

Article

Assimilation of Wheat and Soil States into the APSIM-Wheat Crop Model: A Case Study

Yuxi Zhang ¹, Jeffrey P. Walker ^{1,*}, Valentijn R. N. Pauwels ¹ and Yuval Sadeh ²

¹ Department of Civil Engineering, Monash University, Clayton, VIC 3800, Australia; yuxi.zhang1@monash.edu (Y.Z.); valentijn.pauwels@monash.edu (V.R.N.P.)

² The School of Earth, Atmosphere and Environment, Monash University, Clayton, VIC 3800, Australia; yuval.sadeh@monash.edu

* Correspondence: jeff.walker@monash.edu

Abstract: Optimised farm crop productivity requires careful management in response to the spatial and temporal variability of yield. Accordingly, combination of crop simulation models and remote sensing data provides a pathway for providing the spatially variable information needed on current crop status and the expected yield. An ensemble Kalman filter (EnKF) data assimilation framework was developed to assimilate plant and soil observations into a prediction model to improve crop development and yield forecasting. Specifically, this study explored the performance of assimilating state observations into the APSIM-Wheat model using a dataset collected during the 2018/19 wheat season at a farm near Cora Lynn in Victoria, Australia. The assimilated state variables include (1) ground-based measurements of Leaf Area Index (LAI), soil moisture throughout the profile, biomass, and soil nitrate-nitrogen; and (2) remotely sensed observations of LAI and surface soil moisture. In a baseline scenario, an unconstrained (open-loop) simulation greatly underestimated the wheat grain with a relative difference (RD) of -38.3% , while the assimilation constrained simulations using ground-based LAI, ground-based biomass, and remotely sensed LAI were all found to improve the RD, reducing it to -32.7% , -9.4% , and -7.6% , respectively. Further improvements in yield estimation were found when: (1) wheat states were assimilated in phenological stages 4 and 5 (end of juvenile to flowering), (2) plot-specific remotely sensed LAI was used instead of the field average, and (3) wheat phenology was constrained by ground observations. Even when using parameters that were not accurately calibrated or measured, the assimilation of LAI and biomass still provided improved yield estimation over that from an open-loop simulation.

Keywords: APSIM-Wheat; sequential data assimilation; EnKF; yield forecast



Citation: Zhang, Y.; Walker, J.P.; Pauwels, V.R.N.; Sadeh, Y.

Assimilation of Wheat and Soil States into the APSIM-Wheat Crop Model: A Case Study. *Remote Sens.* **2022**, *14*, 65. <https://doi.org/10.3390/rs14010065>

Academic Editors:
Ephrem Habyarimana
and Nicolas Greggio

Received: 17 November 2021

Accepted: 21 December 2021

Published: 24 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Understanding spatial and temporal variability in crop yield is essential to site-specific management [1]. Physically based crop models simulate daily crop growth based on energy, water, and nutrient exchange with the atmosphere, soil, climate, and field management [1], with such simulations providing insights into the reasons for yield variability [2]. However, a common issue with crop modelling is that they typically require site specific parameters that can only be calibrated from long-term observations over several growing seasons [1,3] or measured using site-specific soil investigation. Thus, for accurate yield estimation and crop monitoring, key input and parameter data, particularly those with large model sensitivity and spatial heterogeneity (e.g., rainfall and soil characteristics) [4–8], must be collected near the study site to avoid considerable uncertainty in the crop model [9–11]. As it is impractical to collect the spatial input and parameter data required due to the high labour and time costs, broad-scale application of these models for yield prediction is difficult [1].

Remote sensing provides vegetation and soil information from space with broad coverage, and a satisfactory resolution and accuracy (root mean squared error) for agricultural purposes at regional scales. Low resolution satellite products such as the Moderate

Resolution Imaging Spectroradiometer (MODIS) leaf area index (LAI) has a pixel size of 500 m and a repeat time of 8 days, with global coverage and a reported accuracy of $0.38 \text{ m}^2/\text{m}^2$ in grass and cereal crops [12]. The Soil Moisture and Ocean Salinity (SMOS; 35–50 km resolution and 2–3 days repeat) and the Soil Moisture Active Passive (SMAP; 36–9 km resolution and 2–3 days repeat) surface soil moisture products have a reported accuracy of $0.04 \text{ m}^3/\text{m}^3$ from validation field experiments [13,14]. Biomass estimation for winter wheat from Landsat Thematic Mapper (TM) and MODIS images has achieved an accuracy of 664 kg/ha [15]. Moreover, high-resolution satellites data such as Sentinel-2 (10 m, 5 days) and Sentinel-1 (5–40 m, 6 days) have been used to map LAI with an accuracy of $0.35\text{--}0.63 \text{ m}^2/\text{m}^2$ in wheat fields [16], and surface soil moisture with an accuracy of $0.08\text{--}0.12 \text{ m}^3/\text{m}^3$ in a vegetated areas [17]. Based on global remote sensing observations, a broad range of vegetation and water indices have been used to estimate crop status, water status and yield using models ranging from simple linear [18–21], quadratic [20], and exponential [22,23] regressions to complex machine learning approaches [24–27].

Fusing remote sensing observations and crop models is expected to take advantage of both the model and remote sensing data to improve estimates of spatial and temporal yield variability [28]. While the model is sensitive to input data accuracy, and the observation accuracy can be affected by the instrument and estimation errors, advanced data assimilation techniques (referred to as data assimilation hereafter) can be used to combine the two sources of information to derive a more accurate estimate of the physical system than any one approach on its own [29]. Accordingly, assimilation methods have been developed to estimate a true system state by carefully estimating and evaluating uncertainties to which model estimates and external observations are subjected [30]. Data assimilation techniques could therefore be used to combine remotely sensed data as observations of the dynamic system with the dynamic model state predictions in order to approach the true states, by reducing the uncertainties from both the model and the external observations.

Data assimilation applications have so far been mainly focused on monitoring food security at regional scale, with the results evaluated using official statistics to support policy-making [31–33]. However, application is required at the field or subfield scale using high-resolution remote sensing data in order to support agricultural decision-making. Only a few studies have explored and validated the application of crop model data assimilation at farm scales [33–35]. Moreover, only limited plant and soil states (usually LAI and/or surface soil moisture) have been tested in these studies, due to the limitations of current remote sensing techniques. Accordingly, Zhang [28] developed an APSIM-EnKF framework and presented a synthetic study to show the potential improvements from assimilating a range of wheat and soil states. This paper expands on that work at experimental plot scale using a field dataset collected over the 2018/19 wheat season. The data assimilation experiments involved in this study were designed to address the following five research questions:

- (1) Which wheat and soil states improve the APSIM yield estimation at field scale?
- (2) Do the remotely sensed observations of model states have a potential to provide spatial variation at subfield scale when assimilated into the one dimensional APSIM-Wheat model?
- (3) At which phenological stage(s) does the wheat and soil states efficiently improve the model yield estimation?
- (4) Does constraining the phenology reduce uncertainties?
- (5) Do the remotely sensed observations of model states have the potential to provide better yield estimation when applied to an uncalibrated model?

2. Methodology

2.1. Dataset

The study site used is a 75×75 m square wheat field, located in the Cora Lynn area of Victoria, Australia. The ground-based dataset was collected in a 2018/19 field experiment described in detail by Zhang [28]. The satellite-based remotely sensed LAI consisted of a

series of 3 m daily LAI images produced by the fusion of Sentinel-2 LAI and PlanetScope surface reflectance, with an accuracy of 0.35–0.63 m²/m² when evaluated using 57 wheat fields (including the study site) across Australia and Israel [16]. This study uses the dataset for model input, wheat and soil state observations to be assimilated into the model, and validation.

The model input includes weather and soil properties. The daily weather dataset used in this study consists of daily rainfall, solar radiation, maximum/minimum temperature, pan evaporation, wind speed, and vapour pressure. The weather dataset was created from the average of several nearby weather stations sourced from the Australian Bureau of Meteorology (BoM, <http://www.bom.gov.au/> (accessed on 16 July 2019)) and the Weather Underground weather station network (WU, <https://www.wunderground.com/> (accessed on 2 March 2019)) to best-match the in situ weather station. The soil property information required for model input was measured in the field during the wheat season at five locations, in six layers throughout the soil profile (0–5, 5–15, 15–25, 25–35, 35–45, 45–55 cm).

The observations of wheat and soil states assimilated into APSIM included ground-measured (ground) and remotely sensed (RS) data. Ground data included: (1) measurements of LAI, biomass (total and separately for leaf and stem), and soil nitrogen in the form of nitrate (NO₃N) in 0–5 cm and 5–15 cm soil layers, from laboratory processing of weekly vegetation and soil samples, (2) in situ station-based daily soil moisture throughout the soil profile (0–5, 5–15, 15–25, 25–35, 35–45, 45–55 cm), and (3) phenology from weekly experimental records and an in situ camera estimated according to Zadoks et al. [36]. The remotely sensed data included (1) high spatiotemporal resolution (3 m, daily) LAI images from a fused Sentinel-2 and PlanetScope dataset [16], and (2) surface soil moisture (field scale by treating the whole field as a single pixel, daily) retrieved from tower-based microwave brightness temperature.

The validation data included the above time series of wheat and soil observations, the date of major phenological stages, and wheat grain yield.

2.2. Model Configuration

The APSIM-EnKF data assimilation framework used in this study was the version developed and described by Zhang [28] (source code available on <https://github.com/yuxi-research/APSIM-EnKF>, published on 6 December 2021). The simulation window was set from 1 July 2018 (Day 0), to 1 March 2019 (Day 243). Winter wheat (variety: RGT Accroc) was sown on 7 August 2018 with a sowing depth of 5 cm. When sowing, fertiliser was applied at the rate of 100 kg/ha using mono-ammonium phosphate (MAP, containing 10% ammonium nitrogen). The row spacing was 0.2 m, and the population was 251 plants/m².

The soil was defined by six layers, with depths of 0–5, 5–15, 15–25, 25–35, 35–45, and 45–55 cm. Key soil parameters controlling the soil moisture dynamics, including the drained upper limit, the lower limit at 15 bars, and the volumetric moisture content of saturated and air-dry soil, were estimated from the maximum and minimum soil moisture during the growing season. Both default and calibrated parameters were used, with the calibrated parameters specified in Table 1. Parameter calibration was performed manually and according to sensitivity: seven parameters in the SoilWat module controlling the rate of infiltration and evapotranspiration were calibrated manually to the station soil moisture data in all six layers, and one parameter in the SoilN module controlling soil nitrogen decomposition was calibrated to the variation of ground-measured soil nitrate in the first two layers. The remaining 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.

