allocated to grain and pod proportional to their individual demand, where D_{head} is the biomass demand of the head. A proportion of the remaining biomass is partitioned to the leaves by a phenology-dependent proportion F_{leaf} as:

$$\triangle \mathrm{Q_{leaf}} = (\triangle \mathrm{Q} - \triangle \mathrm{Q_{head}}) \times \mathrm{F_{leaf}}\ , \tag{3.6}$$

after which the remaining biomass is allocated to the stem as:

$$\triangle \mathrm{Q_{stem}} = \triangle \mathrm{Q} - \triangle \mathrm{Q_{head}} - \triangle \mathrm{Q_{leaf}}\ . \tag{3.7}$$

The daily root biomass increment is simply computed separately as a fixed proportion of the daily above-ground biomass. Re-translocation happens during the grain filling stage; when the daily biomass is insufficient to meet the grain demand, a proportion of biomass in

the pod and stem is re-translocated to the grain.

3.1.1.3 Grain demand

The incoming biomass is allocated to the grain during the grain filling stage to meed the grain demand. The grain demand is determined by the estimated grain number:

$$D_{grain} = N_g \times R_p \times \text{factor}(T,\ N), \tag{3.8}$$

where R_p is the potential rate of grain filling specified as a cultivar parameter, and N_g is the number of grains per plant calculated as:

$$N_g = R_g W_s\ , \tag{3.9}$$

where R_g is the number of grain per gram stem, specified as a cultivar parameter, and W_s is the dry weight of stem at anthesis.

3.1.1.4 Leaf development

The estimation of the leaf area index (LAI) in the wheat module is governed by the prognostic equation:

$$LAI_{t+1} = LAI_t + \triangle LAI\ , \tag{3.10}$$

where the daily increment of LAI is taken from the minimum increment from two approaches: a carbon-limited LAI_c, and a stressed LAI_s:

$$\triangle LAI = \min(\triangle LAI_c,\ \triangle LAI_s)\ , \tag{3.11}$$

where the carbon-limited LAI increment is computed by the increment of the dry leaf weight ($\triangle Q_{leaf}$, computed by Equation 3.6) and a maximum specific leaf area (SLA_{max}), calculated as:

$$\triangle LAI_c = \triangle Q_{leaf} \times SLA_{max}\ . \tag{3.12}$$

The stressed LAI increment ($\triangle LAI_s$) is determined by a leaf-node development approach considering the water and nitrogen stress.

3.1.1.5 Stress factors

APSIM considers stress factors in terms of water and nitrogen, mainly applied to the phenology, photosynthesis, leaf and grain development processes. The computation of these factors is described in the sections below.

Water stress: The water stress of the phenology is calculated as:

$$\text{factor}_{\text{W,pheno}} = \text{h}_{\text{W,pheno}}(\frac{\text{esw}_\text{a}}{\text{esw}_\text{p}}) \ , \tag{3.13}$$

where esw_a and esw_p are the actual and potential extractable soil water, respectively, and $\text{h}_{\text{W,pheno}}(\centerdot)$ is a function relating soil water availability to the soil water stress of the phenology. The water stress of the photosynthesis is calculated as:

$$\text{factor}_{\text{W,photo}} = \frac{\text{W}_\text{u}}{\text{W}_\text{d}} \ , \tag{3.14}$$

where W_u is the total daily water update from the root, and the W_d is the soil water demand of leaf and head. The water stress of the leaf expansion is determined by:

$$\text{factor}_{\text{W,expan}} = \text{h}_{\text{W,expan}}(\frac{\text{W}_\text{u}}{\text{W}_\text{d}}) \ , \tag{3.15}$$

where $\text{h}_{\text{W,expan}}(\centerdot)$ is a function relating soil water availability to the soil water stress of leaf expansion.

Nitrogen stress: The nitrogen stress of the phenology is calculated as:

$$\text{factor}_{\text{N,pheno}} = \text{h}_{\text{N,pheno}} \times \sum_{\text{stem, leaf}} \frac{\text{C}_\text{N} - \text{C}_{\text{N,min}}}{\text{C}_{\text{N,crit}} \times \text{f}_{\text{c,N}} - \text{C}_{\text{N,min}}} \ , \tag{3.16}$$

where $\text{h}_{\text{N,pheno}}$ is a multiplier for the nitrogen deficit effect on the phenology, C_N is the nitrogen concentration of the stem and the leaf, $\text{C}_{\text{N,crit}}$ and $\text{C}_{\text{N,min}}$ are the nitrogen concentration critical and minimum values for the plant parts, respectively, and $\text{f}_{\text{c,N}}$ is a factor for CO_2

effects on the leaf. The nitrogen stress of photosynthesis is calculated as:

$$\text{factor}_{\text{N,photo}} = \text{h}_{\text{N,photo}} \times \sum_{\text{leaf}} \frac{\text{C}_{\text{N}} - \text{C}_{\text{N,min}}}{\text{C}_{\text{N,crit}} \times \text{f}_{\text{c,N}} - \text{C}_{\text{N,min}}} , \tag{3.17}$$

where $\text{h}_{\text{N,photo}}$ is a multiplier for nitrogen deficit effect on photosynthesis. The nitrogen stress of leaf appearance and development is calculated as:

$$\text{factor}_{\text{N,expan}} = \text{h}_{\text{N,expan}} \times \sum_{\text{leaf}} \frac{\text{C}_{\text{N}} - \text{C}_{\text{N,min}}}{\text{C}_{\text{N,crit}} \times \text{f}_{\text{c,N}} - \text{C}_{\text{N,min}}} , \tag{3.18}$$

where $\text{h}_{\text{N,expan}}$ is a multiplier for nitrogen deficit effect on photosynthesis. The nitrogen stress of grain filling is calculated as:

$$\text{factor}_{\text{N,grain}} = \frac{\text{h}_{\text{N,poten}}}{\text{h}_{\text{N,min}}} \times \text{h}_{\text{N,grain}} \times \sum_{\text{stem, leaf}} \frac{\text{C}_{\text{N}} - \text{C}_{\text{N,min}}}{\text{C}_{\text{N,crit}} \times \text{f}_{\text{c,N}} - \text{C}_{\text{N,min}}} , \tag{3.19}$$

where $\text{h}_{\text{N,grain}}$ is a multiplier for nitrogen deficit effect on photosynthesis, while $\text{h}_{\text{N,poten}}$ and $\text{h}_{\text{N,min}}$ are the potential and minimum rate of grain filling, respectively.

3.1.2 Soil modules

The crop module is coupled with soil modules that account for the dynamics of soil water, carbon, nitrogen, organic matter and temperature in multiple layers (7 layers by default). SoilWat, SoilN and SoilTemp are the main soil modules in the APSIM model and are described below according to the APSIM model documentation.

3.1.2.1 SoilWat

SoilWat is a cascading water balance model of APSIM developed from CERES (Jones *et al.* , 1986) and PERFECT (Littleboy *et al.* , 1992). This model calculates infiltration, evapotranspiration, soil evaporation, saturated and unsaturated flow, drainage, backwater and surface ponding in the vertical direction and runon/runoff, and lateral inflow/outflow in the horizontal direction when invoked by APSIM at a daily time step. The soil water scheme balance is shown in Figure 3.2.

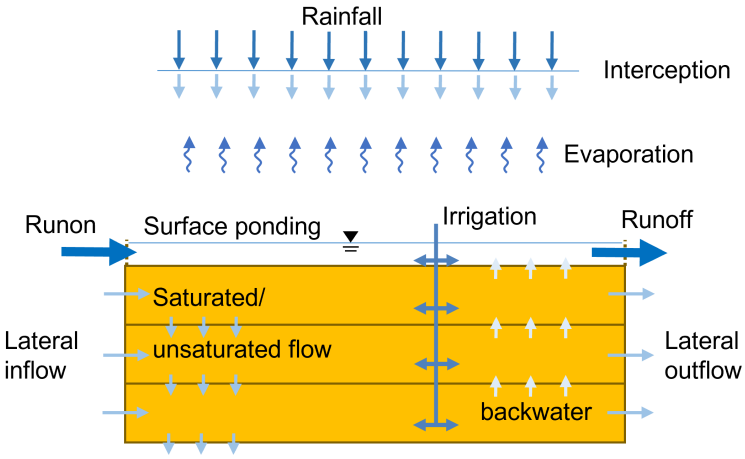

Figure 3.2: The APSIM soil water scheme for an instance of 3 soil layers.

Runoff: Runoff is simulated using the curve number technique of the USDA Soil Conservation Service (Soil Conservation Service, 2004) procedure. This technique estimates the surface runoff of bare soil as a function of daily rainfall using a group of runoff response curves numbered between 0 (no runoff) and 100 (all runoff). The interception is considered by the reduction of curve number with the increase of canopy cover throughout the plant growing process.

Potential evapotranspiration and soil evaporation: The potential evapotranspiration (PET) uses the equilibrium evaporation concept of Priestley & Taylor (1972). Soil evaporation is simulated in two stages, as a combination of two processes described in two crop models, CERES and PERFECT, respectively. In the first stage (the constant stage when $t \leq t_{1,}$), the soil is assumed sufficiently wet to supply the evaporation rate to be equal to the PET. The second stage (the falling rate stage when $t > t_1$) starts when the soil moisture decreases to a certain threshold value, and the evaporation rate is slower than the PET. The two processes are described by two parameters U (from CERES) and ConA (from PERFECT):

$$\sum \mathrm{E_s} = \begin{cases} U \times t, & t \leq t_{1,} \\ U \times t + ConA \times \sqrt{t - t_1}, & t > t_1 \end{cases}, \tag{3.20}$$

where the parameter U is the amount of cumulative evaporation before soil supply decreases below the atmospheric demand (Jones *et al.* , 1986), and the parameter ConA is a regression coefficient that specifies the change of cumulative evaporation in the second stage (Littleboy

et al. , 1992). Different soil evaporation parameters can be specified to summer and winter days switching on a user-specific date.

Saturated water flow: The saturated water flow occurs when the soil water content (SW), or soil moisture (SM), is below the saturated (Sat) soil moisture and above the drained upper limit (DUL). A fraction (specified by SWCON) of the water is allowed to drain into the next deeper layer:

$$\text{Flow}_{\text{sat}} = \text{SWCON} \times (\text{SW} - \text{DUL}) \, . \tag{3.21}$$

Unsaturated water flow: The unsaturated water flow occurs when the soil water content is below the DUL. The movement of soil water depends on the water content gradient between adjacent layers and the diffusivity. The diffusivity is given by:

$$\text{Diffusivity} = \text{DiffusConst} \times \exp(\text{DiffusSlope} \times \theta_{\text{av}}) \, , \tag{3.22}$$

where the DiffusConst and the DiffusSlope are two parameters, and θ_{av} is the average of extractable soil water (the difference between the SW and the permanent wilting point LL) across the two layers. The unsaturated water flow is then calculated by:

$$\text{Flow}_{\text{unsat}} = \text{Diffusivity} \times \text{volumetric soil water gradient} \, . \tag{3.23}$$

Lateral outflow: Lateral outflow is the horizontal flow that occurs when the soil water exceeds DUL, with the soil being on a slope. The lateral outflow is calculated as:

$$\text{Flow}_{\text{lat}} = \text{K}_{\text{lat}} \times \text{d} \times \frac{\text{s}}{\sqrt{1 + \text{s}^2}} \times \frac{\text{L}}{\text{A}} \, , \tag{3.24}$$

where K_{lat} is the lateral conductivity, d is the depth of saturation in the layer (mm), s is the slope (m/m), L is the catchment discharge length (m), and A is the catchment area (m^2).

3.1.2.2 SoilN

The movement of soil organic matter, nitrogen and solute is simulated by a combination of closely linked soil sub-modules (SoilN, SoilOM, Residue, Solute, etc.) that are collect-

ively called the SoilN module in this chapter hereafter. The key features of the SoilN module are described below with a full module description available from Probert *et al.* (1998).

As shown in the SoilN scheme (Figure 3.3), soil organic matter (organic carbon and nitrogen) is simulated as two pools: biom and hum. The biom pool represents the microbial biomass in soil, while the hum pool contains the humic organic matter. Simulation of soil organic matter is based on the carbon flux between pools, while the nitrogen flux is calculated as a proportion of the carbon flux by the carbon-nitrogen ratio of the receiving pool.

Organic matter residue: The residue produced by the plant can be burnt, joined into fresh organic matter (the fom pool) by tillage, or be decomposed into the biom and the hum pools. The scheme of soil organic matter dynamics is shown in Figure 3.3. To initialize the organic matter pools, the initial total carbon is calculated in kg/ha as:

$$\text{total carbon} = \text{fom}_c + \text{oc} \ , \tag{3.25}$$

where the fom_c is the fresh organic matter carbon that is initialized at the beginning of the simulation, and the oc is the organic carbon, calculated as:

$$\text{oc} = \text{biom}_c + \text{hum}_c \ . \tag{3.26}$$

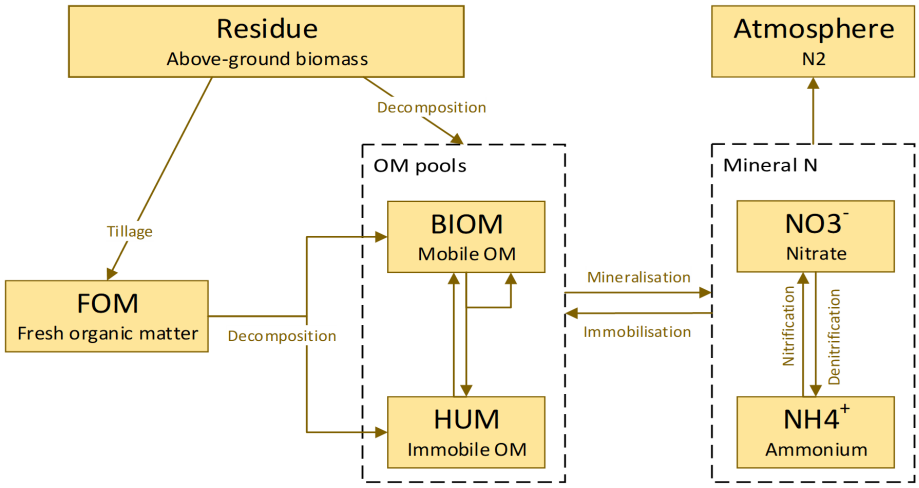

Figure 3.3: APSIM soil organic matter and nitrogen scheme (source: figure reproduced from Probert *et al.* , 1998).

To simulate the slower rate of decomposition of soil organic matter in the deeper soil layer, a part of the biom and hum pool is considered to be non-susceptible to decomposition, determined by two multipliers F_{biom} and F_{inert}, respectively. They are calculated as:

$$inert_c = F_{inert} \times oc \ . \tag{3.27}$$

Since $hum_c = oc - biom_c$, the $biom_c$ is determined by:

$$biom_c = F_{biom} \times \frac{oc - inert_c}{1 + F_{biom}} \ . \tag{3.28}$$

Decomposition: Soil organic matter in all pools are decomposed, and the products of decomposition are received by the biom and the hum pools, as shown in the left panel of Figure 3.3. Generally, the decomposition of each pool is determined by the decay rate and stress factor:

$$Decomposition_{pool} = substance_{pool} \times decay\ rate_{pool} \times factor_{pool}, \tag{3.29}$$

where the subscript pool represents the fom, biom and hum pool, respectively. For the fom pool, the substances are carbohydrate, cellulose or lignin fractions, the stress factor f_{fom} is a function of soil water (W), soil temperature (T) and carbon:nitrogen ratio (CNR), expressed as $factor_{fom}(W, T, CNR)$. For the biom and the hum pools, the substance is the organic carbon and nitrogen to be decomposed, and the stress factor is a function of soil water and soil temperature expressed as $factor_{biom\ or\ hum}(W, T)$. The decomposition rate is individually specified for each pool and substance.

Mineralization and immobilization: Mineralization transforms N in organic matter into mineral N (i.e., ammonium N (NH_4^+), and nitrate N (NO_3^-)), while immobilization is the reverse of mineralization. Mineralization and immobilization occur in all pools, depending on the balance of mineral N released in decomposition and immobilized in microbial synthesis and humidification (Probert *et al.* , 1998). When the availability of mineral N is adequate to supply the N demand for immobilization, the rate of decomposition is slowed.

Nitrification: Nitrification transfers ammonia into nitrates. Nitrification follows Michaelis-Menton kinetics (Godwin & Jones, 1991), depending on NH_4^+ concentration and affected

by sub-optimal water, temperature and pH conditions. Nitrous oxide (N_2O) is emitted as a product of nitrification and is proportional to the nitrided N.

Denitrification: The rate of denitrification is determined by the concentration of NO_3^- and the active carbon in the soil. Nitrous oxide is also the product of denitrification, calculated by the by combining predictions of denitrification with the ratio of N_2 to N_2O emitted during denitrification, according to the model of Del Grosso *et al.* (2000).

3.1.2.3 SoilTemp

The SoilTemp module estimate soil temperature based on the CERES soil temperature model, with the latter developed from EPIC (Williams *et al.* , 1983). Based on simple empirical assumptions, this approach calculates soil temperature with only two additional inputs required: the annual average ambient temperature (TAV) and the annual amplitude in mean monthly temperature (AMP).

The profile of subsurface soil temperature is determined by the change in soil temperature since the hottest day of the year, and a factor that is derived from the depth at which no diurnal variation in temperature is experienced (known as damping depth). For each given layer i, the average soil temperature is governed by:

$$\mathrm{ST(i)} = \mathrm{TAV} + (\frac{\mathrm{AMP}}{2} \times \cos(\mathrm{ALX} + \mathrm{DepthLag}) + \mathrm{TempChange} \times \exp(\mathrm{DepthLag})) \ , \tag{3.30}$$

where ALX is a variable to relate the current day of the year to the time of the hottest day of the year, TempChange is the change in soil temperature since the hottest day of the year, and DepthLag is a depth lag factor that is determined by the accumulated depth of soil layer, and a damping depth, which is further a function of soil bulk density and soil water availability.

3.1.3 State variables and parameters

The state variables of APSIM-Wheat are presented in Table 3.2, while the key parameters and initial conditions of APSIM-Wheat are presented in Tables 3.3 - 3.4.

Table 3.2: Wheat and soil states in the APSIM-Wheat model.

State variable	Unit	P/D[a] ?	Measurements	
			Field observation	Remote sensing
Canopy height	m	P	Tape measure, Ground-based Lidar	Lidar
Canopy cover	-	D	Visual estimation, photographic method	Spectral reflectance
Accumulated thermal time	$^{o}C \cdot d$	P	NA[b]	NA
Phenological stages	-	D	Direct observation	Spectral reflectance
Leaf area index	m^2/m^2	P	Leaf area meter, link to dry biomass	Spectral reflectance
Leaf Wt	g/m^2	P	Sampling	NA
StemWt				
Pod Wt				
GrainWt				
RootWt				
LeafN	g/m^2	P	Sampling	NA
StemN				
PodN				
GrainN				
RootN				
Yield	kg/ha	D	Sampling at harvest	NA
Grain size	g	D	Sampling at harvest	NA
Plant N concentration (stem, leaf, pod, grain)	g/g (N/Wt)	D	Sampling	NA
Grain protein	%	D	Sampling at harvest	NA
Number of nodes	-	P	Direct observation	NA
Number of leaves	-	P	Direct observation	NA
Root length	m	P	NA	NA
Root depth	m	P	Sampling	NA
Soil moisture (surface)	m^3/m^3	P	Sampling	Spectral reflectance, radiometer, radar
Soil moisture (root-zone)	m^3/m^3	P	Sampling	NA
Soil nitrate	kg/ha	P	Sampling	NA
Soil ammonium	kg/ha	P	Sampling	NA

[a] P/D: the state variable is prognostic (P) or diagnostic (D).

[b] NA: not available.

Table 3.3: Key parameters in the Wheat module and management information.

Parameter	Description	Unit	Typical values
Wheat module (cultivar parameters)			
VernSens	Sensitivity to vernalization.	-	1.5-5
PhotopSens	Sensitivity to photoperiod.	-	1.4-5
TT4 (stage 4)	Target TT from End of Juvenile to Floral Initiation.	$^{o}C \cdot$ days	380-400
TT5 (stage 5)	Target TT from Floral Initiation to Flowering.	$^{o}C \cdot$ days	520-630
TT6 (stage 6)	Target TT from Flowering to Start of Grain Filling.	$^{o}C \cdot$ days	80-180
TT7 (stage 7)	Target TT from Start of Grain Filling to End of Grain Filling.	$^{o}C \cdot$ days	580-600
Grains Per Gram Stem	Grain number per gram stem.	grain/g	20-27
Max Grain Size	Maximum grain size.	10^{-3}g	39-65
Potential Grain Filling Rate	Potential grain filling rate.	10^{-3}grain/g/d	1-2.8
Potential Grain Growth Rate	Potential grain growth rate.	10^{-3}grain/g/d	1
Potential Grain N Filling Rate	Potential grain Nitrogen filling rate.	10^{-5}grain/g/d	5.5
Sowing depth	Sowing depth.	mm	30
Row spacing	Row spacing.	mm	250
Population	Plant population.	m^{-2}	120
Fertiliser	Fertiliser amount applied when sowing.	kg/ha	160
Management information			
Sowing depth	Sowing depth.	mm	30
Row spacing	Row spacing.	mm	250
Population	Plant population.	m^{-2}	120
Fertiliser	Fertiliser amount applied when sowing.	kg/ha	160

Table 3.4: Key parameters and initial conditions in the SoilWat and the SoilN modules.

Parameter	Description	Unit	Typical values
SoilWat module			
InitialSW	Initial soil water % full relative to the WheatLL.	%	0-100
BD	Bulk density.	g/cm^3	0.8-1.2
LL15	15 bar lower limit of soil moisture.	m^3/m^3	0.25-0.28
AirDry	Soil moisture of air-dry soil.	m^3/m^3	0.13-0.28
DUL	Drained upper limit of soil moisture.	m^3/m^3	0.45-0.52
SAT	Saturated soil moisture.	m^3/m^3	0.52-0.59
WheatLL	Lower limit of wheat extractable soil moisture.	m^3/m^3	0.26-0.45
ConA (summer/winter)	Drying coefficient for stage 2 soil water evaporation.	-	5
U (summer/winter)	Cumulative soil water evaporation to reach the end of stage 1 soil water evaporation.	mm	5
DiffusConst	Constant in the soil water diffusivity calculation.	mm^2/d	40
DiffusSlope	Effect of soil water storage above the lower limit on soil water diffusivity.	mm^{-1}	16
Salb	The fraction of incoming radiation reflected from bare soil (soil albedo).	-	0.12
CN2Bare	Runoff Curve Number for bare soil with average moisture.	-	73
SWCON	Fractional amount of water above DUL that can drain under gravity per day	d^{-1}	0.3
SoilN module			
RootWt	Initial root weight.	kg/ha	1000
RootCN	Root C:N ratio.	-	40
SoilCN	Soil C:N ratio.	-	12
EnrACoeff	Erosion enrichment coefficient A.	-	7.4
EnrBCoeff	Erosion enrichment coefficient B.	-	0.2
OCTotal	Total organic matter.	%	1.2-0.12
Fbiom (F_{biom})	The biom pool carbon as a fraction of the hum carbon that is subject to decomposition (0-1).	-	0.01-0.04
Finert (F_{inert})	The fraction of the hum pool that is considered to be non-susceptible to decomposition (0-1).	-	0.4-1
pH	pH from either a sample or from analysis (1:5 water).	-	8
SurfaceResidue	Mass of surface residue.	kg/ha	500
ResidueCN	C:N ratio of surface residue.	g/g	100
InitialNO3	Initial soil nitrate nitrogen.	ppm	1
InitialNH4	Initial soil ammonium nitrogen.	ppm	0.1

3.2 Data assimilation and EnKF algorithm

A brief description of the Kalman filter (KF) and the Ensemble Kalman filter (EnKF) algorithms was presented in this section according to Stuart & Zygalakis (2015); Marc (2014).

3.2.1 Filtering problem and Kalman filter

For a dynamic system, a Markov chain passes the state of the system at time step $k-1$ into the next step k according to:

$$\mathbf{x}_k = \mathcal{M}(\mathbf{x}_{k-1}), \tag{3.31}$$

where, $\mathcal{M}(.)$ and $\mathbf{x}$ are the state transfer function and the true states of the system, respectively. At any timestep, the observation of the model states $\mathbf{y}$, and the observation are related to the state vector with an observation operator $\mathcal{H}(\mathbf{x})$, expressed by:

$$\mathbf{y} = \mathcal{H}(\mathbf{x}). \tag{3.32}$$

In a real situation, the true states and transfer function are unknown and so a model is used to approximate the true state and transfer function, with an additional term representing model uncertainty $\mathbf{w}_k$:

$$\hat{\mathbf{x}}_k = \mathrm{M}(\hat{\mathbf{x}}_{k-1}) + \mathbf{w}_k, \tag{3.33}$$

where $\mathrm{M}(.)$ is a nonlinear model to forecast the state, and $\hat{\mathbf{x}}$ is the estimated state of the system. The observation operator is estimated by function $\mathrm{H}(.)$ and an additional term representing observation uncertainty $\mathbf{v}$:

$$\mathbf{y} = \mathrm{H}(\mathbf{x}) + \mathbf{v}. \tag{3.34}$$

The purpose of a filtering problem is to solve the probability distribution of $\hat{\mathbf{x}}$ given the external observation of the system $\mathbf{y}$ to the states so that the estimation of the state $\hat{\mathbf{x}}$ approaches the truth $\mathbf{x}$. The Kalman filter (KF) approach uses a series of linear and Gaussian assumptions for the model and uncertainties, respectively. In the KF, a linear forecasting

model and observation operator is assumed such that:

$$\mathbf{x}_{\mathrm{k}} = \mathbf{M}_{\mathrm{k}}\mathbf{x}_{\mathrm{k}-1} + \mathbf{w}_{\mathrm{k}}, \tag{3.35}$$

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{v}, \tag{3.36}$$

where **M** and **H** are the known state transfer and the observation matrices, respectively. The model and observational uncertainties are assumed unbiased, independent and Gaussian, with the error covariances of model and observation $\mathbf{Q}_{\mathrm{k}}$ and $\mathbf{R}_{\mathrm{k}}$:

$$\begin{cases} \mathrm{E}(\mathbf{w}_{\mathrm{k}}) = 0, \quad \mathrm{E}(\mathbf{w}_{\mathrm{k}} \cdot \mathbf{w}_{\mathrm{l}}^{\boldsymbol{T}}) = \mathbf{Q}_{\mathrm{k}}, (\mathrm{k}, \mathrm{l} = 1, 2, ..., \text{and}\, \mathrm{k} \neq \mathrm{l}) \\ \mathrm{E}(\mathbf{v}_{\mathrm{k}}) = 0, \quad \mathrm{E}(\mathbf{v}_{\mathrm{k}} \cdot \mathbf{v}_{\mathrm{l}}^{\boldsymbol{T}}) = \mathbf{R}_{\mathrm{k}}, (\mathrm{k}, \mathrm{l} = 1, 2, ..., \text{and}\, \mathrm{k} \neq \mathrm{l}). \end{cases} \tag{3.37}$$

Implementation of the KF consists of a forecast and an analysis step. The forecast step is a model forecast that maps the analysis state from the previous timestep $\mathrm{k}-1$ to the next step k. This background state is used as the input to the analysis step, otherwise known as the data assimilation step, where external observations are assimilated into the system. The terms "analysis" and "background" are sometimes denoted by "posterior" and "prior", meaning that the state or error covariance is obtained posterior or prior to the analysis step. The forecast step estimates the background states as a direct model estimation based on the analysis states from the previous step, calculated by:

$$\mathbf{x}_{\mathrm{k}}^{\mathrm{b}} = \mathbf{M}_{\mathrm{k}}\mathbf{x}_{\mathrm{k}-1}^{\mathrm{a}}, \tag{3.38}$$

where the superscripts b and a denote background and analysis quantities, respectively. The error between model estimation and the truth is assumed independent, with the error covariance matrix denoted by **P**. The background error covariance is forecast forward by:

$$\mathbf{P}_{\mathrm{k}}^{\mathrm{b}} = \mathbf{M}_{\mathrm{k}}\mathbf{P}_{\mathrm{k}-1}^{\mathrm{a}}\mathbf{M}_{\mathrm{k}}^{\mathbf{T}} + \mathbf{Q}_{\mathrm{k}}. \tag{3.39}$$

The analysis states are calculated by:

$$\mathbf{x}_{\mathrm{k}}^{\mathrm{a}} = \mathbf{x}_{\mathrm{k}}^{\mathrm{b}} + \mathbf{K}_{\mathrm{k}}(\mathbf{y}_{\mathrm{k}} - \mathbf{H}_{\mathrm{k}}\mathbf{x}_{\mathrm{k}}^{\mathrm{b}}), \tag{3.40}$$

where the **K** is known as the Kalman gain, which is solved by minimizing the analysis error covariance. When optimized, the analysis error covariance $\mathbf{P}_{\mathrm{k}}^{\mathrm{a}}$ is:

$$\mathbf{P}_{\mathrm{k}}^{\mathrm{a}} = (\mathbf{I} - \mathbf{K}_{\mathrm{k}}^{*}\mathbf{H}) \cdot \mathbf{P}_{\mathrm{k}}^{\mathrm{b}}, \tag{3.41}$$

with the Kalman gain $\mathbf{K}_{\mathrm{k}}^{*}$ given by:

$$\mathbf{K}_{\mathrm{k}}^{*} = (\mathbf{P}_{\mathrm{k}}^{\mathrm{b}}\mathbf{H}^{\mathbf{T}}) \cdot (\mathbf{H}\mathbf{P}_{\mathrm{k}}^{\mathrm{b}}\mathbf{H}^{\mathbf{T}} + \mathbf{R}_{\mathrm{k}})^{-1}. \tag{3.42}$$

In summary, equation 3.38 and 3.39 are the forecasting equations of the KF algorithm for states and error covariance, and Equation 3.40 and 3.41 are the analysis equations, with the Kalman gain calculated by Equation 3.42.

3.2.2 Ensemble Kalman filter (EnKF)

The Ensemble Kalman filter (EnKF) is a generalized approach of the Kalman filter, developed to address nonlinear models. The EnKF data assimilation algorithm is based on a Monte Carlo assumption, where an ensemble of stochastic models running can be used to approximate the probability distribution of the state. In the forecast stage, the background state of each ensemble i is calculated as:

$$\mathbf{x}_{\mathrm{k}}^{\mathrm{i,b}} = \mathbf{M}(\mathbf{x}_{\mathrm{k-1}}^{\mathrm{i,a}}). \tag{3.43}$$

The ensemble of model state variables with an ensemble size of N is written as a matrix:

$$\mathbf{X} = [\mathbf{x}^{1},\ \mathbf{x}^{2},\ ...,\ \mathbf{x}^{\mathrm{N}}]. \tag{3.44}$$

The background and analysis state of the ensemble are taken as the background and analysis ensemble means $\mu_{\mathrm{k}}^{\mathrm{a}}$ and $\mu_{\mathrm{k}}^{\mathrm{b}}$ in timestep k, expressed by:

$$\begin{cases} \mu_k^b = \boldsymbol{E}[\mathbf{x}_k^{i,b}] = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_k^{i,b} \\ \mu_k^a = \boldsymbol{E}[\mathbf{x}_k^{i,a}] = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_k^{i,a} \end{cases}, \tag{3.45}$$

for background and analysis states of the i^{th} ensemble $\mathbf{x}_k^i$, respectively.

In the analysis step, the background error covariance is calculated by the equation:

$$\boldsymbol{P}_k^b = \frac{1}{N-1}\mathbf{D}_k^b\mathbf{D}_k^{bT}, \tag{3.46}$$

where $\mathbf{D}_k$ is calculated by all $\mathbf{x}$ as background states in timestep k:

$$\mathbf{D}_k^b = [\mathbf{x}^{1,b} - \mu_k^b, \mathbf{x}^{2,b} - \mu_k^b, \cdots, \mathbf{x}^{N,b} - \mu_k^b]. \tag{3.47}$$

The analysis state of the i^{th} ensemble is calculated by:

$$\mathbf{x}_k^{i,a} = \mathbf{x}_k^{i,b} + \mathbf{K}_k(\mathbf{y}_k + \mathbf{v}_k^i - \mathbf{H}\mathbf{x}_k^{i,b}), \tag{3.48}$$

where $\mathbf{v}_k^i$ is the observation error randomly draw from a known Gaussian distribution $N(0, \mathbf{R}_k)$ and the Kalman gain obtained as:

$$\mathbf{K}_k = \mathbf{P}_k^b\mathbf{H}^T(\mathbf{H}\mathbf{P}_k^b\mathbf{H}^T + \mathbf{R}_k)^{-1}. \tag{3.49}$$

In summary, the EnKF algorithm uses Equation 3.43 and 3.46 in the forecast step and Equation 3.48 in the analysis step.

3.2.3 Evaluation of data assimilation results

The outcomes of the data assimilation were evaluated with the root mean square error (RMSE) of the state variables and the relative difference of yield (RD_{yield}, note that the yield refers to the grain weight at harvest). They are expressed as:

$$\text{RMSE} = \frac{1}{L}\sum_{k=1}^{L}(\mathbf{x}_k^{est} - \mathbf{x}_k^{obs}), \tag{3.50}$$

$$\text{RD}_{yield} = \frac{\text{yield}_{est} - \text{yield}_{obs}}{\text{yield}_{obs}}, \tag{3.51}$$

where L is the total time step. The estimated states $\mathbf{x}_k^{est}$ is the analysis ensemble mean for the assimilation run, and the model estimates for the open-loop run. The $\mathbf{x}_k^{obs}$ is the observed (or true in the synthetic study) states at time step k. The yield_{est} and yield_{obs} are the estimated

and observed (or true in the synthetic study) grain weight in kg/ha at the date of harvest, respectively. The value of RD_{yield} close to zero means that the yield was close the truth. An RD_{yield} from a data assimilation experiment less than that from the open-loop indicates that the data assimilation of external observations contributed to a better yield estimation, compared to no observations assimilated. A negative value indicates that the errors of yield estimation were over-corrected.

3.2.4 Coupling the EnKF with APSIM

The data assimilation framework developed for this project is based on the version APSIMX (APSIM Next Generation, https://www.apsim.info/apsim-next-generation/) under development on GitHub in 2016, when the version APSIM 7.8 was migrated to the new version but the core code for the model physics remained unchanged. This EnKF framework is developed as a built-in module of the model, working cooperatively with an external program to generate and perturb input files for the EnKF. This data assimilation framework is extendable to all future plant modules and can be switched to other state updating data assimilation algorithms with minor modification to the source code. A schematic diagram of the EnKF framework is shown in Figure 3.4.

The APSIM-Wheat model simulates the wheat development processes using a set of inputs and parameters, which contribute to the majority of the background uncertainties. The forecasting equation of APSIM-Wheat can be written as:

$$\mathbf{x}_k = \mathcal{M}(\mathbf{x}_{k-1},\ \mathbf{f}_k,\ \theta), \tag{3.52}$$

where the state vector $\mathbf{x}$ consists of 31 wheat and soil states, $\mathbf{f}_k$ is the time-dependent driving force (weather input for the case of APSIM), and θ is a set of cultivar and soil parameters uniform throughout the simulation window. To start an EnKF data assimilation process, first an ensemble of models are generated. In this study, ensembles were generated by adding Gaussian errors to weather, model parameters and initial conditions. The initialization of the ensembles was prepared with a script exterior the model (Figure 3.4, a set of weather, parameter and initial condition data were perturbed to generate three sets of input data for three ensembles) and is read as input data at the beginning of the model simulation.

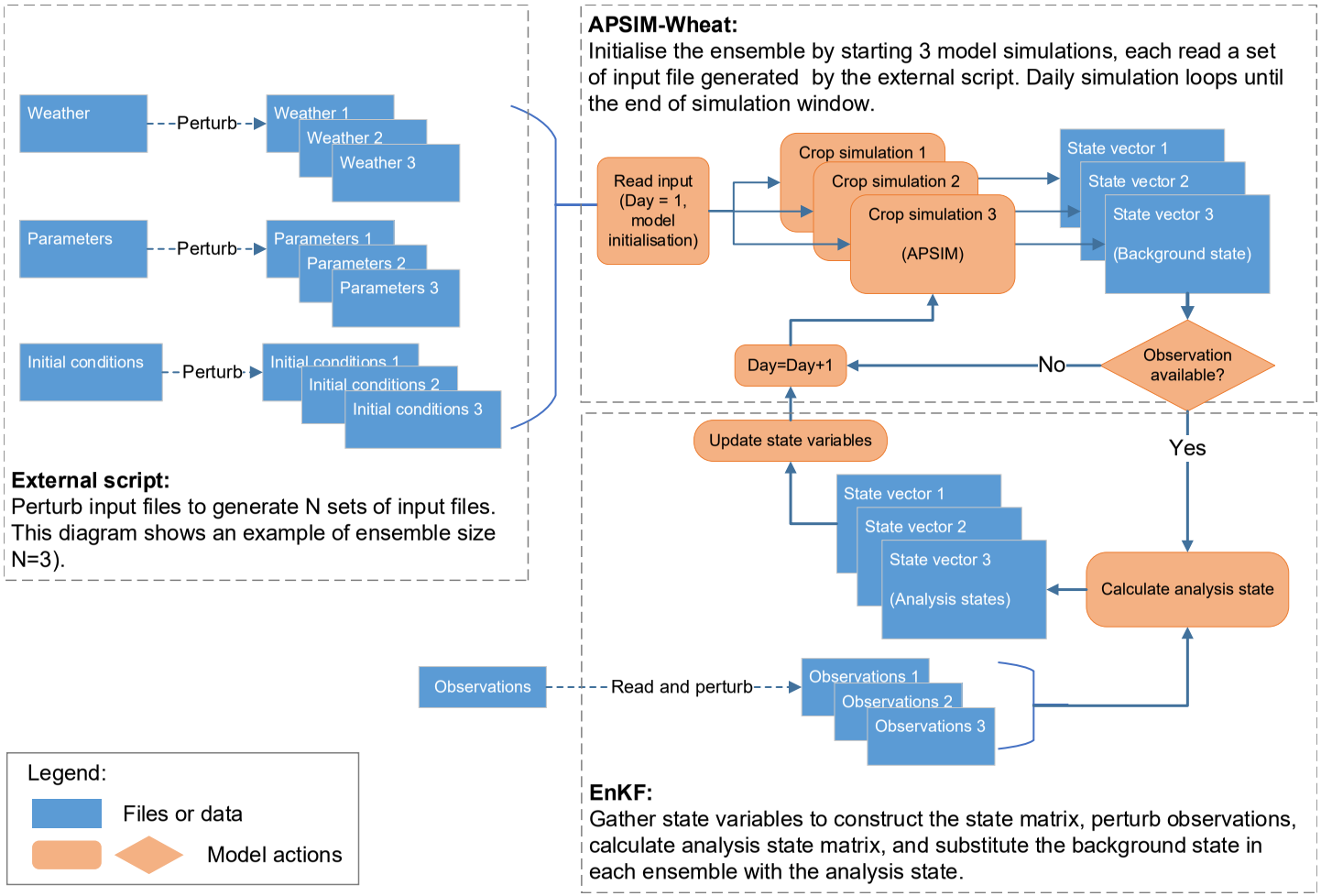

Figure 3.4: Schematic diagram of data assimilation framework coupling EnKF with APSIM. This diagram shows an example with an ensemble size of 3.

After initialization, the ensemble of models are run in parallel and looped at a daily timestep. At the end of each daily simulation, all models are synchronized to make sure that the ensemble of simulations are in the same timestep, after which an observation availability check is performed: if the external observation of the model state is available, the EnKF module is invoked. With the forecast step finished by the daily model simulation, the EnKF module gathers the model state from all ensemble members to construct the background state matrix $\mathbf{X}_k^b$, calculates the background error covariance $\mathbf{P}_k^b$ (Equation 3.46) and the Kalman gain $\mathbf{K}_k$ (Equation 3.49), perturbs observations and calculates the analysis state matrix $\mathbf{X}_k^a$ (Equation 3.48). Once the calculation of the analysis state matrix is finished, the background state variables of all ensemble members are replaced by the analysis state variables. When the state updating process is finished, or external observation is not available, the daily simulation finishes and the model clock moves to the next day.

3.3 Chapter summary

This chapter provided a description of the structure of APSIM wheat and soil modules and an introduction to the EnKF algorithm. Basic characteristics for the estimation of wheat growth is presented, including phenology, photosynthesis, stress factors of water and N, and dynamics of water and organic matter in the soil. For a further understanding of APSIM, it is essential to perform a sensitivity analysis, which is presented in the next chapter. A data assimilation framework for the APSIM model was developed for APSIM-Wheat sub-module using the EnKF data assimilation algorithm, to allow the wheat and soil state variables of the model to be updated by assimilating any external observation to the model system.

Chapter 4

Field experiment and dataset

With the previous chapter introduced the physics of APSIM-Wheat model and summarized the model inputs, wheat and soil states, and outputs, this chapter introduces the dataset collected for this thesis. Input and validation data for APSIM-Wheat were collected in a field experiment from August 2018 to the February 2019 wheat growing season. Weather, soil characteristics and wheat management information were collected as the input data required for APSIM-Wheat. Wheat states and soil variables were collected from both ground and remote sensing measurements throughout the growing season for assimilation into the APSIM model as state variables for improving yield prediction and as validation. Total yield at harvest was also recorded for evaluation of the assimilation results.

4.1 Study site

The site used for this study is located 80 kilometres from Melbourne at Cora Lynn, consisting of an approximately 75×75 m square field with an Antenna Mast Unit (referred to as "AMU tower" or "tower" hereafter, Figure 4.2-a) installed at the south-west corner. The location and layout of the study site is shown in Figure 4.1. The tower consisted of a P-band passive microwave radiometer, TIR, VIS, NIR and SWIR sensors. A weather station (Figure 4.2-b) was installed at the north-west corner of the field. Soil moisture/temperature stations and raingauges were also installed at the site.

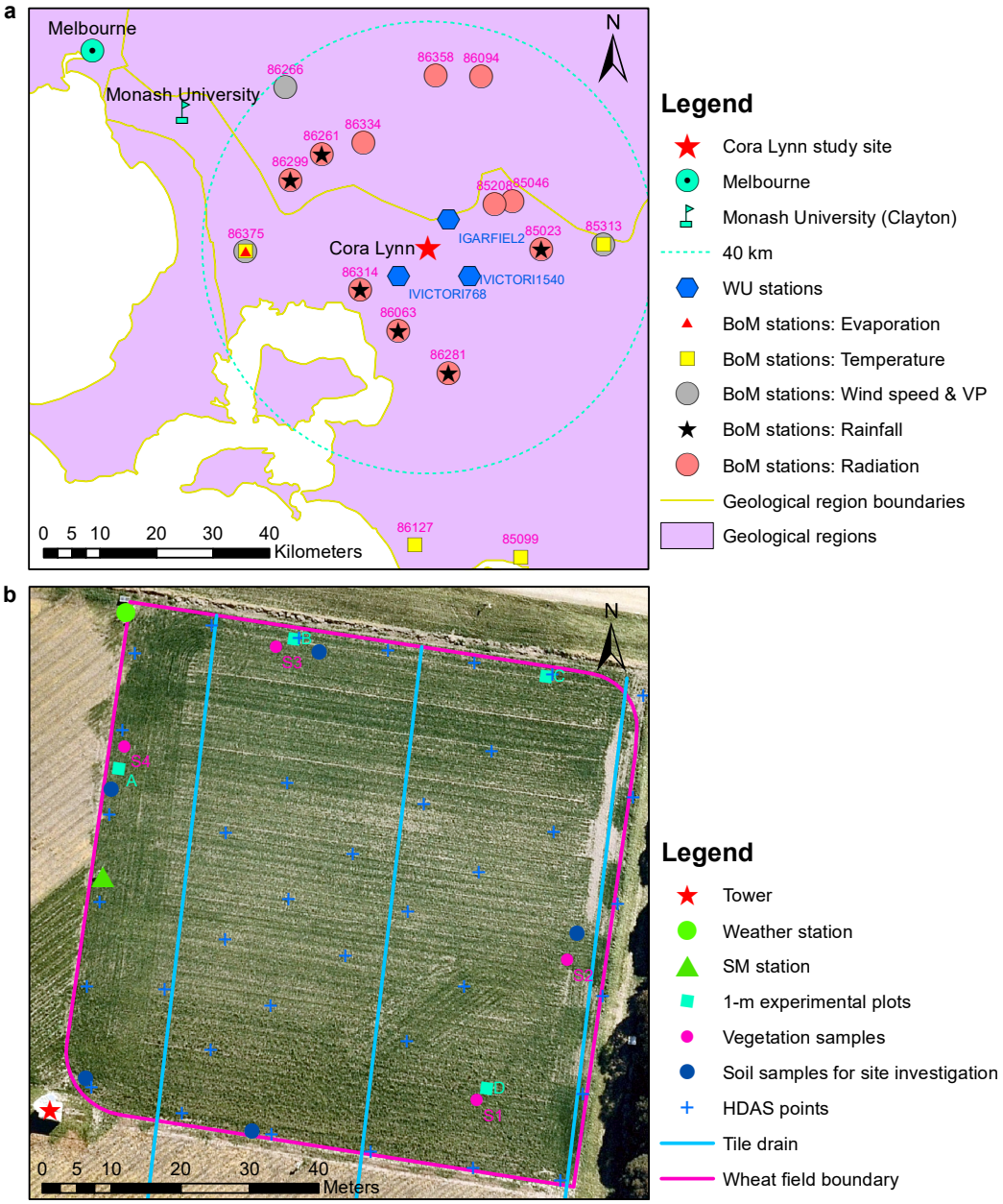

Figure 4.1: Location (a) and layout (b) of the Cora Lynn study site and related data. Field imagery sourced from NearMap (www.nearmap.com).

4.2 Field experiment

Winter wheat seeds (variety: RGT Accroc, produced by Australian Grain & Forage Seeds) were drilled at a 5-cm depth on August 7, 2019 after tillage. Mono-ammonium phosphate (MAP) was evenly applied at 100 kg/ha together with sowing. Emergence was observed on Aug 14. No extra management was applied apart from an irrigation (approx-

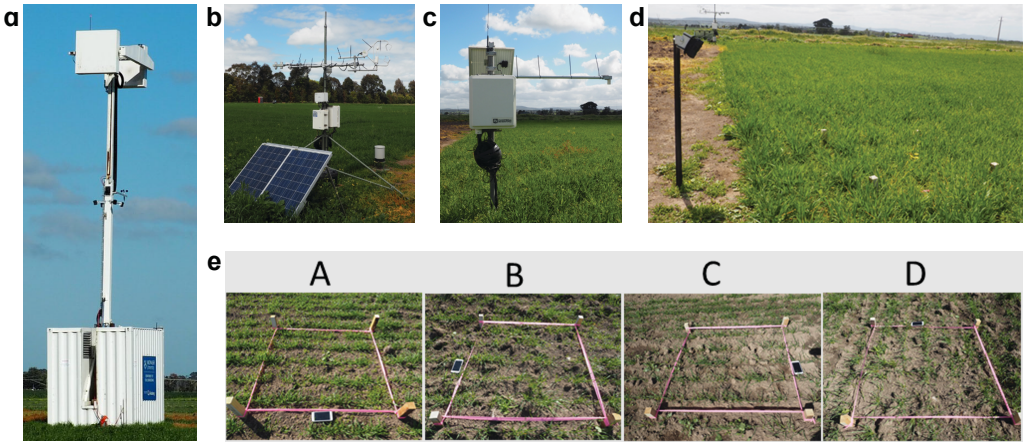

Figure 4.2: In-situ AMU tower (a), Eddy Covariance tower (b), soil moisture station (c), camera (d) and four experimental plots (e).

imately 5 mm) on Nov 16 to maintain wheat growth under unusually dry weather. Four 1×1 m square monitoring areas (plots, labelled A-D, Figure 4.2-e) were marked and circled by pins and tapes, with each plot enclosing 5 rows of wheat at a row spacing 0.2 m. The location of plots (Figure 4.1-b) was carefully selected to ensure that they are uniform with the surrounding area and have no missing plants so that they are representative of the entire paddock, and also to avoid any manual perturbation to the conditions inside the footprint of the tower-based radiometers. A camera (model: Swift 3C High Speed Motion Camera, Figure 4.2-d) was installed 1 m west to the edge of Plot A at the height of 1 m, automatically taking photos of wheat in plot A every 3 hours. Field inspections were scheduled on a weekly basis to obtain wheat height, biomass, LAI, soil moisture and soil nitrogen, and a field investigation conducted to obtain soil properties.

4.2.1 Site characterization

The soil property investigation included the soil texture, permeability, soil water retention curve, bulk density and chemical components of the soil. Table 4.1 provides a summary of the soil characterization from the field investigation.

4.2.1.1 Soil texture and chemical analysis

An intensive soil characterization test was conducted on Sep 21, 2018. Samples were collected at five locations (Figure 4.1-b), each at 6 layers at depths of 0-5, 5-15, 15-25, 25-35,

35-45, 45-55 cm, respectively. To prepare the samples for the texture and chemical analysis, soil samples were dried at room temperature (approx. 25°C) immediately after having been transported to the laboratory. Clay lumps in the soil samples were broken up with pestle and mortar and then separated with a 1.4 mm sieve. Soil particle greater than 1.4 mm was analyzed using a dry sieving method following the procedures of AS 1289.3.6.1:2009 (Standard, 2009), with the fine particle smaller than 1.4 mm analyzed with a diffraction particle size analyzer (model: LS 13 320 Laser Diffraction Particle Size Analyzer) by an external laboratory at the University of Melbourne. To prepare the soil samples for the chemical analysis, a representative sub-sample of approximately 15 grams was separated from each dry soil sample with riffle boxes customized according to AS 1141.2:2015 (Standard, 2015), and then ground with pestle and mortar to pass the 1.4 mm sieve. A full chemical analysis for agricultural soil was conducted by Environmental Analysis Laboratory, Southern Cross University. Key analysis results are shown in Table 4.1.

4.2.1.2 Permeability

Saturated hydraulic conductivity (permeability) was measured in the field with a Guelph permeameter (model: 2800K1 Guelph Permeameter). Measurements were taken on several locations at depths from 10 cm to 50 cm. Surface permeability was measured with a double-ring infiltrometer. The scatter plot of logarithmic permeability to depth was manually fitted to an exponential equation to obtain the permeability in depths 0-5, 5-15, 15-25, 25-35, 35-45, 45-55 cm. The permeability in each layer is shown in Table 4.1.

4.2.1.3 Soil water characteristics curve and bulk density

Undisturbed soil samples were collected to measure bulk density and soil water characteristics (soil moisture at saturation and air-dry, field capacity and permanent wilting point). Samples were collected at depths of 0-5, 5-15, 15-25 cm in the wheat field with a steel ring. Saturated and air-dry soil moisture were obtained by measuring the soil water content of saturated and air-dry soil, respectively. The field capacity and permanent wilting point were estimated by measuring soil water content at suctions of 33 and 1500 kPa (Pollacco, 2008), respectively. The suction at the wet end was measured with a pressure transducer tensiometer (model: T5 Pressure Transducer Tensiometer) connecting to a data logger (model:

Data Logger 6); at the dry end, suction was measured with a WP4C water potential meter (model: WP4C Dew Point Potential Meter). Bulk density was measured by oven drying soil samples of known volume at 105 °C for at least 24 hours. Soil water properties are shown in Table 4.1.

Table 4.1: List of soil properties.

Property	Unit	Depth					
		0-5 cm	5-15 cm	15-25 cm	25-35 cm	35-45 cm	45-55 cm
Soil texture	-	Silty loam	Silty loam	Silty loam	Silty loam	Silty loam	Silty loam
Average weight% of clay/silt/sand	Wt%	18/71/11	21/69/10	18/70/12	20/69/11	16/67/17	17/62/21
Bulk density	g/cm^3	0.72	0.79	0.83	-	-	-
Saturated soil moisture	m^3/m^3	67.32	64.21	64.29	-	-	-
Field capacity	m^3/m^3	33.25	36.57	35.97	-	-	-
Permanent wilting point	m^3/m^3	19.42	19.77	20.69	-	-	-
Air-dry soil moisture	m^3/m^3	5.19	6.21	5.52	-	-	-
Permeability	mm/day	20000	500	53	32	27.7	24
pH	-	5.91	5.798	5.508	5.244	5.248	5.276
Electrical Conductivity	dS/m	0.193	0.166	0.1482	0.1084	0.0826	0.0742
Phosphorus	mg/kg	338.8	360	212.6	60	33.4	32.6
Nitrate Nitrogen	mg/kg	41.3	23.52	17.1	10.2	5.44	4.16
Ammonium Nitrogen	mg/kg	7.5	8.18	5.28	3.3	2.7	2.4
Sulfur	mg/kg	177.96	206.48	123.48	46.44	32.4	34.66
Calcium	mg/kg	61.96	60.42	53.04	37.44	26.54	23.34
Magnesium	mg/kg	19.96	18.66	15.18	12.38	13.22	16.02
Potassium	mg/kg	68.76	61.82	55	50	50	52.4
Sodium	mg/kg	39.96	38.5	35.94	38.6	36.22	37.3
Aluminum	mg/kg	3.05	2.85	1.75	1	1	1
Zinc	mg/kg	3.28	2.68	1.26	0.62	0.64	0.52
Manganese	mg/kg	3.08	2.34	1.42	0.8	0.52	0.32
Iron	mg/kg	188.2	182.6	171.6	153.6	210	131.8
Copper	mg/kg	1.18	1.12	0.9	0.76	1.2	1.14
Boron	mg/kg	0.664	0.706	0.58	0.466	0.382	0.374
Silicon	mg/kg	47.8	50.8	47.8	46	41.6	39.2
Chloride Estimate	equiv. mg/kg	123.4	106.2	95	69.2	52.8	47.6
Total Carbon	%	4.966	4.884	3.854	2.31	1.698	1.13
Total Nitrogen	%	0.476	0.47	0.36	0.184	0.108	0.064
Carbon/Nitrogen Ratio	-	10.6	10.58	11.02	14.22	17.08	18.42

4.2.2 Station measurements

Station measurements include weather data, tower-based optical data and passive microwave data, continuous soil moisture and temperature data and wheat photos.

4.2.2.1 Daily weather dataset

The in-situ weather station (Figure 4.2-b) measures onsite rainfall, incoming radiation, air temperature, evapotranspiration and wind speed. Due to the gaps in the onsite weather data, they were only used for the validation of the online weather station data and selection of the representative stations for Cora Lynn.

The daily weather dataset used in this study was created from the average of several selected weather stations and is shown in Figure 4.3. The weather dataset includes daily rainfall, solar radiation, maximum/minimum temperature, pan evaporation, wind speed and vapour pressure obtained from nearby weather observation stations sourced from the Bureau of Meteorology (BoM) and a personal weather station network through Weather Underground (WU). The BoM stations located within 40 km (except for two air temperature stations) from the study site and three WU stations providing rainfall data that matches the in-situ measurements were marked on the map in Figure 4.1.

4.2.2.2 Tower-based remote sensing data

The tower was specially manufactured by Australian Engineering Solutions and installed at the corner of the Cora Lynn study site. The tower consists of a small shipping container on the ground containing the power and control system, a frame with the instrument housings at the top, and an extendable mast that can lift the instruments to 10 meters. The mast is upright during operating status and hinged at the bottom to allow for lowering for calibrating and loading the instruments from the ground. The top frame rotates in a horizontal and vertical plane such that the instruments can observe the ground with zenith ranging from 0 to 270°.

The Polarimetric P-band Multibeam Radiometer (PPMR) provided the passive microwave observations for this study at both vertical (V) and horizontal (H) polarizations. The PPMR has 4 beams with incidence angles of $\pm 15°$, and $\pm 45°$ when in the horizontal plane, providing the observations at the frequency of 742-752 MHz.

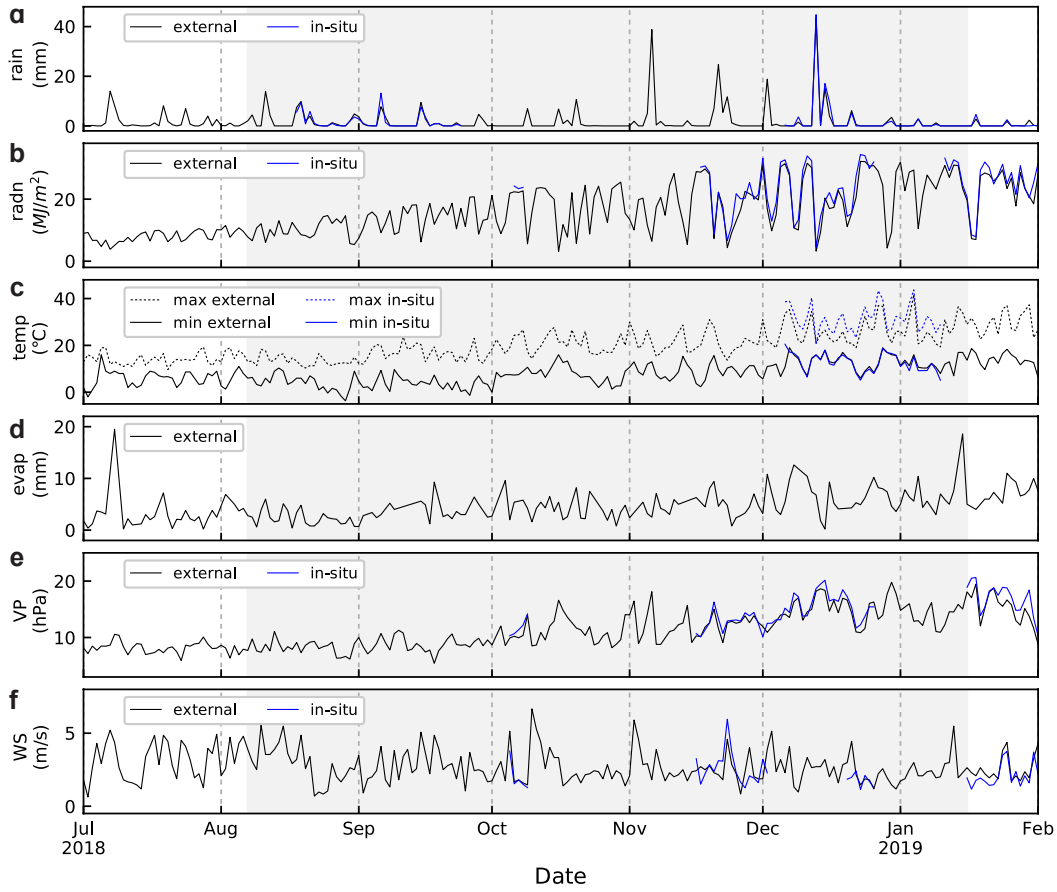

Figure 4.3: Time series of daily weather data in the simulation time period from the in-situ and external weather stations, a: rainfall; b: solar radiation; c: maximum/minimum air temperature; d: pan evaporation; e: vapour pressure, f: wind speed. The shaded area represents the growing season.

4.2.2.3 Continuous soil moisture and temperature

Hydra Probe soil sensors were installed at each soil moisture/temperature stations (Figure 4.2-c) at every 5 cm from the soil surface to 60 cm depth (12 sensors in total), providing continuous measurements of soil moisture and temperature every 15 minutes. The time series of surface soil moisture from the station and weekly sampling is shown in Figure 4.4.

4.2.3 Weekly measurements

Wheat and soil samples were collected weekly. Wheat height and surface soil moisture were measured directly in the field. Biomass, vegetation water content (VWC), LAI and soil nitrogen were measured in the laboratory through destructive sampling.

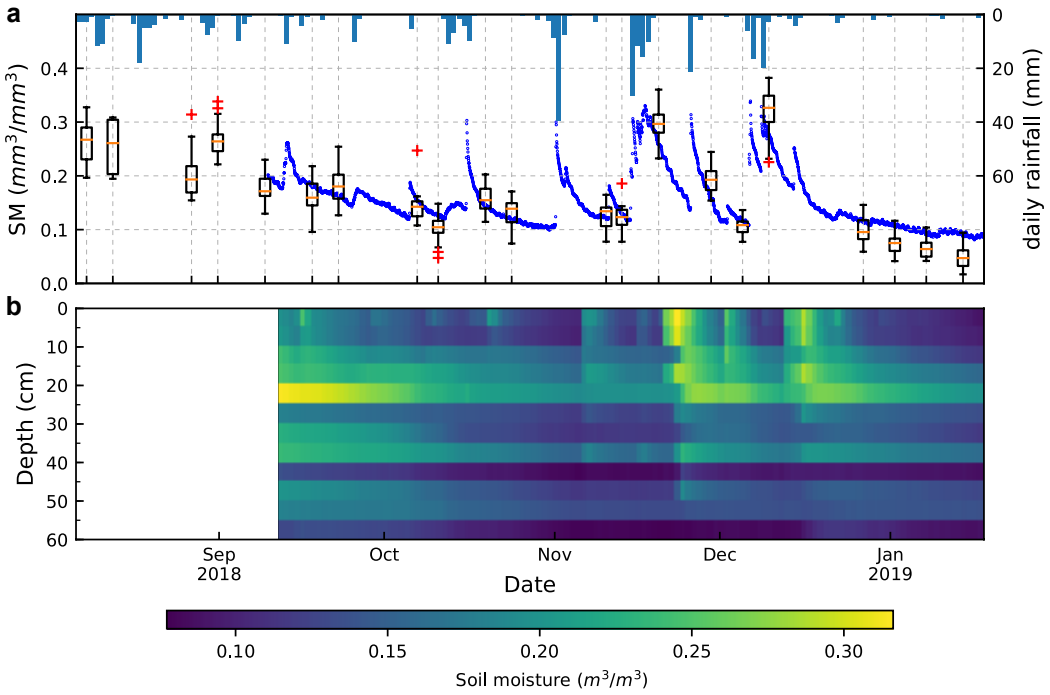

Figure 4.4: HDAS soil moisture in box plot and station 0-5 cm soil moisture at 20-minute time step (a), and station soil moisture of all 12 layers in daily time step (b).

4.2.3.1 Wheat sampling

The overall condition of wheat growth is reflected by wheat height and phenology as shown in Figure 4.5. The phenology was determined according to the Zadoks scale (Zadoks *et al.* , 1974) and subsequently converted to APSIM phenology according to a phenology lookup table in the APSIM documentation (Zheng *et al.* , 2014). Wheat height was estimated with a measuring board (Figure 4.6-a).

Wheat samples were collected weekly from 4 locations by cutting from the base in a 0.5 × 0.5 m square sampling quadrant. It should be noted that the quadrant was placed to include 3 rows of wheat. Therefore, the actual sampling area is 0.3 m 2 and not the area of the sampling quadrant. The location of the samples is marked by S1 to S4 in Figure 4.1. Wheat samples were stored in waterproof bags and sealed with elastic bands immediately after cutting, to stop vapour loss during transportation.

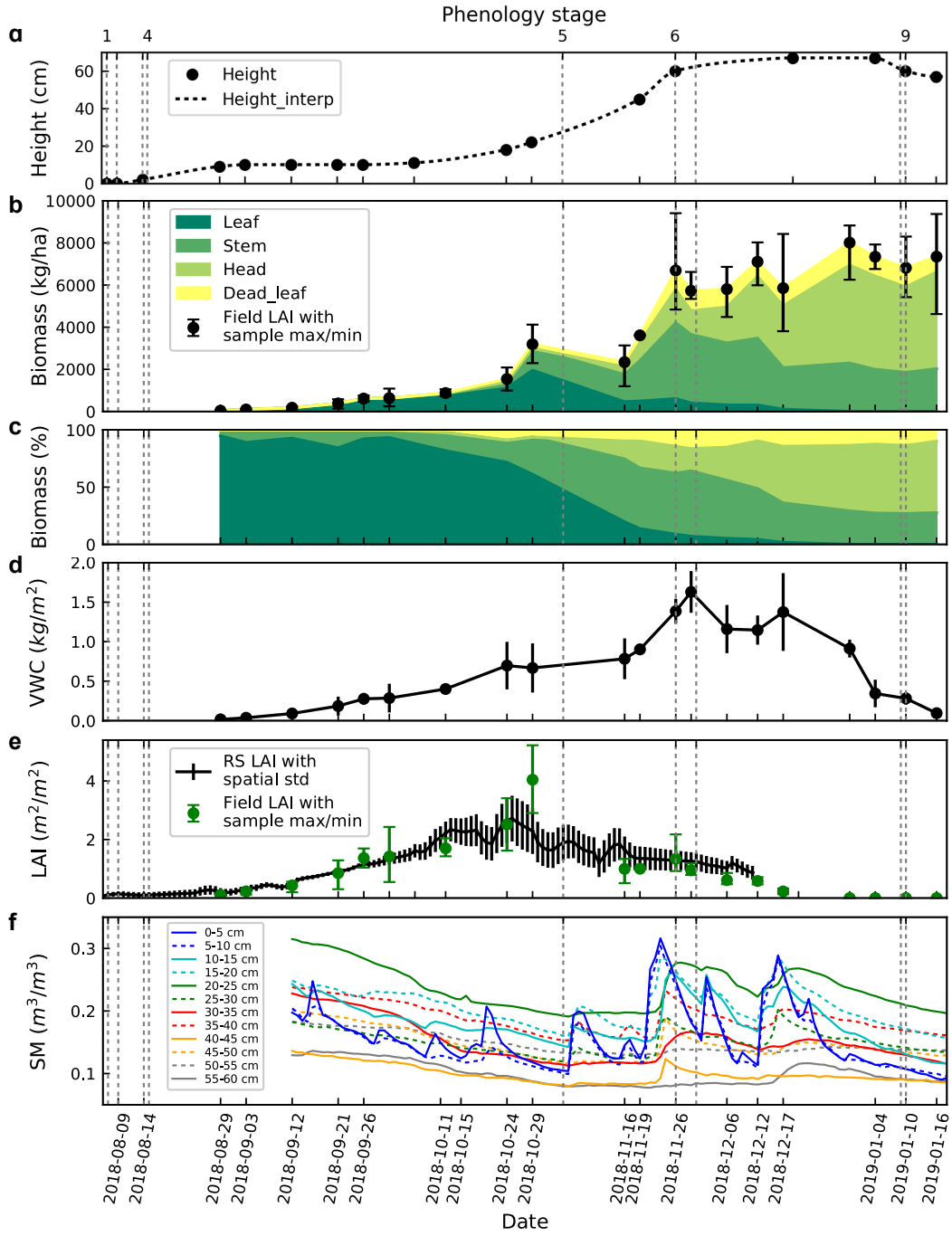

Figure 4.5: Time series of wheat height (a), biomass (b), biomass percent (c), VWC (d), LAI (e), and soil moisture in the growing period.

Figure 4.6: Photo illustration of measurements for wheat height (a), biomass of organs (b), and leaf area (c).

Wheat samples were weighed in the bags after transferring to the laboratory, and the weight of a clean, dry bag was then deducted. The samples were then divided by wheat organs (leaf, stem, spike and dead leaf) to determine the biomass percent of each organ (Figure 4.6-b). Approximately 100 leaves were scanned with a scanner to measure the leaf area (Figure 4.6-c, the two red squares are 20 cm away from each other). Samples were dried in an oven at 60 °C for at least 72 hours. The dry weight of samples, wheat organs cutting from whole plants, and leaves passing scanners were measured with a mass balance. Dry biomass (biomass hereafter) was immediately calculated by dividing the dry plant weight by the sampling area:

$$\text{biomass}_{\text{organ}} = \text{dry weight}_{\text{organ}}/\text{area}, \tag{4.1}$$

where, the subscript organ represents leaf, stem, head, and dead leaf, respectively. The area is the sampling area and equals to 0.3 m^2. LAI was calculated by:

$$\text{LAI} = \text{biomass}_{\text{leaf}} \times \text{SLA}, \tag{4.2}$$

where $\text{biomass}_{\text{leaf}}$ is the leaf biomass in kg/ha. SLA is the specific leaf area in cm^2/g, calculated by dividing the total leaf area of approximately 100 leave blades passing the scanner by their total weight:

$$\text{SLA} = \text{leaf area}/\text{leaf weight}. \tag{4.3}$$

4.2.3.2 Soil sampling

Soil samples were collected at five locations close to the monitoring plots (Figure 4.1-b) for soil nitrogen analysis. Two samples were collected from depths of 0-5 cm and 5-15 cm at each location with an auger and shovel. Careful attention was paid to ensure the soil columns were uniformly taken from the specific depth. These soil samples were prepared and analyzed with the same procedure as the chemical analysis but only with the nitrogen and ammonium analyzed. The time series of soil nitrogen is shown in Figure 4.7.

4.2.3.3 HDAS surface soil moisture

Surface soil moisture was measured with Hydraprobe Data Acquisition System (HDAS) at multiple locations in and around the wheat field. HDAS is a hand-held soil moisture measurement system allowing measuring point-based soil moisture at a depth of 5 cm. HDAS consists of a Hydra Probe sensor and a hand-held computer with built-in GPS. Surface soil moisture data were collected at 37 locations (Figure 4.1-b), with the average of three measurements taken at each location. The HDAS surface soil moisture data is plotted as boxplot together with the 15-min station soil moisture time series in Figure 4.4-a.

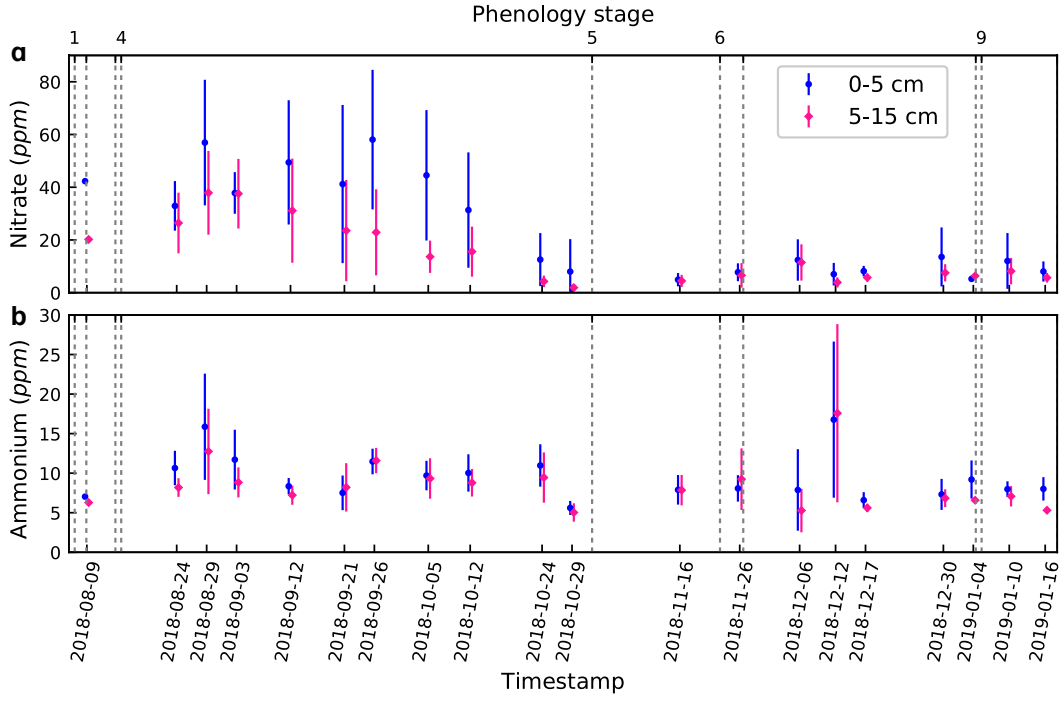

Figure 4.7: Time series of soil nitrogen in the nitrate (a) and ammonium (b) forms.

4.3 Remote sensing dataset

4.3.1 Leaf area index

Leaf area index (LAI) images were obtained from a fused high spatio-temporal resolution remote sensing dataset (Sadeh, 2020). This dataset provides daily vegetation indices and LAI products with 3 m spatial resolution and daily temporal resolution estimated from fused PlanetScope (daily, ~3 m) and Sentinel-2 (5-day, 10 m) spectral reflectance data in visible to near-infrared. The LAI dataset has been validated with in-situ LAI measured from 57 wheat fields (including the Cora Lynn study site) in Australia and Israel, and was found to have good consistency in Green LAI estimation with a coefficient of determination (R^2) of 0.94 and root mean square error (RMSE) of 1.37 throughout the growing season. The fused remote sensing LAI of the field average is shown in Figure 4.5-e. The LAI specific to each experimental plot (A-D, Figure 4.1-b) was obtained by extracting and averaging the remote sensing LAI from several pixels that is close to the plot. The time series of plot-specific LAI in the four experimental plots is shown in Figure 4.8, plotted together with the field LAI measured from the nearest wheat sample (S1, S3, and S4, Figure 4.1-b).

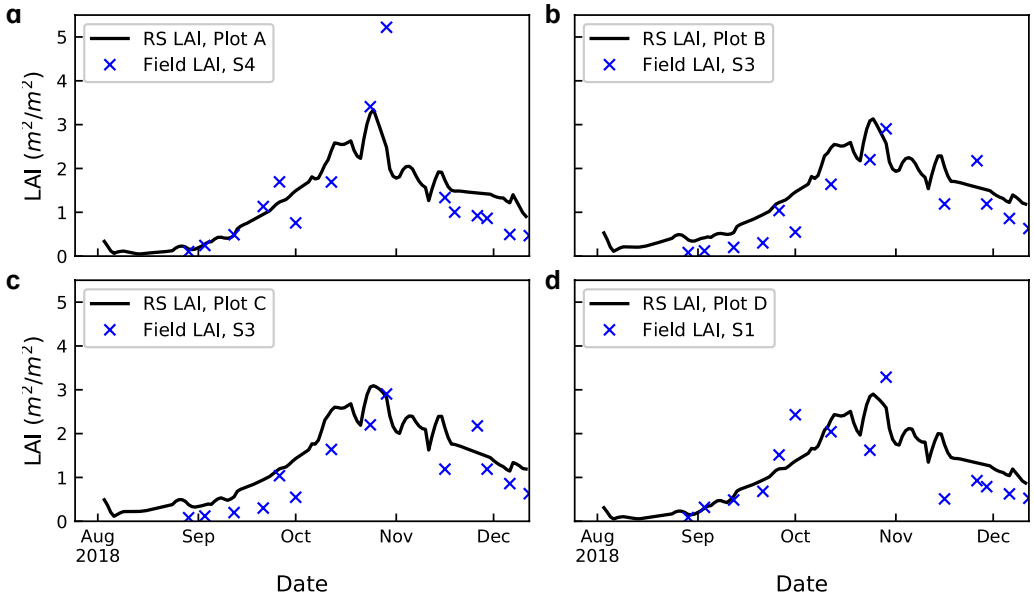

Figure 4.8: Time series of plot-specific remote sensing LAI (RS LAI) and LAI measured from the nearest wheat sample (field LAI) for experimental plots A-D (a-d, respectively).

4.3.2 Near-surface soil moisture

Near-surface soil moisture data was retrieved from the tower-based P-band microwave brightness temperature. The microwave data were obtained from a soil moisture retrieval research team (Boopathi *et al.* , 2018) working synchronously in the study site. This remote sensing surface soil moisture dataset is daily and treats the Cora Lynn wheat field as one pixel, with an R^2 of 0.59 and RMSE of 0.036 m^3/m^3. The soil moisture was retrieved using a linear regression model fitted with the multi-polarization difference index (MPDI) and the 0-5 cm depth soil moisture from the in-situ soil moisture/temperature station. The MPDI calculated is by:

$$\mathrm{MPDI} = (\mathrm{Tb_v} - \mathrm{Tb_h})/(\mathrm{Tb_v} + \mathrm{Tb_h}), \tag{4.4}$$

where the $\mathrm{Tb_v}$ and $\mathrm{Tb_h}$ are the P-band brightness temperature (in K) in the vertical and the horizontal polarization directions. Only the brightness temperature and station soil moisture measured at approximately 6 am every day were used to fit the regression model to reduce the impact of soil temperature during the day. The retrieved remote sensing and the station surface soil moisture is shown in Figure 4.9.

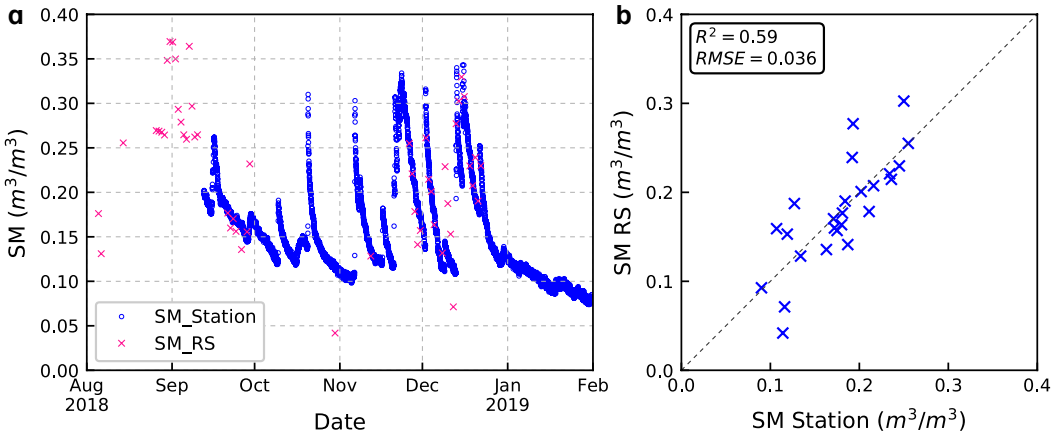

Figure 4.9: Near-surface soil moisture data as: (a) time series from the in-situ station (every 20 mins) and retrieved P-band soil moisture (daily), and, (b) in-situ station data versus retrieved P-band soil moisture data, both collected at 6 am daily.

4.4 Chapter summary

This chapter has provided a brief description of the study site where ground-based and remotely sensed observation of wheat and soil were collected. The dataset includes (1) daily weather data from nearby stations; (2) site characteristics; (3) wheat states including height, biomass of leaf, stem and head from weekly sampling; (4) soil states including soil moisture data from in-situ stations and soil nitrogen from weekly sampling; and (5) remote sensing optical brightness temperature and LAI data.

The weather and site characteristics data were used as input in two case studies (Chapters 7-8). Wheat and soil states collected from ground measurement were used for the assimilation experiments in the ground-based case study (Chapter 7). Remote sensing data were used for the assimilation experiments in the remote sensing-based case study (Chapter 8).

Chapter 5

Sensitivity analysis

In Chapter 3, the structure and physics of the APSIM-Wheat model have been introduced. In the practice of data assimilation, the errors in the estimation of model states and yield usually arise from errors in the weather data, parameters and external observations of wheat and soil states. To further understand these quantities in the APSIM that are most likely to have an impact on yield and thus need to be carefully measured, calibrated, or perturbed in the subsequent data assimilation practice, a sensitivity analysis was performed in this chapter.

5.1 Methodology

5.1.1 Model response to input data

The APSIM-Wheat was coupled with a soil water module (SoilWat), and a soil nitrogen (SoilN) module, with seven soil layers by default. The simulation window is from January 1, 1996 to December 30, 1996, with the sowing date on the day of year (DoY) 131. The dataset used in this sensitivity analysis is a dataset collected in the Dalby area provided by the APSIM model as an example dataset. This dataset includes weather, cultivar parameters, soil parameters, initial states as input data to drive the model. The model state variables are the prognostic variables sequentially estimated by the model. In this sensitivity analysis, each input and state variable data type were tested through a group of simulations by solely change the data with a range of different values (Tables 5.1 - 5.5), with each different value used in one of the simulations, respectively. The range of different values used for the

Table 5.1: Multiplicative factors applied in the variation of weather data. The Sim2 and Sim4 were assumed to be the normal weather conditions, with the Sim1 and Sim5 assumed the extreme weather conditions. For each weather data type, four factors presented in this table were used to run four simulations, respectively, with the time series of each data type multiplied by one factor in each simulation.

Weather data	Unit	Typical inter-annual coef. of variation	Multiplicative factors used in each simulation				
			Sim1	Sim2	Sim3	Sim4	Sim5
Rainfall	mm/day	24%	0.5	0.75	1	1.25	1.5
Radiation	MJ/m^2	22%	0.5	0.75	1	1.25	1.5
Temperature	^oC	15%	0.7	0.85	1	1.15	1.3
Vapour pressure	hPa	25%	0.5	0.75	1	1.25	1.5

test of each data type was within the variation of these data in historical records or their typical ranges. The sensitivity of the model response to the variation of each data type was evaluated by variation of model output regarding the date of anthesis and maturity, maximum LAI (mLAI), biomass yield (mBiomass) and grain yield (referred to as yield hereafter). The determination of input and state variable values used in each simulation is further specified in the remained of this section.

The group of weather variable used in this sensitivity analysis was generated by multiplying each weather data by a range of factors shown in Table 5.1. Two factors assumed to be the variation of weather data under normal conditions were estimated from the typical inter-annual coefficient of variation from 100-year historical records of the dataset. Particularly, another two factors estimated by doubling the variation under the normal condition were used to represent the possible variation in extreme weather conditions.

Cultivar and soil parameters, initial conditions and management information under test are shown in Tables 5.2 - 5.4. The values for cultivar parameters selected to run the model were within the typical range of 12 pre-defined wheat cultivars and existing literature (Ahmed *et al.* , 2016). Soil parameters were selected according to their typical values.

5.1.2 Model response to state variables

The essential requirement that the state variables are worth being measured and assimilated is that the change of their values during the evolution of the model can impact the model outputs. From the perspectively of data assimilation, it is essential to understand how

Table 5.2: Values selected to represent the variation of cultivar parameters and management information. For each parameter type, five values were used to run five simulations, respectively. See Table 3.3 in Chapter 3 for parameter descriptions.

Parameter	Unit	Typical values	Values used in each simulation				
			Sim1	Sim2	Sim3	Sim4	Sim5
VernSens	-	1.5-5	2.5	3	3.5	4	4.5
PhotopSens	-	1.4-5	2.5	3	3.5	4	4.5
TT4 (stage 4)	°C · days	380-400	370	380	390	400	410
TT5 (stage 5)	°C · days	520-630	520	540	560	580	600
TT6 (stage 6)	°C · days	80-180	80	100	120	140	180
TT7 (stage 7)	°C · days	580-600	570	580	590	600	610
Grains Per Gram Stem	grain/g	20-27	22	23	24	25	26
Max Grain Size	10^{-3}g	39-65	40	45	50	55	60
Potential Grain Filling Rate	10^{-3}grain/g/d	1-2.8	1	1.5	2.	2.5	3
Potential Grain Growth Rate	10^{-3}grain/g/d	1	0.8	1	1.2	1.4	1.6
Potential Grain N Filling Rate	10^{-5}grain/g/d	5.5	5	5.2	5.4	5.6	5.8
Sowing depth	mm	30	25	28	30	32	35
Row spacing	mm	250	80	120	160	200	240
Population	m^{-2}	120	80	100	120	140	160
Fertilizer	kg/ha	160	80	120	160	200	240

Table 5.3: Values selected to represent the variation of soil water parameters. For each parameter type, five values were used to run five simulations, respectively. See Table 3.4 in Chapter 3 for parameter descriptions.

Parameter	Unit	Typical values	Values used in each simulation				
			Sim1	Sim2	Sim3	Sim4	Sim5
InitialSW	%	0-100	30	40	50	60	70
BD	g/cm^3	0.8-1.2	0.8	0.9	1	1.1	1.2
LL15	m^3/m^3	0.25-0.28	0.2	0.26	0.28	0.3	0.32
AirDry	m^3/m^3	0.13-0.28	0.2	0.26	0.28	0.3	0.32
DUL	m^3/m^3	0.45-0.52	0.46	0.48	0.5	0.52	0.54
SAT	m^3/m^3	0.52-0.59	0.52	0.54	0.56	0.58	0.6
WheatLL	m^3/m^3	0.26-0.45	0.16-0.35	0.22-0.35	0.26-0.45	0.36-0.45	0.36-0.55
ConA (summer/winter)	-	5	4	4.5	5	5.5	6
U (summer/winter)	mm	5	4	4.5	5	5.5	6
DiffusConst	mm^2/d	40	30	35	40	45	50
DiffusSlope	mm^{-1}	16	8	12	16	20	24
Salb	-	0.12	0.08	0.1	0.12	0.14	0.16
CN2Bare	-	73	50	60	70	80	90
SWCON	d^{-1}	0.3	0.2	0.2-0.3	0.3	0.3-0.4	0.4

Table 5.4: Values selected to represent the variation of soil nitrogen parameters. For each parameter type, five values were used to run five simulations, respectively. See Table 3.4 in Chapter 3 for parameter descriptions.

Parameter	Unit	Typical values	Values used in each simulation				
			Sim1	Sim2	Sim3	Sim4	Sim5
RootWt	kg/ha	1000	600	800	1000	1200	1400
RootCN	-	40	20	30	40	50	60
SoilCN	-	12	8	10	12	14	16
EnrACoeff	-	7.4	5	6	7	8	9
EnrBCoeff	-	0.2	0.1	0.15	0.2	0.25	0.3
OCTotal	%	1.2-0.12	0.08-0.05	1-0.1	1.2-0.1	1.4-0.1	1.6-0.15
Fbiom	-	0.01-0.04	0.02-0.005-	0.01-0.02	0.04-0.01	0.06-0.01	0.06-0.015
Finert	-	0.4-1	0.2-0.5	0.2-1	0.4-1	0.6-1	1
pH	-	8	5	6	7	8	9
SurfaceResidue	kg/ha	500	300	400	500	600	700
ResidueCN	g/g	100	60	80	100	120	140
InitialNO3	ppm	1	0.2	0.6	1	1.4	1.8
InitialNH4	ppm	0.1	0.1	0.3	0.5	0.7	0.9

the perturbation of a specific state variable may impact the model outputs, and at which growing stage the influence was maximized. According to state variables listed in Table 3.2 (Chapter 3), prognostic state variables that have the potential to be observed (either from field measurement or remote sensing) and assimilated into APSIM models were explored.