Table 1. The calibrated soil parameters in the SoilWat and SoilN modules.

Parameter	Description	Layer	Default Value	Calibrated Value
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. Unit in mm ² /day.	-	40	400
SWCON	Fractional amount of water above drain upper limit that can dry under gravity per day. Unit in 1/mm.	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
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

An ensemble of 50 was generated by adding Gaussian distributed random noise to the weather forcings, parameters and the assimilated observations. The random noise applied to each quantity has a mean of zero and a standard deviation being the uncertainty value as shown in Table 2, with types of either additive, meaning that the standard deviation is a constant over time, or multiplicative, meaning that the standard deviation changes with time with at a proportion of the quantity value.

Filter divergence is a common problem in EnKF data assimilation, with the prior error covariance underestimated due to approximations underlying the algorithm. Accordingly, new observations are given less weight at update steps to impact the analysis result, with the forecast diverging from the observations [37]. This problem is usually addressed by appending a model uncertainty term in the state transfer function or multiplying the forecast error covariance by an inflation factor having a value greater than 1. In this study, a covariance inflation factor was used to inflate the ensemble spread, with a value tuned to avoid filter divergence while keeping the ensemble spread small. Further details on this inflation factor are provided in Section 3.2.1.

Table 2. Uncertainties of weather data, model parameters, and state observations to generate the ensembles.

Data	Description	Unit	Uncertainty	
			Type ¹	Std.
Weather (daily)				
Rainfall	Rainfall	mm	M	0.160
Radiation	Radiation	mm	A	1.254
Max/min temperature	Max/min air temperature		A	0.882
Wind speed	Wind speed	m/s	A	0.679
Parameters				
DUL	Drained upper limit	m ³ /m ³	A	0.03
WheatLL	Permanent wilting point	m ³ /m ³	A	0.01
Observations (ground-based measurements)				
LAI	Leaf area index	m ² /m ²	M	0.15
LeafWt	Dry leaf weight	g/m ²	M	0.2
StemWt	Dry stem weight	g/m ²	M	0.2
Biomass	Aboveground dry biomass	g/m ²	M	0.2
SM1	Surface soil moisture (0–5 cm)	m ³ /m ³	A	0.03
SM2, . . . , SM6	Root-zone soil moisture (5–15, 15–25, 25–35, 35–45, 45–55 cm)	m ³ /m ³	A	0.01
NO ₃ N1, NO ₃ N2	Soil nitrogen in the form of nitrate (0–5, 5–15 cm)	kg/ha	M	0.1
Observations (remotely sensed)				
LAI-RS	Leaf area index	m ² /m ²	M	0.1
SM1-RS	Surface soil moisture	m ³ /m ³	A	0.036

¹ Uncertainty type: M = multiplicative, A = additive.

2.3. Data Assimilation Experiments

A baseline data assimilation scenario is first presented by assimilating all wheat and soil states at the field scale. In all subsequent scenarios described in the remainder of this section, the configurations remain the same as the baseline scenario unless specifically specified.

The baseline scenario assimilated states at the field scale at three-day intervals (except for the soil moisture that was assimilated daily). In the baseline scenario, all state observations, except for the soil moisture which was assimilated daily, were averaged and interpolated/resampled to 3 days to create a 3-day field-average time series for each state. The assimilation window was from stage 4 to the middle of stage 7, being the period between the end of juvenile and near the end of grain filling, to avoid model failure when perturbed in early stages and direct update of the grain weight in the late grain-filling stage.

Wheat and soil states were assimilated into the model individually and jointly, with observations obtained from field and remote sensing. The wheat and soil states assimilated into APSIM-Wheat included (1) LAI from ground-based measurements and remote sensing, respectively; (2) total aboveground dry biomass (biomass hereafter) and organ weight (LeafWt and StemWt) from ground-based measurements; (3) soil moisture in the depths of 0–5, 5–15, 15–25, 25–35, 35–45, 45–55 cm measured by in situ soil moisture stations, and 0–5 cm retrieved from brightness temperature by the tower-based passive microwave radiometer; and (4) soil nitrogen in the depths of 0–5 and 5–15 cm measured by destructive sampling and laboratory chemical analysis.

In addition to the baseline scenario, several other scenarios were tested:

- (1) A plot-specific scenario to study the assimilation performance on a spatial scale. Here the assimilation focused on four 1 m × 1 m experimental plots distributed in the field rather than the entire field to explore the impact of high-resolution simulation. The wheat grain yield in each plot was measured by harvesting the grain in the four experimental plots and weighing the dry weight after drying in an oven at 60° for more than 72 h, with the field-average yield calculated as the average yield of the four plots. Only remotely sensed LAI extracted from the nearest nine pixels of the four specific plots were assimilated to estimate the yield in each plot. The simulated yield was validated by the yield data collected in each plot, respectively.
- (2) An observation-limited scenario to explore the optimal assimilation stage. The APSIM-Wheat module describes ten wheat phenological phases: 1—sowing, 2—germination, 3—emergence, 4—end of juvenile, 5—floral initiation, 6—flowering (anthesis), 7—start of grain filling, 8—end of grain filling, 9—maturity, and 10—harvest; with phenological stages being the period between two adjacent phases. Here the study focused on the assimilation of state observations for a single phenological stage among stages 4 (end of juvenile to floral initiation), 5 (floral initiation to flowering), and 6–7 (flowering to end of grain filling).
- (3) A phenology-constrained scenario to minimise the uncertainty from phenology estimation. Here the phenology was also constrained with field observations. However, instead of using the EnKF for constraining the phenology, a more primitive data assimilation method known as direct insertion [29,38,39] was applied. The direct insertion method replaced the modelled phenology with the observed date without weighing their uncertainties. Wheat phenology in the Zadoks phenology scale [36] was estimated through weekly field experimental records and an in situ time-lapse camera. APSIM wheat phenology was then obtained by mapping the Zadoks scale to the APSIM wheat scale through an in-built linear equation provided by the APSIM wheat module documentation.
- (4) An uncalibrated scenario to test the performance of assimilation in a field where the in situ data for model calibration are not available. Here uncalibrated models were used with soil module parameters obtained from the APSIM soil library. In this scenario, fourteen generic soil types measured in the same state of the study site were used to replace the calibrated soil parameters in the baseline scenarios. The parameters include those controlling soil water retention, evapotranspiration, and the percentage of active/inactive nitrogen pool.

The outcomes of the data assimilation experiments 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 dry grain weight at harvest), expressed as:

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

$$\text{RD}_{\text{yield}} = \frac{\text{yield}_{\text{est}} - \text{yield}_{\text{obs}}}{\text{yield}_{\text{obs}}}, \quad (2)$$

where L is the total time step. The estimated states $\mathbf{x}_k^{\text{est}}$ is the analysis ensemble mean for the assimilation while $\mathbf{x}_k^{\text{obs}}$ is the ground observations of the state at time step k . The $\text{yield}_{\text{est}}$ and $\text{yield}_{\text{obs}}$ are the estimated and observed grain weight in kg/ha at the date of harvest, respectively. Therefore, a value of RD_{yield} closer to zero means that the yield estimation is more accurate. An absolute value of RD_{yield} from a data assimilation experiment being less than the open-loop indicates that the data assimilation contributed to a better yield estimation than when no data are assimilated. A negative value indicates that the yield estimation error was overcorrected.

3. Results and Discussion

3.1. Model Calibration

An example of soil module calibration outcomes is shown in Figure 1, as the time series of the estimated soil moisture and ammonium nitrogen before and after the model calibration. After 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 (layers 2–6). The parameter *Fbiom* 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 estimated LAI growing period (further discussed in the results of the soil nitrate assimilation).

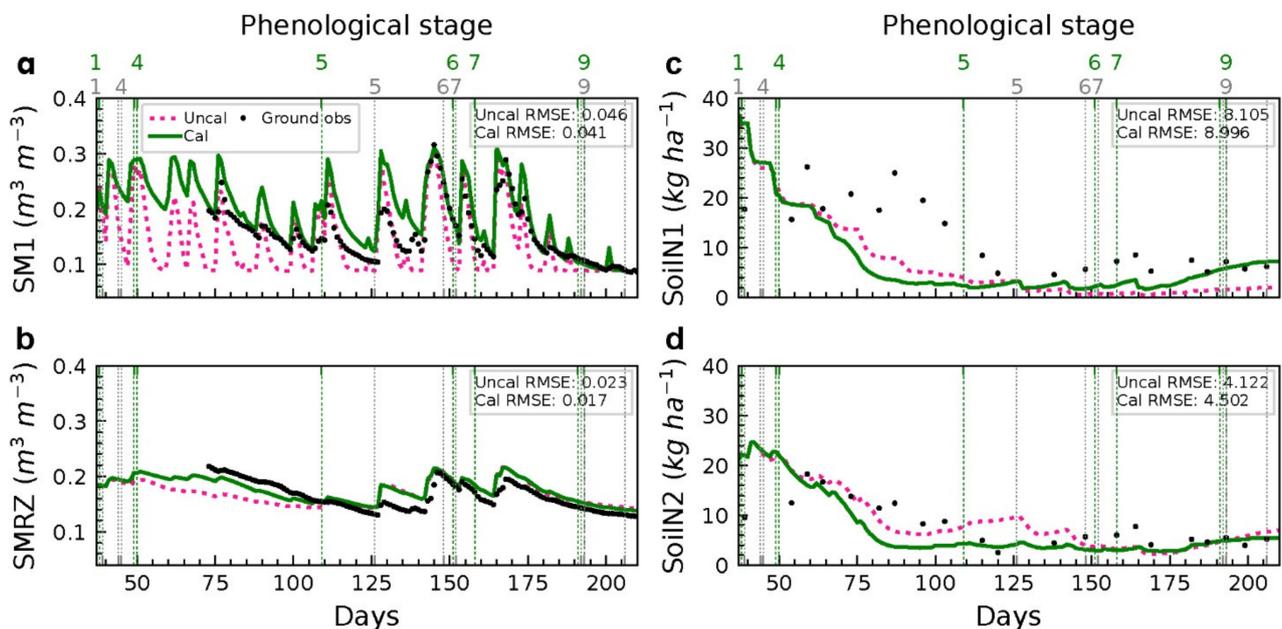


Figure 1. Examples of the estimated soil moisture in the surface layer and root-zone (SM1, SMRZ, (a,b)) and total soil nitrogen in the top two layers (SoilN1, SoilN2, (c,d)) by the calibrated (Cal) and uncalibrated (Uncal) models, respectively. Numbers and vertical lines in grey and green are the observed and estimated phenology, respectively.

3.2. Baseline Scenario

3.2.1. Inflation Factor Selection

The purpose of the inflation factor is to ensure that the ensemble spread is maintained such that the derived error covariance correctly represents the background uncertainty. An ideal scenario is that the ensemble of model state predictions overlaps with the ensemble of observations, with the ensembles approximating the probability distribution of the state using a finite number of ensemble members. With the model estimates and the external observations each providing a probability distribution of the true state according to their own values and uncertainties, the overlay in the two probability distributions identifies those model ensemble members that best agree with the observations when accounting for the uncertainty of each.

Identification of an appropriate inflation factor was a trial-and-error process. Accordingly, the data assimilation run was performed with an initial inflation factor set slightly greater than 1, then the ensemble spread checked. The ensemble spread is too narrow if the state ensemble does not enclose the observation ensemble, with the inflation factor adjusted to a greater value. The adjustment process repeats until an overlay was found between the state and observation ensemble in most assimilation timesteps over the entire

period, while ensuring that the spread was not drastically greater than the spread of the stochastic open-loop run without any inflation.

An inflation factor of 5 was found to be appropriate in the assimilation of wheat states and soil nitrate, except for the biomass assimilation, where a smaller inflation factor of 4 was found. As the inflation factor was only applied when observations were assimilated, a smaller inflation factor of 2 was selected to assimilate soil moisture daily rather than every three days as in the other states.

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 in meteorological data assimilation [40], where the numerical weather prediction models are chaotic, complex, and three-dimensional [41–46]. However, the crop models are relatively simple one-dimensional simulations at a daily time step. In this case, the ensemble spread cannot be maintained with a small inflation factor. For example, Figure 2 shows LAI assimilation results using the selected inflation factor compared to smaller and larger inflation factors. A smaller inflation factor resulted in a limited overlap between the observation and the model ensembles, as found in the ensembles of the five successively assimilated LAI observations starting from day 119 (Figure 2a). In comparison, a larger inflation factor introduced unnecessary noise, making the model less trusted by the data assimilation algorithm (Figure 2c). The EnKF works suboptimally with a larger inflation factor because the Kalman gain is derived to minimise 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.

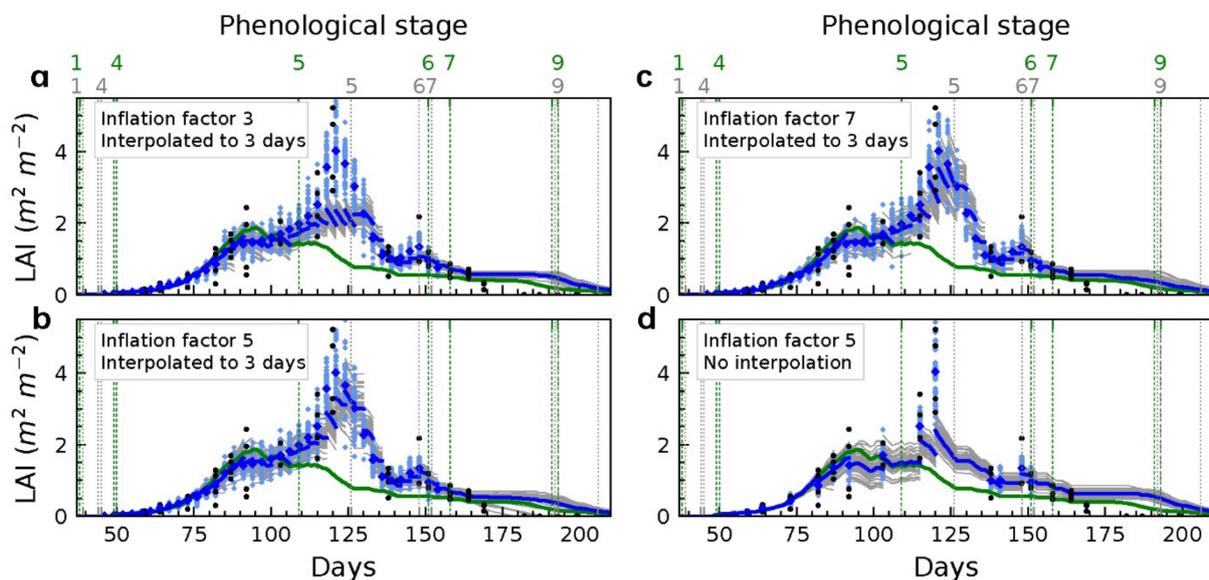


Figure 2. Example of assimilating LAI using a variety of inflation factor values under the interpolation scenario (a–c) and using the selected inflation factor under the baseline scenario with no interpolation (d). The horizontal axis shows the number of days following the start of the simulation. See Figure 3 for legend.

3.2.2. Assimilation of LAI

The year of this study was drier than average, resulting in heavy water stress in the vegetative stages (before phase 5, the floral initiation). With the model calibrated to fit the in situ soil moisture time series, the open-loop simulation greatly underestimated the LAI, organ weight, and wheat yield at harvest.

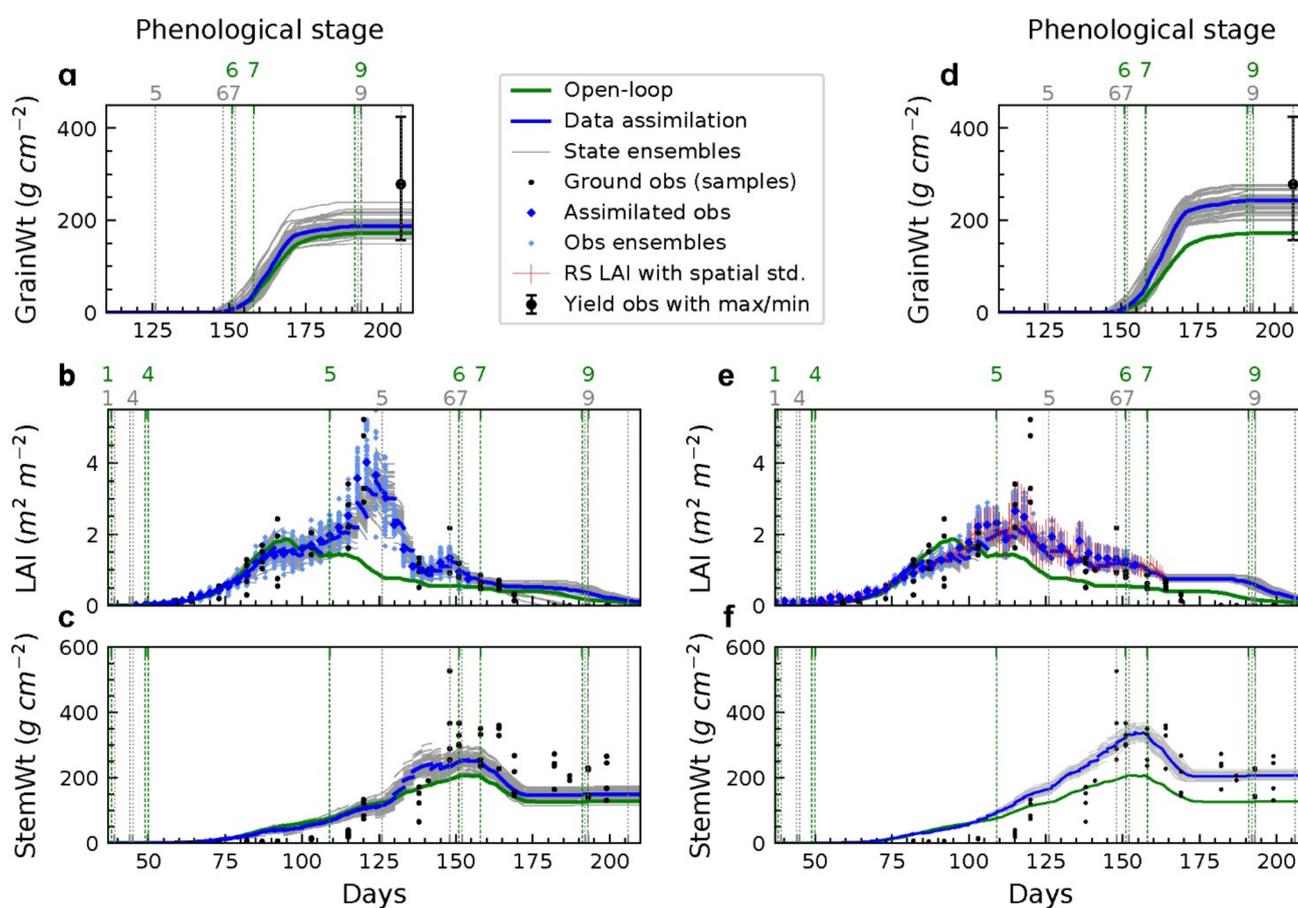


Figure 3. Evolution of GrainWt, LAI, and StemWt with the assimilation of ground measured (a–c) and remotely sensed LAI (d–f). The vertical lines indicate the field-observed phenology in grey and the simulated phenology of the open-loop in green. The legend applies to all figures throughout.

The baseline scenario confirmed the benefit of LAI assimilation into wheat modelling for improved yield estimation. When the ground and RS LAI data were assimilated at intervals of every three days, the estimated LAI time series from both sources was closer to the ground observation than the open-loop (Figure 3). In both assimilation experiments, the development of grain weight in the grain-filling stage (stage 7) became slightly faster with a steeper slope (Figure 3a,d). The measured average yield of the field is 2781.5 kg/ha. The estimated yield was increased from 1716 to 1871 kg/ha (ground LAI) and 2570 kg/ha (RS LAI), reducing the yield RD from -38.3% to -32.7% (ground LAI) and -7.6% (RS LAI). Accordingly, the RMSE of LAI, LeafWt, StemWt, and biomass estimation was reduced (Table 3). Noting that LAI was only assimilated before the early stage of grain filling (approximately day 156 in stage 7), the differences of the grain weight development between the data assimilation and open-loop runs before stage 7 were indirectly caused by more accurate estimation in other states (indicated by reduced RMSEs of wheat states in Table 3) rather than direct updates. Therefore, the assimilation of LAI in stages earlier than the grain-filling corrected errors in states that affect grain-filling and thus led to a better yield estimation.