The sensitivity of model response to the state variables was tested by adding small offsets during the simulation in each major phenology stage (stage 4, 5, 6, and 7, being the end of juvenile to floral initiation, floral initiation to flowering, flowering to start of grain filling, and the grain filling stages, respectively). These state variables included 1) wheat states - leaf area index (LAI), dry weight and nitrogen amount of each plant organ (i.e., leaf, stem, root, pod and grain), canopy height, and root depth; and 2) soil states - soil moisture, nitrate nitrogen and ammonium nitrogen. The state variables were added by a range of offsets, being multiplicatively or additively to the values of the state variables (Table 5.5). In each simulation, one offset was added to a state variable type in a day among DoY 161, 216, 249, and 266, with each in the middle of the major phenology stages, respectively.

Table 5.5: Offset added to modify the state variables during the simulation of the model. For each state variable, a total of sixteen simulations were run, each by adding one of the four offset values to the state variable in a day among the four different stages, respectively.

State variable	Unit	Offset values			Offset type[b]
		Sim1	Sim2	Sim3	
Wheat states					
Leaf area index (LAI)	m^2/m^2	-0.2	0	0.2	M
Weight of plant organs[a]	g/m^2	-0.2	0	0.2	M
Nitrogen amount of plant organs[a]	kg/ha	-0.2	0	0.2	M
Canopy height	m	-0.2	0	0.2	M
Root depth	m	-0.2	0	0.2	M
Accumulative thermal time (TT)	$^oC \cdot$ days	-50	0	50	A
Soil states					
Soil moisture in the surface layer(SM1)	m^3/m^3	-0.05	0	0.05	A
Soil moisture in the first two layers (SM1-SM2)	m^3/m^3	-0.05	0	0.05	A
Soil moisture in all seven layers (SM1-SM7)	m^3/m^3	-0.05	0	0.05	A
Soil nitrate nitrogen (NO3N)	kg/ha	-0.2	0	0.2	M
Soil ammonium nitrogen (NH4N)	kg/ha	-0.2	0	0.2	M

[a] Plant organs include leaf, stem, pod, grain and root.
[b] Offset type: M=multiplicative, A=additive.

5.2 Results

Table 5.6 illustrated the model predicted DoY of flowering (or more generally known as anthesis) and maturity, the mLAI, mBiomass and yield as a result of variations applied to weather. Values presented in these tables are the differences compared to a "reference" simulation, to which no variation was applied. The Normalized Root Mean Square Differences (NRMSD) compared to the reference were calculated for the time series of grain weight (GrainWt), LAI, biomass and surface soil moisture during the growing season. The changes of the flowering and maturity, mLAI, mBiomass and yield in each simulation by applying uncertainties to cultivar, soil parameters, management information and state variables were presented in Figures 5.1 - 5.5.

5.2.1 Impact on phenology

Figure 5.1 shows the date of flowering and maturity as a result of variations applied to the inputs against the simulation labels (Sim 1 to Sim5). All weather input, and only the

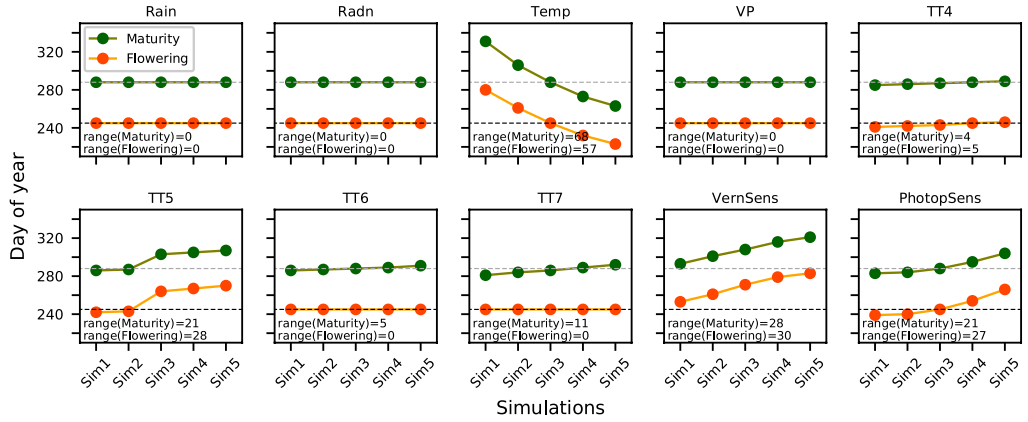

Figure 5.1: Change of the flowering and maturity dates in response to the variation of weather and cultivar parameters. Gridlines are the flowering (grey) and maturity (black) dates of the reference simulation.

cultivar parameters found to impact the phenology development were shown in this chapter, with an extended figure including all input and state variables shown in Figures A.1 - A.4 (Appendix A).

Among the weather inputs, only temperature impacts the development of phenology stages (Table 5.6 and Figure 5.1). The model predicted DoY of anthesis and maturity were impacted by cultivar parameters VernSens, PhotopSens and target thermal time of phenology stages 4 to 7 (Figure 5.1). Soil parameters, and management factors, and state variables were found to have no impact on the model prediction in terms of phenology development.

5.2.2 Impact on mLAI

Figure 5.2 shows the mLAI as a result of variations applied to the inputs against the simulation labels (Sim 1 to Sim5). Only the inputs, parameters and states found to impact the LAI development were shown in this chapter, with an extended figure including all input and state variables shown in Figures A.5 - A.8 (Appendix A).

Radiation influences the mLAI strongly, particularly when the radiation tends to be lower than normal (Table 5.6 and Figure 5.2). The model predicted mLAI is sensitive to changes of cultivar parameters: PhotopSens, TT4 and TT5, as well as soil parameters: InitialSW, DUL, LL15 and WheatLL. As the LAI reaches the maximum on the first few days after the beginning of stages 5, the value of mLAI is most likely to be affected by variations applied on stage 4, and to state variables of LAI and SW1.

Table 5.6: Change of model output in response to the variation of weather input under normal and extreme weather conditions.

			Factors applied								
			Reference	Rainfall		Radiation		Temperature		Vapour pressure	
			-	0.75	1.25	0.75	1.25	0.85	1.15	0.75	1.25
Normal weather conditions		Units	Value	Differences (compared to the reference simulation)							
	Anthesis	DoY	245	0	0	0	0	16	-13	0	0
	Maturity	DoY	288	0	0	0	0	18	-15	0	0
	mLAI	m^2/m^2	6.38	-0.16	0.09	-2.98	0.79	1.24	-1.70	-0.24	0.09
	mYield	kg/ha	3,047	-596	259	474	-564	519	-421	-410	107
	mBiomass	kg/ha	10,317	-1,019	748	-884	-161	1,600	-1,852	-817	487
			NRMSD (%) (compared to the reference simulation)								
	Grain weight		-	12.8	9.2	9.9	11.8	43.9	30.4	9.3	5.9
	LAI		-	4.4	5.5	22.0	18.4	19.8	15.7	4.6	4.2
	Biomass		-	5.0	5.3	16.4	8.4	5.3	6.0	4.4	3.8
	SSM		-	10.3	7.2	6.8	4.3	6.8	5.0	3.3	4.5
Extreme weather conditions				Differences (compared to the reference simulation)							
	Anthesis	DoY	245	0	0	0	0	35	-22	0	0
	Maturity	DoY	288	0	0	0	0	43	-25	0	0
	mLAI	m^2/m^2	6.38	-0.52	0.12	-5.3	0.46	1.68	-2.80	-0.66	0.11
	mYield	kg/ha	3,047	-1,319	953	-1,960	-1,008	915	-761	-828	416
	mBiomass	kg/ha	10,317	-2,379	1,893	-6,988	-775	4,979	-3,111	-1,766	1,177
			NRMSD (%) (compared to the reference simulation)								
	Grain weight	-	27.6	21.8	47.7	20.8	87.1	50.3	18.0	13.0	
	LAI	-	9.9	11.2	39.5	30.5	42.5	26.8	10.2	8.6	
	Biomass	-	11.2	10.9	46.9	13.2	11.1	9.9	9.2	7.7	
	SSM	-	18.8	14.1	19.4	7.8	9.3	9.4	5.9	6.9	

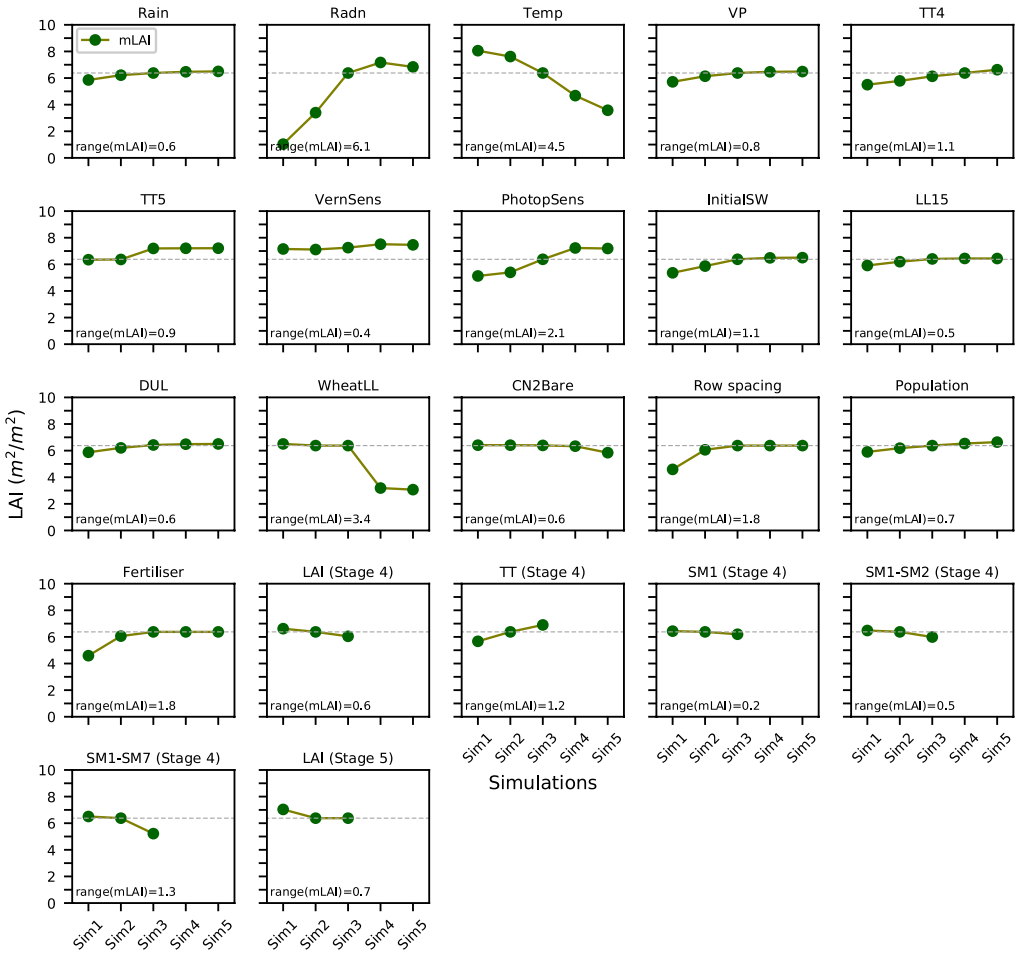

Figure 5.2: Change of the mLAI in response to the variation of input and state variables. The grey gridline is the mLAI of the reference simulation.

5.2.3 Impact on yield and mBiomass

The prediction of yield and mBiomass are impacted by all the weather factors, particularly by rainfall, radiation and temperature (Figure 5.3). High rainfall and relatively low temperature are favourable to the accumulation of yield and biomass, while solar radiation is found to hinder the accumulation of yield and biomass at both extreme weather conditions when it is excessively higher or lower than normal (Table 5.6). It should be noted that the negative impact caused by some disadvantageous weather conditions is stronger than the benefits gained from those advantageous conditions. For instance, rainfall and vapour pressure have a stronger negative impact on yield and biomass when they are lower than normal, comparing to the benefits gained from higher-than-normal conditions (Table 5.6).

PhotopSens, VernsSens, TT4, TT5, TT7 and the Potential Grain Filling Rate are most

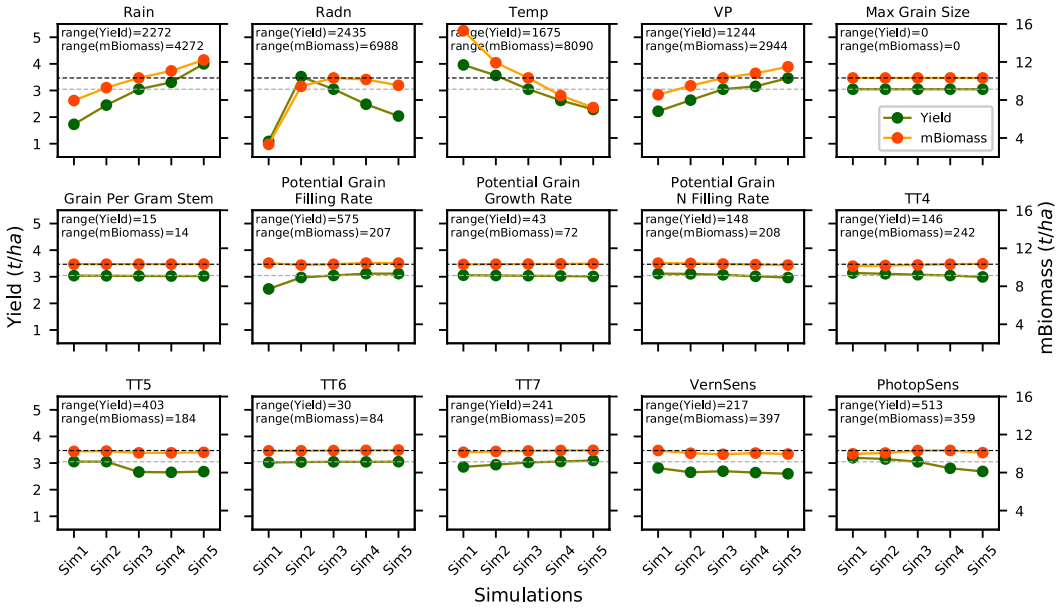

Figure 5.3: Change of the yield and mBiomass in response to the variation of weather and cultivar parameters. Gridlines are the yield (grey) and mBiomass (black) of the reference simulation.

influential on yield and biomass among all cultivar parameters, while the TT6, the Potential Grain Growth Rate and the Potential Grain N Filling Rate have a slight impact on the model estimation (Figure 5.3). These parameters are classified as two groups by the way they impact the development of the plant. One group, including PhotopSens, VernSens and the target thermal time of phenology stages 4 to 7, impact the time length that allows biomass to accumulate by affecting the phenology development of wheat. Another group are those who directly influence the rate of grain development, involving Potential Grain Filling and Growth rate and Potential Grain N Filling Rate.

According to Figure 5.4, the biomass and yield estimation is highly sensitive to parameters related to soil water conditions, particularly the initial soil moisture conditions (InitialSW) and those controlling the wheat extractable water (DUL, LL15 and WheatLL). Management information such as plant row spacing and fertilizer amount applied with sowing also impact the biomass and yield estimation.

The change of state variables have an impact on yield and mBiomass, and this impact varies with the multiplicative factors been applied to, as well as the phenology stage at which the state variable was changed. According to Figures 5.5 - 5.6, among all states under discussion, the variation of the soil moisture has the strongest impact on the development

of yield and biomass. This impact caused by the soil water of sub-surface outweighs that of the surface, and is maximized when changes were applied at stages 4 to 5. The reduction of LAI at stage 5 was found to decrease yield and biomass, while no significant increase in yield and biomass was found when LAI was raised.

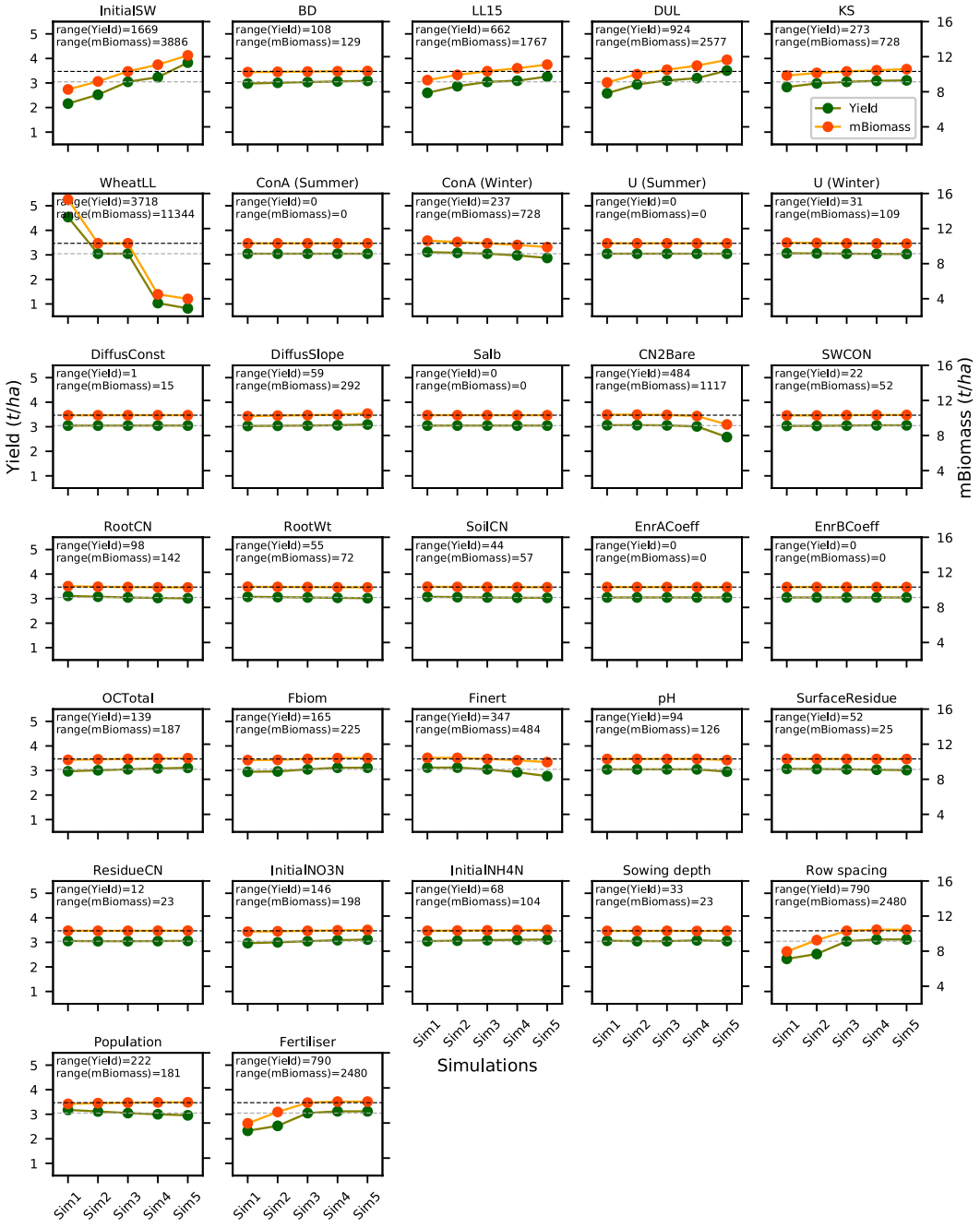

Figure 5.4: Change of the yield and mBiomass in response to the variation of soil parameters and management information. See Figure 5.3 for legend.

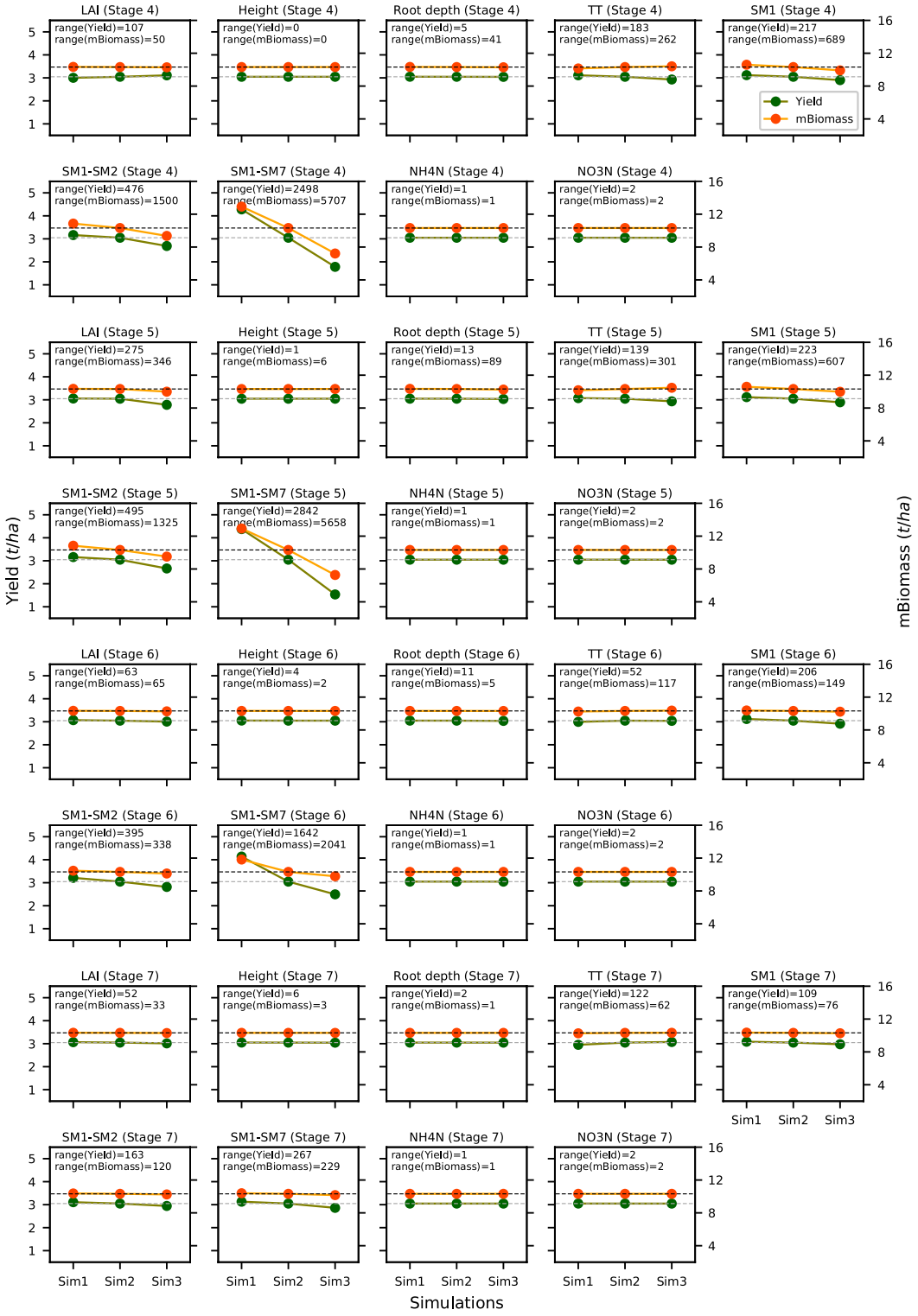

Figure 5.5: Change of the yield and mBiomass in response to the wheat and soil states in different phenology states. See Figure 5.3 for legend.

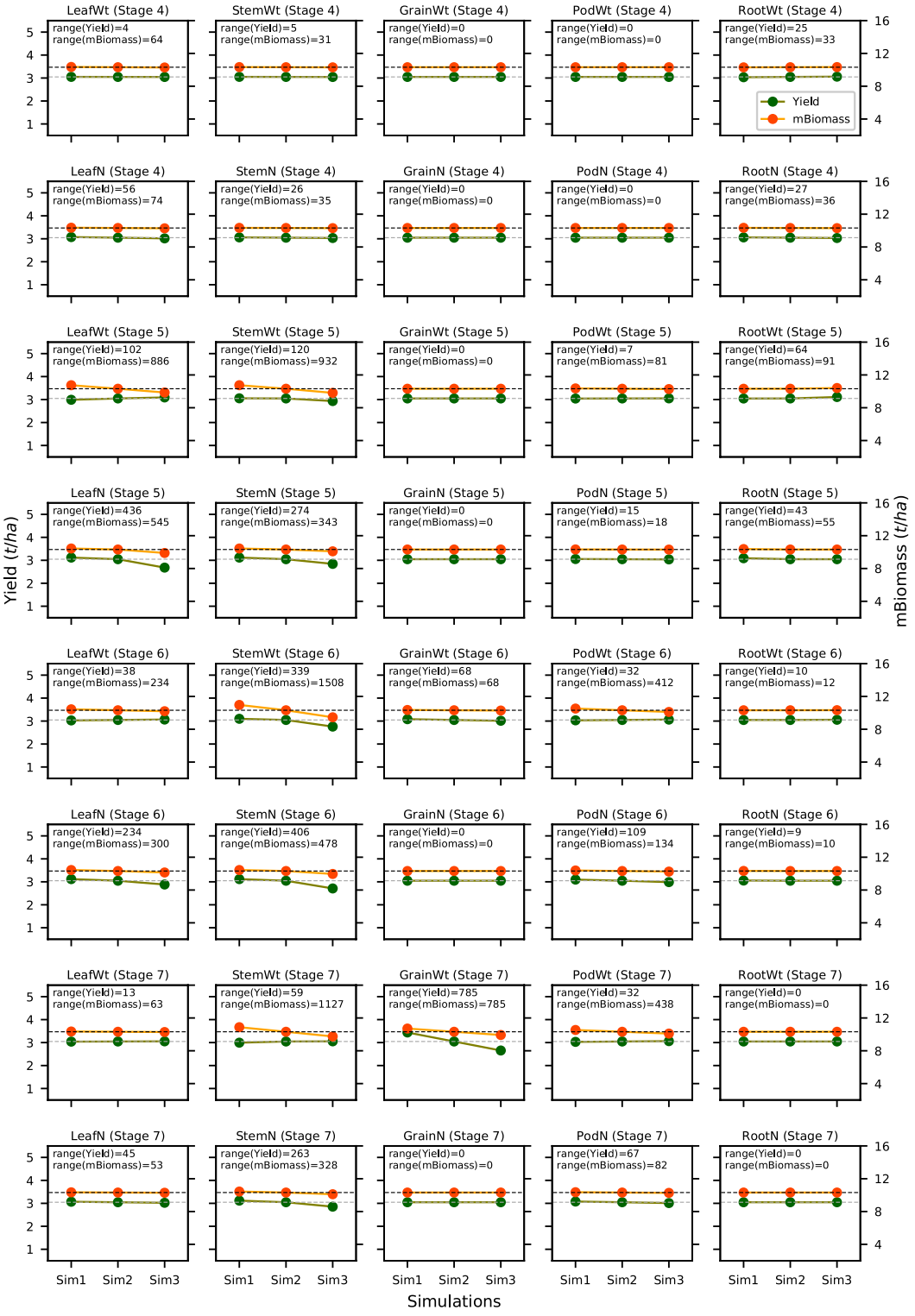

Figure 5.6: Change of the yield and mBiomass in response to the wheat states in different phenology states. See Figure 5.3 for legend.

Figure 5.6 shows that the estimation of yield and mBiomass were also sensitive to changes of weight and nitrogen amount of leaf and stem. More specifically, the estimation was sensitive to the weight of leaf on stage 5 and stem on stages 5 to 7, and the nitrogen of leaf on stages 5 to 6 and stem on stages 5 to 7.

5.3 Discussions

5.3.1 Uncertainties of APSIM yield estimation

While it is well acknowledged in the context of hydrological modelling that the primary uncertainties are originated from weather forcing, parameters and model physics (or sometimes referred as model structure) (Turner *et al.* , 2008; Di Baldassarre & Montanari, 2009; McMillan *et al.* , 2012; Li *et al.* , 2014), there exist slight differences for APSIM crop modelling. The results of the analysis show that the uncertainty of an APSIM simulation is primarily sourced from some of the weather forcing (including rainfall, radiation and temperature), cultivar parameters (including VernSens, PhotopSens and target thermal time of stage 4-7), soil properties (including InitialSW, DUL, LL15 and WheatLL) and management factors (fertilizer and row spacing). The imperfection of model structure is another source of error but is not considered in this analysis. Thus, in the implementation of data assimilation using Ensemble Kalman filter (EnKF), they are recommended to be carefully perturbed for ensemble generation. As the overall impact on model estimation resulted from the variations of weather forcing is more significant than that from parameters, indicating that the accuracy of model estimation is more likely to be affected by the uncertainties sourced from weather forcing than the characteristics of wheat cultivars, particular attention should be given to weather forcing for ensemble generation.

5.3.2 Impact of thermal accumulation

Temperature is the primary driving force of phenology development, as the development of phenology is controlled by the accumulation of temperature over time. Thus, the variation of temperature affects LAI, biomass and yield in a way by changing the time length that allows them to accumulate. To illustrate this, cues can be found from the analysis of temperature and some cultivar parameters: under the condition that air temperature is relatively

low (multiplied by factors 0.8 and 0.5), the plant spends longer time in each growing stages, leading to significantly higher estimates of LAI, biomass and yield (Table 5.6). The change of thermal accumulation also explains the impact on the accumulation of matters caused by some cultivar parameters, whose variation either impacts the accumulation of thermal time, or determines whether the demand for accumulated thermal is met for the plant to transient to the next growing stage (Figure 5.3).

5.3.3 Impact of soil water availability

As of result of soil water deficit being a key source of environmental stress acting on phenology, photosynthesis and leaf development through a multiplicative factor ranges from 0 to 1, the lack of water produces heavy stress on plant growth, while the surplus amount of soil water has limited benefits to yield and biomass. When the same amount of variation is applied to soil water content at different layers, a more significant influence on mLAI, yield and biomass is produced by the variation of bottom layers. This is caused by the nature of soil water content at different depths, that soil water content is less dynamic in deep layers than in surface layers. Thus, even a tiny variation of soil water on bottom layers forces the model to act differently. Therefore, the accuracy of soil water in bottom layers are more important than those of the top layers.

The high sensitivity of model estimation to soil water also indicated that the model is likely to be affected by water-related parameters, including initial soil water, DUL, LL15, and WheatLL. This also corresponds to the result that yield and biomass were changed when variations were applied to these parameters (Table 5.4). Therefore, water-related soil parameters are necessary to be carefully measured (if available) or calibrated before using the model.

5.3.4 Cultivar parameters

It should be noted that the values of cultivar parameters are more likely to be acquired from calibration rather than measured through long-term sampling and laboratory tests from field experiments. In this case, although yield and biomass are affected by cultivar parameters strongly, the sensitivity of the model to cultivar parameters should be considered in two aspects, through both the differences of parameters among varies cultivars and the possible

variation among different plant individuals of the same cultivar. This is because parameters may vary greatly between different cultivars, but they differ less among plants of the same cultivar, which is the case in reality, where usually only one cultivar is planted in an area. As an outcome of such consideration, the sensitivity analysis should reveal which parameter should be carefully calibrated by testing possible parameter values taken from a range of cultivars, as well as indicating which parameters, if well-calibrated, should be carefully perturbed through applying variations that are limited to the possible uncertainties of plant individuals of a certain cultivar. As the values of cultivar parameters were selected according to a range of different cultivars, this analysis discusses more the sensitivity of APSIM to different cultivars, indicating those parameters which play a vital role in the development of yield and thus should be carefully calibrated. The assessment of the uncertainties of these parameters is further discussed in the synthetic study in Chapter 6.

This analysis found that the model predicted yield is most likely to be affected by VernSens, PhotopSens, target thermal time, and Potential Grain Filling Rate out of eleven parameters. However, this result is, to some extent, against the conclusion of existing sensitivity analysis in literature, where Zhao *et al.* (2014) applied variations of 50% to the cultivar parameter of APSIM-Wheat and concluded that Grains Per Gram Stem, Max Grain Size and Potential Grain Filling Rate are most important among the eleven cultivar parameters. This discrepancy can be explained by the assumption that the variations of cultivar parameters were limited to the range of a group of cultivars whose cultivar parameters were already known, and thus more realistic to represent the possible variations among cultivars. For example, an intuitive conclusion can be made that the yield is greatly improved when increasing the number of grain per gram stem by 50%, while, however, such difference in terms of the grain number is much smaller among realistic cultivars.

5.3.5 Field management

The management information that necessary for crop modelling in APSIM includes plant spacing, population and the applied fertilizer amount. It is worth noting that the plant spacing and population may affect the yield differently. For example, it is more likely to harvest more products per unit area with a higher plant density, which explain the significant difference caused by the variation of plant row spacing and plant population.

5.3.6 An exception of mLAI prediction

While the date selected for state variables to be changed in stage 5 was a couple of days after LAI reaches the peak, the mLAI is supposed to be affected only when state variable was changed on the date selected at stage 4. However, an exception was observed when the modification of LAI at stages 5 increased the mLAI. This is explained by the fact that although LAI was deceasing on the date, it was modified at stage 5, the difference between the LAI and the peak was still small. When LAI on that day was multiplied with a factor 1.2, the modified LAI value exceeded the peak and thus became the new maximum.

5.4 Chapter summary

The sensitivity analysis presented in this chapter explored the response of APSIM model results to variations of weather forcing, cultivar parameters, soil parameters and state variables. Table 5.7 gives an overview of the results given by this analysis. A table of import state variables and how they can be measured is shown in Table 5.8. This chapter lay the foundation for experiments of data assimilation of state variables into APSIM.

Table 5.7: Major and minor factors impacting the key APSIM-Wheat model outputs found in the sensitivity analysis.

Data type	Impact on the key model outputs[a]	
	Major	Minor
Weather	Rainfall, temperature, solar radiation	Vapour pressure
Cultivar parameters	VernSens, PhotopSens, TT4, TT5, TT7, Potential Grain Filling Rate	TT6, Grains Per Gram Stem, Max Grain Size, Potential Grain Growth Rate, Potential Grain N Filling Rate
Soil parameters and initial conditions	InitialSW, DUL, LL15, WheatLL	All other parameters.
Management	Row spacing, fertilizer applied at sowing	Population

[a] The division of major and minor impact is based on the range of yield and maturity date in the sensitivity analysis. For weather input, range(Yield) > 1,500 kg/ha is considered to have a major impact. For parameters and management information, range(Yield) > 500 kg/ha or range(Maturity) > 10 days is considered major.

Table 5.8: Sensitivity of the APSIM-Wheat model to the wheat and soil states.

State variable	Unit	P/D[a]?	Measurements		Sensitivity[c]
			Field observation	Remote sensing	
Canopy height	m	P	Tape measure, Ground-based Lidar	Lidar	Little
Canopy cover	-	D	Visual estimation, photographic method	Spectral reflectance	-
Accumulated thermal time	°C · d	P	NA[b]	NA	Medium
Phenological stages	-	D	Direct observation	Spectral reflectance	-
Leaf area index	m^2/m^2	P	Leaf area meter, link to dry biomass	Spectral reflectance	Medium
LeafWt	g/m^2	P	Sampling	NA	Medium
StemWt					Medium
PodWt					Low
GrainWt					Medium
RootWt					Low
LeafN	g/m^2	P	Sampling	NA	Medium
StemN					Medium
PodN					Medium
GrainN					Little
RootN					Low
Yield	kg/ha	D	Sampling at harvest	NA	-
Grain size	g	D	Sampling at harvest	NA	-
Plant N concentration (stem, leaf, pod, grain)	g/g (N/Wt)	D	Sampling	NA	-
Grain protein	%	D	Sampling at harvest	NA	-
Number of nodes	-	P	Direct observation	NA	Little
Number of leaves	-	P	Direct observation	NA	Little
Root length	m	P	NA	NA	-
Root depth	m	P	Sampling	NA	Little
Soil moisture (surface)	m^3/m^3	P	Sampling	Spectral reflectance, radiometer, radar	Medium
Soil moisture (root-zone)	m^3/m^3	P	Sampling	NA	High
Soil nitrate	kg/ha	P	Sampling	NA	Little
Soil ammonium	kg/ha	P	Sampling	NA	Little

[a] P/D: the state variable is prognostic (P) or diagnostic (D).

[b] NA: not available.

[c] The division of sensitivity levels is based on the values of the range of yield fall in the most sensitive stages (except the GrainWt in the grain filling stage that directly determines the yield). High: more than 500 kg/ha; Medium: 100 - 500 kg/ha; Low: 10 - 100 kg/ha; Little: less than 10 kg/ha.

Chapter 6

Synthetic study

The previous chapter presented a sensitivity analysis on the response of APSIM-Wheat modelling to a range of small disturbances applied to weather input, cultivar parameters, soil parameters and state variables. The sensitivity analysis provided information on which weather data, parameters and initial conditions of the APSIM-Wheat model need to be carefully perturbed in the implementation of data assimilation experiments, and which state variable have the potential to be assimilated to help the model to make a better estimation on other state variables and thus crop yield. To give a well-rounded illustration of the potential for improving APSIM-Wheat yield estimation using data assimilation techniques, the assimilation performance of all possible prognostic state variables is tested here. As some of these state variables are difficult to observe, a synthetic study is used to gain this understanding.

6.1 Methodology

6.1.1 Experimental set-up

A typical synthetic experiment consists of three components: a "truth" generation run, an open-loop run, and a data assimilation run (Curnel *et al.* , 2011) whereby the synthetic true states and observations, the stochastic model simulation without assimilating observations as a reference, and the simulation with assimilating observations are created, respectively. The procedure is shown in Figure 6.1 and described further below.

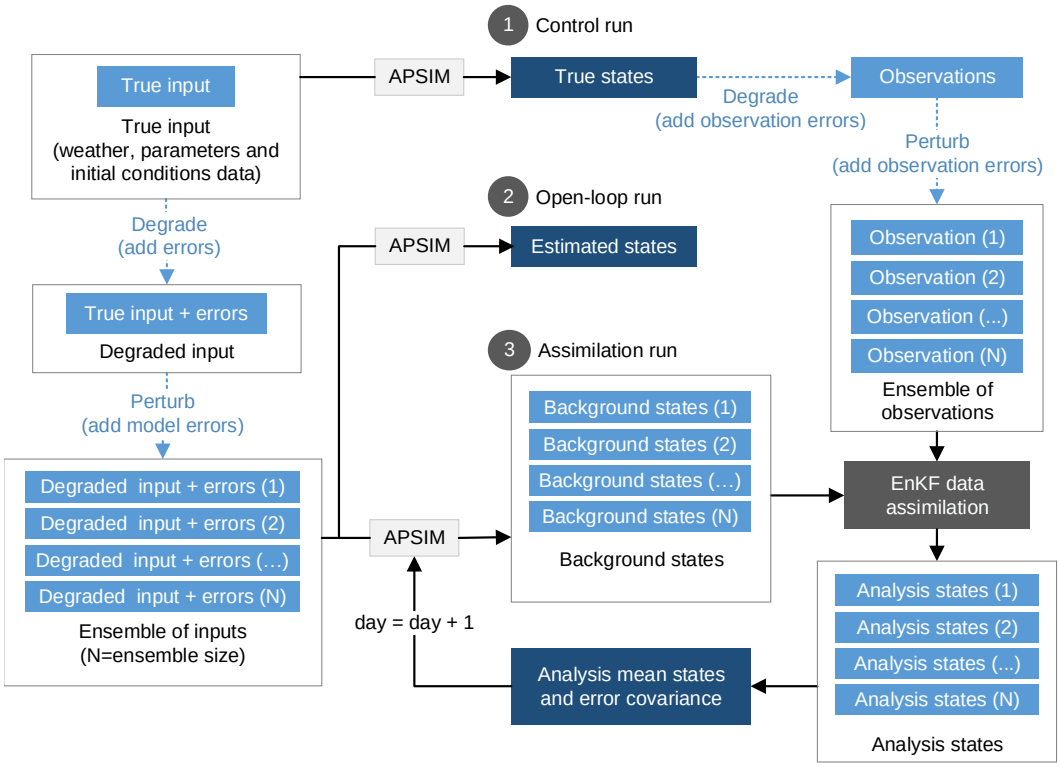

Figure 6.1: Schematic of synthetic study.

6.1.1.1 Truth generation run

The truth generation run aims to provide "error-free" state estimation for use as 1) observations to assimilate and 2) to validate model outputs. This is achieved by running the APSIM model under an ideal scenario where the input data are accurately known ("true" input) and the model is assumed to perfectly represent the real physics of crop growth. The true input is therefore a collection of weather, parameter and initial condition data which is assumed to be the truth, although not known perfectly in reality. Model state variables obtained by running the perfect model with true input is denoted "truth". Same as the sensitivity analysis, the simulation window is taken as a whole year from January 1, 1996 to December 30, 1996 in this synthetic study, using a set of APSIM example weather data and parameters. The example weather input is provided in the APSIM software package as an example dataset, including rainfall, temperature, solar radiation, and vapour pressure. The date of sowing was set to day of year (DoY) 131 (May 10) after a 5-day rainfall event. The cultivar parameters, soil parameters and initial conditions used as model input are presented in Tables 6.1 and 6.2.

Table 6.1: Soil parameters and initial conditions used as model input.

Soil parameter	Unit	Values used in	Layer						
			1	2	3	4	5	6	7
Soil depth (depth of each soil layer)	cm		0-15	15-30	30-60	60-90	90-120	120-150	150-200
Initial SM (initial soil moisture)[a]	m^3/m^3	Truth	0.25						
		Open-loop	0.24						
		Assimilation (std.)[b]	0.1	0.09	0.08	0.07	0.06	0.06	0.06
WheatLL (lower limit of soil moisture that is extractable by the plant)	m^3/m^3	Truth	0.10	0.11	0.12	0.13	0.14	0.14	0.14
		Open-loop	0.09	0.10	0.11	0.12	0.13	0.13	0.13
		Assimilation (std.)[b]	0.02						
DUL (drained upper limit)	m^3/m^3	Truth	0.36	0.35	0.34	0.33	0.32	0.32	0.32
		Open-loop	0.38	0.37	0.36	0.35	0.34	0.34	0.34
		Assimilation (std.)[b]	0.02						
AirDry (soil moisture of air-dry soil)	m^3/m^3	All	WheatLL - 0.02[c]						
SAT (soil moisture of saturated soil)	m^3/m^3	All	DUL + 0.1[d]						
BD (bulk density)	g/cm^3	All	1.3	1.35	1.4	1.45	1.5	1.5	1.5

[a] This initial SM is the volumetric soil moisture, being equivalent to the parameter InitialSW mentioned in Chapters 3 and 5 that was defined as the plant available soil water content relative to the WheatLL. No perturbation was applied to the initial SM because a warm-up period of four months before sowing was included in the simulation window to allow the spread of ensemble soil moisture to be sufficiently large.