Table 3. Relative difference (RD) and root mean square error (RMSE) of yield and wheat states from the assimilation of wheat states in the baseline, plot-specific, and phenology-constrained scenarios.

Assimilated State (s)	Scenario	Yield (kg/ha)	RD% of Yield	RMSE			
				LAI m ² /m ²	LeafWt kg/ha	StemWt kg/ha	Biomass kg/ha
Open-loop	Baseline	1716	−38.3	0.91	49.5	93.5	150.4
Ground LAI	Baseline	1871	−32.7	0.40	29.2	83.0	114.5
RS LAI	Baseline	2570	−7.6	0.56	34.8	81.7	75.5
Ground biomass	Baseline	2520	−9.4	0.75	44.7	34.1	78.4
Ground LeafWt and StemWt	Baseline	2289	−17.7	0.69	40.5	35.7	98.0
Ground SM1	Baseline	1407	−49.4	1.07	55.9	126.4	186.8
RS SM1	Baseline	1618	−41.8	0.94	50.8	100.7	159.2
Ground SM1-2	Baseline	1451	−47.8	1.08	56.2	126.7	187.0
Ground SM1-6	Baseline	1242	−55.4	1.06	54.8	132.8	195.8
Ground NO ₃ N1	Baseline	1660	−40.3	1.04	54.9	105.4	166.4
Ground NO ₃ N2	Baseline	2016	−27.5	0.95	51.1	87.5	144.0
Ground NO ₃ N1-2	Baseline	2061	−25.9	0.97	52.6	84.8	139.9
RS LAI (Plot A)	Plot-specific	2652	−4.6	0.50	29.3	93.6	76.3
RS LAI (Plot B)	Plot-specific	2685	−3.5	0.56	30.9	105.2	80.2
RS LAI (Plot C)	Plot-specific	2819	1.3	0.57	31.5	111.2	82.3
RS LAI (Plot D)	Plot-specific	2574	−7.5	0.58	35.1	93.3	78.5
Open-loop	Phenology-constrained	1931	−30.6	0.86	47.4	85.1	149.4
Ground LAI	Phenology-constrained	2232	−19.7	0.40	28.6	63.9	108.5
Ground biomass	Phenology-constrained	2779	−0.1	0.82	43.3	38.2	73.4
Ground LeafWt and StemWt	Phenology-constrained	2224	−20.0	0.52	33.0	32.4	108.4

Although both the ground and RS LAI showed improvements in yield estimation, the performance of ground LAI was disappointingly worse than for RS LAI (Figure 3 and Table 3) with the ground measurements expected to be more accurate than from remote sensing. However, the degraded results can be explained by two reasons. First, the result from a later phenological limited scenario showed that the ground LAI assimilated in stages 4 and 5 provided much better performance than assimilating in all stages (further discussed in the results of observation-limited scenario). While wheat leaves at the phenological stage 6–7 are in the middle to late senescence stage, the dried leaves make the ground LAI measurement from destructive sampling difficult and less accurate. Meanwhile, the model is more sensitive to changes of LAI in stage 6–7 than in other stages [28]. Although the assimilation of stage 6–7 showed the best efficiency in a synthetic study [28] where the synthetic LAI observations were assumed more accurate, it also diverged greatly with an inaccurate LAI observation. Second, the time series of the remotely sensed LAI is more smoothly varying than the ground measured LAI. The updates in stage 5, particularly days 120 to 150, persistently drove the StemWt to a greater value that was closer to the observations. In the model physics, the grain-filling rate is proportional to the StemWt value on the day of flowering. Thus, the assimilation of remotely sensed LAI resulted in a

greater grain-filling rate in stage 7 than using the ground measured LAI, and consequently, a greater yield that is closer to the observation.

3.2.3. Assimilation of Biomass

Two schemes were applied to the assimilation of field biomass: (1) the assimilation of total biomass as an observation of the sum of aboveground organ weight (Figure 4 left panel); and (2) the joint assimilation of individual organ weight, namely, LeafWt and StemWt (Figure 4, right panel). Both assimilation schemes provided more accurate model estimation for wheat states and yield at harvest compared to the open-loop, according to the reduced RD in the yield and RMSE in the LAI, LeafWt, StemWt, and biomass estimates (Table 3). Moreover, the performance of biomass assimilation in the two schemes both provided reduced absolute RD in the yield and RMSE in the other wheat states than the assimilation of ground-measured LAI (Table 3). The reason that the scheme of assimilating total biomass provided better performance than using organ biomass or LAI (which is highly correlated to the LeafWt) is that it provides more information on the state variables in the system, and it allocates the weight to the observations of these wheat states through the error covariance calculated by the EnKF, meaning it avoided internal inconsistency caused by assimilating multiple state variables.

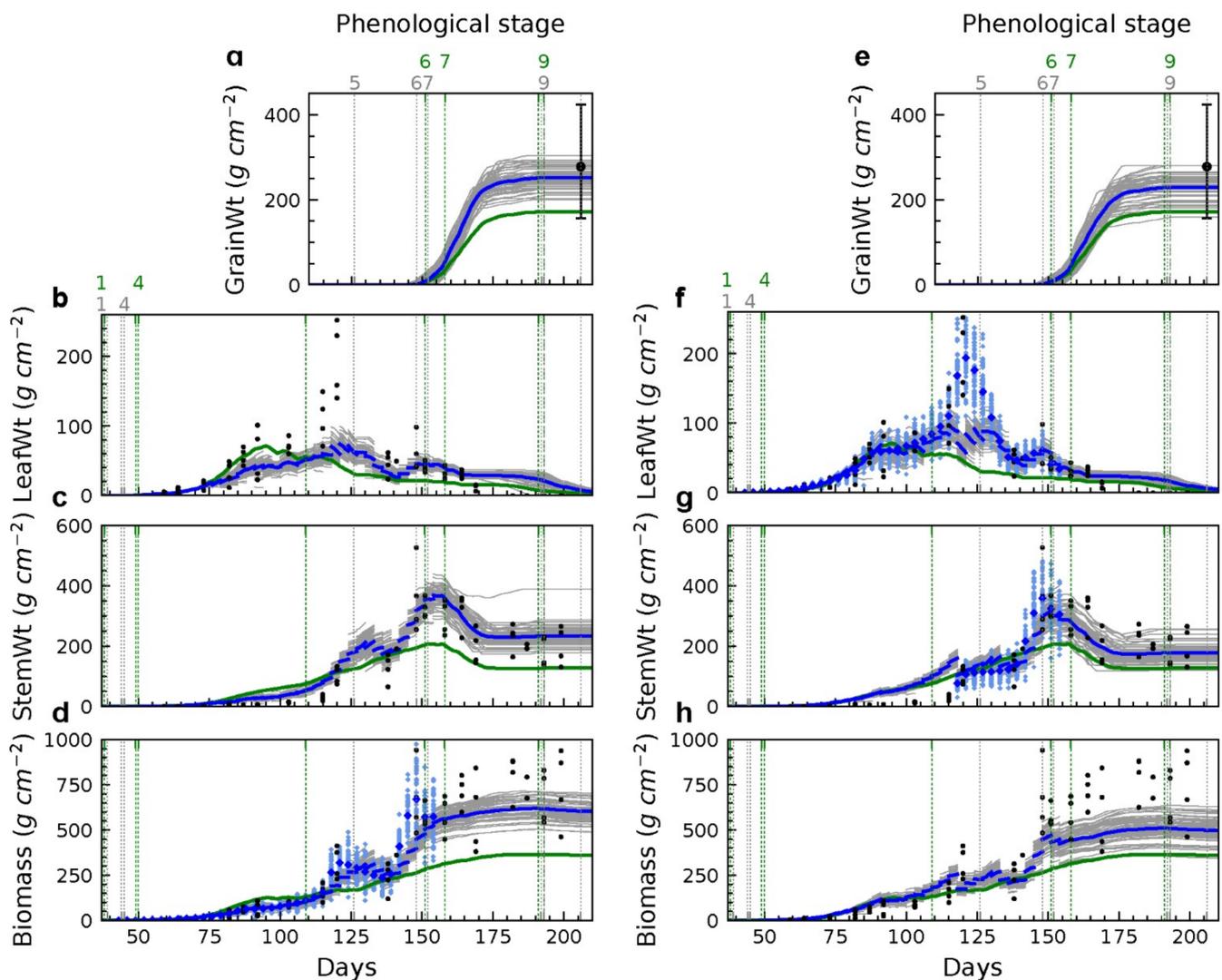


Figure 4. Evolution of GrainWt, LeafWt, StemWt, and biomass with the assimilation of biomass alone (a–d) and the joint assimilation of LeafWt and StemWt (e–h), respectively. See Figure 3 for legend.

In practice, the scheme of assimilating total biomass is also more applicable than the organ weight. First, Table 3 shows that the assimilation of total biomass provided more accurate yield estimation than individual organ weight. Second, measuring organ weight requires extra work than total biomass measurements, requiring cutting and partitioning the plant organs manually. In contrast, total biomass is less labour-consuming when measured from destructive sampling, and it can be estimated from remote sensing often with empirical regression techniques [15,47–50]. Third, solely constraining the biomass of one of the wheat organs may lead the other biomass states to be erroneous (particularly in the grain-filling stage). For example, state updating can increase the stem biomass by increasing the biomass in other organs because the error covariance shows a positive correlation among all the organ weight states. However, in the process of wheat development, the leaf starts to wither when the stem starts to grow quickly, so the increased StemWt should imply a decreased LeafWt in an actual situation, which could be opposite to the state updating result.