[b] The value presented is the standard deviation (std.) of the random term added to the respective variable of the open-loop run in producing the ensemble for the data assimilation run.

[c] AirDry was assumed 0.02 lower than the selected WheatLL value.

[d] SAT was assumed 0.1 higher than the perturbed DUL value.

Table 6.2: Cultivar parameters used as model input.

Cultivar parameter	Unit	Selected values used in		
		Truth	Open-loop	Data assimilation (std.)[a]
VernSens (vernalization sensitivity)	-	2	2.1	0.1
PhotopSens (photoperiod sensitivity)	-	3.5	3.6	0.1
TT4 (target thermal time in stage 4)	°C · day	400	410	20
TT5 (target thermal time in stage 5)	°C · day	580	600	30
TT6 (target thermal time in stage 6)	°C · day	120	125	6
TT7 (target thermal time in stage 7)	°C · day	590	610	30
Potential Grain Filling Rate	10^{-3} · grain/g/d	2	2.1	0.2
Potential Grain Growth Rate	grain/g/d	1	1.1	0.1
Potential Grain N (nitrogen) Filling Rate	10^{-5} · grain/g/d	5.5	5	0.5

[a] The value presented is the standard deviation (std.) of the random term added to the respective variable of the open-loop run in producing the ensemble for the data assimilation run.

6.1.1.2 Open-loop run

It should be noted that the term “open-loop” is in contrast to a “closed-loop” data assimilation scenario, meaning that there is no external constraint on the simulation results. In a typical synthetic study, both the open- and the closed-loop scenarios describe a situation where the uncertainties of weather data, parameters, initial conditions and imperfection of the model physics are considered to contribute to the uncertainties of the model states.

The simulation of both scenarios is based on the same single “degraded” input dataset generated from the truth to mimic a realistic situation, where all model input and parameters are not known accurately. The degradation of weather data was performed by drawing a single sample from a Gaussian distribution characterized by weather and parameter values as the mean and their uncertainties as the standard deviation. The values of soil and cultivar parameters were selected to have a discrepancy with the truth representing uncertainty (Tables 6.1 and 6.2). Starting from the single degraded input dataset, the open-loop run was performed as an ensemble of stochastic simulations with different perturbed input datasets generated by adding random noise to the degraded input. The approach for ensemble generation is described in Section 6.1.2.1 in detail. The ensemble of perturbed simulations was run in parallel, with the resultant model states and output taken as the mean of all ensemble members. As a result of degradation, a discrepancy in the model state was produced by the open-loop run relative to the truth.

6.1.1.3 Data assimilation run

The open-loop run gives an imitation of a realistic situation where input data suffers from uncertainties. It also gives the errors in the estimated model state and output compared to the truth. The data assimilation scenario uses the same ensemble of simulations with the same input data as the open-loop, but with the synthetic observations of model state variables assimilated during the model evolution. Successful implementation of data assimilation can be inferred if the estimation error from the open-loop is reduced with the assimilation of external observations. This synthetic study primarily focused on the performance of wheat yield estimation.

The synthetic observations were generated by applying observational uncertainties to the true model states. In the implementation of the EnKF data assimilation run, model ensembles were perturbed using the same approach as for the open-loop, while observations were additionally perturbed. The EnKF algorithm plays the role of updating state variables at each observation time step by merging the model simulation (known as the background or prior states) with the external observations to deliver a set of posterior states. This updating process accounts for the ensemble error covariance given by the probability distribution of both model simulated states (background uncertainties) and their observations (observational uncertainties). The estimation of the background and observational uncertainties is specified in Section 6.1.2.

6.1.2 Ensemble generation

Accurate estimation of background and observational uncertainties is vital for the successful implementation of data assimilation (Pellenq & Boulet, 2004; Ines *et al.* , 2013). While it is well acknowledged in hydrology that the primary source of uncertainty is weather forcing, parameters and model physics (sometimes referred as model structure), it is slightly different for crop modelling: being the weather forcing, cultivar parameters, soil properties and field management for the case of APSIM-Wheat modelling according to the sensitivity analysis in Chapter 5. The estimation of uncertainty sourced from the model structure usually requires an ensemble of models (Li *et al.* , 2014), however only one model was employed in this thesis, and so this uncertainty was neglected throughout. Particularly, in this

specific synthetic study, the model was assumed perfect and used to generate the truth. Field management factors, including plant density, irrigation and fertilization, were assumed to be well-known with little variation across a field, and so their uncertainties were neglected. To summary, the quantities that were perturbed are:

1. weather forcing, including rainfall, radiation, and temperature;

2. cultivar parameters, including sensitivity to photoperiod and vernalization, target accumulative thermal time for phenology stages 4-7, potential rate of grain filling, grain growth and nitrogen filling; and,

3. soil parameters, including initial soil moisture, drain upper limit, and permanent wilting point.

The determination of corresponding uncertainties is discussed in the remainder of this subsection.

6.1.2.1 Weather forcing uncertainties

The weather forcing was perturbed with the method described by Turner *et al.* (2008). This method avoids the bias caused by constraining the value randomly drawn from the Gaussian distribution not to exceed the range where they are physically realistic. According to Turner *et al.* (2008), weather data can be classified into three types: "unrestricted", "restricted", and "semi-restricted", according to whether the range of data is unrestricted, restricted at both ends, or only restricted by an upper or lower bound. The forcing vector at time step k for the ith ensemble is generated by:

$$f_k^i = f_k + \zeta_k^i + \beta^i, \tag{6.1}$$

where f_k is the observed point forcing data at time step k, ζ_k^i is a Gaussian error with zero mean and standard deviation of σ_1, and β^i represents an offset taken as a single sample from the Gaussian distribution with zero mean and standard deviation of σ_2. Unrestricted means that the data is not physically constrained and therefore will not be beyond its normal range when the perturbation is applied. For the unrestricted data type (e.g., air temperature), the

standard deviation σ_1 and σ_2 are calculated by:

$$\sigma_1 = \xi, \tag{6.2}$$

$$\sigma_2 = \chi. \tag{6.3}$$

The semi-restricted data type has a lower boundary f_{min} to constrain the minimum value, including rainfall, radiation and evaporation that must be larger than or equal to zero. The standard deviations are calculated by:

$$\sigma_1 = (f_k - f_{min})\xi, \tag{6.4}$$

$$\sigma_2 = (f_k - f_{min})\chi, \tag{6.5}$$

where ξ and χ are constants. The values of ξ and χ used in this study were taken from Turner *et al.* (2008) as presented in Table 6.3.

6.1.2.2 Parameter and initial condition uncertainties

A discussion about the uncertainty of APSIM cultivar parameters is rarely found in the literature (Zhao *et al.* , 2014). Therefore, the parameters are assumed well-known or well-calibrated and their uncertainties were limited to around 5% of the parameter values (Tables 6.2), so that the ensemble spread produced by the perturbation of parameters is large enough but not excessively large. Initial soil moisture has a strong impact on model state and yield estimation as concluded in the sensitivity analysis, and should therefore be carefully perturbed. The soil moisture was initialized on the first day of simulation (DoY 1), and a 4-month warm-up period was given to allow the soil moisture reach an equilibrium state before the sowing date (DoY 131). By the date of sowing, suitable ensemble spread was generated for soil moisture, and therefore further perturbation of initial soil moisture at sowing was not required.

6.1.2.3 Observational uncertainties

Observational uncertainties are the result of instrument inaccuracy and imperfect retrieval algorithms. The techniques for measuring LAI and surface soil moisture from remote

Table 6.3: Uncertainty estimation of weather forcing according to Turner *et al.* (2008).

Data	Unit	Restriction type	ξ	χ
Rainfall	mm	Semi-restricted with lower bound	0.25	0.25
Radiation	MJ/m^2	Semi-restricted with lower bound	0.864	0.864
Temperature	°C	Unrestricted	1.4	0.6

sensing are mature, and so their observational uncertainties (Table 6.4) were aligned with those for remote sensing products and based on several validation experiments in the literature. For example, the MODIS LAI product was reported to have an uncertainty of 0.38 m^2/m^2 in grass and cereal crop areas (Tan *et al.* , 2005), and SMOS soil moisture products were reported to have an accuracy of 0.04 m^3/m^3 (Kerr *et al.* , 2012). The observational uncertainties of other state variables, including nitrogen amount of leaf, stem and pod, and dry weight of leaf and stem whose measurements currently require field sampling and laboratory analysis, were assumed to be proportional to their estimated values from the model.

In realistic situations, the measurement of observations is usually not always available at the same time. For example, LAI observations from satellite missions are available every 8 or 16 days (e.g., MODIS and Landsat) and can depend on cloud cover, while surface soil moisture is usually available every 2 to 4 days (e.g., SMOS and SMAP), depending on the latitude. Thus, the synthetic observation data used here were assumed to be available at a specified time interval, by sampling the observations from a continuous daily time series with specified acquisition intervals.

6.1.3 Data assimilation

The performance of data assimilation can depend on the assimilation set-up. For example, the EnKF outcome is affected by the assimilated state variables, ensemble size, assimilation frequency, and the timing of phenology stage when observations are available. The data assimilation experiments presented in this chapter include ensemble size determination, the assimilation of single and multiple state variable types, the assimilation of state variables in different observation availability scenarios, and assimilation when the phenology is constrained.

Table 6.4: Wheat and soil state variables of APSIM included in the synthetic data assimilation.

State variable(s)	Description	Unit	Frequency	Uncertainty
Wheat states				
LAI	Leaf area index	m^2/m^2	8 days	0.4
LeafWt	Leaf weight	g/cm^2	7 days	5% of the state values
LeafN	Leaf nitrogen	g/cm^2	7 days	5% of the state values
StemWt	Stem weight	g/cm^2	7 days	5% of the state values
StemN	Stem nitrogen	g/cm^2	7 days	5% of the state values
PodN	Pod nitrogen	g/cm^2	7 days	5% of the state values
Soil states				
SM1, SM2, …, SM7	Volumetric soil moisture in layer 1, 2, …, 7	m^3/m^3	3 days	0.03
NO3N1, NO3N2, …, NO3N7	Soil nitrogen in the form of nitrate in layer 1, 2, …, 7	kg/ha	7 days	5% of the state values
NH4N1, NH4N2, …, NH4N7	Soil nitrogen in the form of ammonium in layer 1, 2, …, 7	kg/ha	7 days	5% of the state values

6.1.3.1 Ensemble size determination

The ensemble size is the number of ensemble members, which is ideally as large as possible to accurately represent the probability distribution of model states. However, a large ensemble size results in a high computational cost, particularly when the model accounts for spatial variability with multiple pixels. The determination of ensemble size is therefore a trade-off between more statistical samples and computational efficiency. This experiment aimed to determine the minimum ensemble size required for assimilation experiments with APSIM-Wheat. Six open-loop runs were tested, having ensemble sizes of 10, 20, 50, 100, 200 and 400, respectively.

6.1.3.2 Assimilation of single and multiple state variables

This experiment aimed to explore the potential of improving yield estimation by constraining all prognostic wheat and soil state variables in APSIM. The sensitivity analysis in Chapter 5 has demonstrated that the accuracy of several state variables in APSIM-Wheat may have an impact on APSIM-Wheat yield estimation. Based on the sensitivity analysis, the wheat and soil states focused on in this synthetic data assimilation experiment are presented in Table 6.4.

Wheat and soil are two separate modules in APSIM, and the impact of assimilating state

variables is therefore discussed according to those two separate groups. The assimilation of single wheat states included LAI, LeafWt and LeafN, StemWt, StemN, and PodN, while soil states included soil moisture, ammonium and nitrites in the top, medium and bottom layers (layers 1, 4, and 7, respectively) and all 7 soil layers. The combined data assimilation was conducted by assimilating several wheat and soil states together. The frequency and uncertainties of state variables presented in Table 6.4 apply to all assimilation experiments in this chapter, and is referred to as the "reference" configuration for observation availability hereafter. The reference availability of soil states is during the whole simulation period, while wheat states were only taken to be available during phenology phase 4 (end of juvenile) to phase 7 (end of grain filling), because the assimilation of wheat states in early stages can cause the model failure when the wheat states were accidentally updated to a value close to zero.

6.1.3.3 Assimilation of observations with different repeat interval and accuracy

The availability of state observations usually varies in different wheat phenology stages, with different frequency and accuracy depending on the measurement schedule and techniques. This experiment aimed to explore the minimum requirement of observation availability with regard to interval, stages and accuracy such that the data assimilation resulted in a more correct yield estimation.

The discussion of phenology stages is limited to the three long stages: stage 4, the stage from the end of juvenile to floral initiation, stage 5, the stage of floral initiation to flowering (or anthesis), and stage 6-7 (discussed as a composite because stage 6 is very short), the stage from flowering to the end of grain filling. The state variables observations were assumed to be available for a range of frequencies and phenology stages, being every 4, 8, 12, and 16 days, and stages 4, 5, and 6-7, respectively. Particularly, the assimilation of soil states before the sowing date was also tested.

The observation accuracy impact was tested by assimilating a group of synthetic observations degraded to have a set of uncertainty levels ranging from 0 to 7, with the uncertainty values defined in Table 6.5. The degraded observations of LAI and SM1 were generated by adding Gaussian errors with fixed standard deviation to the truth (i.e., the random noise was additive), while the uncertainties of other state variables were assumed multiplicative,

Table 6.5: Uncertainty values applied to wheat and soil state variables in each uncertainty level.

State variable(s)	Additive/ Multiplicative?	Uncertainty level							
		0	1	2	3	4	5	6	7
LAI	A	0	0.1	0.3	0.5	1	1.5	2	2.5
SM1	A	0	0.01	0.02	0.05	0.1	0.15	0.2	0.25
Other states	M	0	2%	5%	10%	15%	20%	30%	50%

generated as Gaussian errors with the standard deviation values proportional to the value of the true state variables at each time step.

6.1.3.4 Assimilation when phenology is constrained

This experiment was based on a simple assumption that all phenology stages are observable and accurately measured, to provide an understanding of the role of phenology in APSIM-Wheat modelling. While the model internally simulates the development of phenology stages, this approach was disabled for this assimilation experiment. Instead, the phenology dates of all ensembles were uniformly set to the observations obtained from the truth run.

6.2 Results and discussion

6.2.1 Ensemble size

Figure 6.2 shows the standard deviation (std.) of the ensemble for LAI and soil moisture in the surface layer (SM1) from open-loop simulations with different ensemble size. With the increase in ensemble size, the standard deviation of ensembles tended to become more stable. The standard deviation of LAI became stable when the ensemble reached 50 members. Although the difference of standard deviation estimated for SM1 from different ensemble sizes was generally small, the standard deviation of soil moisture was also found to become more stable when the ensemble size reached 50. This indicates that the ensemble size can be reduced to 50 without significantly degrading the representation of model uncertainty distribution using a finite number of ensembles. Therefore, an ensemble size of 50 was selected as being adequate to represent the probability distribution of the stochastic APSIM-Wheat model, and was selected for the following data assimilation experiments.

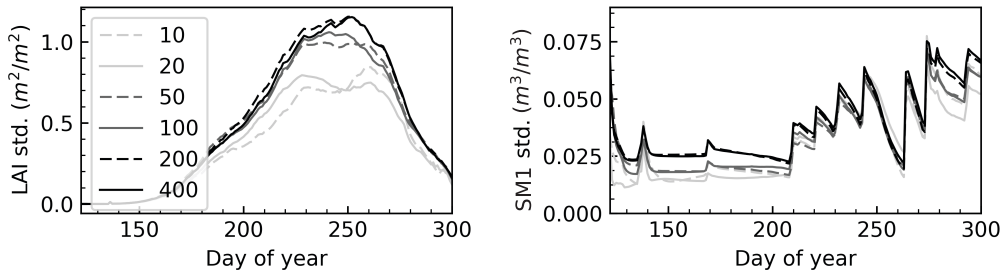

Figure 6.2: Ensemble standard deviation of LAI (a) and SM1 (b) estimation from open-loop simulations with different ensemble size.

6.2.2 Assimilation of state variable types

6.2.2.1 Correlation between wheat and soil states

A strong correlation was found to exist within each group of state variables. The assimilation of any of the states in the wheat or soil group was able to correct the estimation errors of other states in the same group, but the impact on states in the other group was negligible. This is found from the root mean square error (RMSE) of all state variables (Tables 6.6 and 6.7) as a result of assimilating a single state variable type, respectively. With the assimilation of a single wheat state type, the RMSE of all wheat state variables were reduced compared to the open-loop simulation, but no distinct reduction was found in the RMSE of soil states. Correspondingly, the assimilation of any of the soil moisture, ammonium and nitrate state variables were able to give a better estimation of the rest of the soil state types and in all 7 layers, but with little impact on the wheat state variables.

The same phenomenon was found in the time series of GrainWt, LAI and SM1 shown in Figures 6.3 - 6.4, where the assimilation results of LAI and SM1 were presented, as two representative state variables among each group. In the LAI assimilation experiment, the time series of LAI were updated to approach the truth with the introduction of external LAI observations into the model (Figure 6.3-b), but the values of soil moisture (Figure 6.3-c/-d, SM1 in the near-surface and SM2 in the root-zone) were only slightly changed at the grain-filling stage (phenology stage 7). In the SM1 assimilation experiment, the posterior LAI was only slightly different from the prior values at each time step the SM1 observations were assimilated (Figure 6.4). As a result, LAI was slightly better estimated during the leaf growth stage, but was over-estimated afterwards when the leaves withered.

The weak correlation between wheat and soil states can be explained by the weak link

Table 6.6: Relative difference (RD) of yield and root mean square error (RMSE) of wheat and soil states with the assimilation of a single type of state variables.

Assimilated state	RD of yield (%)	RMSE of states (compared to the truth)													
		LAI	Grain Wt	Leaf Wt	Stem Wt	Leaf N	Stem N	Pod N	SM1 ($\cdot 10^{-2}$)	SM4 ($\cdot 10^{-2}$)	SM7 ($\cdot 10^{-2}$)	NH4N1	NH4N4 ($\cdot 10^{-4}$)	NO3N1	NO3N4
Open-loop	10.1	0.5	44	22.8	100.7	0.8	0.59	0.16	4.9	1.9	1.6	0.12	7	3.9	1.1
LAI	0.8	0.32	41.1	13.8	62.1	0.55	0.6	0.15	5	1.8	1.5	0.12	4	2.4	1.1
GrainWt	9.9	0.36	27.8	15.6	60.5	0.48	0.3	0.11	4.9	2	1.7	0.12	5.9	3.9	1.1
LeafWt	-2.6	0.19	35.7	8.4	55.1	0.35	0.58	0.14	5	1.6	1.7	0.12	3.3	2.8	1.1
StemWt	1.7	0.34	22.5	14.3	62.7	0.39	0.3	0.12	4.5	1.3	1.6	0.11	5.5	3.7	1.1
LeafN	3.8	0.3	21	13	74.8	0.27	0.27	0.14	4.6	1.4	1.5	0.11	3.9	2.4	1.1
StemN	11.8	0.41	36.5	17.6	57.2	0.58	0.27	0.12	4.8	1.9	1.6	0.12	4	2.7	1.1
PodN	11.4	0.41	35.6	17.5	67.2	0.61	0.36	0.08	4.9	2	1.5	0.12	5.3	3.9	1.1
SM1	0.8	0.55	42.9	23.8	82.3	0.72	0.57	0.18	3.8	0.6	1.2	0.05	5.3	3.6	1.1
SM4	7.4	0.64	41.5	27.7	93	0.82	0.57	0.17	3.8	0.6	1.2	0.1	5.3	3.9	0.8
SM7	7.5	0.5	43	21.5	96.6	0.78	0.61	0.16	4.8	1.5	1.4	0.12	5.7	3.2	0.4
NH4N1	-2	0.58	32.4	25.6	83.4	0.59	0.5	0.19	3.3	0.7	1.3	0.05	3.8	3.5	1.2
NH4N4	12.7	0.82	43.6	35.7	102.4	1.03	0.58	0.19	4.8	1.5	1.9	0.11	2.4	3.3	1.1
NH4N7	11.3	0.56	43.5	24.2	100.6	0.86	0.62	0.16	4.7	1.7	1.8	0.12	7.2	4.9	0.8
NO3N1	-5.5	0.42	43.9	18.8	78.7	0.54	0.59	0.18	3.3	1	1.3	0.08	4.4	1.5	0.3
NO3N4	-2.2	0.54	39	23.8	54.4	0.8	0.52	0.15	4.5	1.6	2.7	0.1	5.1	3.2	0.2
NO3N7	7.8	0.49	40.9	21.3	90.6	0.78	0.61	0.15	3.5	1.5	1.6	0.09	5.6	4	0.6

Table 6.7: RD of yield and RMSE of wheat and soil states with the assimilation of combined types of state variables.

Assimilated states	RD of yield (%)	RMSE of states (compared to the truth)													
		LAI	Grain Wt	Leaf Wt	Stem Wt	Leaf N	Stem N	Pod N	SM1 ($\cdot 10^{-2}$)	SM4 ($\cdot 10^{-2}$)	SM7 ($\cdot 10^{-2}$)	NH4N1	NH4N4 ($\cdot 10^{-4}$)	NO3N1	NO3N4
Open-loop	10.1	0.5	44	22.8	100.7	0.8	0.59	0.16	4.9	1.9	1.6	0.12	7	3.9	1.1
Assimilation of multiple wheat states															
LAI, LeafWt	4.9	0.21	33.9	9.1	51.4	0.39	0.54	0.15	4.7	1.5	1.7	0.11	3.4	2.6	1.1
LAI, StemWt	7.2	0.29	26.4	12.5	49.6	0.35	0.24	0.11	5.4	1.9	1.6	0.13	3.5	3.9	1.1
LAI, LeafWt, StemWt	6.1	0.2	27.2	8.4	42.4	0.32	0.36	0.12	5.1	2.1	1.6	0.12	3.1	3.1	1.1
LeafWt, StemWt	-1	0.19	25.7	8.1	42.9	0.27	0.4	0.12	4.9	1.6	1.5	0.11	2.1	2.7	1
Assimilation of multiple soil states															
SM1, NH4N1	1.85	0.5	34.1	21.9	82.3	0.64	0.51	0.17	1.3	0.7	1.1	0.05	4.3	6.6	1.1
SM1, NO3N1	-2.99	0.24	32.1	10.7	47.4	0.36	0.44	0.13	1.3	0.8	1.6	0.05	3.2	1.2	0.9
NO3N1, NH4N1	0.2	0.58	32.8	25.6	76.6	0.58	0.45	0.19	3.2	0.8	1.7	0.05	3.4	1.3	0.5
Assimilation of mixed wheat and soil states															
LAI, SM1	-5	0.3	41.9	13.3	51	0.48	0.58	0.17	1.3	0.7	1.5	0.05	4.5	2.2	1
LAI, SM1, NH4N1	-3.9	0.24	34.1	10.4	47.2	0.37	0.47	0.13	1.4	0.7	2.1	0.05	3.2	4.4	1.6
LAI, SM1, NO3N1	-3.6	0.25	36.4	11	45.1	0.42	0.49	0.14	1.6	0.8	1.6	0.06	3.4	1.2	0.6
LAI, NO3N1, NH4N1	-9.6	0.22	39.2	10	43.3	0.31	0.47	0.15	3.4	0.9	1.8	0.05	3	1.4	0.5

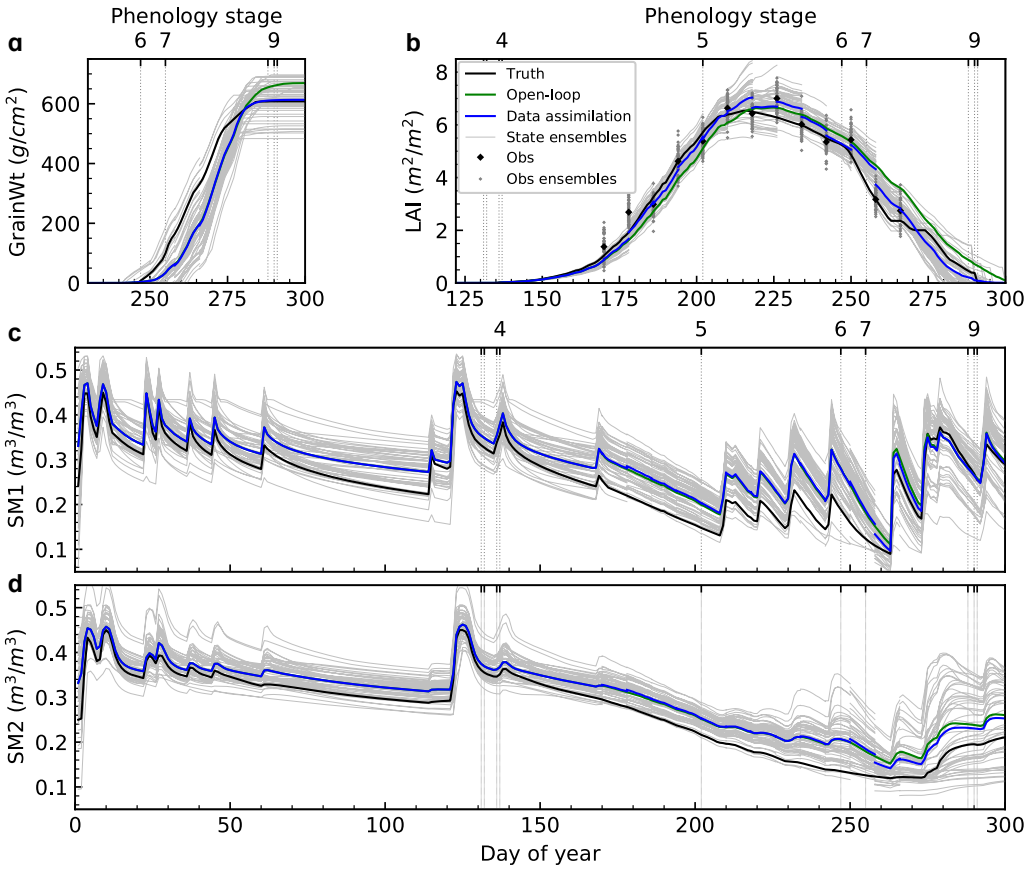

Figure 6.3: Evolution of GrainWt (a), LAI (b), SM1 (c) and SM2 (d) in the LAI-assimilation experiment. Legend applies to all subsequent figures in this chapter.

between the wheat and soil modules. The two modules were primarily linked by extractable water and nutrients, and the mechanism of water and nutrients to impact wheat growth is accumulative and slow. Thus the abrupt change of states in one module does not immediately affect the value of states in the other module. The sensitivity analysis also showed that the change of states either in the wheat or the soil group only affect the other group in stage 6-7, giving a weak correlation of errors between the two groups. Therefore, the states in one group are not significantly affected when assimilating states in the other group.

6.2.2.2 Impact on soil states among layers

The assimilation of soil states in one of the soil layers can correct errors of the same states in other soil layers through the correlation of errors. As shown in Table 6.6, the RMSE of soil moisture, ammonium and nitrite in the layer where the observation was assimilated is significantly less than the open-loop, with adjacent layers receiving stronger impact than

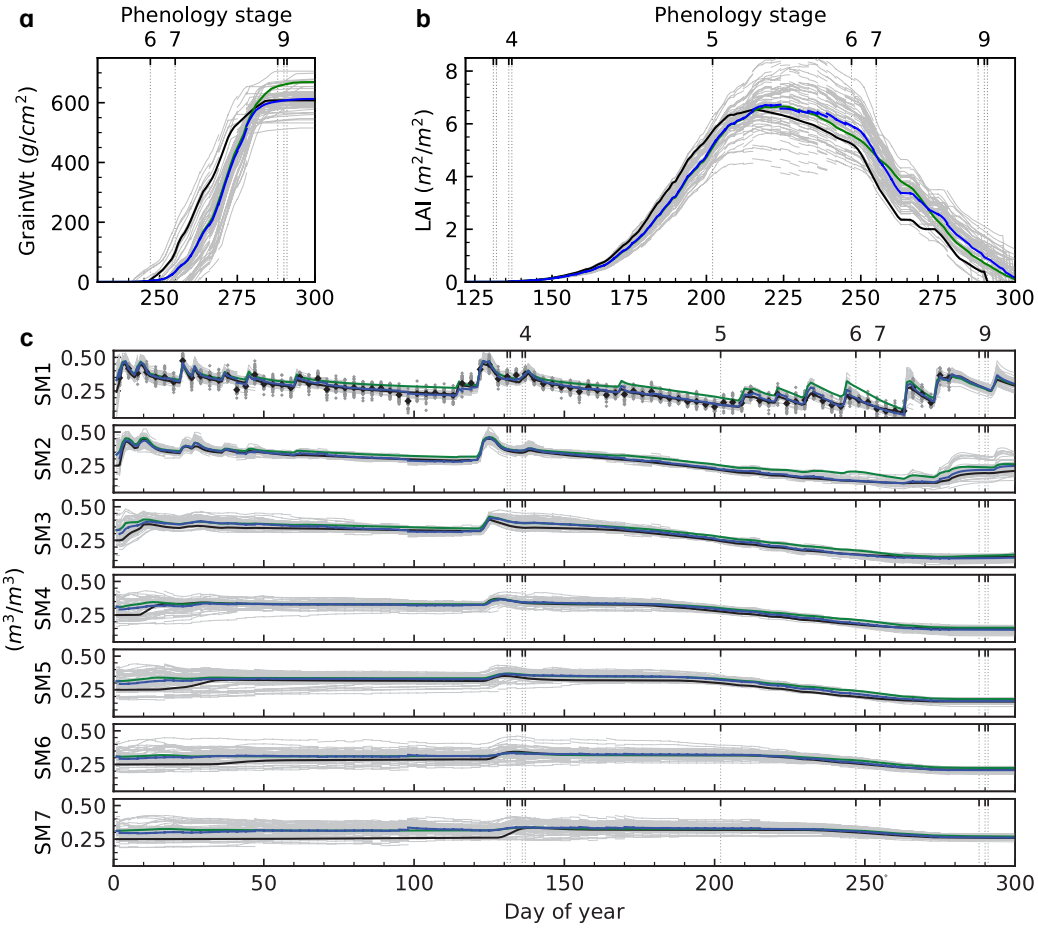

Figure 6.4: Evolution of GrainWt (a), LAI (b) and soil moisture in all soil layers (c) in the SM1-assimilation experiment. See Figure 6.3 for legend.

those far from the layer where observations were assimilated. For an instance of SM1 assimilation, SM1 was well fitted to the truth over the whole simulation time span (Figure 6.5). This improvement of soil moisture estimation was found in all layers and at nearly all stages, with deeper layers receiving a smaller less correction of soil moisture than top layers (Figure 6.4).

6.2.2.3 Impact on GrainWt and yield estimation

By assimilating wheat and soil states over the whole growing season, LAI, LeafWt and StemWt among wheat states were found in Table 6.5 to improve yield estimation. The error of the estimated yield caused by uncertainties was partially corrected by the assimilation of these states, as found in the relative difference of yield reduced from 10.1% in the open-loop to 0.8%, -2.6% and 1.7%, respectively (Table 6.6).

The assimilation of any soil state contributed to a better yield estimation, and the assimilation of soil states in top layers was more effective than bottom layers in improving yield estimation. Nearly all relative difference (RD) of yield given by the assimilation of soil states in the different soil layers was found to be less than the open-loop result, indicating a better yield estimation with data assimilation. It is intuitive to understand that errors of yield estimation caused by the rainfall uncertainty could be corrected by updating soil moisture. The better yield estimation caused by assimilating soil nitrogen states is likely to be associated with the soil moisture, which is the key state variable to affect APSIM yield estimation according to the previous sensitivity analysis, and is also better estimated due to the strong association among state variables in the soil group.

A better time series of GrainWt is not necessarily linked to a better yield at harvest, and vice versa. An example is in the assimilation of some wheat state types (e.g. GrainWt, StemN, PodN) which were found to give a better time series of estimated GrainWt. However, the yield at harvest remained nearly unchanged from the open-loop. Another clue is found in the soil moisture assimilation: although the yield at harvest was well-fitted to the truth, no direct update of GrainWt was found with the assimilation of those soil states during the grain filling stages. This phenomenon is likely due to the mechanism of how APSIM models the grain filling. The grain demand is determined by StemWt at flowering, cultivar parameters and stress factors affected by temperature and nitrogen, and that once it is fulfilled, the daily incoming biomass is not allocated to grain but to other organs. Therefore, even if GrainWt is accumulated daily, the yield at harvest is constrained by a maximum GrainWt under the impact of other factors.

6.2.2.4 Assimilation of combined states

The combined assimilation of multiple wheat state types improved the estimation of all wheat states and gave a more correct yield at harvest. This was found in the assimilation of the combination of LAI, LeafWt and StemWt, were the RMSE of all wheat states and the RD of yield were lower than the open-loop (Table 6.7), indicating a better-estimated wheat state combination and a partially corrected yield. Similar to the combination of wheat states, the assimilation of multiple soil state types improved the estimation of all soil states and gave a more correct yield at harvest, as the RMSE of all soil states were reduced, and the yield was

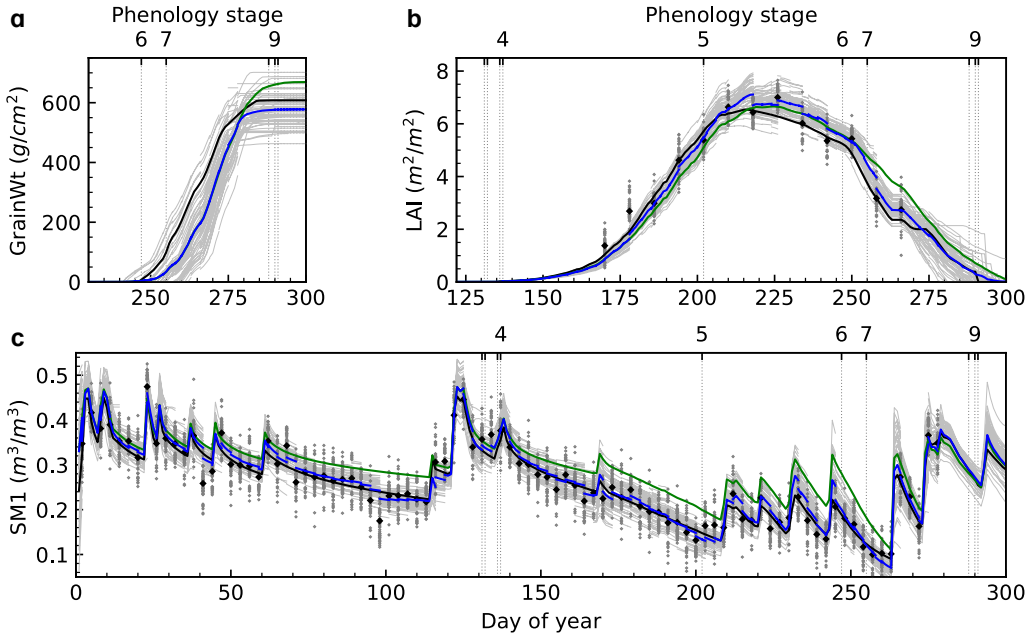

Figure 6.5: As for Figure 6.3 but for the combined LAI/SM1 assimilation experiment. See Figure 6.3 for legend.

more close to the truth (Table 6.7).

The combined assimilation of mixed wheat and soil state types improved the estimation of almost all the wheat and soil states, but the yield was sometimes over-corrected. For instance, the combined assimilation of LAI and SM1 on the same date resulted in the RMSE of all state variables being smaller than the open-loop, but the RD of yield resulting from the combined assimilation was lower than zero (Table 6.7). This over-correction indicates a conflict between the state variables in the wheat and soil groups: when the assimilation of one group of states impacts the estimated yield, but the states in the other group almost remained unchanged. Therefore, solely assimilating either wheat or soil states led to a lowered yield estimation that approached the truth, while the combined assimilation seemed to amplify this reduction and give an underestimated yield at harvest. The cause of this conflict is probably twofold. Firstly, when LAI was assimilated while other wheat states were not constrained, the assimilation of LAI may push other wheat states in a wrong direction that leads to worse yield estimation. Secondly, although data assimilation was able to constrain and better estimate some states, the uncertainties caused by some cultivar parameters cannot be cancelled because their parameters directly control the grain filling process that is not affected by state variables.

6.2.3 Impact of observation availability

6.2.3.1 Impact of observation availability in various growth stages

The wheat phenology stage when observations are assimilated is critical for the assimilation of wheat states. Table 6.8 shows the relative difference of yield by assimilating single state variable types at different frequencies and stages. It is found that the assimilation of wheat states, including LAI, GrainWt, LeafWt, StemWt, LeafN and StemN contributed to better yield estimation when they were assimilated in specific stages. Generally, stage 6-7 is the optimal stage for the assimilation of wheat states, second to which is stage 5, while the assimilation of wheat states in stage 4 had nearly no improvement for yield estimation. This is further discussed below.

Assimilating wheat states in early stages (stage 4) neither corrected wheat state variables in late stages nor improved yield estimation. All the wheat states assimilation experiments failed when assimilated in stage 4 only, as indicated by the RD of yield being higher than the open-loop. The reason for poorer yield estimation is that the assimilation at a very early stage was not able to give correct state variables in medium stages (at around stage 5) and thus had no benefit to the late stages where the grain starts to develop.

Observation of LAI, LeafWt, LeafN and StemN benefited yield estimation when assimilated in stage 5, the only stage where assimilation of StemN was successful. The RD of yield given by assimilating these wheat states in stage 5 was slightly lower than for the open-loop, with the time series of soil states remaining unchanged. This indicates that the discrepancy of yield in each ensemble caused by uncertainties before the grain filling stage can be reduced by constraining the wheat states. It can be concluded that more correct wheat states before flowering and grain filling stage can improve the yield estimation.

Stage 6-7 (flowering to the end of grain filling) was the optimal stage for the assimilation of LAI, LeafWt, and StemWt. Significantly lower RD of yield was found when these state variables were assimilated in stage 6-7, compared to the open-loop and the data assimilation in stage 5. Both stages 5 and 6-7 are optimal for LeafN. However, PodN, and StemN when assimilated every 4 days did not benefit yield estimation in stage 6-7.

Soil states were less sensitive to phenology states and could be assimilated anytime during the whole growing stages after sowing, with observations in stages close to grain

Table 6.8: Relative difference of yield with the assimilation of state variables at various frequency and at different stages.

Assimilated state	Before sowing	Stage 4				Stage 5				Stage 6-7				
	4 d	4 d	8 d	12 d	16 d	4 d	8 d	12 d	16 d	2 d	4 d	8 d	12 d	16 d
	RD of yield (%)													
Open-loop	10.1													
LAI	-	11.3	12	12.1	12.2	7.3	7.2	5.6	6.5	-	-2.1	1.4	0.8	0.7
GrainWt	-	-	-	-	-	-	-	-	-	2.4	9.3	12.4	9.5	11.9
LeafWt	-	10.7	12.3	11.8	11.8	8	6.5	7.3	6.9	-	0.9	2.9	2.3	5.8
StemWt	-	12.2	12.5	12.2	12	12	11.7	11.7	12	-	-5.3	4.4	6.8	6.3
LeafN	-	10.7	11.2	11.6	11.4	8.2	7.1	7.7	6.9	-	6	8.6	7.2	-1.5
StemN	-	11.9	11.5	11	11.5	9.1	9.8	9	9.2	-	12.7	5.6	6.4	6.1
PodN	-	-	-	-	-	-	-	-	-	-	14.9	14.8	16.4	9.5
SM1	10.2	5.3	6.3	7	7.2	2	6.1	5.8	8	-	2.8	1	1.1	5
NH4N1	8.9	4.4	6	3.8	3.4	-1.4	-1.1	0.2	1.1	-	-0.2	14	14.4	9
NO3N1	11	9.1	10.7	11.1	10.7	0.8	1.2	3.4	2	-	4.8	6.5	5.9	7.9

filling giving the better yield estimation. Table 6.8 suggests that assimilating SM1 in stage 6-7 gives the lowest RD of yield, followed by assimilating in stage 5 and 4, respectively. Nevertheless, the observations of soil ammonium and nitrate in the first layer (NH4N1 and NO3N1) were more useful in stage 5, compared to stage 4 and 6-7

The RD of yield values given by assimilation of soil states before sowing showed no difference from the open-loop. This is due to the fertilization activity and continuous rainfall at around the sowing date. The soil nitrogen before sowing was small and sharply increased at sowing because of fertilization, and the continuous rainfall led to a nearly saturated soil profile, and thus the uncertainties of soil moisture after saturation depended only on the uncertainties of rainfall and the soil water retention properties in this case. Although the soil moisture was in a more correct state after updating with its observations, the states are reset after being extremely wet or dry, and thus, the effect of SM assimilation is minimalized.

6.2.3.2 Impact of observation availability in various repeat interval

The influence of data assimilation frequency was not as critical when observations were assimilated in the optimal stage. Table 6.8 shows the RD of yield when the observation was available at different frequencies at the optimal stages. Overall, an interval of 16 days provided an improved yield estimation when states were assimilated in the optimal stage. According to Table 6.8, when GrainWt, LeafWt, and StemWt were individually assimilated in stage 6-7, yield was clearly better estimated when the frequency of observations changed from 16 to 4 days. This suggested that the more frequent the observations in this stage, the better the yield estimated. Nevertheless, there is no point in collect intensive observations of wheat states in stage 5 as the assimilation of LAI, LeafWt, LeafN and StemN are not likely to affect the yield forecast by the assimilation frequencies at this stage. However, the assimilation frequency was found to have a clear impact on SM1 assimilation through all the growing stages, as a shorter observation interval gave a lower RD of yield for SM1 in all stages in Table 6.8. The assimilation of GrainWt was found to benefit yield estimation in stage 6-7, only if the observations were available at least every 4 days.

Table 6.9: Relative difference of yield with the assimilation of state variables in various uncertainty levels (defined in Table 6.5) in the optimal stage.

Assimilated state	Assimilated stage (optimal)	Observational uncertainty level (when assimilated in the optimal stage)							
		0	1	2	3	4	5	6	7
		RD of yield (%)							
Open-loop		10.1							
LAI	Stage 6-7	-2.6	5	1.2	2.6	5.8	7.5	8.7	10
LeafWt	Stage 6-7	-8.4	-6.2	-4.2	1.9	-2.2	1.1	-11.4	7.1
StemWt	Stage 6-7	-3.6	-0.2	-6.5	3.8	7.6	8.6	6.7	10.1
LeafN	Stage 5[a]	8	8.2	9	3.8	8.9	-	-	-
StemN	Stage 5	10.4	8	10.5	10.8	10.2	-	-	-
SM1	Stage 6-7	3.1	2	3.5	5.8	6.6	6.4	10.3	10.6
NH4N1	Stage 5	-0.8	0.1	0.4	1.1	-0.6	-	-	-
NO3N1	Stage 5	6.5	6.5	9	8.7	9.4	-	-	-

[a] Both stage 5 and 6-7 are optimal for LeafN. Here an exampling of assimilating LeafN in stage 5 was presented.