3.2.4. Assimilation of Soil Moisture

The assimilation of soil moisture in the first soil layer (SM1) from in situ stations (ground) and remotely sensed microwave radiometer (RS) provided a slightly worse estimation in wheat states and yield than the open-loop (Figure 5). The RMSE of LAI and total biomass (Table 3) increased from 0.91 m²/m² and 150.4 g/cm² (open-loop) to 1.07 m²/m² and 186.8 g/cm² (ground) and 0.94 m²/m² and 159.5 g/cm² (RS), with the absolute value of yield relative difference increased from 38.3% to 49.4% (ground) and 41.8% (RS). The assimilation of in situ soil moisture in soil layers 1–2 (SM1–2) and soil layers 1–6 (SM1–6) provided similar degradation to performance in the estimation of yield and other states than the open-loop (Tables 3 and 4). Thus, the assimilation of soil moisture did not benefit yield or model state estimation.

The result that assimilating SM did not improve yield estimation contradicts a previous synthetic study [28], which showed improvement in wheat states and yield estimation when soil moisture was assimilated. This can be explained by the fact that the synthetic study assumed perfect physics in the water balance model and attributed all uncertainty to the model parameters and initial conditions, which was accounted for by perturbing parameters and initial conditions when generating ensembles. However, in reality, the uncertainty from the model physics, which the algorithm cannot consider, cannot be neglected. Moreover, in this case, the available period of soil moisture observation was not long enough to solve all parameters in the model, particularly in subsurface layers whose soil moisture dynamics are reduced.

The assimilation of SM1 resulted in a lower SM1 level, as the SM1 observations were smaller than the model estimates with a reduced RMSE in SM1. Some correction was made to the SM estimates in the first two layers, but such correction was not found in deeper layers, as shown in the RMSE values of SM in each layer (Table 4).

Although the assimilation of SM1 pushed the estimated SM1 to approach the station observation, this improvement over other simulations only remains until the soil is completely wet or dry again. When the soil moisture becomes very wet (or dry), and the soil moisture is at the upper (or lower) limit of soil moisture determined by the soil moisture characteristics, the soil cannot become wetter (or dryer) with any water flowing into (or out of) the system. Therefore, even if the soil moisture is updated by data assimilation in a timestep, it resets once it reaches the upper or lower limit. The effect of data assimilation is thus cancelled.

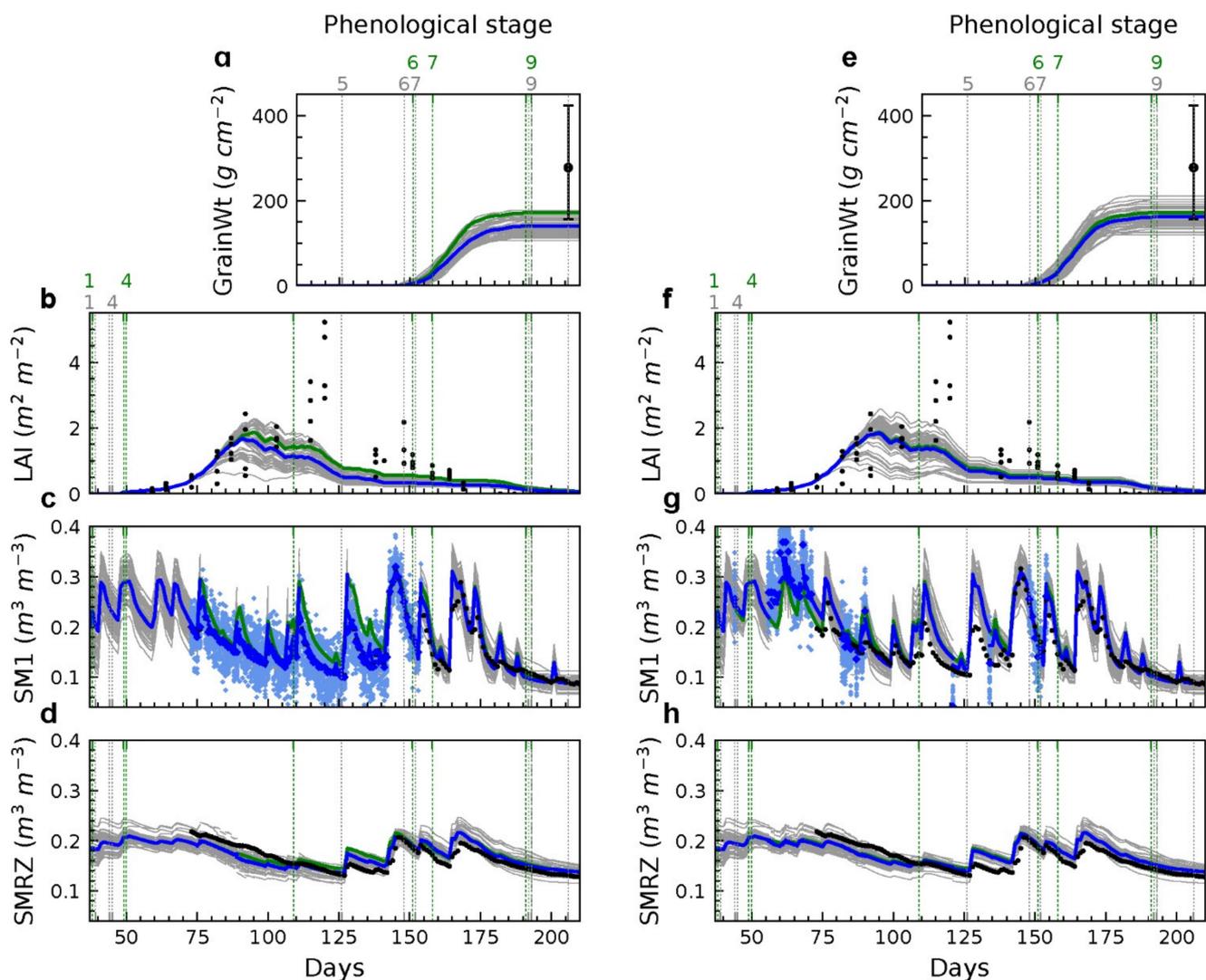


Figure 5. Evolution of GrainWt, LAI, and soil moisture in the first layer and root-zone (SM1, SMRZ) with the assimilation of SM1 from ground measurements (a–d) and remote sensing (e–h). See Figure 3 for legend.

3.2.5. Assimilation of Soil Nitrate

The model provided a poor performance in estimating ammonium-nitrogen; thus the assimilation of soil nitrogen was limited to nitrate-nitrogen (NO₃N). Measurements of soil nitrate in the first two soil layers (NO₃N1, NO₃N2) were assimilated individually and collectively into the model. When assimilated collectively, a higher grain and plant nitrogen (Grain N, PlantN) level was found (Figure 6), as the NO₃N1 and NO₃N2 observations were higher than the model estimates. A higher level of plant organ nitrogen was also found in the individual assimilation of NO₃N in each soil layer. A higher yield estimation was given by the individual assimilation of NO₃N1 and the collective assimilation of NO₃N1 and NO₃N2 (Table 3), but the grain-filling rate remained unchanged from the open-loop. This result indicates that the increased yield in the NO₃N assimilation was not caused by a higher grain-fill rate associated with greater plant biomass, which is only slightly higher than the open-loop, but rather by the relief of nitrogen stress because of a higher plant nitrogen level.

Table 4. Relative difference (RD) and root mean square error (RMSE) of yield and soil states from the assimilation of soil moisture in layer 1 (SM1), layers 1–2 (SM1–2), and layers 1–6 (SM1–6), and soil nitrate-nitrogen in layer 1 (NO₃N1), layer 2 (NO₃N2), and layers 1–2 (NO₃N1–2), respectively.

Assimilated State (s)	Yield (kg/ha)	RD% of Yield	RMSE						
			LAI	SM1	SM1	SM3	SM4	SM5	SM6
			m ² /m ²	m ³ /m ³					
Open-loop	1716	−38.3	0.91	0.041	0.026	0.024	0.026	0.016	0.018
Ground SM1	1407	−49.4	1.07	0.025	0.021	0.028	0.025	0.016	0.017
RS SM1	1618	−41.8	0.94	0.038	0.025	0.026	0.026	0.016	0.018
Ground SM1–2	1451	−47.8	1.08	0.025	0.020	0.028	0.025	0.017	0.018
Ground SM1–6	1242	−55.4	1.06	0.027	0.015	0.010	0.017	0.011	0.012
Ground NO ₃ N1	1660	−40.3	1.04	0.039	0.026	0.036	0.023	0.015	0.018
Ground NO ₃ N2	2016	−27.5	0.95	0.040	0.027	0.028	0.026	0.016	0.019
Ground NO ₃ N1–2	2061	−25.9	0.97	0.041	0.025	0.031	0.031	0.020	0.020
RS LAI	2570	−7.6	0.56	0.044	0.027	0.032	0.029	0.021	0.022