6.2.3.3 Impact of data quality

Large uncertainties in observation did not have a negative impact on data assimilation if they were correctly estimated in the error covariances. As presented in Table 6.9, when the uncertainty level increased (changed from level 0 to 7), no distinct deterioration was found in the performance of yield estimation, with the RD of yield estimated by assimilating LeafWt, LeafN, StemN, NO4N1 and NO3N1 being lower than the open-loop (10.1%). However, LAI and soil moisture were found to benefit the yield estimation when the uncertainty increased to 2 m^2/m^2, and 0.15 m^3/m^3, respectively. The StemWt improved the yield estimation when the relative uncertainty increased to 30%. This is explained by how the EnKF algorithm deals with uncertainties, with less weight given to observations when the observational uncertainties are larger than the ensemble spread, and thus the posterior states are more close to the model estimation at each step the state variables are updated. Therefore, observations with larger uncertainties are trusted less and thus have less impact on the posterior states.

6.2.4 Assimilation when phenology is constrained

By direct-insertion of known phenology dates into the model, the uncertainties of all wheat state variables and yield were reduced. As shown in Figure 6.6, the spread of LAI

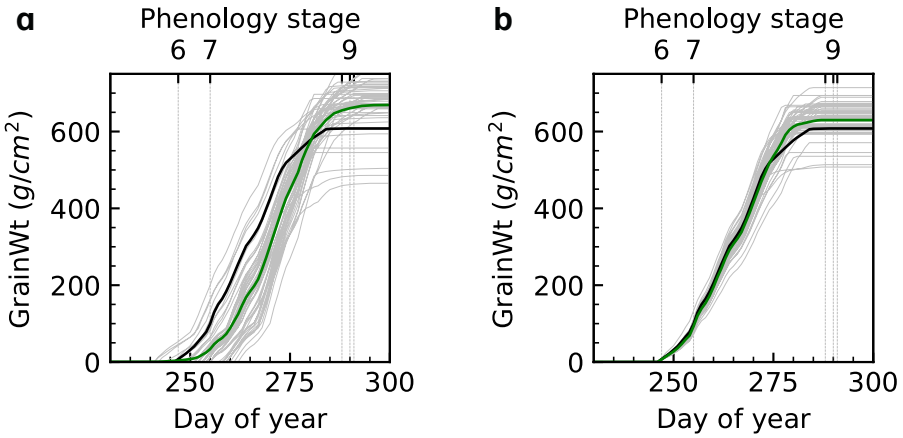

Figure 6.6: The GrainWt estimated in open-loop with: (a) phenology unknown, and (b) phenology constrained. See Figure 6.3 for legend.

and GrainWt of the open-loop from known phenology was clearly smaller than that from an unknown phenology. This significant reduction of ensemble spread in wheat states was due to the mechanism of the APSIM modelling phenology. As described in Chapter 3, the phenology shifts when the thermal time (the accumulation of temperature over time), reaches a target value, with the fraction of daily biomass increment partitioned to each wheat organ depending on the phenology stage. Thus, uncertainty in the temperature leads to a discrepancy of phenology among ensembles, with the different time length of each phenology stage determining the timespan of organs to develop and accumulate biomass. This effect is eliminated when the errors of phenology are completely corrected by the direct insertion of true phenology dates. Therefore, the assimilation of phenology was found to significantly reduce the errors of estimated state variables and yield caused by the uncertainties of temperature.

6.2.5 Data availability from remote sensing

This synthetic study demonstrated the potential of assimilating LAI, wheat organ weight, and soil moisture in improving yield estimation of APSIM-Wheat. Table 6.10 summarized the key state variables that improved yield estimation in this study, with the assimilation interval and observation accuracy. The observations of these wheat and soil states in the APSIM-Wheat model can be obtained from remote sensing, field sampling, and laboratories analysis. LAI and surface soil moisture can easily be obtained from current remote sensing techniques (e.g., Landsat, MODIS), while sub-surface soil moisture and phenology

Table 6.10: Summary of the key assimilated states, with minimum assimilation interval and observation accuracy that improved yield estimation.

State	Assimilation stages		Minimum assimilation interval and observation accuracy	
	Optimal	Sub-optimal	Internal	Accuracy
LAI	Stage 6-7	Stage 5	16 days	2 m^2/m^2
GrainWt	Stage 6-7	None	4 days	-
LeafWt	Stage 6-7	Stage 5	16 days	-
LeafN	Stage 5, 6-7	None	16 days	-
StemWt	Stage 6-7	None	16 days	30% (relative)
StemN	Stage 5	None	16 days	-
SM1	Stage 6-7	Stage 4 & 5	16 days	0.15 m^3/m^3
NH4N1	Stage 5	Stage 4	16 days	-
NO3N1	Stage 5	Stage 6-7	16 days	-

currently requires direct observation in the field. Techniques to detect crop phenology from remote sensing only focuses on the dates of green-up, peak LAI and senescence, based on the photosynthesis of plant leaves, with the detection of booting and anthesis date in the reproductive stage unavailable (e.g., Reed *et al.* , 1994; Zhang *et al.* , 2003; Vina *et al.* , 2004; Sakamoto *et al.* , 2005; Boschetti *et al.* , 2009; You *et al.* , 2013). To obtain nitrogen and biomass amount of plant organs, field sampling and laboratory analysis currently need to be employed.

The existing remote sensing LAI products from satellites satisfy the requirement of LAI assimilation for APSIM. Several studies validating LAI products with ground measurements reported the uncertainties of LAI products from a range of satellites. For instance, MODIS LAI products have been reported in several studies to have an uncertainty (RMSE) of 0.38 m^2/m^2 in grass and cereal crop areas (Tan *et al.* , 2005), 0.63 m^2/m^2 in broadleaf forest (Serbin *et al.* , 2013), 0.66 m^2/m^2 in biomes including grass, crop, shrubs, savannas and forest (Yang *et al.* , 2006), and 1.14 m^2/m^2 at global scale (Weiss *et al.* , 2007), respectively. These uncertainties are in the range of 0-2 m^2/m^2, beyond which nearly no correction was found in the APSIM yield estimation by updating LAI. Clear improvement of yield estimation was found when the uncertainty of LAI was below 1 m^2/m^2, which is satisfied by the MODIS observational uncertainties according to validation experiments (Tan *et al.* , 2005; Serbin *et al.* , 2013; Fensholt *et al.* , 2004). Therefore, the existing remote sensing technology can provide LAI remote sensing data with sufficient quality for the success of LAI data

assimilation for the APSIM model.

The existing remote sensing surface soil moisture products from satellites satisfies the requirement of SM1 assimilation for APSIM. The goals of the Soil Moisture and Ocean Salinity (SMOS) mission was to deliver soil moisture maps with accuracy (unbiased RMSE) of 0.04 m^3/m^3 (Kerr *et al.* , 2012), and the soil moisture product has been validated to meet the mission requirement (Jackson *et al.* , 2012; Sanchez *et al.* , 2012; Al Bitar *et al.* , 2012). The Soil Moisture Active Passive (SMAP) mission was found to meet the designed accuracy of 0.04 m^3/m^3 for the passive and 0.06 m^3/m^3 for the active product (Colliander *et al.* , 2017; Chen *et al.* , 2017). According to Table 6.9, soil moisture uncertainty less than 0.15 m^3/m^3 had an improvement on yield estimation. Thus, the existing remote sensing technology can provide surface soil moisture remote sensing data with sufficient quality for the success of SM1 data assimilation for the APSIM model.

6.3 Chapter summary

This chapter presented a synthetic study exploring the potential of assimilating all wheat and soil state variables found to be important from a sensitivity analysis into APSIM for an improved wheat growth and yield estimation. The state variables that improved yield estimation included LAI, GrainWt, LeafWt, StemWt, LeafN, and StemN and all soil states. Phenology stage 6-7 was found to be the optimal assimilation stage for LAI, LeafWt, StemWt, LeafN, and soil moisture, and stage 5 was optimal for LeafN, StemN and soil nitrogen. Among these state variables, the observation requirements of LAI and near-surface soil moisture are satisfied by the current remote sensing techniques. Accordingly, these results support the hypothesis of this thesis, that assimilation of APSIM state variables has the potential to improve wheat monitoring and yield estimation, laying the foundation for assimilating real-world data into APSIM in the subsequent two chapters.

Chapter 7

Case study 1 - assimilation with ground measurements

The previous chapter has demonstrated the potential of assimilating all prognostic wheat and soil states into APSIM-Wheat in a synthetic study, identifying several states that have the potential to improve yield estimation. In the synthetic study, the uncertainties in the system were assumed from specific known sources, being Gaussian and unbiased. However, in realistic situations, the uncertainties in the model and observations may not follow such assumption, sometimes being difficult to identity and estimate. Therefore, the performance of data assimilation in the APSIM model needs to be further validated with realistic datasets. This chapter further assimilates observations of wheat and soil observations that have been directly obtained from field measurement into the APSIM-Wheat model, including variables that can be remotely sensed and those that can only be observed in the field or laboratory. The outcome of this study informs what is included in the remote sensing case study of Chapter 8.

7.1 Study site and dataset

The study site is the 75×75 m square wheat field located in Cora Lynn, Victoria, Australia, as described in detail in Chapter 4. This case study uses the weather and soil properties as the model input, and field-observed wheat and soil information as the external state observations and validation data.

Weather and soil parameters are the basic input data for APSIM. The weather forcing includes daily rainfall, solar radiation, daily maximum and minimum air temperature, vapour pressure, pan evaporation and wind speed. The uncertainties of rainfall, radiation and temperature estimated from several nearby weather stations were used to perturb the weather forcing. Soil properties and chemical composition were from the field measurements described in Chapter 4.

Field-observed wheat and soil data including ground measured LAI, biomass (total and separately for leaf and stem), soil moisture of each layer, soil nitrogen in the form of nitrate (NO3N) in the top two layers were the state variable observations assimilated into APSIM. The model output was evaluated with independent observations including phenology and average yield collected from four experimental plots.

7.2 Methodology

7.2.1 Model setup and calibration

The physically-based APSIM-Wheat model was used to simulate wheat growth with a daily time step, accounting for the interaction of the plant with the environment. The wheat module considers winter wheat phenology by ten phases (or nine stages, noting that the phenology stage is the period between two adjacent phases) as described in Chapter 3, controlled by the temperature and day length, and affected by vernalization and stress factors. The biomass accumulation is based on a simple radiation use efficiency light utilization approach. Total daily incoming biomass is allocated to the wheat organs (leaf, stem, pod and grain) with a proportion that varies with the phenology stage. The model considers plant extractable water and soil nitrogen as two stress factors to the growth of wheat. A water balance model is coupled with the wheat module to simulate soil water movement and estimate plant extractable water. The model considers rainfall-runoff, evapotranspiration, infiltration, unsaturated flow, saturated flow, and lateral flow processes. The soil nitrogen module uses three organic matter pools (fresh, hum and biom pools) to simulate the soil carbon and nitrogen transformation in the processes of decomposition, nitrification/denitrification, and mineralization/immobilization.

In this case study, the simulation window was set from June 1, 2018 to March 1, 2019,

Table 7.1: The calibrated soil parameters in the SoilWat and SoilN modules.

Parameter	Description	Layer	Default value	Calibrated value
SoilWat module				
ConA summer ($ConA_{summer}$)	A regression coefficient that specifies the change of cumulative evaporation in the second stage of soil evaporation calculation. Applied to summer days.	-	5	3.8
U summer (U_{summer})	Amount of cumulative evaporation before soil supply decreases below the atmospheric demand. Applied to summer days.	-	5	3.8
ConA winter ($ConA_{winter}$)	Same as the $ConA_{summer}$ but for winter days.	-	5	1.5
U winter (U_{winter})	Same as the U_{summer} but for winter days.	-	5	1.5
SummerDate	Start date for switching to summer parameters for soil water evaporation.	-	1-Nov	19-Nov
DiffusConst	Parameter in the soil water diffusivity calculation.	-	40	400
SWCON	Fractional amount of water above drain upper limit that can dry under gravity per day.	1	0.3	0.02
		2	0.3	0.1
		3	0.3	0.1
		4	0.3	0.1
		5	0.3	0.3
		6	0.3	0.3
SoilN module				
Fbiom (F_{biom})	The biom pool carbon as a fraction of the hum carbon that is subject to decomposition.	1	0.04	0.1
		2	0.04	0.05
		3	0.02	0.03
		4	0.02	0.02
		5	0.02	0.01
		6	0.02	0.01

with sowing date on August 7, 2018. Seed was drilled to 5 cm depth with 100 kg/ha mono-ammonium phosphate (MAP, containing10% ammonium nitrogen) fertilizer applied. The row spacing was 0.2 m and the population was 251 plants/m^2.

The soil was described by 6 layers, with depths of 0-5, 5-15, 15-25, 25-35, 35-45, 45-55 cm. Several parameters were calibrated (Table 7.1) due to the high sensitivity of the model to soil states: several parameters in the SoilWat module were calibrated manually to fit the station soil moisture data in all six layers, and a parameter in the SoilN module controlling soil nitrogen dynamics was calibrated manually to fit the variation trend of ground measured soil nitrate in the first two layers. The retained parameters were set to measured values in the first four layers, with deeper layers assumed to be the same as those in the fourth layer.

7.2.2 Data assimilation setup

The Ensemble Kalman filter (EnKF) data assimilation algorithm was coupled with APSIM. The algorithm is described in Chapter 3 so only the specifics for this case study are included here. With **x** being the state vector of APSIM, it consists of 31 state variables and is expressed by:

$$\mathbf{x} = [\text{LAI, SM1, ..., SM6, LeafWt, StemWt, GrainWt, PodWt, RootWt, LeafN, StemN, GrainN, PodN, RootN, Height, RootDepth, NO3N1, ..., NO3N6, ..., NH4N1, ..., NH4N6}]^{\mathrm{T}}, \tag{7.1}$$

where, LAI is the leaf area index, SM is the soil moisture (volumetric soil water content), and NO3N and NH4N are the soil nitrogen in the form of nitrate (NO3) and ammonium (NH4), respectively, with numbers 1-6 denoting the number of soil layers. The Wt and N are the dry weight and nitrogen of the wheat leaf, stem, grain and pod, respectively. The state matrix consists of state vectors of ensemble members expressed by:

$$\mathbf{X} = [\mathbf{x}^1,\ \mathbf{x}^2,\ ...,\ \mathbf{x}^{\mathrm{N}}],$$

where N is the ensemble size (50 in this study), and the superscripts 1, 2, ..., N refer to the i^{th} ensemble member. Ensembles were generated by perturbing weather inputs, model parameters and observations. In this study, the model forecast was stochastic, with the underestimation of wheat states and yield caused by the uncertainties of weather and parameter inputs and the model physics. The forecasting equation is a Markov chain that transits the states from step $\mathrm{k}-1$ to k, expressed by:

$$\mathbf{x}_{\mathrm{k}}^{\mathrm{b}} = \mathcal{M}(\mathbf{x}_{\mathrm{k}-1}^{\mathrm{a}},\ \mathbf{f}_{\mathrm{k}} + \mathbf{w}_{\mathrm{k}},\ \theta), \tag{7.2}$$

where the superscript b and a denotes that the quantity is prior or posterior to the analysis step, with the corresponding states and error covariance known as "background" and "analysis" quantities, respectively. The $\mathbf{w}_{\mathrm{k}}$ is the weather uncertainties, assumed independent and non-biased, randomly drawn from a Gaussian distribution with the mean of zero and the standard deviation as listed in Table 7.2. Uncertainties of the model physics were not

considered in this study. In this study, the observation vector **y** was taken as:

$$\mathbf{y} = [\text{LAI},\ \text{SM1},\ ...,\ \text{SM6},\ \text{LeafWt},\ \text{StemWt},\ \text{NO3N1},\ \text{NO3N2},\ \text{NH4N1},\ \text{NH4N2}]^{\text{T}}. \tag{7.3}$$

The observation is linked to model states with a linear observation operator H according to:

$$\mathbf{y}_{\text{k}} = \mathbf{H}\mathbf{x}_{\text{k}} + \mathbf{v}_{\text{k}}, \tag{7.4}$$

where, **H** is a matrix mapping the state vector to the observation vector with elements of 0 and 1 because the observations are the state directly, and the $\mathbf{v}_{\text{k}}$ is the observational uncertainties following the Gaussian distribution $\text{N}(0,\ \mathbf{R})$. The observational error covariance **R** is a diagonal matrix with each element on the main diagonal being the variance of each observation uncertainty (Table 7.2) and the off-diagonal elements being zero. It should be noted that only the key weather data, parameters and observations listed in Table 7.2 were perturbed, selected by their impact on model output according to the sensitivity analysis in Chapter 5.

Filter divergence is a common problem with the implementation of EnKF data assimilation. The prior error covariance could be underestimated due to the approximations underlying the algorithm, with the new observations given less weight to impact the analysis result, which therefore diverges from the observations (Anderson & Anderson, 1999). This is usually addressed by appending a term of model uncertainties in Equation 7.2, or by multiplying the background error covariance by an inflation factor greater than 1. In this study, a covariance inflation factor λ was applied to inflate the ensemble spread, with a value tuned to avoid filter divergence while keeping the ensemble spread small. Explanation of this inflation factor is described further in a later section.

7.2.3 Data assimilation scenarios

State variables were assimilated into APSIM-Wheat in three scenarios according to the state types: 1) a baseline scenario applied to the soil states where only a single type of observation was assimilated at the frequency of data availability; 2) an interpolation scenario applied to the wheat states where the field-observed state variables were interpolated to every

Table 7.2: Uncertainties of weather, parameters and observations.

Data type	Unit	Uncertainty		Data type	Unit	Uncertainty	
		type[a]	std.			type[a]	std.
Weather (daily)				Observations			
Rainfall	mm	M	0.160	LAI	m^2/m^2	M	0.15
Radiation	mm	A	1.254	LeafWt	g/m^2	M	0.2
Max/min temperature	mm	A	0.882	StemWt	g/m^2	M	0.2
Wind speed	mm	A	0.679	Biomass	g/m^2	M	0.2
Parameters				SM1	m^3/m^3	A	0.03
DUL	m^3/m^3	A	0.03	SM2, ..., SM6	m^3/m^3	A	0.01
LL	m^3/m^3	A	0.01	NO3N1, NO3N2	kg/ha	M	0.1

[a] Uncertainty: M=multiplicative, A=additive.
[b] NA: not available.

3 days before being assimilated; and, 3) a combined scenario where multiple observation types were assimilated together. Moreover, assimilation of the single type of state variables was tested with observations in different phenology stages and frequencies.

For each scenario, the assimilation of observations stopped at before the grain-filling stage (stage 7) for two reasons. First, the development of GrainWt is strongly impacted by the assimilation of wheat states, which causes an immediate update due to the strong correlation among wheat states (further discussed in the results of wheat states assimilation), making the validity of yield estimation strongly reliant on the accuracy of just a few wheat state observations that have been assimilated in the grain-filling stage. However, the field measurement of some states (e.g., LAI and LeafWt) are less accurate at this stage when leaves are withering and the LAI is close to zero. Second, from the perspective of guiding field management, it is favourable to understand the yield variability at early growing stages. Using data collected at a late growing stage does not help farmers to improve profit when the opportunity to alter inputs such as irrigation and fertilization has already passed.

7.3 Results and discussion

7.3.1 Model calibration and inflation factor selection

The soil parameters in the SoilWat and SoilN modules were calibrated manually, using the field observation of daily station soil moisture data in all 6 layers and the weekly soil

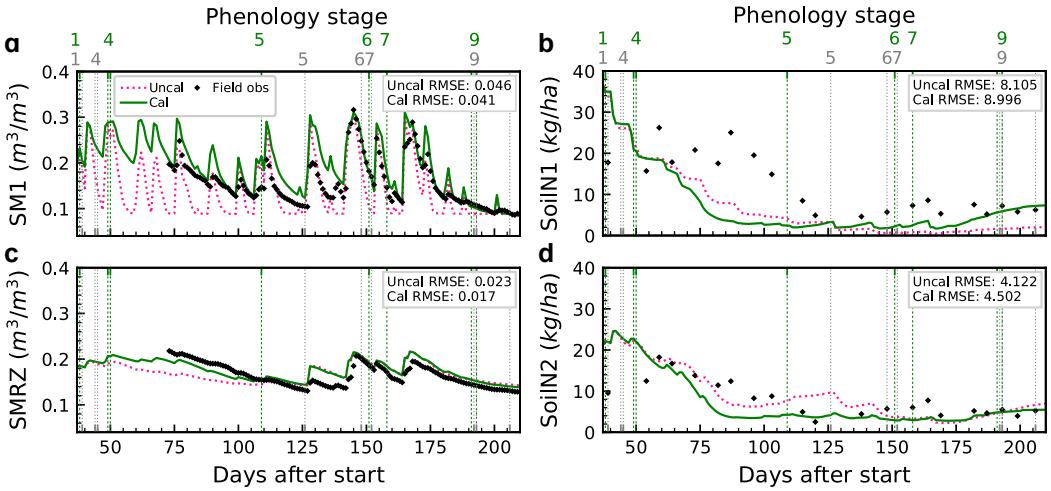

Figure 7.1: Examples of the estimated soil moisture in the surface layer (a) and root-zone (c), and soil nitrogen in the top two layers (b, d) by the calibrated (Cal) and uncalibrated (Uncal) model, respectively. Numbers and gridlines in grey and green are the observed and estimated phenology, respectively.

nitrate-nitrogen data in the top two layers. For the calibration of SoilWat module, with the soil water characteristics (DUL, LL, Sat, AirDry) estimated by the maximum and minimum soil moisture through the growing season and applied to both the calibrated and uncalibrated models, only the parameters controlling the rate of infiltration and evapotranspiration were calibrated. In the SoilN module, only one parameter F_{biom}, the fraction controlling the biom pool carbon as a fraction of the hum carbon that is subject to decomposition, were calibrated. The parameter values before (by default) and after the calibration are shown in Table 7.1.

An example of the outcome of soil modules calibration is shown in Figure 7.1, as the time series of the estimated soil moisture and ammonium nitrogen before and after the model calibration. After a basic calibration, the soil moisture and nitrogen followed the measured field data with a reduced RMSE of soil moisture in the surface layer (layer 1) and the root-zone (layer 2-6). The parameter F_{biom} in the SoilN module was adjusted to fit the soil nitrogen (as the sum of nitrogen in the ammonium and nitrate) dynamics. However, the estimation of soil nitrogen only fitted the observations in the late growing season due to a delay of the LAI growing period (further discussed in results of soil nitrate assimilation).

The inflation factor used in this study was tuned by a trial-and-error to ensure that the ensemble maintains an error covariance representative of the background uncertainties, and may vary with the assimilation frequency and the state variable type. The inflation factor

was 5 in the assimilation of wheat states, except for the biomass assimilation, where a smaller inflation factor of 4 was used. The inflation factor was 2 in the assimilation of soil states, because the inflation factor was applied only when observations were assimilated, and the observations of soil states (primarily soil moisture observed daily) were more frequently assimilated than the wheat states (every 3 days). In the joint assimilation of mixed wheat and soil states, an inflation factor of 5 was used.

It should be noted that while the inflation factor of 5 is very large relative to the commonly used inflation factor values that are usually slightly greater than 1 (Petrie & Dance, 2010) in the domain of meteorology data assimilation, where the numerical weather prediction models are chaotic and more complex, considering two or three dimensions. However, the crop models are relatively simple and simulate one or two dimensional at a daily time step. In this case, the ensemble spread cannot be maintained with a small inflation factor. For example, Figure 7.2 shows an example of LAI assimilation using the selected inflation factor compared to smaller and larger inflation factors. A smaller inflation factor caused little overlap between the observation and the model ensembles, as found in the ensembles of the 5 successively assimilated LAI observations starting from day after start (DaS) 119 (Figure 7.2-b), while a larger inflation factor brought in unnecessary noise, making the model less trusted by the data assimilation algorithm (Figure 7.2-c). The EnKF works sub-optimally with a large inflation because the Kamlan Gain is driven to minimize the model error without accounting for the inflation factor. An alternative method is to add noise to the model states at each time step, but this approach sometimes causes model failure when errors are added in the growth stage, accidentally driving the wheat states to be near zero, and is also less computational effective.

7.3.2 Assimilation of wheat states

LAI, LeafWt, and StemWt are direct observations of model states, and thus their observation are linked with the model states with a value of 1 in the observation operator **H**. Biomass (total above-ground dry biomass in g/m^3) is the sum of above-ground organ weight (LeafWt, StemWt, PodWt and GrainWt). Therefore, the organ weight is mapped to the biomass observations through:

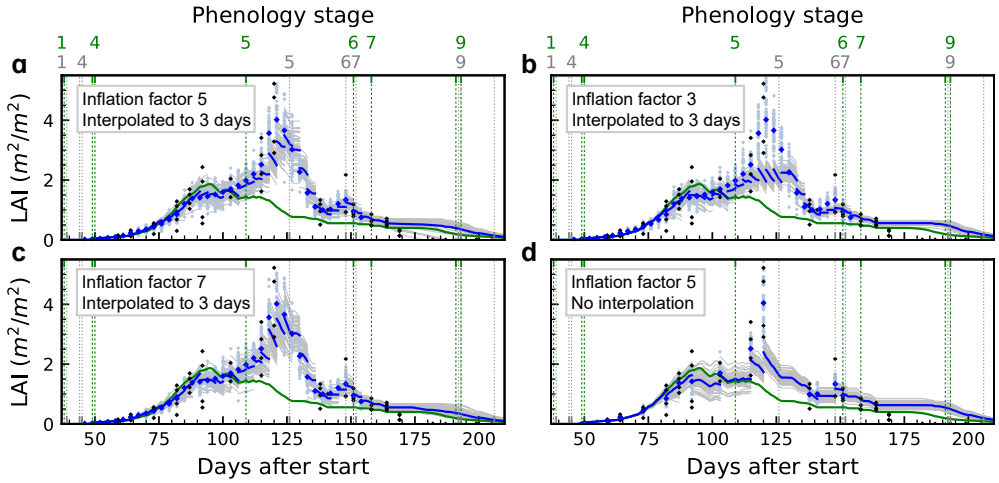

Figure 7.2: Example of assimilating LAI using a variety of inflation factor values under the interpolation scenario (a-c) and using the selected inflation factor under the baseline scenarios (d). See Figure 7.3 for legend.

$$[\text{biomass}] = [1,\ 1,\ 1,\ 1][\text{LeafWt},\ \text{StemWt},\ \text{PodWt},\ \text{GrainWt}]^{\text{T}} + \mathbf{v}_{\text{k}}^{\text{biomass}}, \tag{7.5}$$

where the matrix $[1,\ 1,\ 1,\ 1]$ is a sub-matrix of the observation operator **H** in regard to biomass and organ weight.

The interpolation scenario (scenario 2) was applied to the wheat state assimilation to avoid filter divergence due to the sharp increase of biomass and LAI during DaS 70 and 130 makes the observations differ greatly from the model estimation. In the case of the baseline assimilation scenario (scenario 1), little overlap was found between the probability distributions of the model states and observations represented by their respective ensembles (Figure 7.2-d), suggesting that the value of the observations is in the tail of the Gaussian probability distribution of the model states with a close-to-zero probability, which is unrealistic.

7.3.2.1 Assimilation of LAI

The growing season of this study was drier than average, resulting in heavy water stress in the vegetative stages (i.e. before the floral initiation phase in stage 5). Even with the model calibrated to the station soil moisture time series, the open-loop failed to correctly represent the growing trend of the LAI and the organ weight, giving a significant underestimation of

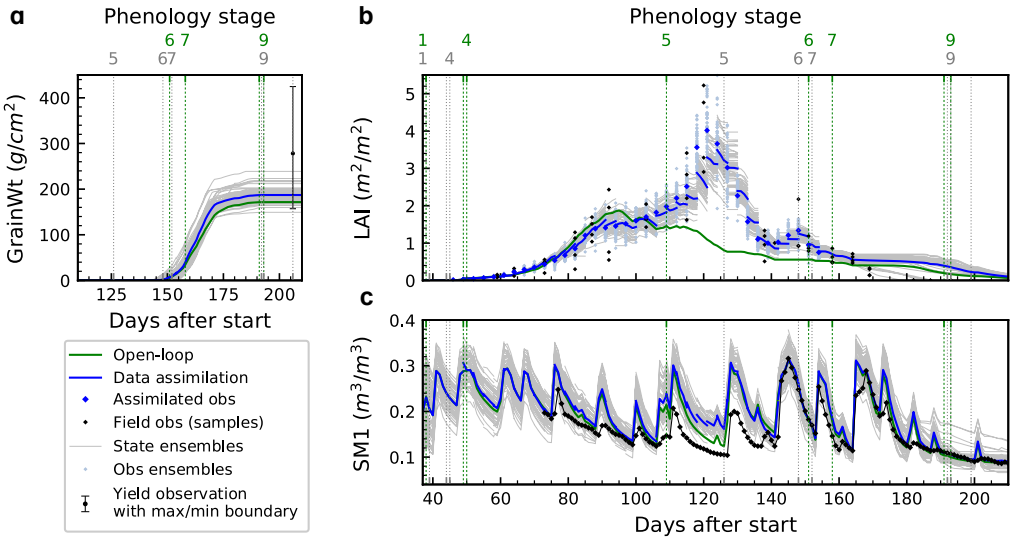

Figure 7.3: Evolution of GrainWt (a), LAI (b) and SM1 (c) with the assimilation of LAI. The open-loop is the ensembles mean of the open-loop states, and data assimilation is the ensemble mean of the analysis states. Numbers and gridlines in grey and green are the field-observed and open-loop estimated phenology, respectively. Legend applies to all subsequent figures in this chapter.

the yield at harvest (the underestimated yield under drought was also reported in Asseng *et al.* , 1998). However, with the interpolated 3-day LAI assimilated into the model, wheat states (e.g., LAI in Figure 7.3-b) were clearly increased relative to the open-loop to close to the field observations. Consequently, with the LAI updated in the vegetative stages, the development of grain weight in the grain-filling stage (stage 7) became slightly faster with a steeper slope, increasing the yield (grain weight at harvest) from 1,716 kg/ha to 1,871 kg/ha and reducing the RD of the yield compared to the observed amount from -38.3% to -32.7 % (Figure 7.3-a). Therefore, by giving a more correct LAI state in the vegetative stages, the development of grain weight in the reproductive stages (i.e. floral initiation to maturity, stage 5-9) was affected, giving a more correct yield estimation at harvest. Noting that no LAI data were assimilated in the grain filling stage, this result demonstrated that assimilation of LAI state was able to push the model to provide a more accurate estimation in the other state variables (indicated by a reduced RMSE of wheat states in Table 7.3), leading to a better yield estimation.

The assimilation of LAI showed little impact on soil moisture in the surface layer. As found in 7.3-c, clear update in the soil moisture were only found on DaS 83, 106 and 116 to 122. This concurs with the results from the synthetic study, showing that the assimilation of

state variables in either the wheat or the soil group alone had little effect on updating states in the other group, because states in these two groups are not directly associated.

7.3.2.2 Assimilation of biomass

Two schemes were applied to the assimilation of biomass: 1) the assimilation of biomass individually as an observation of the sum of above-ground organ weight (Equation 7.5), and, 2) the assimilation of each wheat organ (i.e., LeafWt and StemWt) independently. The evolution of the time series of wheat states from the schemes is shown in Figures 7.4 and 7.5.

Both assimilation schemes were able to provide a more accurate model estimation for wheat states and yield at harvest. Compared to the open-loop, better LAI and biomass estimation were found in the RMSE of all wheat organs (Table 7.3) in both data assimilation scenarios. The wheat states estimated from the total biomass assimilation gave a better estimation for wheat states and yield.

The yield was also better estimated by both biomass assimilation scenarios. The RD of the yield from the two scenarios was -9.4% and -17.7 % respectively, both being closer to zero than the open-loop. However, the result showed that the assimilation of total biomass gave a better yield estimation than that of organ weight.

The scheme of assimilating total biomass was found to be more applicable than the organ weight for several reasons. First, the assimilation of organ weight did not provide a more accurate estimation than using total biomass. Second, the observations are time-consuming to make as the plant needs to be cut manually in the laboratory. Third, although total biomass in this case was measured by drying and weighing the destructive sampling, it can also be measurable from remote sensing data. Fourth, solely constraining the biomass of one of the wheat organs may lead the other biomass states to be wrong (particularly in the grain-filling stage). For example, increasing the stem biomass by state updating can lead to increased biomass in other organs because the error covariance shows a positive association among all the organ weight states. However, in the process of wheat development, leaf withering happens when the stem starts to grow quickly, so the increased StemWt should imply a decreased LeafWt in the real situation, which could be opposite to the state updating result.

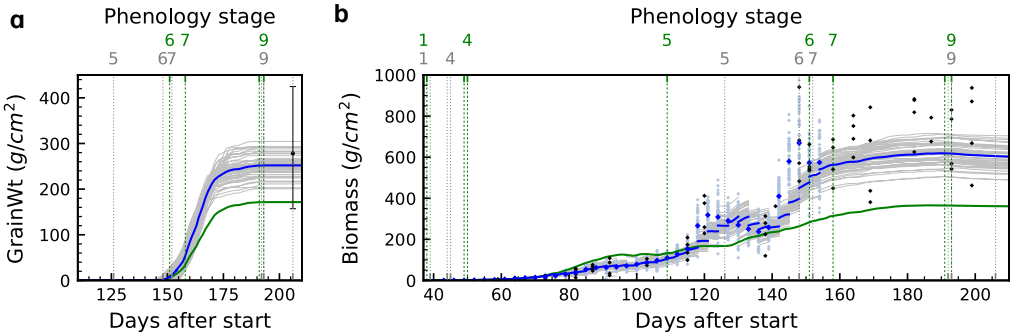

Figure 7.4: Evolution of GrainWt (a) and biomass (b) with the assimilation of biomass. See Figure 7.3 for legend.

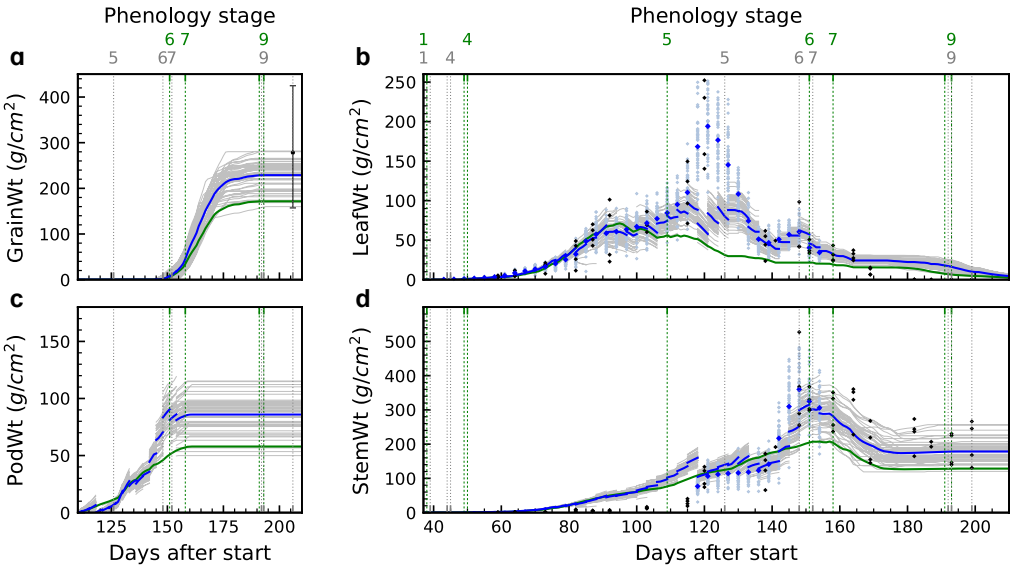

Figure 7.5: Evolution of GrainWt (a), LeafWt (b), PodWt (c), and StemWt (d) with the joint assimilation of LeafWt and StemWt. See Figure 7.3 for legend.

Table 7.3: Statistics of yield and wheat states from the assimilation of wheat states.

Assimilated state(s)	Phenology[a]	Yield (kg/ha)	RD% of yield	RMSE			
				LAI	LeafWt	StemWt	Biomass
Open-loop	Est	1,716	-38.3	0.91	49.5	93.5	150.4
LAI	Est	1,871	-32.7	0.40	29.2	83.0	114.5
Total biomass	Est	2,520	-9.4	0.75	44.7	34.1	78.4
Organ weight (leaf and stem)	Est	2,289	-17.7	0.69	40.5	35.7	98.0
Open-loop	Obs	1,931	-30.6	0.86	47.4	85.1	149.4
LAI	Obs	2,232	-19.7	0.40	28.6	63.9	108.5
Total biomass	Obs	2,779	-0.1	0.82	43.3	38.2	73.4
Organ weight (leaf and stem)	Obs	2,224	-20.0	0.52	33.0	32.4	108.4

[a] Phenology is estimated by the model (Est) or constrained by the field observations (Obs).

7.3.3 Assimilation of soil states

Although the soil states are weakly linked to the wheat states, any bias existing in the soil moisture caused by the EnKF approximation that using a limited number of ensembles can change the wheat state through the error covariance matrix due to the model's high sensitivity to the soil moisture. Therefore, the assimilation of soil states sometimes causing a distinct difference between the posterior and prior wheat states and making the model go wrong (e.g., LAI could be reduced to zero and the plant is thus considered dead). To avoid such problem, the assimilation of soil sates should not be updating wheat states directly when no wheat states are constrained by observations. Therefore, in the experiments of soil states assimilation, only soil states were included in the assimilation state vector **x**, expressed by:

$$\mathbf{x} = [\text{SM1},\ ...,\ \text{SM6},\ \text{NO3N1},\ ...,\ \text{NO3N6},\ ...,\ \text{NH4N1},\ ...,\ \text{NH4N6}]^{\text{T}}. \tag{7.6}$$

7.3.3.1 Assimilation of SM1

According to the time series of SM across the soil profile (Figure 7.6), the assimilation of SM1 (soil moisture in the first layer) pushed only the estimated SM1 to approach the station observation. However, this effect only remains a few timesteps: when the soil moisture becomes very wet (or dry) and the soil moisture is at the upper (or lower) boundary of soil moisture determined by the soil moisture characteristics, the soil cannot get wetter (or dryer) with any water increment flowing into (or out) the system. Therefore, even if the soil moisture is updated by data assimilation in a timestep, the soil moisture is reset once it reaches the upper or lower boundary, and the effect of data assimilation is thus cancelled.

Knowledge or a basic calibration for soil properties is essential before the model can be used, as the model is highly sensitive to soil moisture. According to the sensitivity analysis in Chapter 5, the model gives a significant difference in the estimation of yield when the soil moisture is constrained by a different set of soil water parameters, due to the different level of soil moisture estimation. With a basic calibration of soil properties, the assimilation of SM1 showed a slightly better soil moisture estimation in the top two layers (Figure 7.6) with reduced RMSE from 0.041 to 0.025 and 0.026 to 0.021, respectively. No correction

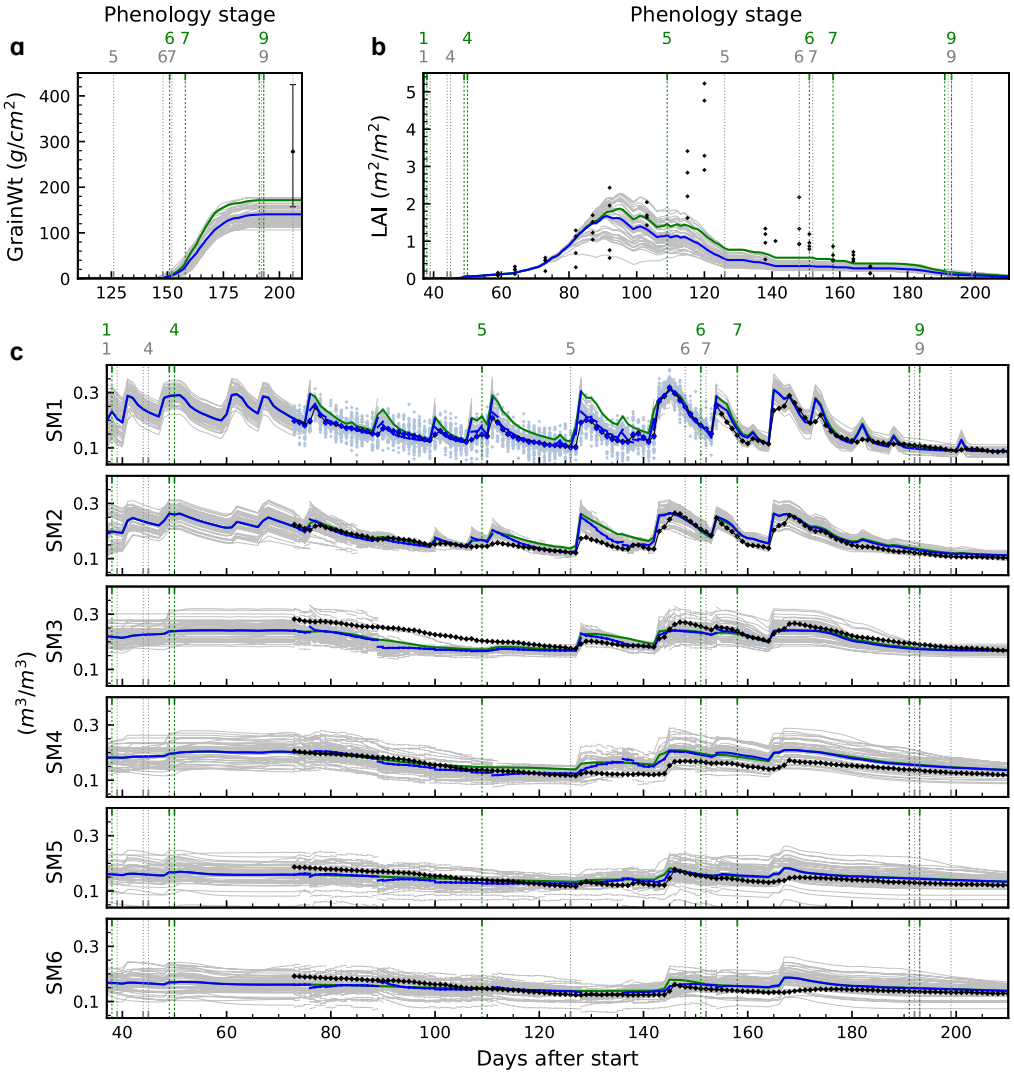

Figure 7.6: Evolution of GrainWt (a), LAI (b), and soil moisture profile (c) with the assimilation of SM1. See Figure 7.3 for legend.

was found in the soil moisture of the bottom layers (3 to 6): when the soil moisture in those layers was updated, the posterior state values quickly returned to the prior values after a few timesteps (e.g., DaS 76 and 89).