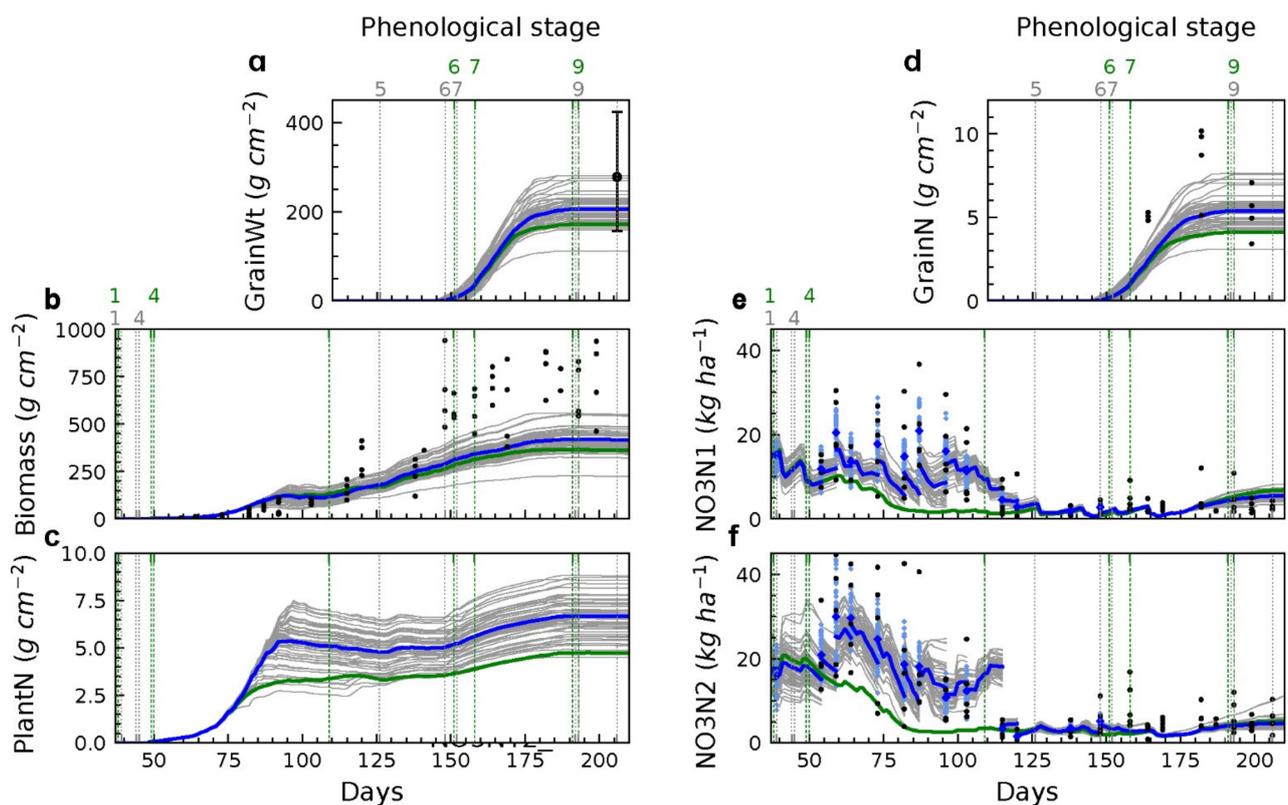


Figure 6. Evolution of GrainWt, biomass, PlantN, GrainN, NO₃N1, and NO₃N2 (a–f) with the joint assimilation of soil nitrate in the first two layers. See Figure 3 for legend.

3.3. Plot-Specific Scenario

Figure 7 shows data pairs of yield estimated by RS LAI assimilation in the plot-specific scenario and from ground observations in each experimental plot labelled A–D and the field average. The model was initialised and driven by spatially homogeneous initial conditions, parameters, and weather forcings across the field. At the field scale, the mean yield was estimated by assimilating the average LAI of the field (baseline scenario), while at the plot level, it was estimated by assimilating the LAI of the four plots (plot scale). The

yield estimates in the plots were evaluated against the yield observations collected in each plot, with mean yield being evaluated by the average yield of the four plots. A huge yield range (1926.4 kg/ha) was presented across the four experimental plots due to poor field management: unevenly sowed seeds, fail to treat weeds, delayed irrigation in the drought, etc. Results show that at the field level (black dots with a grey horizontal line), the estimated yield in each plot was uniform when LAI was assimilated at the field scale and no spatial heterogeneity information was included in the model settings and the assimilated LAI observations. In the plot-specific scenario, different plot-specific LAI data were assimilated into the spatially homogeneous model at the plot-specific scale, resulting in different yield estimation. Moreover, the data pairs in plots B and D more closely approached the 1:1 line, while the remaining data pairs were only slightly further from the line. Therefore, the spatially distributed LAI data brought spatial information into the model and further improved the spatial estimation of yield.

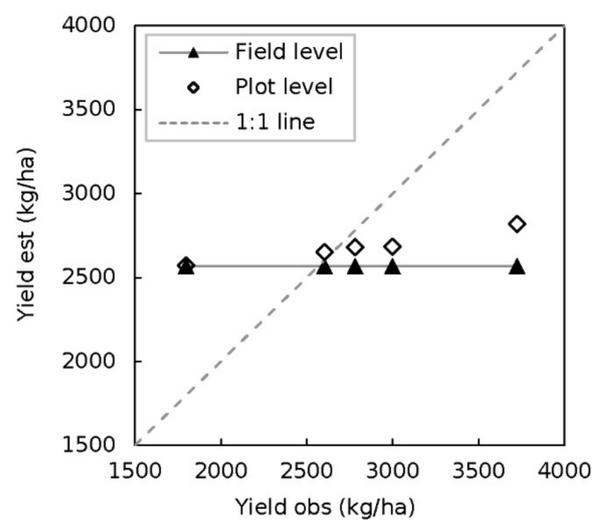


Figure 7. Yield estimates by assimilating LAI at the field (solid triangle) and plot-specific (open diamond) scales.

The results from the plot-specific scenario imply that even when the available data only allow spatially homogeneous simulation across the field, high spatial resolution LAI images can provide subfield spatial variation to the model. However, these results are limited to the data collected at only four experimental plots during one growing season, meaning that this result must be further evaluated with more validation experiments.

3.4. Observation-Limited Scenario

Figure 8 shows the RD of yield from different assimilation experiments in the observation-limited scenario. These experiments included assimilation of ground measurements of LAI, biomass, LeafWt and StemWt, soil moisture in the first soil layer (SM1), first two soil layers (SM1–2), all six soil layers (SM1–6), soil nitrate-nitrogen in the first (NO₃N1), second (NO₃N2) and first two soil layers (NO₃N1–2), and remotely sensed LAI and SM1.

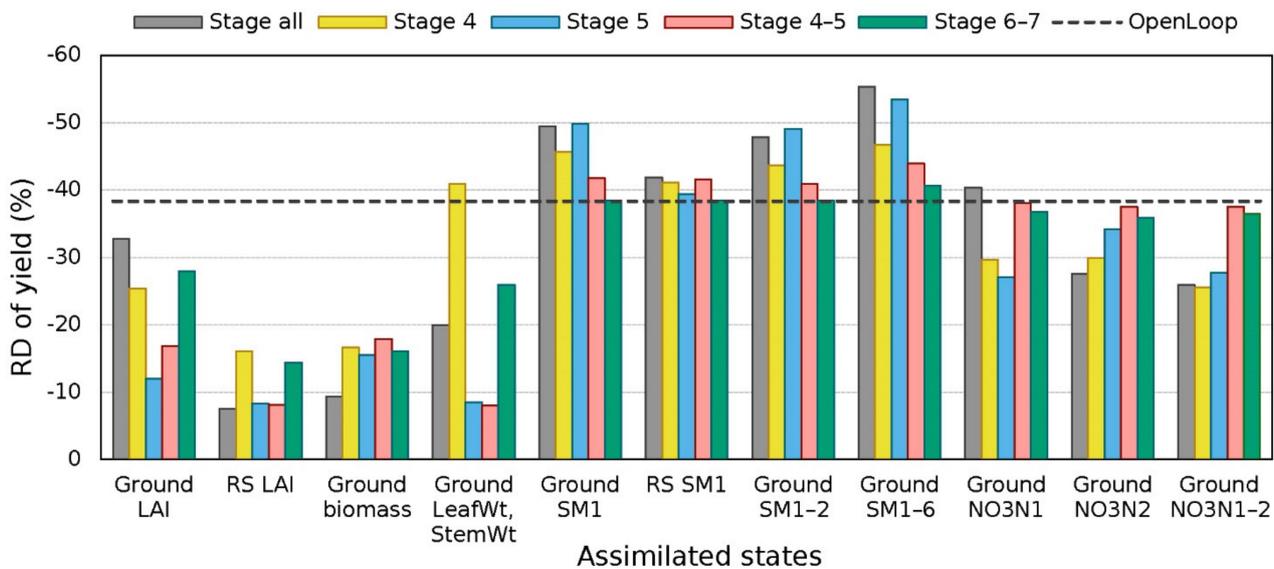


Figure 8. Relative difference of yield estimate by assimilating various state variables in different phenological stages, compared to the open-loop scenario. Stage 4: end of juvenile to floral initiation; stage 5: floral initiation to flowering; stage 6: flowering to start of grain-filling; stage 7: start of grain-filling to end of grain-filling.

The results show that the assimilation of LAI (ground and RS) and the joint assimilation of LeafWt and StemWt in stages 5 and 4–5 (vegetative stage) provided more accurate yield estimates than in other phenological stages. Biomass observations in every stage improved yield estimation, while the best yield estimation was provided when observations were available over the whole season. Soil nitrate-nitrogen (NO₃N1, NO₃N2, and NO₃N1–2) contributed to better yield estimation in stages 4 and 5. The observations of SM did not benefit yield estimation when assimilated in any stage.

3.5. Phenology-Constrained Scenario

In the open-loop run of the baseline scenario, the model provided phenology estimates within ± 5 days from the field observations for all phenological phases except phase 5 (floral initiation), when the estimated timing was 18 days ahead of the observations (Figure 9a). The reason is that the drought in this season caused a delay in phenology development, but the impact of water stress on phenology development is not described in the model physics. Another reason is that the wheat cultivar grown in this study may have a different requirement for the target accumulative thermal time in stages 4 and 5 from the parameters used in the model. Thus, it requires different periods to complete these stages.

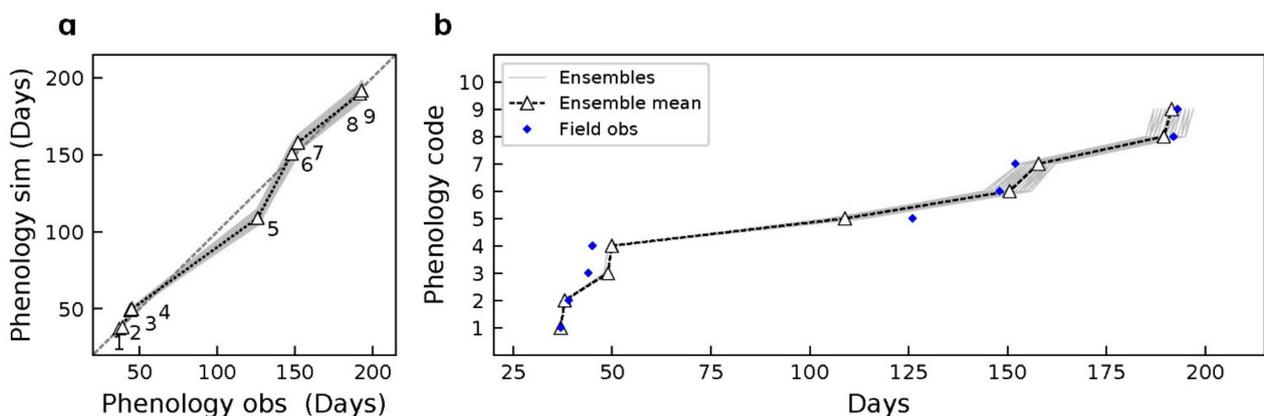


Figure 9. Phenological stage changes from model estimates versus field observations (a) and time series of phenology from model estimates and field observations (b).