Compared to the open-loop, the assimilation of SM1 gave a slightly worse estimation in wheat states and yield, with a reduced SM1 level caused by the assimilation of lower SM1 observations (Figure 7.6). The RMSE of LAI and total biomass increased from 0.74 (open-loop) to 0.87, with the underestimation of yield increased from 38.3% to 49.4%. This result was also found in the assimilation of SM into other soil layers (Table 7.4). Thus, it can be concluded that the assimilation of SM1 did not give an improvement in wheat states

Table 7.4: Relative difference (RD) and root mean square error (RMSE) of yield and soil states from the assimilation of soil moisture in layer 1, layers 1-2 and layers 1-6, respectively with soil parameter calibration.

Assimilated state(s)	Model calibrated?	Yield (kg/ha)	RD% of yield	RMSE						
				LAI	SM1	SM1	SM3	SM4	SM5	SM6
Open-loop	Yes	1,716	-38.3	0.91	0.041	0.026	0.024	0.026	0.016	0.018
SM1	Yes	1,660	-40.3	1.04	0.039	0.026	0.036	0.023	0.015	0.018
SM1-2	Yes	2,016	-27.5	0.95	0.040	0.027	0.028	0.026	0.016	0.019
SM1-6	Yes	2,061	-25.9	0.97	0.041	0.025	0.031	0.031	0.020	0.020
Open-loop	No	2,226	-20.0	0.87	0.106	0.058	0.036	0.036	0.029	0.020
SM1	No	2,018	-27.4	0.80	0.074	0.048	0.033	0.032	0.022	0.008
SM1-2	No	1,939	-30.3	0.83	0.074	0.049	0.034	0.032	0.020	0.010
SM1-6	No	2,006	-27.8	0.89	0.075	0.039	0.021	0.028	0.020	0.011

or yield estimation in over and above that from using a calibrated model.

It should be noted that the errors of the estimated SM caused by wrong soil water parameters in an uncalibrated model could not be cancelled even when the SM was assimilated on a daily basis. According to Table 7.4, although the uncalibrated soil parameters seemingly provided a higher (thus more correct) yield estimation due to a higher level of soil moisture, the assimilation of SM in an uncalibrated model did not give a more correct soil moisture state compared to the open-loop from using a calibrated model. Furthermore, SM observation best improved model performance in terms of SM state estimation and yield prediction when they are used to obtain a calibrated soil water parameter set for the model, rather than being assimilated as external observations to update model states.

7.3.3.2 Assimilation of soil nitrate

Measurements of soil nitrate in the first two layers were assimilated individually and collectively into the model. With a higher level of soil nitrate observation in the first two layers assimilated into the model, a higher plant nitrogen level was found in the organs (Figure 7.7-c and -e). A higher level of plant organ nitrogen was also found in the individual assimilation of soil nitrate in each layer. A higher yield estimation was given by the individual assimilation of soil nitrate in layer 2 and collective assimilation of soil nitrate in the first two layers, but the grain-fill rate remained unchanged from the open-loop. This shows that the increased yield was not because of a higher grain-fill rate caused by a higher plant biomass, but by the relief of nitrogen stress because of a higher plant nitrogen level.

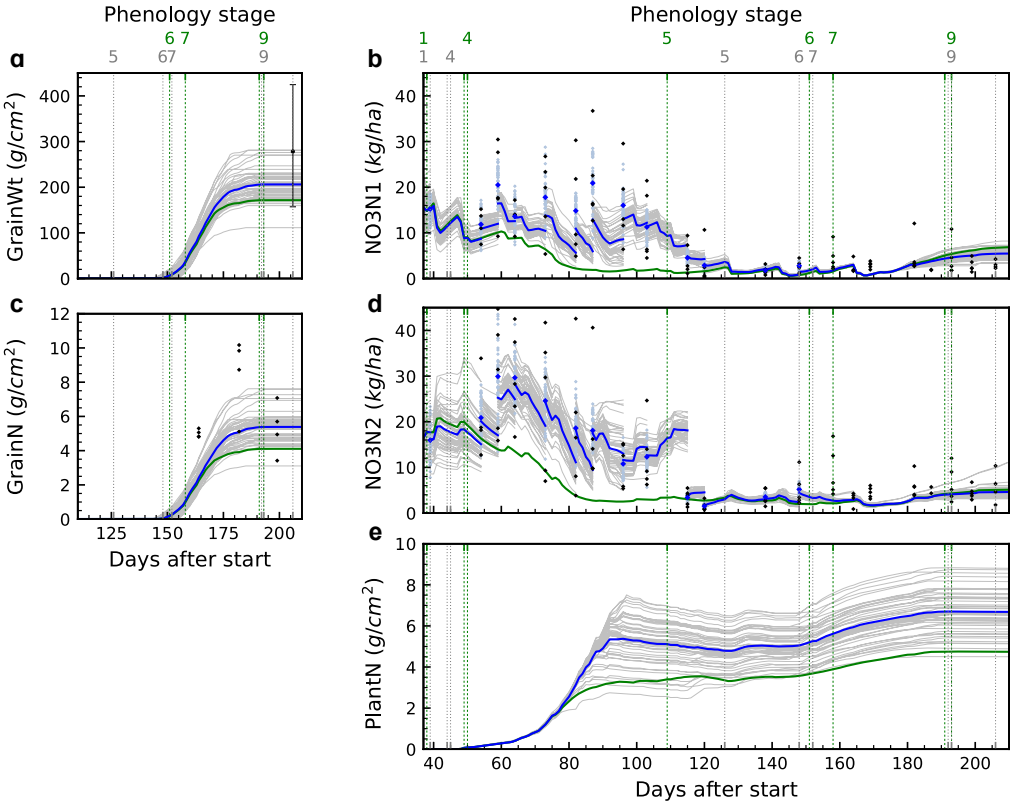

Figure 7.7: Evolution of GrainWt and grain nitrogen (a, c), soil nitrate in layer 1 and 2 (b, d), and total plant nitrogen (e) with the assimilation of soil nitrate in top two layers. See Figure 7.3 for legend.

7.3.4 Data assimilation when observations are limited

7.3.4.1 Impact of assimilation in different phenology stages

This experiment explored the assimilation of state variables at different intervals and phenology stages. In each assimilation experiment, the state variables were assimilated in three long stages 4 (end of juvenile to floral initiation), 5 (floral initiation to flowering), 4-5, 6-7 (flowering to end of grain filling) and all stages, respectively. The RD of yield from each experiment is shown in Figure 7.8, with the result showing that the assimilation of LAI and biomass observations in the vegetative stage gave more accurate yield estimation than assimilation in the other stages. The observations of SM did not benefit yield estimation when assimilated in any individual phenology stages. The optimal stage for soil nitrate assimilation was found to be stages 4 and 5.

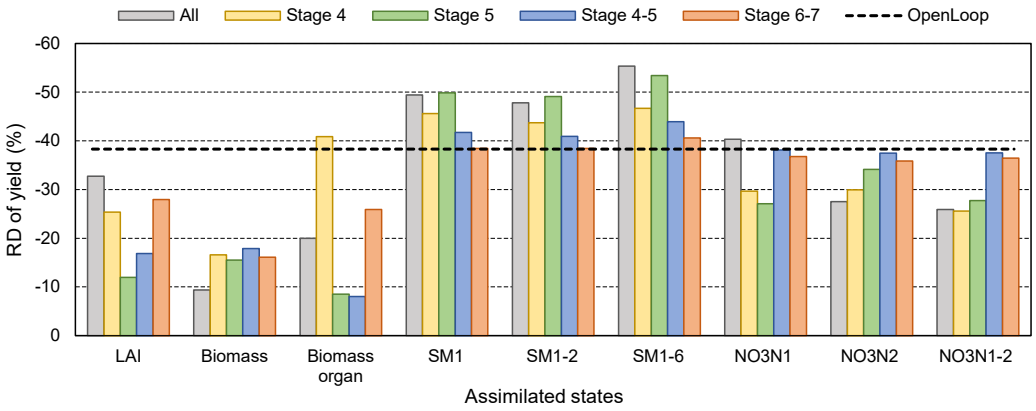

Figure 7.8: Relative difference of yield estimate by assimilating various state variables in different phenology stages relative to the open-loop scenario. Stage 4: end of juvenile to floral initiation; stage 5: floral initiation to flowering; stage 6: flowering to start of grain-filling; stage 7: start to end of grain-filling.

7.3.4.2 Constraining model phenology to observation

This experiment explored the assimilation of phenology through a simple direct-insertion data assimilation approach by setting the model phenology date to the observed value. The date of wheat phenology was determined from the weekly field visit record and in-situ camera photos in Zadoks phenology scale (Zadoks *et al.* , 1974), and mapped to the APSIM wheat phenology scale through an in-built linear equation provided by the APSIM wheat module documentation. Overall, the model showed good consistency in phenology estimation with field observation for all but stage 5 (Figure 7.9). The observed delay in the beginning date of stage 5 was not well-estimated by the model, probably because of the inadequate vernalization due to a late sowing and water deficit in the early stages.

By directly inserting the observations of phenology date into the model, the underestimation of yield was reduced from 38.3% to 30.6% according to Table 7.3 (open-loop when phenology determined by observation), with a slightly reduced RMSE in all estimated wheat states. When further assimilating wheat states (LAI, biomass and organ weight) into the model together with the phenology, the underestimation of yield was further reduced. The yield estimated by the LAI and biomass assimilation (Table 7.3 for RD of yield), and the wheat state estimation (Table 7.3 for RMSE) by the LAI and the organ weight assimilation, showed improved accuracy when the phenology was determined by observation compared to that determined by model simulation.

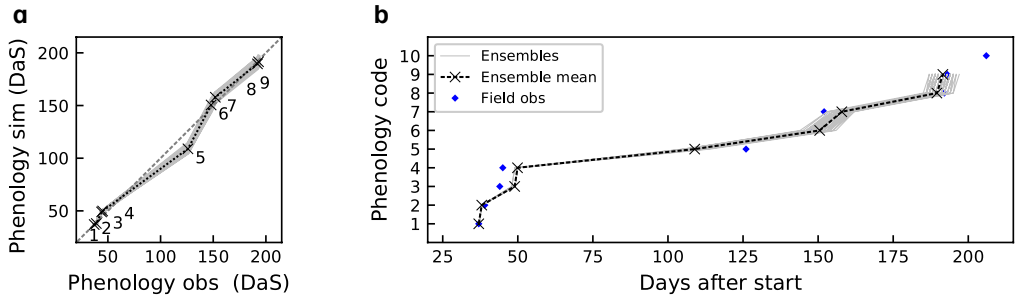

Figure 7.9: Phenology stages of field observation and model estimation in days after start (DaS): (a) the phenology of model simulation versus field observation, where a clear delay of phenology phase 5 (floral initiation) was found in the observation; and (b) change of phenology from model simulation and field observation over time.

7.3.4.3 Data assimilation when using an uncalibrated model

This experiment explored the assimilation of wheat states into an uncalibrated model using properties of different soil types. Fourteen soil types from the APSIM soil library (measured in Victoria, Australia) were used to replace the soil property measurements and calibrated soil parameters (including those controlling soil water retention, evapotranspiration, and the percent of active/inactive nitrogen pool shown in Tables B.1 - B.2 in Appendix B) used in earlier parts of this study, while retaining the initial soil nitrogen consistent with observations. The open-loop from 8 of the 14 uncalibrated models showed a significant underestimation of yield, being even lower than that from the calibrated model (Figure 7.10 grey bars). With the assimilation of LAI (yellow bars) and biomass (green bars) into the uncalibrated models individually, the underestimation of yield was reduced in 10 and 14 out of the 14 models, respectively. This result showed that the assimilation of wheat states, especially the biomass, is a practical way of improving APSIM-Wheat yield estimation even in a wheat field where soil properties cannot be accurately measured or calibrated. The blue bars in Figure 7.10 showed that the estimated yield from all 14 models was near zero when the SM1-2 were assimilated, representing a model failure due to a distinct mismatch in soil moisture between the observations and model estimation. Therefore, the assimilation of soil moisture is not recommended for APSIM-Wheat if the soil module is uncalibrated.

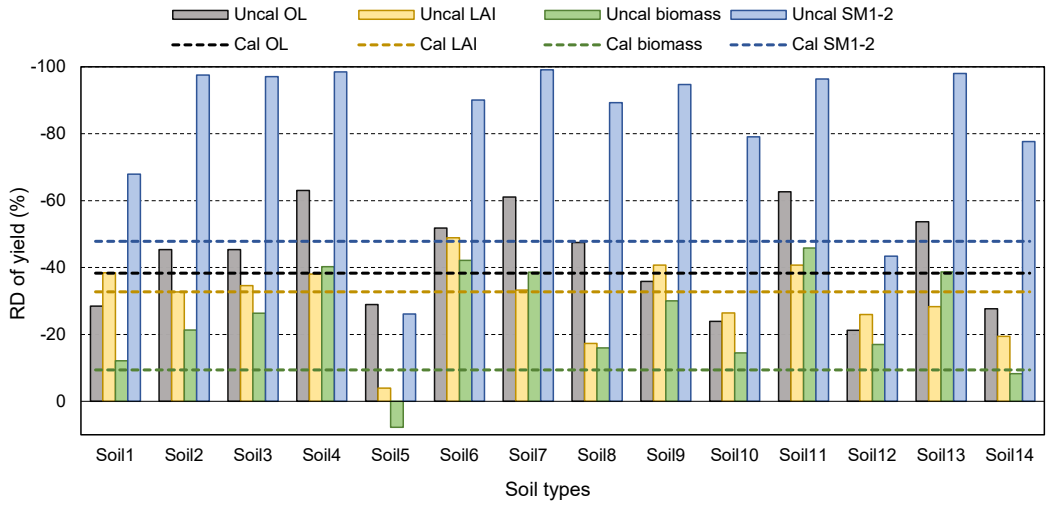

Figure 7.10: Relative difference of yield estimated by assimilating LAI, biomass or near-surface soil moisture with an uncalibrated model using 14 types of soil properties, as compared to using the calibrated model. Legend: Uncal and Cal represent that the data was assimilated into an uncalibrated or a calibrated model. OL is the open-loop, and LAI, biomass and SM1-2 are the assimilation of LAI, biomass, and SM in the top two layers, respectively.

7.4 Chapter summary

This chapter presented the data assimilation case study for the Cora Lynn area with the dataset collected in the 2018 wheat season. Based on the knowledge gained from the previous synthetic study, the ground-measured wheat and soil states assimilated in this experiment include LAI, biomass, organ weight of leaf and stem, soil moisture in 6 soil layers from depth 0 to 55 cm, and soil nitrogen in the top two layers from depth 0 to 15 cm. These state variables were assimilated into the APSIM-Wheat model individually and collectively. Assimilation of observations during different phenology stages was tested to provide an understanding of which stages have the greatest impact for improved yield prediction. In addition to the data assimilation process, a simple scenario of constraining the phenology stages to real observations was tested.

The data assimilation experiments showed that both biomass and LAI provided improved yield estimation through data assimilation and are observable from remote sensing, making them promising for future practices of data assimilation. The assimilation of these wheat states benefited yield estimation when the observations were collected before the anthesis completed (stage 4 and 5). Although a basic calibration to soil parameters is essential

before the model can be used, the assimilation of these states is capable of correcting the errors of yield estimation by uncalibrated models. By constraining phenology, the assimilation of these variables was found to further reduce the underestimation of yield. However, the assimilation of soil moisture did not improve yield estimation in this study. The state variables found to improve yield estimation in this chapter is the basis of the data assimilation experiment with remote sensing data in Chapter 8.

Chapter 8

Case study 2 - assimilation of remotely sensed data

The results of the synthetic study in Chapter 6 and case study 1 in Chapter 7 have demonstrated that assimilating direct field observations of LAI, total biomass, organ weight and phenology into APSIM-Wheat had the greatest impact on improving model estimates of wheat states and yield prediction compared to an open-loop simulation. The soil moisture was found to have no impact on the yield estimation when assimilated in the case study 1. This chapter extended that work to the assimilation of observations obtained only by optical and microwave remote sensing methods. Among these state variables, remotely sensed LAI and surface soil moisture were obtained specific to the study area, and were thus assimilated into the APSIM-EnKF data assimilation framework in this case study.

8.1 Dataset, model and data assimilation setup

This case study uses the same base field dataset as the previous case study described in Chapter 7. The key difference between case study 2 and case study 1 is the source of the observational data that is assimilated. Further details are described below.

The remote sensing data assimilated herein include: (1) high spatio-temporal resolution (3 m, daily) LAI images from a fused dataset, and (2) surface soil moisture (field scale by treating the whole field as a single pixel, daily) retrieved from tower-based brightness temperature. The time series of LAI, and retrieved surface soil moisture from the tower are

those shown in Chapter 4.

The model settings remain unchanged from the previous study, including the composition of sub-modules, sowing windows, management rules, soil layers, initial conditions, and calibrated model parameters. Among the data assimilation settings, state vectors, input and parameter uncertainties and the ensemble generation methods remain unchanged. However, the observations vector $\mathbf{y}$ was reduced to:

$$\mathbf{y} = [\text{LAI},\ \text{SM1}]^{\text{T}}, \tag{8.1}$$

with LAI and SM1 being the remotely sensed dataset described above, with estimated uncertainties of 10% of LAI observation, and 0.036 m^3/m^3, respectively. The prior error covariance was multiplied with an ensemble inflation factor to inflate the ensemble spread using the same values as in the previous case study.

8.2 Data assimilation scenarios

The remote sensing data were assimilated into the APSIM-Wheat model according to two strategies, by assimilating each type of remote sensing data (either LAI or SM1) individually and jointly. Moreover, different data assimilation scenarios were applied to the assimilation of each state type. Specifically, when solely assimilating the high spatio-temporal resolution LAI images, the scenarios applied included:

1. field-averaged remotely sensed LAI was assimilated at an interval of 3 days by integrally treating the paddock as a whole (baseline scenario);

2. a plot-specific scenario where the LAI extracted from the nearest pixels of the 4 specific experimental plots were assimilated to estimate the yield in each plot;

3. an observation-limited scenario, assuming that the availability of LAI observations is limited to a single phenology stage (among stages 4, 5, and 6-7);

4. an uncalibrated scenario assuming that the model is uncalibrated due to the absence of site-specific data so that all the soil properties were simply assumed with the soil types previously measured in nearby locations provided by the model.

In the experiments of solely assimilating SM1, or jointly assimilation LAI and SM1, only a baseline scenario was applied, where LAI was assimilated every 3 days and SM1 assimilated daily at the field scale.

8.3 Results and discussion

8.3.1 Assimilation of LAI

8.3.1.1 Baseline scenario

While the previous chapter demonstrated the yield improvement by assimilating ground-measured LAI into the APSIM-Wheat model, it must be noted that the remotely sensed LAI data used here underestimated the LAI compared to the field observations, but was less noisy (Figure 4.5-e in Chapter 4). However, by assimilating the underestimated remotely sensed LAI data, the estimated yield increased to 2570 kg/ha, with a relative difference (RD) of yield improved from -38.3% to -7.6%. Accordingly, the estimation of LAI and organ biomass were improved, with the RMSE values shown in Table 8.1. Therefore, this baseline scenario has further confirmed the benefit of LAI assimilation into wheat modelling for an improved yield estimation.

Table 8.1: Relative difference (RD) of yield and root mean square error (RMSE) of wheat states from the data assimilation of remotely sensed LAI in the baseline and the plot-specific scenarios, compared to the result of assimilating ground-measured LAI in the previous case study in Chapter 7 .

Assimilated observation(s) [a]	Yield (kg/ha)	RD of yield [b] (%)	RMSE			
			LAI	LeafWt	StemWt	Biomass
Open-loop	1,716	-38.3	0.91	49.5	93.5	150.4
RS[a] LAI (field mean)	2,570	-7.6	0.56	34.8	81.7	75.5
RS LAI (plot A)	2,652	-4.6	0.50	29.3	93.6	76.3
RS LAI (plot B)	2,685	-3.5	0.56	30.9	105.2	80.2
RS LAI (plot C)	2,819	1.3	0.57	31.5	111.2	82.3
RS LAI (plot D)	2,574	-7.5	0.58	35.1	93.3	78.5
Field LAI	1,871	-32.7	0.40	29.2	83.0	114.5
RS SM1	1,618	-41.8	0.94	50.8	100.7	159.2
RS LAI and RS SM1	2,340	-15.9	0.56	34.6	68.7	84.4

[a] Observations assimilated: RS=remotely sensed, Field=Field-measured.
[b] The relative difference of yield was calculated against the plot-average yield.

By comparing the time series of GrainWt obtained from data assimilation using remotely

sensed LAI (Figure 8.1-b) with that from assimilating ground-measured LAI (Figure 8.1-d), it was found that the assimilation of remotely sensed LAI resulted in higher (thus more accurate in this case) yield estimation than using field LAI data. The higher yield estimated by the remotely sensed LAI could be explained by the StemWt, whose value at phenology phase 6 (anthesis) directly determines the growing rate of grain weight in the subsequent grain-filling stage according to the model physics: by comparing the time series of StemWt the two experiments (Figure 8.1-c and -f, respectively), a higher StemWt value at phenology phase 6 was found in the assimilation of remotely sensed LAI.

The reason why the assimilation of remotely sensed LAI had better improvement on the model yield estimation than the field LAI is two-fold. First, the uncertainty used in the remote sensing observations was smaller than the ground measurements: 10% for the remotely sensed LAI and 15% for the field LAI, as the time series of remotely sensed LAI is less noisy than the field LAI. Note that the data assimilation algorithm gives greater emphasis to observations of smaller uncertainty. Second, although estimation of the LAI peak was higher in the field LAI assimilation, the StemWt did not become higher than that in the remotely sensed LAI assimilation, because the higher observational uncertainty in the field LAI made the wheat states less impacted by the assimilation. Soon after the LAI peak, when the ensemble spread of StemWt was more extensive, the field LAI observations showed a quick reduction, causing the StemWt to be reduced by the assimilation of field LAI during Day after Start (DaS) 140 and 150. Consequently, the StemWt in phase 6 showed a lower estimation from the field LAI assimilation as compared to remotely sensed LAI assimilation.

8.3.1.2 Plot-specific scenario

This scenario compared the difference between assimilating remotely sensed LAI at the field and plot scales. By assimilating remotely sensed LAI data at the plot-specific level, the evolution of the estimated GrainWt for the four experimental plots was presented in Figure 8.2. The yield estimated at the plot-specific level was validated with the yield collected in the four experimental plots, while that estimated at the field level was evaluated against the averaged yield of the four plots.

Figure 8.3-a shows the yield estimated by the LAI assimilation versus the yield measured in each experimental plot. The mean yield at the plot-specific level was estimated by

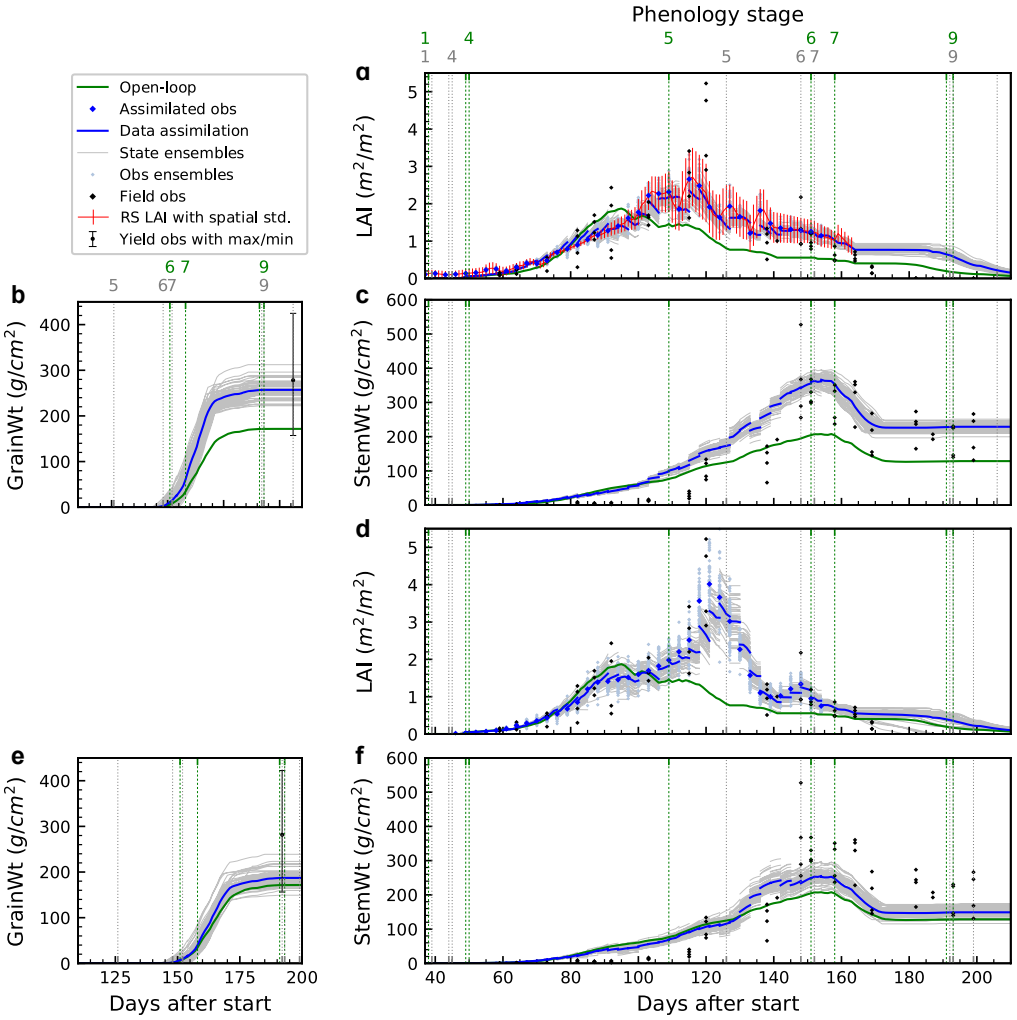

Figure 8.1: Evolution of LAI, GrainWt and StemWt with the assimilation of remotely sensed LAI (a-c) and for comparison field LAI (d-f). The vertical lines display the field-observed phenology in grey, and the simulated phenology of the open-loop in green. The legend applies to all subsequent figures in this chapter.

assimilating the average LAI of the four plots, while that at the field level was estimated by assimilating the average LAI of all pixels in the field. Figure 8.3-b shows the yield estimated by the assimilation of plot-specific LAI compared to the yield observations in each plot, respectively.

Figure 8.3-a shows that the estimated yield in each plot was uniform when LAI was assimilated at the field scale. With the plot-specific LAI assimilated into the model whose settings and inputs were uniform for all four plots, the estimated yield specific to each plot shows a spatial differentiation, with the plot-specific yield in plots C and D slightly closer to the 1:1 line. This result indicates that the high spatial resolution LAI image could bring

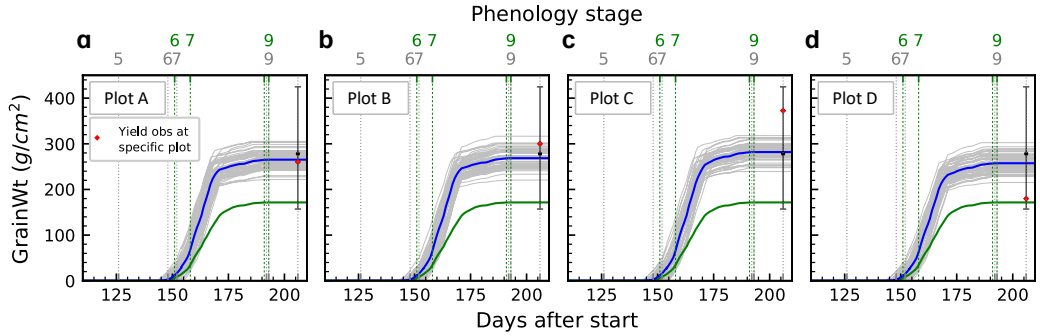

Figure 8.2: Evolution of GrainWt with the assimilation of pixel-specific remotely sensed LAI at the four plots (A to D). The red dots represent the pixel-specific yield collected in each plot, while the black dots with the error bar represent the average yield and the maximum/minimum yield of all subsamples (note that each wheat sample collected at the experimental plots contains 5 subsamples).

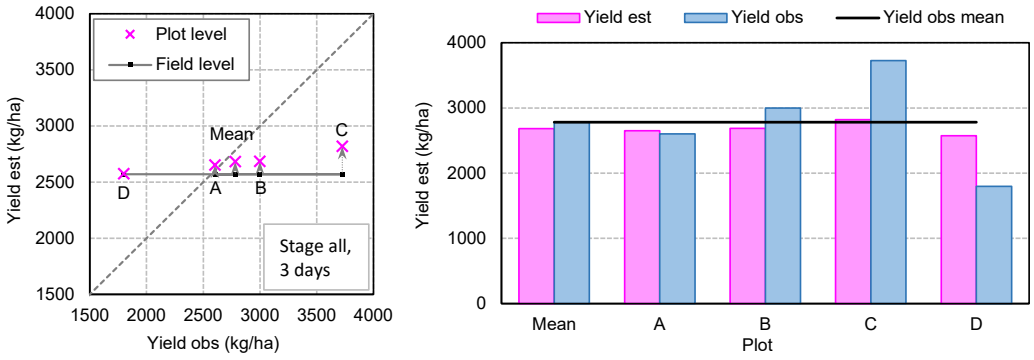

Figure 8.3: (a) Yield estimates by assimilating LAI at the field and plot level. The black dots represent the yield by the assimilation of LAI at the field scale, and pink crosses (linked with arrows) represent that at the plot-specific scale. (b) Yield estimates by assimilating the plot-specific LAI compared to the plot-specific yield observations.

spatial variation to the model even when the model inputs are assumed uniform across the field.

The result of the plot-specific scenario implies that even when the available data only allows the model to simulate at the field scale, high spatial resolution LAI images can provide sub-field spatial variation to the model, allowing the yield to be estimated at pixel resolution. Therefore, although the usage of crop models are usually limited by the difficulties of collecting input data, especially for estimating the spatial variation of yield at the sub-field scale, a uniform model running at the field scale can likely provide an estimate of yield with sub-field variation, given that the remote sensing data are sub-field resolution. Note that this scenario was only evaluated with data collected at four experimental plots and in one

growing season, meaning that more experiments are required to support the validity of this conclusion.

8.3.1.3 Observation-limited scenario

This scenario explored the assimilation of LAI in different phenology stages. A group of data assimilation experiments was conducted, by assimilating the LAI in stages 4, 5, 4-5, 6-7 and all stages. The RD of yield from each experiment is shown in Figure 8.4 (grey bars), compared to using field LAI measured from the previous data assimilation case study (yellow bars), with a value closer to zero representing a more accurate yield estimation. The results showed that the assimilation of remotely sensed LAI in stage 5 provided the most accurate yield estimation compared to assimilating in other stages, and with comparable accuracy to the assimilation over the whole growth stage. This result is consistent with the result of the previous case study using ground LAI measurements.

8.3.1.4 Uncalibrated scenario

This scenario explored the assimilation of wheat states into an uncalibrated model using different soil property types. This study used the same set of 14 uncalibrated models setup in earlier parts of this thesis (Chapter 7) by replacing the measured and calibrated soil parameters with the properties of 14 soil types taken from the APSIM soil library. The RD of yield is shown in Figure 8.5.

The open-loop from 8 out of the 14 uncalibrated models showed a substantial underestimation of yield. However, with the assimilation of remotely sensed LAI (Figure 8.5 green bars) into the uncalibrated models the underestimation of yield was reduced in all 14 cases, with two models been over-corrected. By comparing to the assimilation of field LAI (yellow bars), the assimilation of remotely sensed LAI provided a more accurate yield estimation in 11 out of the 14 models. This result showed that the assimilation of both ground- and remote sensing-based LAI improved APSIM-Wheat yield estimation, even in a wheat field where soil properties cannot be accurately measured or calibrated.

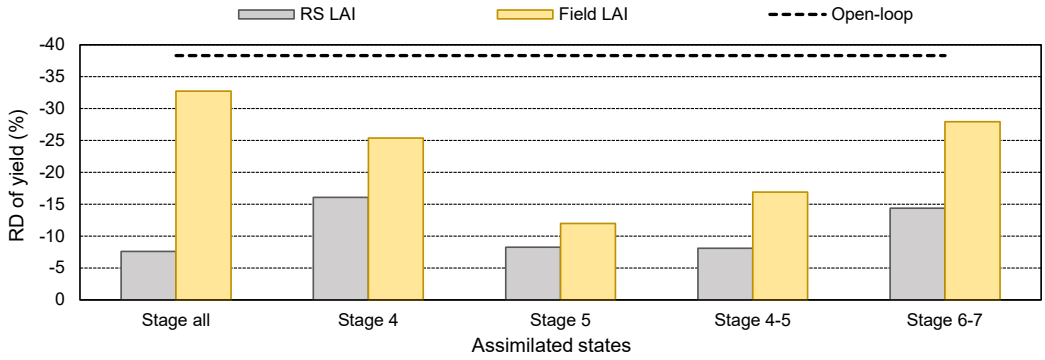

Figure 8.4: Relative difference of yield estimate by assimilating remotely sensed (RS) and field-measured (Field) LAI in different phenology stages and relative to the open-loop. Stage 4: end of juvenile to floral initiation; stage 5: floral initiation to flowering; stage 6: flowering to start of grain filling; stage 7: start to end of grain filling.

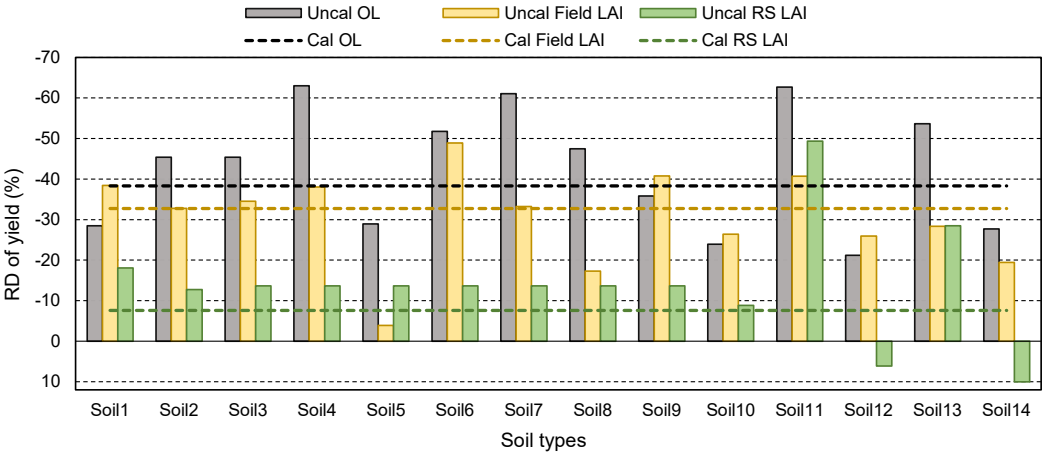

Figure 8.5: Relative difference of yield estimated by assimilating LAI with an uncalibrated model using 14 soil property types, as compared to using the calibrated model. Legend: Uncal and Cal represent that the data were assimilated into an uncalibrated or a calibrated model. OL is the open-loop.

8.3.2 Assimilation of SM1

Figure 8.6 shows the time series of GrainWt, LAI and soil moisture in the near-surface and root-zone as a result of the joint assimilation of remotely sensed SM1. It was found that the assimilation of SM1 had little impact on the development of GrainWt, LAI and wheat biomass. Some correction was made to the estimation of soil moisture in the first two layers, but a similar correction was not found in deeper layers, as shown in the RMSE values of SM in each layer (Table 8.2). This result showed that the assimilation of SM1 did not help improve yield estimation given a well-calibrated model, which is consistent with the previous case study that assimilated in-situ SM1 observations.

Table 8.2: Relative difference of yield and root mean square error of soil states from the data assimilation of soil moisture in layer 1.

Assimilated state(s)	Yield (kg/ha)	RD of yield (%)	RMSE						
			LAI	SM1	SM2	SM3	SM4	SM5	SM6
Open-loop	1,716	-38.3	0.91	0.041	0.026	0.024	0.026	0.016	0.018
LAI (baseline scenario)	2,570	-7.6	0.56	0.044	0.027	0.032	0.029	0.021	0.022
SM1	1,618	-41.8	0.94	0.038	0.025	0.026	0.026	0.016	0.018
LAI and SM1	2,340	-15.9	0.56	0.047	0.028	0.025	0.032	0.019	0.021

8.3.3 Joint assimilation of LAI and SM1

Figure 8.7 shows the time series of GrainWt, LAI and soil moisture in the near-surface and root-zone as a result of the joint assimilation of LAI and SM1. It was found that when the LAI and SM1 were assimilated jointly, the yield and wheat states were better estimated relative to the open-loop, but this improvement was not as distinct as solely assimilating LAI (Table 8.1). The joint assimilation provided a more erroneous soil moisture estimation, with a higher RMSE in all layers relative to the open-loop (Table 8.2). This is due to the assimilation of LAI that increased the soil moisture values during DaS 95 to 120 (Figure 8.7-c), making it further overestimated.

8.4 Chapter summary

This chapter presented a data assimilation case study for the Cora Lynn study area using the dataset collected in the 2018 wheat season with only remotely sensed observation assimilated. Based on the knowledge gained from the previous synthetic study and case study 1 (assimilation with ground measurements), the remotely sensed data assimilated in this experiment included high spatio-temporal resolution LAI images from a fused Sentinel-2 and PlanetScope dataset, and surface soil moisture from a tower-based radiometer. These state variables were assimilated into the APSIM-Wheat model under different scenarios, accounting for different scales, availability of observations and model parameters. The results have shown that the assimilation of remotely sensed LAI is promising in yield estimation, particularly when assimilated in the phenology stage 5 (floral initiation to flowering). The assimilation of remotely sensed LAI also helped improve yield estimation in uncalibrated

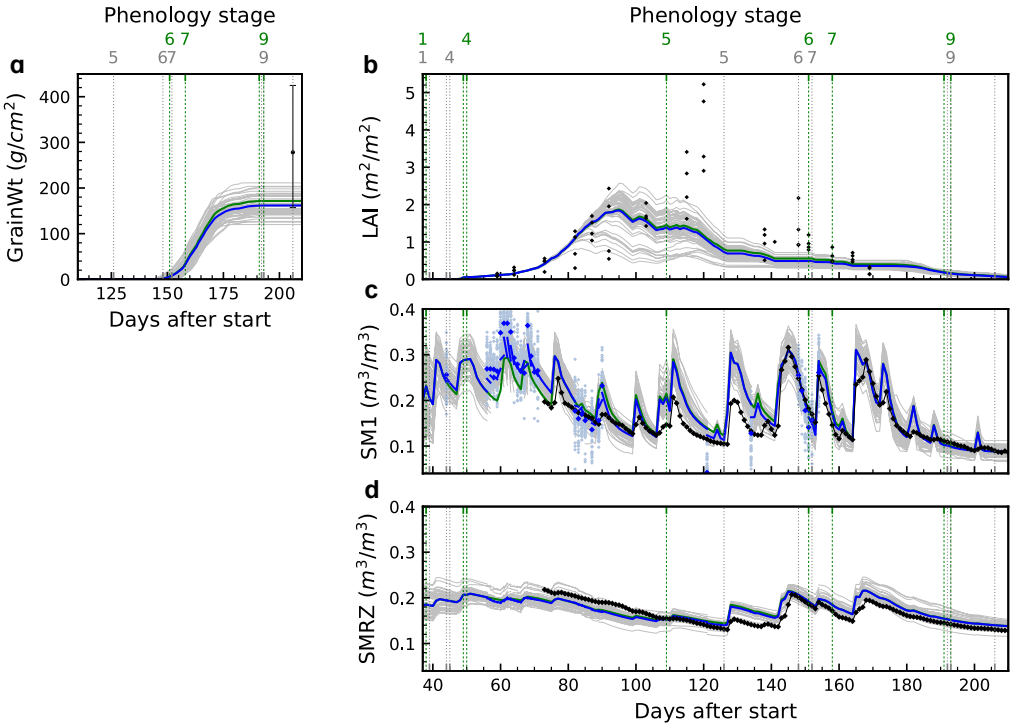

Figure 8.6: Evolution of grain weight (a) LAI (b), SM1 (c) and root-zone soil moisture (d) with the assimilation of SM1.

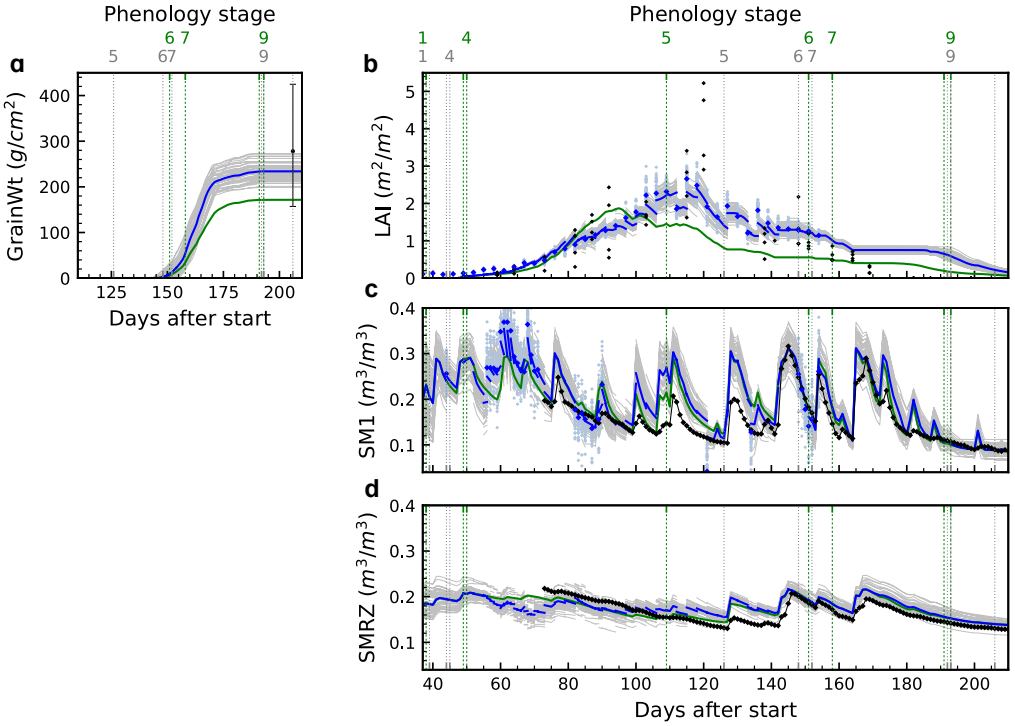

Figure 8.7: Evolution of grain weight (a) LAI (b), SM1 (c) and root-zone soil moisture (d) with the joint assimilation of LAI and SM1.

models, introduce spatial information into a uniform model, and consequently reduce the strong dependence of the APSIM-Wheat model on the soil information usually unavailable. However, similar to the previous case study, the assimilation of remotely sensed SM1 did not improve yield estimation. The joint assimilation of both LAI and SM1 did not perform better than solely assimilating LAI as a result of assimilating more erroneous remotely sensed SM1.

While ultimately this case study used only LAI and SM, it was proceeded by a sensitivity analysis (Chapter 5) and extensive assimilation study incorporating a wide range of possible variables (Chapters 6 - 7). The possibility of data assimilation being applied in a spatially distributed way was explored using a unique model to Australia.

Chapter 9

Conclusions and future work

The main contribution of this research was to comprehensively and systematically explore the assimilation of all potential state variables into the crop model APSIM-Wheat. Through a series of data assimilation experiments, progressive understanding of the potential to improve yield estimation by assimilating external observations into the model was achieved. This chapter summarizes the main conclusions of this thesis and provides recommendations for future research.

9.1 Conclusions

9.1.1 Sensitivity analysis

The sensitivity analysis tested the model response to any uncertainty in the weather input, soil parameters, cultivar parameters, management information and state variables in each phenology stage. The conclusions of the sensitivity analysis are listed below.

1. Weather input data, and soil parameters related to soil water and sowing information were found to be the primary source of uncertainties in yield estimation for APSIM.

2. Among the weather inputs to APSIM, rainfall, radiation and temperature were found to have vital impacts on the model simulation of wheat states and yield. The temperature was found to strongly impact the phenology because it plays a vital role in phenology development and photosynthesis.

3. Cultivar parameters that govern phenology development have a strong impact on the

leaf and grain development by affecting the time length over which the biomass accumulates, and the way that biomass is partitioned to each plant organs (which depends on the growing stage). These parameters are VernSens, PhotopSens and target thermal time of stage 4 to 7. Additional to the impact of phenology, the yield estimation is also subject to parameters that directly control the rate of grain development. These parameters are the Potential Grain Filling Rate, Potential Grain Growth Rate, and Potential Grain N Filling Rate.

4. The wheat states and yield are highly sensitive to soil parameters controlling the soil water dynamics. These parameters, including DUL, LL15, and WheatLL, need to be carefully measured or calibrated before using the model.

5. LAI, soil moisture, plant weight and nitrogen, phenology, and soil nitrogen are important state variables. Their impact on model evolution varies with growth stage. The accuracy of these state variables has the potential to result in accurate yield estimation. More specifically, the important state variables and the growth stage when they become important are listed below:

   (a) LAI in stage 4 and 5;

   (b) LeafWt and LeafN in stage 5 and 6; with the impact in stage 6 outweighing that in stage 5;

   (c) StemWt and StemN in stage 5, 6, and 7, with the impact in stage 6 outweighing that in stage 5 and 7;

   (d) PodWt and PodN in stage 6;

   (e) GrainWt in stage 7;

   (f) RootWt in stage 5;

   (g) Accumulative thermal time in all stages; and,

   (h) SM in all stages, with the impact in stage 4 and 5 outweighing that in stage 6 and 7 and in root-zone outweighing that in the near-surface.

9.1.2 Synthetic study

In the synthetic study, all prognostic model state variables were assimilated into the APSIM-Wheat model for their impact on yield estimation using a synthetically created data-set. The conclusions are listed below.

1. An ensemble size of 50 was satisfactory for the EnKF algorithm to assimilate wheat and soil states into the APSIM-Wheat model.

2. The state variables that have the potential to improve yield estimation are: (i) wheat states - LAI, GrainWt, LeafWt, StemWt, LeafN and StemN; and (ii) soil state - all water and nitrogen states. There were certain wheat phenology stages where assimilating theses state variables provided the greatest impact on improving yield estimation and these are often where degradation in the open-loop simulation could occur. Specifically:

   (a) Phenology stage 5 (floral initiation to flowering) was found to be the optimal assimilation stage for LeafN, StemN, soil nitrogen, and was sub-optimal for LAI, LeafWt, and soil moisture.

   (b) Phenology stage 6-7 (flowering to end of grain filling) was found to be the optimal assimilation stage for LAI, LeafWt, StemWt, LeafN, and was sub-optimal for soil nitrate-nitrogen.

3. The direct insertion of phenology reduced the uncertainties of yield and state variables and thus led to a better yield estimation.

4. Current remote sensing techniques provide LAI and SM1 with adequate accuracy for these variables to have a positive impact when assimilated.

9.1.3 Case study 1 - assimilation with ground measurements

In the first case study, ground-measured wheat and soil states were assimilated into the APSIM-Wheat model for improving yield estimation using an in-situ data set collected in the field. The conclusions are listed below.

1. The assimilation of wheat and soil states successfully improved the yield estimation of APSIM-Wheat, using field observations of LAI, total biomass, and organ weight of leaf and stem. However, little improvement was found when using in-situ soil moisture and field observations of soil nitrate.

2. The assimilation of biomass provided the best yield estimation among the assimilation of all individual wheat states. Both biomass and LAI provided improved yield estimation through data assimilation and are observable from remote sensing. Thus they are promising for future implementation of operational data assimilation.

3. Pcoor performance of assimilating soil moisture and soil nitrate was obtained. However, this does not imply that these data are not useful in wheat modelling and yield estimation. On the contrary, they are essential in calibrating parameters related to soil moisture and nitrogen estimation. With the wheat growth primarily subject to water and nitrogen stress and the model highly sensitive to soil moisture and initial nitrogen, a correct estimation of soil water and nitrogen states is essential for wheat modelling.

4. The assimilation of wheat states reduced uncertainties caused by unknown model parameters (e.g. wheat cultivar type, soil properties). With the assimilation of wheat states, a more accurate yield estimation was provided by the model even with uncalibrated soil and cultivar parameters.

5. The assimilation of wheat states benefited yield estimation with the observations collected before the phenology phase 6 (flowering). With the information of wheat states collected in stages 4 and 5, the assimilation of wheat states helped the model to provide an improved yield estimation even without any calibration.

9.1.4 Case study 2 - assimilation of remotely sensed data

In the second case study, remotely sensed wheat and soil states were assimilated into the APSIM-Wheat model for improving yield estimation using remotely sensed observations in place of in-situ LAI and surface soil moisture. The conclusions are listed below.

1. The assimilation of remotely sensed LAI improved the yield estimation of APSIM-Wheat at both the field scale and the plot-specific scale, with the yield estimation

outperforming that using field observations. The optimal phenology stage for assimilating LAI was found to be stage 5.

2. The spatial information provided by the fine-resolution LAI has the potential to provide sub-field spatial variation to the model.

3. Similarly to the previous case study, little improvement in yield estimation was found when assimilating remotely sensed surface soil moisture.

4. The assimilation of remotely sensed LAI into an uncalibrated model reduced the uncertainty caused by unknown model parameters and provided a more accurate yield estimation.

9.2 Recommendations for future work

The perspectives for future research are listed below.

1. Application of the APSIM-EnKF framework is not only limited to the wheat yield estimation tested in this case, but is also applicable for near real-time crop monitoring for crop and soil status. Moreover, it can be extended to over 20 other crop types (i.g., maize, barley, sorghum, cotton, etc.) available in the APSIM model and adapted to other state-updating algorithms (i.g., KF, PF, etc.) with only minor modification to the source code. Consequently, the data assimilation framework can be used to monitor over 20 crop types, with the flexibility of using different data assimilation algorithms.

2. The APSIM-EnKF data assimilation framework presented in this thesis is one-dimensional. A two-dimensional data assimilation framework can be developed to assimilate fine-resolution images to cover the spatial variability and provide precision wheat states and yield estimation specific to pixels.

3. In the first case study (using field observation), biomass was found to provide better yield estimation than LAI. This result was not further evaluated using remotely sensed data due to the lack of a suitable dataset. Future research may focus on wheat biomass retrieval from remote sensing and its subsequent assimilation.

4. The first case study found that constraining wheat phenology could partially cancel the errors caused by the uncertainty in model input and parameters (i.e., temperature and the target thermal time) that control phenology development. Future efforts can be put on the assimilation of remotely sensed phenology information.

5. The case studies have shown that the spatial information provided by the fine-resolution LAI has the potential to provide sub-field spatial variation to the model. Nevertheless, limited by the datasets available for this study, it could not be conclusively shown crop models uncertainty in estimating the spatial variation of wheat development can be solved by assimilating remote sensing data into the model. Consequently, results should be further evaluated using datasets for a range of geographic locations.

6. With more observations becoming available in the future, this work can be expanded to an agricultural monitoring service platform to provide agricultural monitoring and forecasting.

References

Abrahamsen, P., & Hansen, S. 2000. Daisy: an open soil-crop-atmosphere system model. *Environmental modelling & software*, **15**(3), 313–330.

Ahmed, M., Akram, M. N., Asim, M., Aslam, M., Hassan, F.-u., Higgins, S., Stöckle, C. O., & Hoogenboom, G. 2016. Calibration and validation of APSIM-Wheat and CERES-Wheat for spring wheat under rainfed conditions: Models evaluation and application. *Computers and Electronics in Agriculture*, **123**, 384–401.

Al Bitar, A., Leroux, D., Kerr, Y. H., Merlin, O., Richaume, P., Sahoo, A., & Wood, E. F. 2012. Evaluation of SMOS soil moisture products over continental US using the SCAN/SNOTEL network. *IEEE Transactions on Geoscience and Remote Sensing*, **50**(5), 1572–1586.

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**(12), 2741–2758.

Arnold Jr, C. P., & Dey, C. H. 1986. Observing-systems simulation experiments: Past, present, and future. *Bulletin of the American Meteorological Society*, **67**(6), 687–695.

Asseng, S., Keating, B., Fillery, I., Gregory, P., Bowden, J., Turner, N., Palta, J., & Abrecht, D. 1998. Performance of the APSIM-wheat model in Western Australia. *Field Crops Research*, **57**(2), 163–179.

Asseng, S., Van Keulen, H., & Stol, W. 2000. Performance and application of the APSIM Nwheat model in the Netherlands. *European journal of agronomy*, **12**(1), 37–54.

Asseng, S., Bar-Tal, A., Bowden, J., Keating, B. A., Van Herwaarden, A., Palta, J., Huth, N., & Probert, M. E. 2002. Simulation of grain protein content with APSIM-Nwheat. *European Journal of Agronomy*, **16**(1), 25–42.

Bai, T., Zhang, N., Mercatoris, B., & Chen, Y. 2019. Improving Jujube Fruit Tree Yield Estimation at the Field Scale by Assimilating a Single Landsat Remotely-Sensed LAI into the WOFOST Model. *Remote Sensing*, **11**(9).

Ban, H. Y., Ahn, J. B., & Lee, B. W. 2019. Assimilating MODIS data-derived minimum input data set and water stress factors into CERES-Maize model improves regional corn yield predictions. *PLoS One*, **14**(2), e0211874.

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**(2), 82–91.

Batchelor, W. D., Basso, B., & Paz, J. O. 2002. Examples of strategies to analyze spatial and temporal yield variability using crop models. *European Journal of Agronomy*, **18**(1), 141–158.

Boopathi, N., Ye, N., Wu, X., Walker, J. P., Rao, Y., Jackson, T. J., Kerr, Y., Kim, E., McGrath, A., & Yeo, I.-Y. 2018. Towards soil moisture retrieval using tower-based P-band radiometer observations. *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, 1407–1410.

Boschetti, M., Stroppiana, D., Brivio, P. A., & Bocchi, S. 2009. Multi-year monitoring of rice crop phenology through time series analysis of MODIS images. *International Journal of Remote Sensing*, **30**(18), 4643–4662.

Bouman, B. 1995. Crop modelling and remote sensing for yield prediction. *NJAS wageningen journal of life sciences*, **43**(2), 143–161.

Bouman, B., Van Diepen, C., Vossen, P., & van Der Wal, T. 1997. Simulation and systems analysis tools for crop yield forecasting. *Application of systems approaches at the farm and regional levels*, 325–340.

Brisson, N., Gary, C., Justes, E., Roche, R., Mary, B., Ripoche, D., Zimmer, D., Sierra, J., Bertuzzi, P., & Burger, P. 2003. An overview of the crop model STICS. *European Journal of agronomy*, **18**(3), 309–332.

Chapman, S. C. 2008. Use of crop models to understand genotype by environment interactions for drought in real-world and simulated plant breeding trials. *Euphytica*, **161**(1-2), 195–208.

Chen, Q., Zeng, J., Cui, C., Li, Z., Chen, K.-S., Bai, X., & Xu, J. 2017. Soil moisture retrieval from SMAP: a validation and error analysis study using ground-based observations over the little Washita watershed. *IEEE Transactions on Geoscience and Remote Sensing*, **56**(3), 1394–1408.

Chen, Y., Zhang, Z., & Tao, F. 2018. Improving regional winter wheat yield estimation through assimilation of phenology and leaf area index from remote sensing data. *European Journal of Agronomy*, **101**, 163–173.

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.

Clevers, J. G. P. W., & Van Leeuwen, H. J. C. 1996. Combined use of optical and microwave remote sensing data for crop growth monitoring. *Remote Sensing of Environment*, **56**(1), 42–51.

Colliander, A., Cosh, M. H., Misra, S., Jackson, T. J., Crow, W. T., Chan, S., Bindlish, R., Chae, C., Collins, C. H., & Yueh, S. H. 2017. Validation and scaling of soil moisture in a semi-arid environment: SMAP validation experiment 2015 (SMAPVEX15). *Remote Sensing of Environment*, **196**, 101–112.

Curnel, Y., de Wit, A. J., Duveiller, G., & Defourny, P. 2011. Potential performances of remotely sensed LAI assimilation in WOFOST model based on an OSS Experiment. *Agricultural and Forest Meteorology*, **151**(12), 1843–1855.

De Lannoy, G. J., Reichle, R. H., Houser, P. R., Pauwels, V. R., & Verhoest, N. E. 2007.

Correcting for forecast bias in soil moisture assimilation with the ensemble Kalman filter. *Water Resources Research*, **43**(9).

De Wit, A. d., & Van Diepen, C. 2007. Crop model data assimilation with the Ensemble Kalman filter for improving regional crop yield forecasts. *Agricultural and Forest Meteorology*, **146**(1), 38–56.

Dee, D. P. 2005. Bias and data assimilation. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, **131**(613), 3323–3343.

Dee, D. P., & Da Silva, A. M. 1998. Data assimilation in the presence of forecast bias. *Quarterly Journal of the Royal Meteorological Society*, **124**(545), 269–295.

Del Grosso, S., Parton, W., Mosier, A., Ojima, D., Kulmala, A., & Phongpan, S. 2000. General model for N2O and N2 gas emissions from soils due to dentrification. *Global Biogeochemical Cycles*, **14**(4), 1045–1060.

Delécolle, R., Maas, S., Guérif, M., & Baret, F. 1992. Remote sensing and crop production models: present trends. *ISPRS Journal of Photogrammetry and Remote Sensing*, **47**(2-3), 145–161.

Dente, L., Satalino, G., Mattia, F., & Rinaldi, M. 2008. Assimilation of leaf area index derived from ASAR and MERIS data into CERES-Wheat model to map wheat yield. *Remote sensing of Environment*, **112**(4), 1395–1407.

Di Baldassarre, G., & Montanari, A. 2009. Uncertainty in river discharge observations: a quantitative analysis. *Hydrology and Earth System Sciences*, **13**(6), 913.

Doherty, J. 1994. PEST: a unique computer program for model-independent parameter optimisation. *Water Down Under 94: Groundwater/Surface Hydrology Common Interest Papers; Preprints of Papers*, 551.

Doherty, J., Brebber, L., & Whyte, P. 1994. PEST: Model-independent parameter estimation. *Watermark Computing, Corinda, Australia*, **122**, 336.

Donatelli, M., van Evert, F. K., Rutgers, B., Trevisan, M., Sorce, A., Balderacchi, M., Di Guardo, A., Ewert, F., Leffelaar, P., & Meuter, E. 2007. *Agricultural Production and Externalities Simulator (APES) prototype to be used in Prototype 1 of SEAMLESS-IF*. Report.

Donatelli, M., Russell, G., Rizzoli, A. E., Acutis, M., Adam, M., Athanasiadis, I. N., Balderacchi, M., Bechini, L., Belhouchette, H., & Bellocchi, G. 2010. *A component-based framework for simulating agricultural production and externalities*. Springer. Pages 63–108.

Dong, T., Liu, J., Qian, B., Zhao, T., Jing, Q., Geng, X., Wang, J., Huffman, T., & Shang, J. 2016. Estimating winter wheat biomass by assimilating leaf area index derived from fusion of Landsat-8 and MODIS data. *International Journal of Applied Earth Observation and Geoinformation*, **49**, 63–74.

Doorenbos, J. 1975. Guidelines for predicting crop water requirements. *Food and Agriculture organization. Rome, Irrig. Drainage pap.*, **24**.

Dorigo, W. A., Zurita-Milla, R., de Wit, A. J., Brazile, J., Singh, R., & Schaepman, M. E. 2007. A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. *International journal of applied earth observation and geoinformation*, **9**(2), 165–193.

Duan, Q., Gupta, V. K., & Sorooshian, S. 1993. Shuffled complex evolution approach for effective and efficient global minimization. *Journal of optimization theory and applications*, **76**(3), 501–521.

Eberhart, R., & Kennedy, J. 1995. Particle swarm optimization. *Proceedings of the IEEE international conference on neural networks*, **4**, 1942–1948.

Eitzinger, J., Trnka, M., Hösch, J., Žalud, Z., & Dubrovský, M. 2004. Comparison of CERES, WOFOST and SWAP models in simulating soil water content during growing season under different soil conditions. *Ecological Modelling*, **171**(3), 223–246.

Evensen, G. 2003. The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean dynamics*, **53**(4), 343–367.

Fang, H., Liang, S., Hoogenboom, G., Teasdale, J., & Cavigelli, M. 2008. Corn-yield estimation through assimilation of remotely sensed data into the CSM-CERES-Maize model. *International Journal of Remote Sensing*, **29**(10), 3011–3032.

Farré, I., Robertson, M., Walton, G., & Asseng, S. 2002. Simulating phenology and yield response of canola to sowing date in Western Australia using the APSIM model. *Australian Journal of Agricultural Research*, **53**(10), 1155–1164.

Fensholt, R., Sandholt, I., & Rasmussen, M. S. 2004. Evaluation of MODIS LAI, fAPAR and the relation between fAPAR and NDVI in a semi-arid environment using in situ measurements. *Remote sensing of Environment*, **91**(3), 490–507.

Fischer, A., Kergoat, L., & Dedieu, G. 1997. Coupling satellite data with vegetation functional models: review of different approaches and perspectives suggested by the assimilation strategy. *Remote Sensing Reviews*, **15**(1-4), 283–303.

Godwin, D., & Jones, C. A. 1991. Nitrogen dynamics in soil-plant systems. *Modeling plant and soil systems*, 287–321.

Guo, C., Tang, Y., Lu, J., Zhu, Y., Cao, W., Cheng, T., Zhang, L., & Tian, Y. 2019. Predicting wheat productivity: Integrating time series of vegetation indices into crop modeling via sequential assimilation. *Agricultural and Forest Meteorology*, **272-273**, 69–80.

Guyot, G. 1990. Optical properties of vegetation canopies. *Optical properties of vegetation canopies.*, 19–43.

Guérif, M., & Duke, C. 1998. Calibration of the SUCROS emergence and early growth module for sugar beet using optical remote sensing data assimilation. *European journal of agronomy*, **9**(2), 127–136.

Holzworth, D. P., Huth, N. I., deVoil, P. G., Zurcher, E. J., Herrmann, N. I., McLean, G., Chenu, K., van Oosterom, E. J., Snow, V., & Murphy, C. 2014. APSIM–evolution towards a new generation of agricultural systems simulation. *Environmental Modelling & Software*, **62**, 327–350.

Houser, P. R., De Lannoy, G. J., & Walker, J. P. 2010. *Land surface data assimilation*. Springer. Pages 549–597.

Huang, J., Tian, L., Liang, S., Ma, H., Becker-Reshef, I., Huang, Y., Su, W., Zhang, X., Zhu, D., & Wu, W. 2015a. Improving winter wheat yield estimation by assimilation of the leaf area index from Landsat TM and MODIS data into the WOFOST model. *Agricultural and Forest Meteorology*, **204**, 106–121.

Huang, J., Ma, H., Su, W., Zhang, X., Huang, Y., Fan, J., & Wu, W. 2015b. Jointly assimilating MODIS LAI and ET products into the SWAP model for winter wheat yield estimation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **8**(8), 4060–4071.

Huang, J., Sedano, F., Huang, Y., Ma, H., Li, X., Liang, S., Tian, L., Zhang, X., Fan, J., & Wu, W. 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.

Huang, J., Ma, H., Sedano, F., Lewis, P., Liang, S., Wu, Q., Su, W., Zhang, X., & Zhu, D. 2019. Evaluation of regional estimates of winter wheat yield by assimilating three remotely sensed reflectance datasets into the coupled WOFOST–PROSAIL model. *European Journal of Agronomy*, **102**, 1–13.

Ines, A. V., 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.

Jackson, T. J., Bindlish, R., Cosh, M. H., Zhao, T., Starks, P. J., Bosch, D. D., Seyfried, M., Moran, M. S., Goodrich, D. C., & Kerr, Y. H. 2012. Validation of Soil Moisture and Ocean Salinity (SMOS) soil moisture over watershed networks in the US. *IEEE Transactions on Geoscience and Remote Sensing*, **50**(5), 1530–1543.

Jacquemoud, S., & Baret, F. 1990. PROSPECT: A model of leaf optical properties spectra. *Remote sensing of environment*, **34**(2), 75–91.

Jamet, C., & Loisel, H. 2013. Data Assimilation Methods. *Surface Ocean-Lower Atmosphere Processes*, 303–317.

Jamieson, P., Porter, J., Goudriaan, J., Ritchie, J. v., Van Keulen, H., & Stol, W. 1998a. A comparison of the models AFRCWHEAT2, CERES-Wheat, Sirius, SUCROS2 and SWHEAT with measurements from wheat grown under drought. *Field Crops Research*, **55**(1-2), 23–44.

Jamieson, P., Semenov, M., Brooking, I., & Francis, G. 1998b. Sirius: a mechanistic model of wheat response to environmental variation. *European Journal of Agronomy*, **8**(3), 161–179.

Jansson, P.-E., & Karlberg, L. 2010. COUP manual: Coupled Heat and Mass Transfer Model for Soileplanteatmosphere Systems. **2017**.

Jin, H., Li, A., Wang, J., & Bo, Y. 2016a. Improvement of spatially and temporally continuous crop leaf area index by integration of CERES-Maize model and MODIS data. *European Journal of Agronomy*, **78**, 1–12.

Jin, X., Kumar, L., Li, Z., Xu, X., Yang, G., & Wang, J. 2016b. Estimation of Winter Wheat Biomass and Yield by Combining the AquaCrop Model and Field Hyperspectral Data. *Remote Sensing*, **8**(12).

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.

Jones, C. A., Kiniry, J. R., & Dyke, P. 1986. *CERES-Maize: A simulation model of maize growth and development*. Texas A& M University Press.

Jones, J. W., Hoogenboom, G., Porter, C. H., Boote, K. J., Batchelor, W. D., Hunt, L., Wilkens, P. W., Singh, U., Gijsman, A. J., & Ritchie, J. T. 2003. The DSSAT cropping system model. *European journal of agronomy*, **18**(3), 235–265.

Kanamitsu, M. 1989. Description of the NMC global data assimilation and forecast system. *Weather and Forecasting*, **4**(3), 335–342.

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.

Kayad, A. G., Al-Gaadi, K. A., Tola, E., Madugundu, R., Zeyada, A. M., & Kalaitzidis, C. 2016. Assessing the spatial variability of alfalfa yield using satellite imagery and ground-based data. *PloS one*, **11**(6), e0157166.

Keating, B., Scientific, C., Meinke, H., Probert, M., Huth, N., & Hills, I. 2001. NWheat: documentation and performance of a wheat module for APSIM.

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., McLean, G., Verburg, K., Snow, V., Dimes, J. P., Silburn, M., Wang, E., Brown, S., Bristow, K. L., Asseng, S., Chapman, S., McCown, R. L., Freebairn, D. M., & Smith, C. J. 2003. An overview of APSIM, a model designed for farming systems simulation. *European Journal of Agronomy*, **18**(3-4), 267–288.

Kennedy, J., & Eberhart, R. 1995. Particle swarm optimization. *Proceedings of ICNN'95-International Conference on Neural Networks*, **4**, 1942–1948.

Kergoat, L., Fischer, A., Moulin, S., & Dedieu, G. 1995. Satellite measurements as a constraint on estimates of vegetation carbon budget. *Tellus B: Chemical and Physical Meteorology*, **47**(1-2), 251–263.

Kerr, Y. H., Waldteufel, P., Richaume, P., Wigneron, J. P., Ferrazzoli, P., Mahmoodi, A., Al Bitar, A., Cabot, F., Gruhier, C., & Juglea, S. E. 2012. The SMOS soil moisture retrieval algorithm. *IEEE Transactions on Geoscience and Remote Sensing*, **50**(5), 1384–1403.

Kumar, S. V., Reichle, R. H., Harrison, K. W., Peters-Lidard, C. D., Yatheendradas, S., & Santanello, J. A. 2012. A comparison of methods for a priori bias correction in soil moisture data assimilation. *Water Resources Research*, **48**(3).

Launay, M., & Guerif, M. 2005. Assimilating remote sensing data into a crop model to improve predictive performance for spatial applications. *Agriculture, ecosystems & environment*, **111**(1), 321–339.

Li, H., Kalnay, E., Miyoshi, T., & Danforth, C. M. 2009. Accounting for model errors in ensemble data assimilation. *Monthly Weather Review*, **137**(10), 3407–3419.

Li, H., Chen, Z. X., Wu, W. B., Jiang, Z. W., Liu, B., & Hasi, T. 2015a. Crop Model Data Assimilation with Particle Filter for Yield Prediction Using Leaf Area Index of Different Temporal Scales. *2015 Fourth International Conference on Agro-Geoinformatics*.

Li, H., Jiang, Z.-w., Chen, Z.-x., Ren, J.-q., Liu, B., & Hasituya. 2017a. Assimilation of temporal-spatial leaf area index into the CERES-Wheat model with ensemble Kalman filter and uncertainty assessment for improving winter wheat yield estimation. *Journal of Integrative Agriculture*, **16**(10), 2283–2299.

Li, H., Chen, Z., Liu, G., Jiang, Z., & Huang, C. 2017b. Improving Winter Wheat Yield Estimation from the CERES-Wheat Model to Assimilate Leaf Area Index with Different Assimilation Methods and Spatio-Temporal Scales. *Remote Sensing*, **9**(3), 190.

Li, Y., Ryu, D., Western, A. W., Wang, Q., Robertson, D. E., & Crow, W. T. 2014. An integrated error parameter estimation and lag-aware data assimilation scheme for real-time flood forecasting. *Journal of hydrology*, **519**, 2722–2736.

Li, Z., Jin, X., Zhao, C., Wang, J., Xu, X., Yang, G., Li, C., & Shen, J. 2015b. Estimating wheat yield and quality by coupling the DSSAT-CERES model and proximal remote sensing. *European Journal of Agronomy*, **71**, 53–62.

Linker, R., & Ioslovich, I. 2017. Assimilation of canopy cover and biomass measurements in the crop model AquaCrop. *Biosystems Engineering*, **162**, 57–66.

Lisson, S., & Cotching, W. 2011. Modelling the fate of water and nitrogen in the mixed vegetable farming systems of northern Tasmania, Australia. *Agricultural Systems*, **104**(8), 600–608.

Littleboy, M., Silburn, D., Freebairn, D., Woodruff, D., Hammer, G., & Leslie, J. 1992. Impact of soil erosion on production in cropping systems. I. Development and validation of a simulation model. *Soil Research*, **30**(5), 757–774.

Liu, D., Mishra, A. K., & Yu, Z. 2019. Evaluation of hydroclimatic variables for maize yield estimation using crop model and remotely sensed data assimilation. *Stochastic Environmental Research and Risk Assessment*, **33**(7), 1283–1295.

Liu, P., Bongiovanni, T., Monsivais-Huertero, A., Judge, J., Steele-Dunne, S., Bindlish, R., & Jackson, T. J. 2016. Assimilation of active and passive microwave observations for improved estimates of soil moisture and crop growth. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **9**(4), 1357–1369.

Luo, Q., & Kathuria, A. 2013. Modelling the response of wheat grain yield to climate change: a sensitivity analysis. *Theoretical and applied climatology*, **111**(1-2), 173–182.

Lægdsmand, M. 2011. FASSET documentation: Crop simulation model.

Ma, G., Huang, J., Wu, W., Fan, J., Zou, J., & Wu, S. 2013a. Assimilation of MODIS-LAI into the WOFOST model for forecasting regional winter wheat yield. *Mathematical and Computer Modelling*, **58**(3), 634–643.

Ma, H., Huang, J., Zhu, D., Liu, J., Su, W., Zhang, C., & Fan, J. 2013b. Estimating regional winter wheat yield by assimilation of time series of HJ-1 CCD NDVI into WOFOST–ACRM model with Ensemble Kalman Filter. *Mathematical and Computer Modelling*, **58**(3), 759–770.

Ma, Y., Wang, S., Zhang, L., Hou, Y., Zhuang, L., He, Y., & Wang, F. 2008. Monitoring winter wheat growth in North China by combining a crop model and remote sensing data. *International Journal of Applied Earth Observation and Geoinformation*, **10**(4), 426–437.

Maas, S. 1988a. Using satellite data to improve model estimates of crop yield. *Agronomy Journal*, **80**(4), 655–662.

Maas, S. J. 1988b. Use of remotely-sensed information in agricultural crop growth models. *Ecological modelling*, **41**(3-4), 247–268.

Maestrini, B., & Basso, B. 2018. Predicting spatial patterns of within-field crop yield variability. *Field crops research*, **219**, 106–112.

Marc, B. 2014. *Introduction to the principles and methods of data assimilation in the geosciences*.

McMillan, H., Krueger, T., & Freer, J. 2012. Benchmarking observational uncertainties for

hydrology: rainfall, river discharge and water quality. *Hydrological Processes*, **26**(26), 4078–4111.

Monteith, J. L. 1986. How do crops manipulate water supply and demand? *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, **316**(1537), 245–259.

Mosleh, M. K., Hassan, Q. K., & Chowdhury, E. H. 2015. Application of remote sensors in mapping rice area and forecasting its production: a review. *Sensors (Basel)*, **15**(1), 769–91.

Moulin, S., Bondeau, A., & Delecolle, R. 1998. Combining agricultural crop models and satellite observations: from field to regional scales. *International Journal of Remote Sensing*, **19**(6), 1021–1036.

Nagarajan, K., Judge, J., Monsivais-Huertero, A., & Graham, W. D. 2012. Impact of assimilating passive microwave observations on root-zone soil moisture under dynamic vegetation conditions. *IEEE Transactions on Geoscience and Remote Sensing*, **50**(11), 4279–4291.

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**(5).

Nendel, C., Berg, M., Kersebaum, K., Mirschel, W., Specka, X., Wegehenkel, M., Wenkel, K., & Wieland, R. 2011. The MONICA model: testing predictability for crop growth, soil moisture and nitrogen dynamics. *Ecological Modelling*, **222**(9), 1614–1625.

Noori, O., & Panda, S. S. 2016. Site-specific management of common olive: Remote sensing, geospatial, and advanced image processing applications. *Computers and Electronics in Agriculture*, **127**, 680–689.

Novelli, F., Spiegel, H., Sandén, T., & Vuolo, F. 2019. Assimilation of Sentinel-2 Leaf Area Index Data into a Physically-Based Crop Growth Model for Yield Estimation. *Agronomy*, **9**(5).

Olesen, J. E., Petersen, B. M., Berntsen, J., Hansen, S., Jamieson, P., & Thomsen, A. 2002. Comparison of methods for simulating effects of nitrogen on green area index and dry matter growth in winter wheat. *Field Crops Research*, **74**(2), 131–149.

Olioso, A., Inoue, Y., Ortega-Farias, S., Demarty, J., Wigneron, J.-P., Braud, I., Jacob, F., Lecharpentier, P., Ottlé, C., & Calvet, J.-C. 2005. Future directions for advanced evapotranspiration modeling: Assimilation of remote sensing data into crop simulation models and SVAT models. *Irrigation and Drainage Systems*, **19**(3-4), 377–412.

Otter, S., & Ritchie, J. T. 1985. *Validation of the CERES-wheat model in diverse environments*. Springer. Pages 307–310.

Palosuo, T., Kersebaum, K. C., Angulo, C., Hlavinka, P., Moriondo, M., Olesen, J. E., Patil, R. H., Ruget, F., Rumbaur, C., & Takáč, J. 2011. Simulation of winter wheat yield and its variability in different climates of Europe: a comparison of eight crop growth models. *European Journal of Agronomy*, **35**(3), 103–114.

Panda, S. S., Hoogenboom, G., & Paz, J. O. 2010. Remote sensing and geospatial technological applications for site-specific management of fruit and nut crops: a review. *Remote Sensing*, **2**(8), 1973–1997.

Pannkuk, C., Stockle, C., & Papendick, R. 1998. Evaluating CropSyst simulations of wheat management in a wheat-fallow region of the US Pacific Northwest. *Agricultural systems*, **57**(2), 121–134.

Patel, H., GG–SHROFF, J. P., V–SHEKH, A. V., & RP–BHATT, B. 2010. Calibration and validation of CERES-wheat model for wheat in middle Gujarat region. *Journal of Agrometeorology*, **12**(1), 114–117.

Paustian, M., & Theuvsen, L. 2017. Adoption of precision agriculture technologies by German crop farmers. *Precision agriculture*, **18**(5), 701–716.

Pauwels, V., Verhoest, N. E., De Lannoy, G. J., Guissard, V., Lucau, C., & Defourny, P. 2007. Optimization of a coupled hydrology–crop growth model through the assimilation of observed soil moisture and leaf area index values using an ensemble Kalman filter. *Water Resources Research*, **43**(4).

Pellenq, J., & Boulet, G. 2004. A methodology to test the pertinence of remote-sensing data assimilation into vegetation models for water and energy exchange at the land surface. *Agronomie*, **24**(4), 197–204.

Penman, H. 1956. Evaporation: an introductory survey. *Neth. J. Agric. Sci.*, **4**, 9–29.

Petrie, R. E., & Dance, S. L. 2010. Ensemble-based data assimilation and the localisation problem. *Weather*, **65**(3), 65–69.

Plant, R. E. 2001. Site-specific management: the application of information technology to crop production. *Computers and Electronics in Agriculture*, **30**(1), 9–29.

Plummer, S. 2000. Perspectives on combining ecological process models and remotely sensed data. *Ecological Modelling*, **129**(2), 169–186.

Pollacco, J. A. P. 2008. A generally applicable pedotransfer function that estimates field capacity and permanent wilting point from soil texture and bulk density. *Canadian Journal of Soil Science*, **88**(5), 761–774.

Porter, J., Jamieson, P., & Wilson, D. 1993. Comparison of the wheat simulation models AFRCWHEAT2, CERES-Wheat and SWHEAT for non-limiting conditions of crop growth. *Field Crops Research*, **33**(1-2), 131–157.

Price, W. L. 1977. A controlled random search procedure for global optimisation. *The Computer Journal*, **20**(4), 367–370.

Priestley, C., & Taylor, R. 1972. On the assessment of surface heat flux and evaporation using large-scale parameters. *Monthly weather review*, **100**(2), 81–92.

Probert, M., Dimes, J., Keating, B., Dalal, R., & Strong, W. 1998. APSIM's water and nitrogen modules and simulation of the dynamics of water and nitrogen in fallow systems. *Agricultural Systems*, **56**(1), 1–28.

Prévot, L., Chauki, H., Troufleau, D., Weiss, M., Baret, F., & Brisson, N. 2003. Assimilating optical and radar data into the STICS crop model for wheat. *Agronomie*, **23**(4), 297–303.

Raes, D., Steduto, P., Hsiao, T. C., & Fereres, E. 2009. AquaCrop the FAO crop model to simulate yield response to water: II. Main algorithms and software description. *Agronomy Journal*, **101**(3), 438–447.

Reed, B. C., Brown, J. F., VanderZee, D., Loveland, T. R., Merchant, J. W., & Ohlen, D. O. 1994. Measuring phenological variability from satellite imagery. *Journal of vegetation science*, **5**(5), 703–714.

Reichle, R. H. 2008. Data assimilation methods in the Earth sciences. *Advances in Water Resources*, **31**(11), 1411–1418.

Reichle, R. H., McLaughlin, D. B., & Entekhabi, D. 2002. Hydrologic data assimilation with the ensemble Kalman filter. *Monthly Weather Review*, **130**(1), 103–114.

Robertson, M. 2006. *What do we mean by spatial and temporal variability*.

Robertson, M., Carberry, P., Huth, N., Turpin, J., Probert, M. E., Poulton, P., Bell, M., Wright, G., Yeates, S., & Brinsmead, R. 2002. Simulation of growth and development of diverse legume species in APSIM. *Australian Journal of Agricultural Research*, **53**(4), 429–446.

Rodell, M., Houser, P., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C., Arsenault, K., Cosgrove, B., Radakovich, J., & Bosilovich, M. 2004. The global land data assimilation system. *Bulletin of the American Meteorological Society*, **85**(3), 381–394.

Rouse, D. 1988. Use of crop growth-models to predict the effects of disease. *Annual Review of Phytopathology*, **26**(1), 183–201.

Rötter, R. P., Palosuo, T., Kersebaum, K. C., Angulo, C., Bindi, M., Ewert, F., Ferrise, R., Hlavinka, P., Moriondo, M., & Nendel, C. 2012. Simulation of spring barley yield in different climatic zones of Northern and Central Europe: a comparison of nine crop models. *Field Crops Research*, **133**, 23–36.

Sadeh, Y. 2020. Coupling Fused High Spatio-Temporal Resolution Remote Sensing Data and Crop Modelling to Predict Wheat Yield at the Field Scale.

Sakamoto, T., Yokozawa, M., Toritani, H., Shibayama, M., Ishitsuka, N., & Ohno, H. 2005. A crop phenology detection method using time-series MODIS data. *Remote sensing of environment*, **96**(3), 366–374.

Salo, T. J., Palosuo, T., Kersebaum, K. C., Nendel, C., Angulo, C., Ewert, F., Bindi, M., Calanca, P., Klein, T., & Moriondo, M. 2016. Comparing the performance of 11 crop simulation models in predicting yield response to nitrogen fertilization. *The Journal of Agricultural Science*, **154**(7), 1218–1240.

Sanchez, N., Martínez-Fernández, J., Scaini, A., & Perez-Gutierrez, C. 2012. Validation of the SMOS L2 soil moisture data in the REMEDHUS network (Spain). *IEEE Transactions on Geoscience and Remote Sensing*, **50**(5), 1602–1611.

Semenov, M., Wolf, J., Evans, L., Eckersten, H., & Iglesias, A. 1996. Comparison of wheat simulation models under climate change. II. Application of climate change scenarios. *Climate Research*, 271–281.

Serbin, S. P., Ahl, D. E., & Gower, S. T. 2013. Spatial and temporal validation of the MODIS LAI and FPAR products across a boreal forest wildfire chronosequence. *Remote Sensing of Environment*, **133**, 71–84.

Service, Soil Conservation. 2004. *Chapter 10: Estimation of Direct Runoff from Storm Rainfall*. Book Section 10.

Shaw, R., Lark, R., Williams, A., Chadwick, D., & Jones, D. 2016. Characterising the within-field scale spatial variation of nitrogen in a grassland soil to inform the efficient design of in-situ nitrogen sensor networks for precision agriculture. *Agriculture, Ecosystems & Environment*, **230**, 294–306.

Silvestro, P., Pignatti, S., Pascucci, S., Yang, H., Li, Z., Yang, G., Huang, W., & Casa, R. 2017. Estimating Wheat Yield in China at the Field and District Scale from the Assimilation of Satellite Data into the Aquacrop and Simple Algorithm for Yield (SAFY) Models. *Remote Sensing*, **9**(5).

Singh, A., Goyal, V., Mishra, A., & Parihar, S. 2013. Validation of CropSyst simulation model for direct seeded rice–wheat cropping system. *Current science*, 1324–1331.

Son, N., Chen, C., Chen, C., Chang, L., Duc, H., & Nguyen, L. 2013. Prediction of rice crop yield using MODIS EVI− LAI data in the Mekong Delta, Vietnam. *International journal of remote sensing*, **34**(20), 7275–7292.

Standard, Australian. 2009. Methods for testing soils for engineering purposes. Method 3.6.1: Soil classification tests-Determination of the particle size distribution of a soil-Standard method of analysis by sieving. AS 1289.3.6.1:2009.

Standard, Australian. 2015. Methods for sampling and testing aggregates. Method 2: Basic testing equipment. AS 1141.2:2015.

Stuart, A., & Zygalakis, K. 2015. *Data Assimilation: A Mathematical Introduction*. Report. Oak Ridge National Laboratory (ORNL), Oak Ridge, TN (United States).

Stöckle, C. O., Donatelli, M., & Nelson, R. 2003. CropSyst, a cropping systems simulation model. *European journal of agronomy*, **18**(3), 289–307.

Tan, B., Hu, J., Zhang, P., Huang, D., Shabanov, N., Weiss, M., Knyazikhin, Y., & Myneni, R. B. 2005. Validation of Moderate Resolution Imaging Spectroradiometer leaf area index product in croplands of Alpilles, France. *Journal of Geophysical Research: Atmospheres*, **110**(D1).

Thorburn, P. J., Probert, M. E., & Robertson, F. A. 2001. Modelling decomposition of sugar cane surface residues with APSIM–Residue. *Field Crops Research*, **70**(3), 223–232.

Thorburn, P. J., Biggs, J. S., Collins, K., & Probert, M. 2010. Using the APSIM model to estimate nitrous oxide emissions from diverse Australian sugarcane production systems. *Agriculture, Ecosystems & Environment*, **136**(3-4), 343–350.

Thorp, K., Wang, G., West, A., Moran, M., Bronson, K., White, J., & Mon, J. 2012. Estimating crop biophysical properties from remote sensing data by inverting linked radiative transfer and ecophysiological models. *Remote Sensing of Environment*, **124**, 224–233.

Timsina, J., & Humphreys, E. 2006. Performance of CERES-Rice and CERES-Wheat models in rice–wheat systems: A review. *Agricultural Systems*, **90**(1-3), 5–31.

Turner, M., Walker, J., & Oke, P. 2008. Ensemble member generation for sequential data assimilation. *Remote Sensing of Environment*, **112**(4), 1421–1433.

Van Diepen, C. v., Wolf, J., Van Keulen, H., & Rappoldt, C. 1989. WOFOST: a simulation model of crop production. *Soil use and management*, **5**(1), 16–24.