By running the open-loop with the phenology constrained by ground observations, the underestimation of yield was reduced (absolute RD of yield reduced from 38.3% to 30.6%) with a slight reduction of RMSE in all estimated wheat states (Table 3). Furthermore, when assimilating wheat states (LAI, biomass or organ weight) with the phenology constrained, the underestimation of yield was further reduced. The yield estimated by the LAI and biomass assimilation (RD of yield, Table 3), and the wheat state estimation (RMSE, Table 3) 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.

3.6. Uncalibrated Scenario

The open-loop simulation from 8 out of the 14 uncalibrated models showed a substantial underestimation of yield (Figure 10, grey bar). When the ground LAI or biomass was assimilated into the uncalibrated models, the underestimation of yield was reduced in 10 and 14 models, respectively. When using RS LAI, the underestimation of yield was reduced in all 14 cases compared to the open-loop and in 11 cases compared to the ground LAI but with two models being overcorrected. When assimilating in situ soil moisture in the top two layers (ground SM1–2), the estimated yield from all 14 models was near zero, representing model failure due to a distinct mismatch in soil moisture between the observations and model estimation.

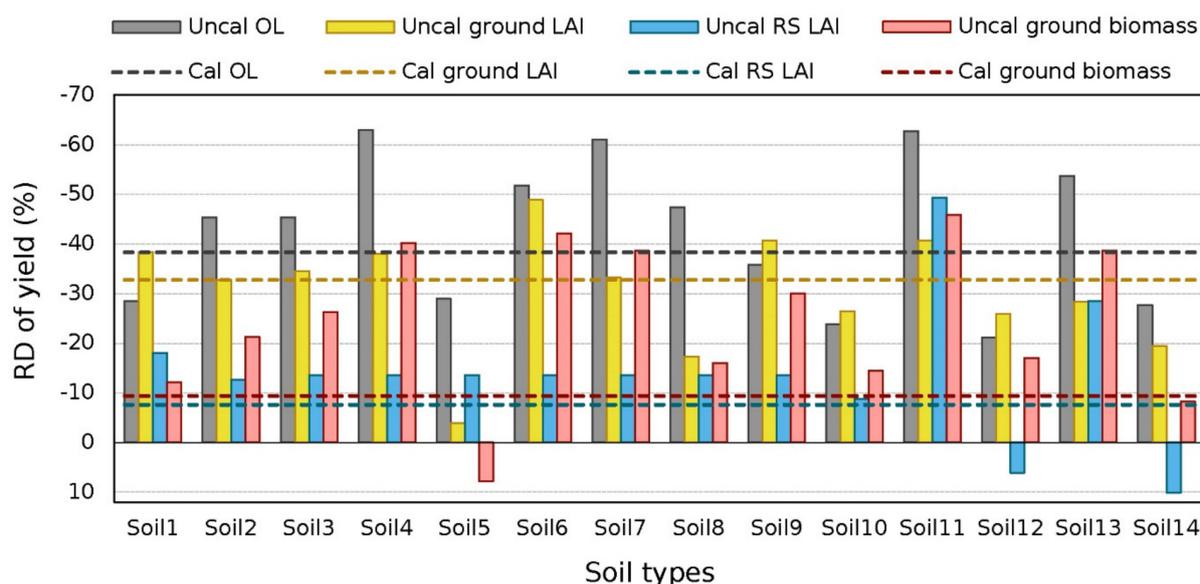


Figure 10. Relative difference of yield estimated by assimilating remotely sensed (RS) LAI, ground measured LAI, and ground measured biomass with an uncalibrated model using 14 default soil property types, as compared to using the calibrated model. Uncal and Cal represent those observations that were assimilated into uncalibrated or calibrated versions of the model. OL is the open-loop and LAI and biomass are the assimilation of leaf area index and biomass, respectively.

The result illustrates that the assimilation of wheat states, especially the biomass, reduced uncertainties caused by incorrect model parameters (cultivar-specific parameters, soil properties). Thus, assimilating wheat states is expected to improve APSIM-Wheat yield estimation even in a wheat field where soil properties have not been accurately measured or calibrated. In particular, remotely sensed LAI provided the best improvement in yield estimation.

4. Limitations and Perspectives

The main and fundamental limitation of this study is that it was tested only on one small experimental field over one growing season due to the difficulty of collecting a wide range of data required to drive the model and test different observations for assimilation.

Secondly, for best performance, the method needs site-specific calibration and soil measurements (both time-consuming and expensive) and a quantitative description of uncertainties from diverse sources. Finally, although the ground-measured biomass outperformed other states in improving yield estimation, the observations of biomass from remote sensing were not obtained and tested for assimilation due to a lack of a suitable dataset.

The application of the APSIM-EnKF framework is not limited to wheat yield estimation. It is also applicable for near real-time crop monitoring for crop and soil status, including over 20 other crop types (corn, barley, sorghum, cotton, etc.) available in the APSIM system. It can also be adapted to other state-updating algorithms (Kalman filter, particle filter, etc.) with minor modifications to the source code. 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.

5. Conclusions

This paper presented a data assimilation case study for the Cora Lynn area with the 2018/19 wheat season dataset. In a baseline scenario, the ground-measured and remotely sensed wheat and soil states were assimilated, including LAI, biomass, organ weight of leaf and stem, soil moisture in six soil layers from depth 0 to 55 cm, and soil nitrogen in the top two soil layers from depth 0 to 15 cm. These results were compared to scenarios where the state variables were assimilated into the APSIM-Wheat model under different spatial scales, availability of observations, and model parameters.

In the baseline scenario, the data assimilation showed that biomass and LAI, particularly biomass, provided improved yield estimation. Being observable from remote sensing makes them promising future data assimilation practices. Poor performance was found from assimilating soil moisture and soil nitrate. However, this does not imply that soil moisture and nitrogen data are unimportant for wheat modelling and yield estimation. Rather, that in this case they were already well estimated. Accordingly, they are essential in calibrating parameters related to soil moisture and nitrogen estimation where such parameters are not already well defined. With 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. Importantly, although a basic calibration to soil parameters is recommended, uncalibrated scenarios showed that the assimilation of biomass and LAI reduced the errors of yield estimates caused by parameter uncertainties (wheat cultivar type, soil properties) in uncalibrated models. Consequently, the heavy dependence of the APSIM-Wheat model on the soil information that is usually unavailable, particularly in spatial scale, can be relieved.

A plot-specific scenario showed the potential to provide subfield spatial variation in a spatially homogeneous model when high-resolution LAI was assimilated. Moreover, an observation-limited scenario showed that the assimilation of wheat states (LAI, biomass, LeafWt, and StemWt) benefited yield estimation the most when the observations were collected before the anthesis was completed (stages 4 and 5). Furthermore, a phenology-constrained scenario found that the uncertainty from model inputs (e.g., temperature) and parameters (e.g., target thermal time of phenological stages) that control phenology development can be partially reduced by direct insertion, with further improvement on yield estimates.

Author Contributions: Conceptualization, Y.Z. and J.P.W.; methodology, Y.Z., J.P.W. and V.R.N.P.; software, Y.Z.; validation, Y.Z.; formal analysis, Y.Z.; investigation, Y.Z.; resources, Y.Z., J.P.W. and V.R.N.P.; data curation, Y.Z. and Y.S.; writing—original draft preparation, Y.Z.; writing—review and editing, Y.Z., J.P.W., V.R.N.P. and Y.S.; visualization, Y.Z.; supervision, J.P.W. and V.R.N.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The meteorological data are available from <http://www.bom.gov.au/> (accessed on 16 July 2019) and the Weather Underground weather station network data are available from <https://www.wunderground.com> (accessed on 2 March 2019). The field validation data presented in this study are openly available in the P-band Radiometer Inferred Soil Moisture (PRISIM) website at <https://www.prism.monash.edu/index.html> (archived on 22 April 2021).