Van Ittersum, M., Howden, S., & Asseng, S. 2003. Sensitivity of productivity and deep drainage of wheat cropping systems in a Mediterranean environment to changes in CO 2, temperature and precipitation. *Agriculture, Ecosystems & Environment*, **97**(1), 255–273.

Vazifedoust, M., Van Dam, J., Bastiaanssen, W., & Feddes, R. 2009. Assimilation of satellite data into agrohydrological models to improve crop yield forecasts. *International Journal of Remote Sensing*, **30**(10), 2523–2545.

Vina, A., Gitelson, A. A., Rundquist, D. C., Keydan, G., Leavitt, B., & Schepers, J. 2004. Monitoring maize (Zea mays L.) phenology with remote sensing. *Agronomy Journal*, **96**(4), 1139–1147.

Weiss, M., Baret, F., Garrigues, S., & Lacaze, R. 2007. LAI and fAPAR CYCLOPES global products derived from VEGETATION. Part 2: validation and comparison with MODIS collection 4 products. *Remote sensing of Environment*, **110**(3), 317–331.

Wendling, U., Schellin, H.-G., & Thomä, M. 1991. Bereitstellung von täglichen Informationen zum Wasserhaushalt des Bodens für die Zwecke der agrarmeteorologischen Beratung. *Zeitschrift für Meteorologie*, **41**(6), 468–475.

Whelan, B. 2006. *Yield variability due to weeds, insects or disease, and to some abiotic factors and past management*. Pages 30–32.

Whelan, B., & Palmer, A. 2006. *Impact of soil attributes on crop variability*. Pages 21–29.

Wiegand, C. L., Richardson, A. J., Jackson, R. D., Pinter, P. J., Aase, J. K., Smika, D. E., Lautenschlager, L. F., & McMurtrey, J. 1986. Development of agrometeorological crop model inputs from remotely sensed information. *IEEE transactions on geoscience and remote sensing*, 90–98.

Williams, J., Renard, K., & Dyke, P. 1983. EPIC: A new method for assessing erosion's effect on soil productivity. *Journal of Soil and water Conservation*, **38**(5), 381–383.

Wolf, J., Evans, L., Semenov, M., Eckersten, H., & Iglesias, A. 1996. Comparison of wheat simulation models under climate change. I. Model calibration and sensitivity analyses. *Climate Research*, 253–270.

Xie, Y., Wang, P., Sun, H., Zhang, S., & Li, L. 2017. Assimilation of Leaf Area Index and Surface Soil Moisture With the CERES-Wheat Model for Winter Wheat Yield Estimation Using a Particle Filter Algorithm. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **10**(4), 1303–1316.

Xiong, W., Conway, D., Holman, I., & Lin, E. 2008. Evaluation of CERES-Wheat simulation of Wheat Production in China. *Agronomy Journal*, **100**(6), 1720–1728.

Yang, W., Tan, B., Huang, D., Rautiainen, M., Shabanov, N. V., Wang, Y., Privette, J. L., Huemmrich, K. F., Fensholt, R., & Sandholt, I. 2006. MODIS leaf area index products: From validation to algorithm improvement. *IEEE Transactions on Geoscience and Remote Sensing*, **44**(7), 1885–1898.

You, X., Meng, J., Zhang, M., & Dong, T. 2013. Remote Sensing Based Detection of Crop Phenology for Agricultural Zones in China Using a New Threshold Method. *Remote Sensing*, **5**(7), 3190–3211.

Zadoks, J. C., Chang, T. T., & Konzak, C. F. 1974. A decimal code for the growth stages of cereals. *Weed research*, **14**(6), 415–421.

Zhang, W., Liu, W., Xue, Q., Chen, J., & Han, X. 2013. Evaluation of the AquaCrop model for simulating yield response of winter wheat to water on the southern Loess Plateau of China. *Water Sci Technol*, **68**(4), 821–8.

Zhang, X., Friedl, M. A., Schaaf, C. B., Strahler, A. H., Hodges, J. C. F., Gao, F., Reed, B. C., & Huete, A. 2003. Monitoring vegetation phenology using MODIS. *Remote Sensing of Environment*, **84**(3), 471–475.

Zhang, Y., Feng, L., Wang, E., Wang, J., & Li, B. 2012. Evaluation of the APSIM-Wheat

model in terms of different cultivars, management regimes and environmental conditions. *Canadian Journal of Plant Science*, **92**(5), 937–949.

Zhao, G., Bryan, B. A., & Song, X. 2014. Sensitivity and uncertainty analysis of the APSIM-wheat model: Interactions between cultivar, environmental, and management parameters. *Ecological Modelling*, **279**, 1–11.

Zhao, Y., Chen, S., & Shen, S. 2013. Assimilating remote sensing information with crop model using Ensemble Kalman Filter for improving LAI monitoring and yield estimation. *Ecological modelling*, **270**, 30–42.

Zheng, B., Chenu, K., Alastair, D., & Scott, C. 2014. *The APSIM-Wheat Module*. Report.

Appendix A

Sensitivity analysis results

A.1 Impact on phenology

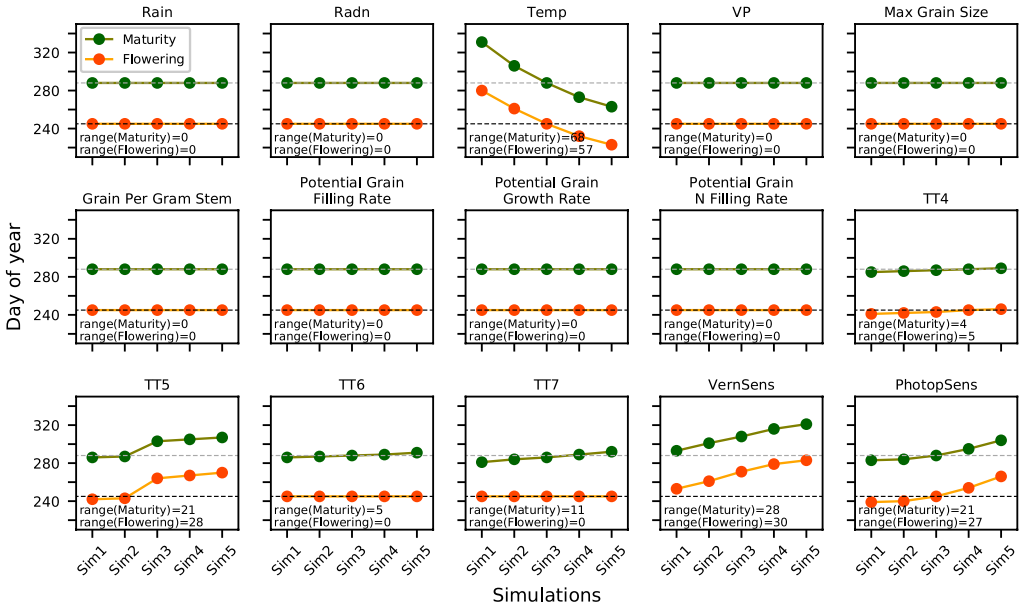

Figure A.1: Change of the flowering and maturity dates in response to the variation of weather and cultivar parameters. Gridlines are the flowering (grey) and maturity (black) dates of the reference simulation.

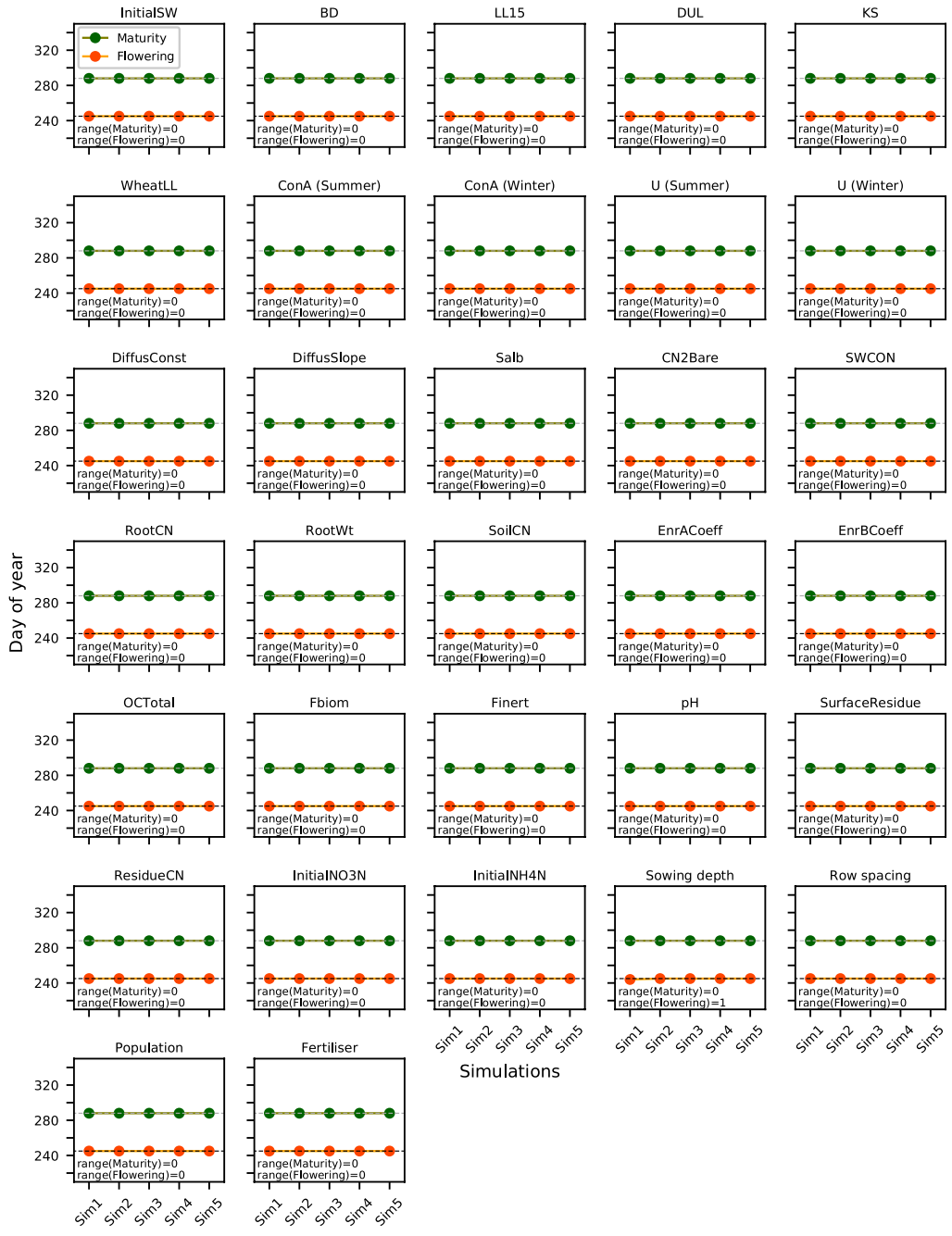

Figure A.2: Change of the flowering and maturity dates in response to the variation of soil parameters and management information. Gridlines are the flowering (grey) and maturity (black) dates of the reference simulation.

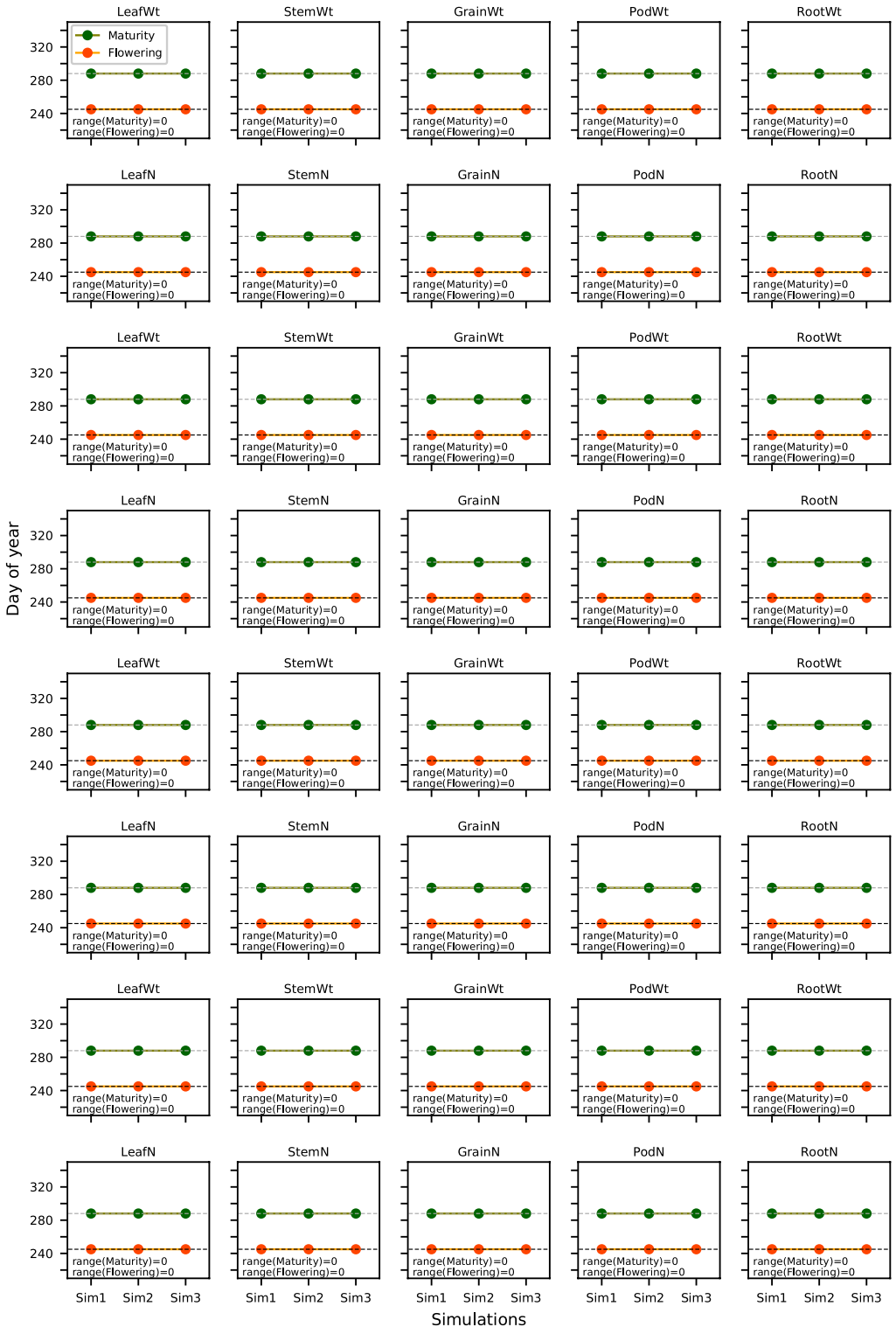

Figure A.3: Change of the flowering and maturity dates in response to the wheat states in different phenology states. Gridlines are the flowering (grey) and maturity (black) dates of the reference simulation.

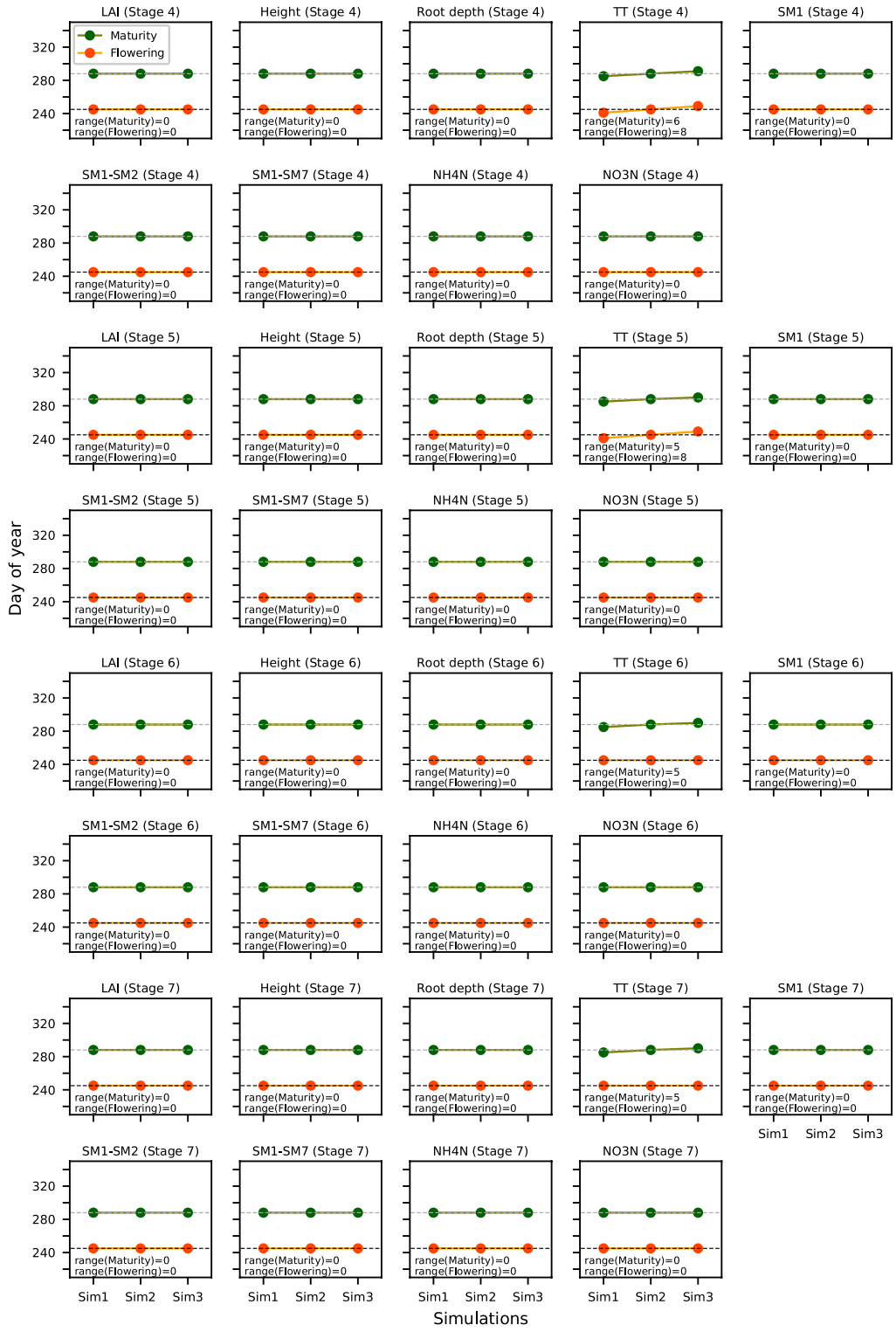

Figure A.4: Change of the flowering and maturity dates in response to the wheat and soil states in different phenology states. Gridlines are the flowering (grey) and maturity (black) dates of the reference simulation.

A.2 Impact on mLAI

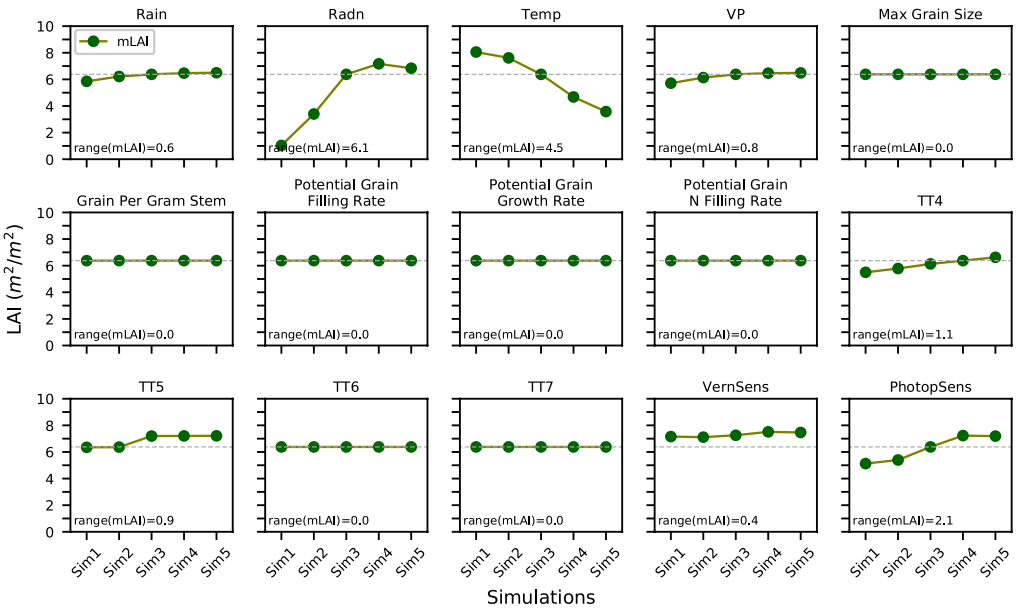

Figure A.5: Change of the mLAI in response to the variation of weather and cultivar parameters. The grey gridline is the mLAI of the reference simulation.

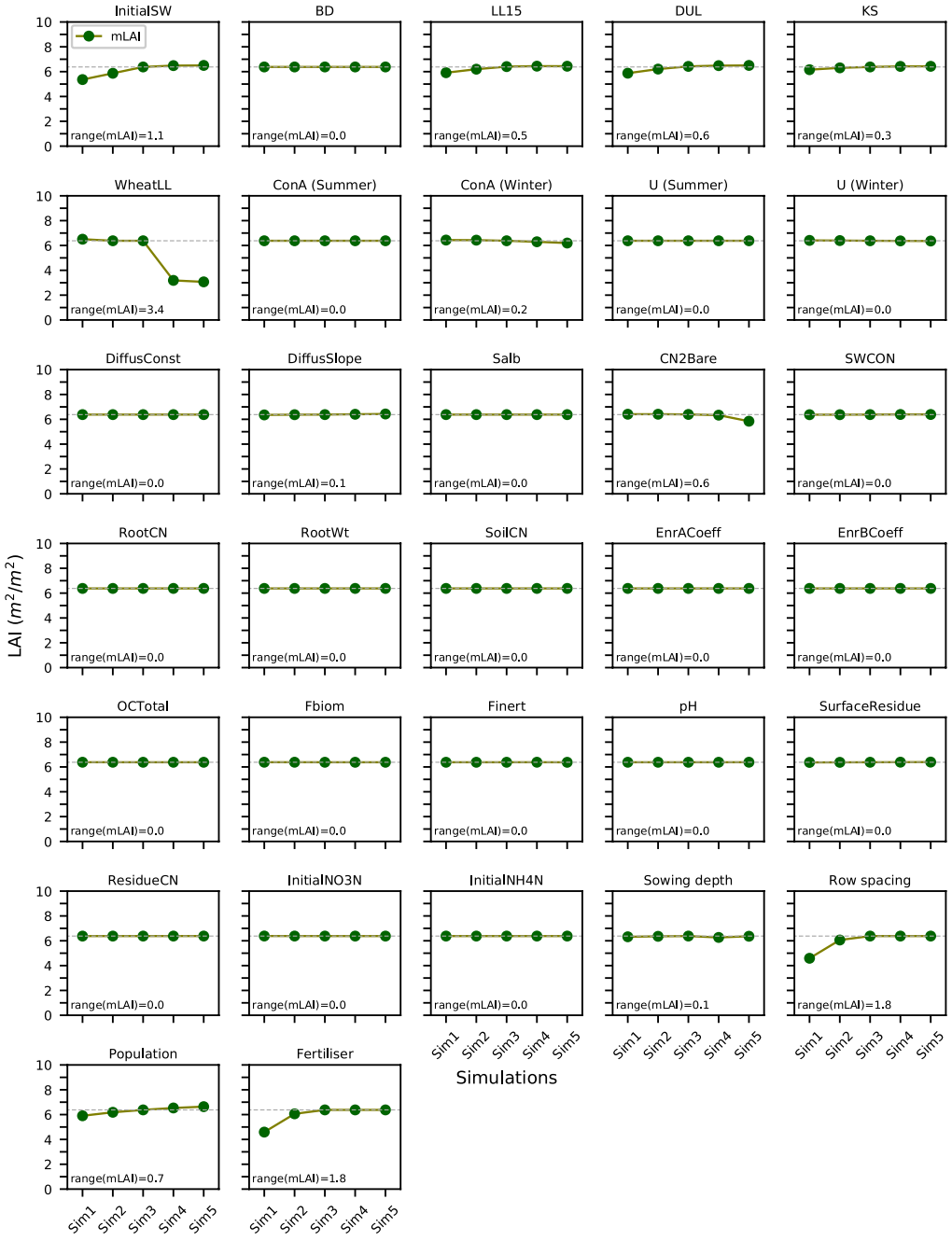

Figure A.6: Change of the mLAI in response to the variation of soil parameters and management information. The grey gridline is the mLAI of the reference simulation.

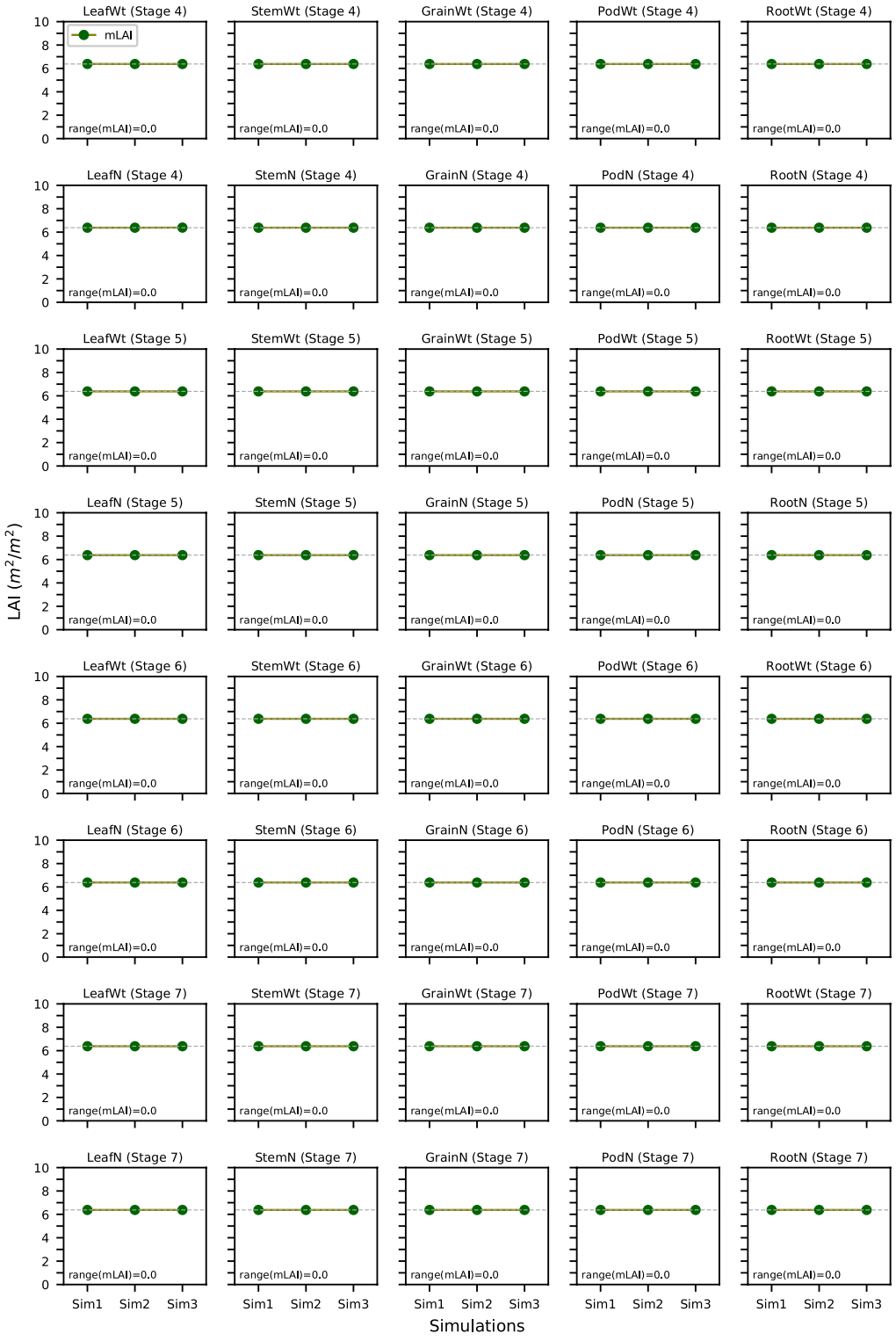

Figure A.7: Change of the mLAI in response to the wheat states in different phenology states. The grey gridline is the mLAI of the reference simulation.

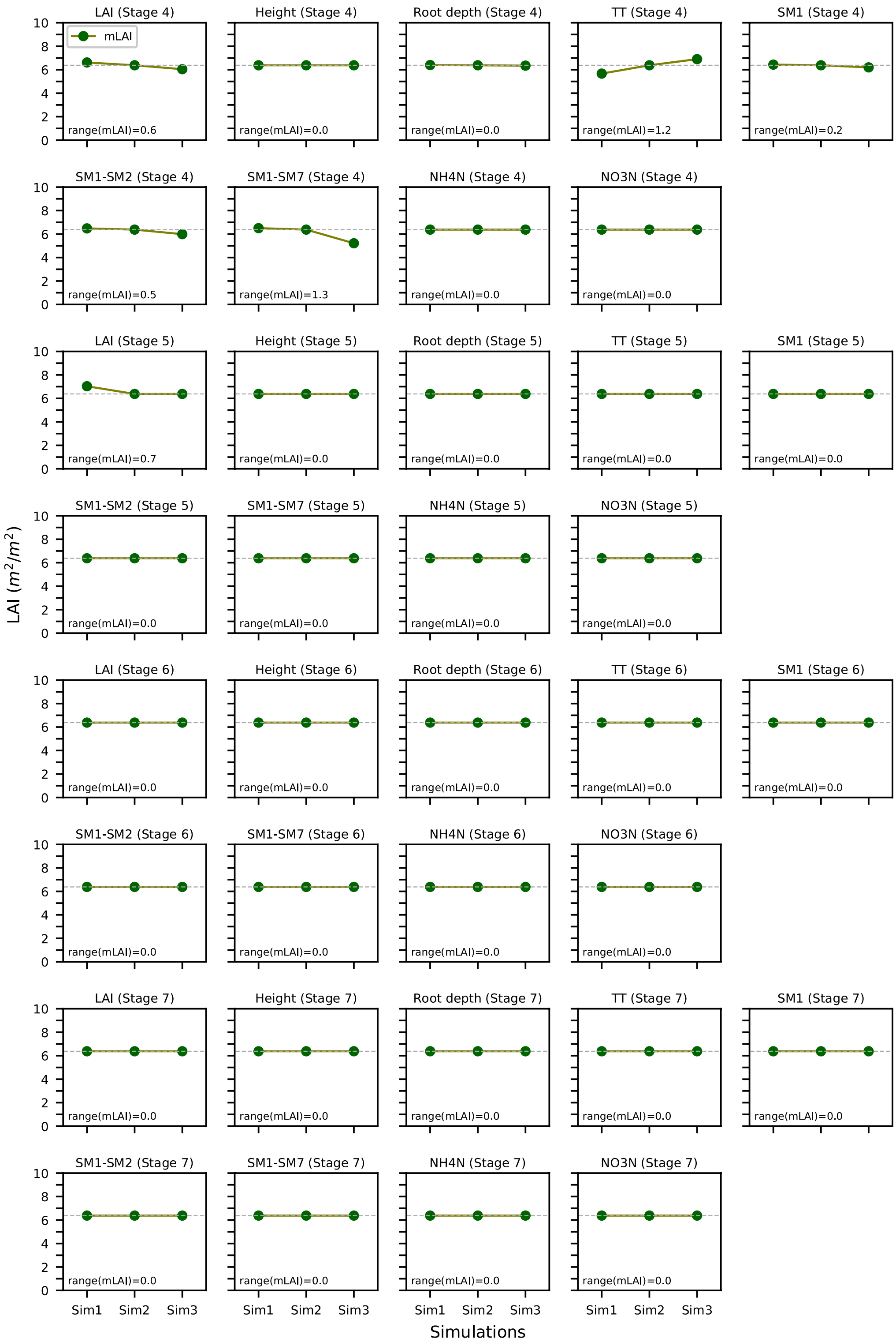

Figure A.8: Change of the mLAI in response to the wheat and soil states in different phenology states. The grey gridline is the mLAI of the reference simulation.

Appendix B

Soil parameters used in the uncalibrated model

Table B.1: Soil information and parameters of the fourteen soil types used in the case studies. Data were collected in Victora, Australia, provided by the APSIM model.

Soil No.	SoilType	Site	Latitude	Longitude	Diffus-Const	Diffus-Slope	Salb	CNBare
Soil1	Loam	Smeaton	-37.332	143.949	88	35	0.13	73
Soil2	Loam	Clunes	-37.291	143.787	88	35	0.13	73
Soil3	Clayey sand	Inverleigh	-38.098	144.058	88	35	0.12	73
Soil4	Clay	Inverleigh	-38.098	144.058	40	16	0.12	73
Soil5	Ironstone	Woolbrook Homestead	-38.039	144.025	88	35	0.13	73
Soil6	Clay	Little River	-37.966	144.553	40	16	0.12	73
Soil7	Clay loam	Lexton	-37.273	143.514	88	35	0.13	73
Soil8	Sandy clay loam	Yalla-Y-Poora	-37.281	142.928	88	35	0.13	73
Soil9	Clay	Murdeduke Homestead	-38.191	143.742	40	16	0.12	73
Soil10	Clay loam	Lismore	-37.915	143.335	88	35	0.13	73
Soil11	Brown Sodosol	Lake Bolac	-37.785	142.806	88	35	0.13	73
Soil12	Brown Sodosol	Westmere	-37.686	142.967	88	35	0.13	84
Soil13	Brown Sodosol	Westmere	-37.686	142.967	88	35	0.13	84
Soil14	Clay	Dunkeld	-37.649	142.342	88	35	0.13	73

Table B.2: Soil parameters of the fourteen soil types used in the case studies. Data were collected in Victora, Australia, provided by the APSIM model.

Soil	Depth	BD	AirDry	LL15	DUL	SAT	SWCON	OCTotal	Fbiom	Finert
No.	cm	g/cm^3	m^3/m^3	m^3/m^3	m^3/m^3	m^3/m^3	1/day	%	-	-
Soil 1	0-5	1.47	0.09	0.17	0.35	0.41	0.5	0.77	0.04	0.4
	5-15	1.59	0.14	0.17	0.29	0.37	0.5	0.62	0.02	0.5
	15-25	1.59	0.14	0.17	0.29	0.37	0.5	0.62	0.02	0.5
	25-35	1.59	0.14	0.17	0.29	0.37	0.5	0.62	0.02	0.5
	35-45	1.62	0.17	0.17	0.31	0.36	0.3	0.38	0.02	0.7
	45-55	1.62	0.17	0.17	0.31	0.36	0.3	0.38	0.02	0.7
Soil 2	0-5	1.36	0.1	0.2	0.41	0.46	0.5	0.77	0.04	0.4
	5-15	1.38	0.18	0.22	0.38	0.45	0.5	0.62	0.02	0.5
	15-25	1.38	0.18	0.22	0.38	0.45	0.5	0.62	0.02	0.5
	25-35	1.38	0.18	0.22	0.38	0.45	0.5	0.62	0.02	0.5
	35-45	1.32	0.22	0.22	0.42	0.47	0.3	0.38	0.02	0.7
	45-55	1.32	0.22	0.22	0.42	0.47	0.3	0.38	0.02	0.7
Soil 3	0-5	1.42	0.06	0.12	0.24	0.44	0.5	1.55	0.04	0.4
	5-15	1.32	0.17	0.21	0.35	0.47	0.3	1.24	0.02	0.5
	15-25	1.32	0.17	0.21	0.35	0.47	0.3	1.24	0.02	0.5
	25-35	1.32	0.17	0.21	0.35	0.47	0.3	1.24	0.02	0.5
	35-45	1.39	0.21	0.21	0.4	0.45	0.3	0.78	0.02	0.7
	45-55	1.39	0.21	0.21	0.4	0.45	0.3	0.78	0.02	0.7
Soil 4	0-5	1.26	0.11	0.22	0.45	0.5	0.3	3.23	0.04	0.4
	5-15	1.29	0.24	0.3	0.44	0.49	0.3	2.58	0.02	0.5
	15-25	1.29	0.24	0.3	0.44	0.49	0.3	2.58	0.02	0.5
	25-35	1.29	0.24	0.3	0.44	0.49	0.3	2.58	0.02	0.5
	35-45	1.27	0.3	0.3	0.44	0.49	0.3	1.62	0.02	0.7
	45-55	1.27	0.3	0.3	0.44	0.49	0.3	1.62	0.02	0.7
Soil 5	0-5	1.58	0.09	0.17	0.21	0.37	0.6	1.2	0.04	0.4
	5-15	1.58	0.09	0.17	0.21	0.37	0.6	1.2	0.04	0.4
	15-25	1.79	0.11	0.14	0.19	0.3	0.4	0.96	0.02	0.5
	25-35	1.79	0.11	0.14	0.19	0.3	0.4	0.96	0.02	0.5
	35-45	1.33	0.21	0.21	0.36	0.47	0.3	0.6	0.02	0.7
	45-55	1.33	0.21	0.21	0.36	0.47	0.3	0.6	0.02	0.7
Soil 6	0-5	1.21	0.1	0.2	0.33	0.51	0.3	0.77	0.04	0.4
	5-15	1.21	0.1	0.2	0.33	0.51	0.3	0.62	0.02	0.5
	15-25	1.31	0.2	0.25	0.4	0.48	0.3	0.62	0.02	0.5
	25-35	1.31	0.2	0.25	0.4	0.48	0.3	0.62	0.02	0.5
	35-45	1.23	0.28	0.28	0.45	0.5	0.3	0.38	0.02	0.7
	45-55	1.23	0.28	0.28	0.45	0.5	0.3	0.38	0.02	0.7
Soil 7	0-5	1.2	0.13	0.25	0.44	0.52	0.5	0.77	0.04	0.4
	5-15	1.05	0.21	0.26	0.4	0.57	0.3	0.62	0.02	0.5
	15-25	1.05	0.21	0.26	0.4	0.57	0.3	0.62	0.02	0.5
	25-35	1.05	0.21	0.26	0.4	0.57	0.3	0.62	0.02	0.5
	35-45	1.14	0.25	0.25	0.49	0.54	0.3	0.38	0.02	0.7
	45-55	1.14	0.25	0.25	0.49	0.54	0.3	0.38	0.02	0.7

Table B.2: (Continue).

Soil No.	Depth cm	BD g/cm^3	AirDry m^3/m^3	LL15 m^3/m^3	DUL m^3/m^3	SAT m^3/m^3	SWCON 1/day	OCTotal %	Fbiom -	Finert -
Soil 8	0-5	1.02	0.08	0.15	0.31	0.58	0.5	3.31	0.04	0.4
	5-15	1.06	0.19	0.24	0.31	0.57	0.3	2.65	0.02	0.5
	15-25	1.06	0.19	0.24	0.31	0.57	0.3	2.65	0.02	0.5
	25-35	1.06	0.19	0.24	0.31	0.57	0.3	2.65	0.02	0.5
	35-45	1.3	0.33	0.33	0.43	0.48	0.3	1.65	0.02	0.7
	45-55	1.3	0.33	0.33	0.43	0.48	0.3	1.65	0.02	0.7
Soil 9	0-5	1.23	0.1	0.2	0.37	0.51	0.3	1.2	0.04	0.4
	5-15	1.23	0.1	0.2	0.37	0.51	0.3	1.2	0.04	0.4
	15-25	1.17	0.16	0.2	0.38	0.53	0.3	0.96	0.02	0.5
	25-35	1.17	0.16	0.2	0.38	0.53	0.3	0.96	0.02	0.5
	35-45	1.11	0.2	0.2	0.41	0.55	0.3	0.6	0.02	0.7
	45-55	1.11	0.2	0.2	0.41	0.55	0.3	0.6	0.02	0.7
Soil 10	0-5	1.32	0.09	0.17	0.3	0.47	0.5	1.2	0.04	0.4
	5-15	1.32	0.09	0.17	0.3	0.47	0.5	1.2	0.04	0.4
	15-25	1.35	0.14	0.17	0.31	0.46	0.5	0.96	0.02	0.5
	25-35	1.35	0.14	0.17	0.31	0.46	0.5	0.96	0.02	0.5
	35-45	1.35	0.18	0.18	0.41	0.46	0.3	0.6	0.02	0.7
	45-55	1.35	0.18	0.18	0.41	0.46	0.3	0.6	0.02	0.7
Soil 11	0-5	1.45	0.11	0.22	0.33	0.38	0.5	0.65	0.04	0.4
	5-15	1.45	0.11	0.22	0.33	0.38	0.5	0.65	0.04	0.4
	15-25	1.35	0.22	0.27	0.41	0.46	0.5	0.65	0.02	0.5
	25-35	1.35	0.22	0.27	0.41	0.46	0.5	0.65	0.02	0.5
	35-45	1.16	0.34	0.34	0.48	0.53	0.5	0.33	0.02	0.7
	45-55	1.16	0.34	0.34	0.48	0.53	0.5	0.33	0.02	0.7
Soil 12	0-5	1.6	0.07	0.13	0.26	0.36	0.5	1	0.03	0.4
	5-15	1.5	0.11	0.14	0.3	0.4	0.3	0.5	0.02	0.6
	15-25	1.5	0.11	0.14	0.3	0.4	0.3	0.5	0.02	0.6
	25-35	1.5	0.11	0.14	0.3	0.4	0.3	0.5	0.02	0.6
	35-45	1.3	0.22	0.22	0.4	0.48	0.2	0.25	0.01	0.9
	45-55	1.3	0.22	0.22	0.4	0.48	0.2	0.25	0.01	0.9
Soil 13	0-5	1.45	0.11	0.22	0.37	0.42	0.5	0.5	0.02	0.6
	5-15	1.45	0.11	0.22	0.37	0.42	0.5	0.5	0.02	0.6
	15-25	1.45	0.11	0.22	0.37	0.42	0.5	0.5	0.02	0.6
	25-35	1.3	0.18	0.23	0.4	0.48	0.3	0.25	0.01	0.9
	35-45	1.3	0.18	0.23	0.4	0.48	0.3	0.25	0.01	0.9
	45-55	1.3	0.18	0.23	0.4	0.48	0.3	0.25	0.01	0.9
Soil 14	0-5	0.93	0.04	0.07	0.25	0.62	0.4	1.05	0.04	0.4
	5-15	1.3	0.16	0.2	0.4	0.48	0.4	0.45	0.04	0.4
	15-25	1.28	0.2	0.2	0.44	0.49	0.3	0.3	0.02	0.6
	25-35	1.28	0.2	0.2	0.44	0.49	0.3	0.3	0.02	0.6
	35-45	1.33	0.2	0.2	0.4	0.47	0.3	0.3	0.02	0.7
	45-55	1.33	0.2	0.2	0.4	0.47	0.3	0.3	0.02	0.7