Acknowledgments: We thank Monash University for the Faculty of Engineering International Post-graduate Research Scholarship to cover the doctoral tuition fees and for the solid and persistent support in completing the research. Acknowledgement is made to the APSIM Initiative, which takes responsibility for quality assurance and a structured innovation programme for APSIM's modelling software, which is provided free for research and development use (see www.apsim.info (accessed on 19 April 2017) for details).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Batchelor, W.D.; Basso, B.; Paz, J.O. Examples of strategies to analyze spatial and temporal yield variability using crop models. *Eur. J. Agron.* **2002**, *18*, 141–158. [[CrossRef](#)]
2. Palosuo, T.; Kersebaum, K.C.; Angulo, C.; Hlavinka, P.; Moriondo, M.; Olesen, J.E.; Patil, R.H.; Ruget, F.; Rumbaur, C.; Takáč, J. Simulation of winter wheat yield and its variability in different climates of Europe: A comparison of eight crop growth models. *Eur. J. Agron.* **2011**, *35*, 103–114. [[CrossRef](#)]
3. Mosleh, M.K.; Hassan, Q.K.; Chowdhury, E.H. Application of remote sensors in mapping rice area and forecasting its production: A review. *Sensors* **2015**, *15*, 769–791. [[CrossRef](#)] [[PubMed](#)]
4. Kaufmann, R.K.; Snell, S.E. A biophysical model of corn yield: Integrating climatic and social determinants. *Am. J. Agric. Econ.* **1997**, *79*, 178–190. [[CrossRef](#)]
5. McMillan, H.; Krueger, T.; Freer, J. Benchmarking observational uncertainties for hydrology: Rainfall, river discharge and water quality. *Hydrol Process* **2012**, *26*, 4078–4111. [[CrossRef](#)]
6. Dzotsi, K.A.; Basso, B.; Jones, J.W. Development, uncertainty and sensitivity analysis of the simple SALUS crop model in DSSAT. *Ecol. Model.* **2013**, *260*, 62–76. [[CrossRef](#)]
7. Wallach, D.; Thorburn, P.J. Estimating uncertainty in crop model predictions: Current situation and future prospects. *Eur. J. Agron.* **2017**, *88*, A1–A7. [[CrossRef](#)]
8. Maestrini, B.; Basso, B. Predicting spatial patterns of within-field crop yield variability. *Field Crops Res.* **2018**, *219*, 106–112. [[CrossRef](#)]
9. Kuang, B.; Mahmood, H.S.; Quraishi, M.Z.; Hoogmoed, W.B.; Mouazen, A.M.; van Henten, E.J. Chapter four—Sensing Soil Properties in the Laboratory, In Situ, and On-Line: A Review. In *Advances in Agronomy*; Sparks, D.L., Ed.; Academic Press: Cambridge, MA, USA, 2012; Volume 114, pp. 155–223.
10. Zhao, G.; Siebert, S.; Enders, A.; Rezaei, E.E.; Yan, C.; Ewert, F. Demand for multi-scale weather data for regional crop modeling. *Agric. For. Meteorol.* **2015**, *200*, 156–171. [[CrossRef](#)]
11. Jin, C.; Xiao, X.; Wagle, P.; Griffis, T.; Dong, J.; Wu, C.; Qin, Y.; Cook, D.R. Effects of in-situ and reanalysis climate data on estimation of cropland gross primary production using the Vegetation Photosynthesis Model. *Agric. For. Meteorol.* **2015**, *213*, 240–250. [[CrossRef](#)]
12. Tan, B.; Hu, J.; Zhang, P.; Huang, D.; Shabanov, N.; Weiss, M.; Knyazikhin, Y.; Myneni, R.B. Validation of Moderate Resolution Imaging Spectroradiometer Leaf Area Index Product in Croplands of Alpilles, France. *J. Geophys. Res. Atmos.* **2005**, *110*. Available online: <https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2004JD004860> (accessed on 21 October 2020). [[CrossRef](#)]
13. Sanchez, N.; Martínez-Fernández, J.; Scaini, A.; Perez-Gutierrez, C. Validation of the SMOS L2 soil moisture data in the REMEDHUS network (Spain). *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 1602–1611. [[CrossRef](#)]
14. Colliander, A.; Jackson, T.J.; Bindlish, R.; Chan, S.; Das, N.; Kim, S.; Cosh, M.; Dunbar, R.; Dang, L.; Pashaian, L. Validation of SMAP surface soil moisture products with core validation sites. *Remote Sens. Environ.* **2017**, *191*, 215–231. [[CrossRef](#)]
15. Bao, Y.; Gao, W.; Gao, Z. Estimation of winter wheat biomass based on remote sensing data at various spatial and spectral resolutions. *Front. Earth Sci. China* **2009**, *3*, 118. [[CrossRef](#)]
16. Sadeh, Y.; Zhu, X.; Dunkerley, D.; Walker, P.J.; Zhang, Y.; Rozenstein, O.; Manivasagam, V.S.; Chenu, K. Fusion of Sentinel-2 and PlanetScope time-series data into daily 3 m surface reflectance and wheat LAI monitoring. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *96*, 102260. [[CrossRef](#)]
17. Pulvirenti, L.; Squicciarino, G.; Cenci, L.; Boni, G.; Pierdicca, N.; Chini, M.; Versace, C.; Campanella, P. A surface soil moisture mapping service at national (Italian) scale based on Sentinel-1 data. *Environ. Model. Softw.* **2018**, *102*, 13–28. [[CrossRef](#)]
18. Patel, N.; Ravi, N.; Navalgund, R.; Dash, R.; Das, K.; Patnaik, S. Estimation of rice yield using IRS-1A digital data in coastal tract of Orissa. *Remote Sens.* **1991**, *12*, 2259–2266. [[CrossRef](#)]

19. Li, Y.; Liao, Q.; Li, X.; Liao, S.; Chi, G.; Peng, S. Towards an operational system for regional-scale rice yield estimation using a time-series of Radarsat ScanSAR images. *Int. J. Remote Sens.* **2003**, *24*, 4207–4220. [[CrossRef](#)]
20. Son, N.; Chen, C.; Chen, C.; Chang, L.; Duc, H.; Nguyen, L. Prediction of rice crop yield using MODIS EVI– LAI data in the Mekong Delta, Vietnam. *Int. J. Remote Sens.* **2013**, *34*, 7275–7292. [[CrossRef](#)]
21. Campos, I.; González-Gómez, L.; Villodre, J.; Calera, M.; Campoy, J.; Jiménez, N.; Plaza, C.; Sánchez-Prieto, S.; Calera, A. Mapping within-field variability in wheat yield and biomass using remote sensing vegetation indices. *Precis. Agric.* **2019**, *20*, 214–236. [[CrossRef](#)]
22. Nuarsa, I.W.; Nishio, F.; Hongo, C. Relationship between rice spectral and rice yield using MODIS data. *J. Agric. Sci.* **2011**, *3*, 80. [[CrossRef](#)]
23. Nuarsa, I.W.; Nishio, F.; Hongo, C. Rice yield estimation using Landsat ETM+ data and field observation. *J. Agric. Sci.* **2011**, *4*, 45. [[CrossRef](#)]
24. Pantazi, X.E.; Moshou, D.; Alexandridis, T.; Whetton, R.L.; Mouazen, A.M. Wheat yield prediction using machine learning and advanced sensing techniques. *Comput. Electron. Agric.* **2016**, *121*, 57–65. [[CrossRef](#)]
25. Kayad, A.; Sozzi, M.; Gatto, S.; Marinello, F.; Pirotti, F. Monitoring within-field variability of corn yield using Sentinel-2 and machine learning techniques. *Remote Sens.* **2019**, *11*, 2873. [[CrossRef](#)]
26. Wang, Y.; Zhang, Z.; Feng, L.; Du, Q.; Runge, T. Combining multi-source data and machine learning approaches to predict winter wheat yield in the conterminous united states. *Remote Sens.* **2020**, *12*, 1232. [[CrossRef](#)]
27. Kamir, E.; Waldner, F.; Hochman, Z. Estimating wheat yields in Australia using climate records, satellite image time series and machine learning methods. *ISPRS J. Photogramm. Remote Sens.* **2020**, *160*, 124–135. [[CrossRef](#)]
28. Zhang, Y. Towards improved crop growth and yield estimation: Observation constrained wheat modelling. Doctoral Thesis, Monash University, Victoria, Australia, 2020. [[CrossRef](#)]
29. Houser, P.R.; De Lannoy, G.J.; Walker, J.P. Hydrologic Data Assimilation. In *Approaches to Managing Disaster—Assessing Hazards, Emergencies and Disaster Impacts*; Tiefenbacher, J., Ed.; IntechOpen: London, UK, 2012; pp. 41–64.
30. Walker, J.P.; Houser, P.R. Hydrologic data assimilation. In *Advances in Water Science Methodologies*; CRC Press: Boca Raton, FL, USA, 2005; pp. 45–68.
31. Ines, A.V.; Das, N.N.; Hansen, J.W.; Njoku, E.G. Assimilation of remotely sensed soil moisture and vegetation with a crop simulation model for maize yield prediction. *Remote Sens. Environ.* **2013**, *138*, 149–164. [[CrossRef](#)]
32. Zhao, Y.; Chen, S.; Shen, S. Assimilating remote sensing information with crop model using Ensemble Kalman Filter for improving LAI monitoring and yield estimation. *Ecol. Model.* **2013**, *270*, 30–42. [[CrossRef](#)]
33. Huang, J.; Sedano, F.; Huang, Y.; Ma, H.; Li, X.; Liang, S.; Tian, L.; Zhang, X.; Fan, J.; Wu, W. Assimilating a synthetic Kalman filter leaf area index series into the WOFOST model to improve regional winter wheat yield estimation. *Agric. For. Meteorol.* **2016**, *216*, 188–202. [[CrossRef](#)]
34. Vazifedoust, M.; Van Dam, J.; Bastiaanssen, W.; Feddes, R. Assimilation of satellite data into agrohydrological models to improve crop yield forecasts. *Int. J. Remote Sens.* **2009**, *30*, 2523–2545. [[CrossRef](#)]
35. Guo, C.; Tang, Y.; Lu, J.; Zhu, Y.; Cao, W.; Cheng, T.; Zhang, L.; Tian, Y. Predicting wheat productivity: Integrating time series of vegetation indices into crop modeling via sequential assimilation. *Agric. For. Meteorol.* **2019**, *272–273*, 69–80. [[CrossRef](#)]
36. Zadoks, J.C.; Chang, T.T.; Konzak, C.F. A decimal code for the growth stages of cereals. *Weed Res.* **1974**, *14*, 415–421. [[CrossRef](#)]
37. Anderson, J.L.; Anderson, S.L. A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Weather Rev.* **1999**, *127*, 2741–2758. [[CrossRef](#)]
38. Bernard, R.; Vauclin, M.; Vidal-Madjar, D. Possible use of active microwave remote sensing data for prediction of regional evaporation by numerical simulation of soil water movement in the unsaturated zone. *Water Resour. Res.* **1981**, *17*, 1603–1610. [[CrossRef](#)]
39. Jackson, T.; Schugge, T.; Nicks, A.; Coleman, G.; Engman, E. Soil moisture updating and microwave remote sensing for hydrological simulation/La remise à jour de l'état d'humidité des sols en vue de la simulation hydrologique. *Hydrol. Sci. J.* **1981**, *26*, 305–319. [[CrossRef](#)]
40. Petrie, R.E.; Dance, S.L. Ensemble-based data assimilation and the localisation problem. *Weather* **2010**, *65*, 65–69. [[CrossRef](#)]
41. Elsner, J.; Tsonis, A. Nonlinear prediction, chaos, and noise. *Bull. Am. Meteorol. Soc.* **1992**, *73*, 49–60. [[CrossRef](#)]
42. Bell, M.J.; Forbes, R.M.; Hines, A. Assessment of the FOAM global data assimilation system for real-time operational ocean forecasting. *J. Mar. Syst.* **2000**, *25*, 1–22. [[CrossRef](#)]
43. Buizza, R. Chaos and weather prediction—A review of recent advances in Numerical Weather Prediction: Ensemble forecasting and adaptive observation targeting. *Il Nuovo Cim. C* **2001**, *24*, 273–302.
44. Cosgrove, B.A.; Lohmann, D.; Mitchell, K.E.; Houser, P.R.; Wood, E.F.; Schaake, J.C.; Robock, A.; Marshall, C.; Sheffield, J.; Duan, Q. Real-Time and Retrospective Forcing in the North American Land Data Assimilation System (NLDAS) Project. *J. Geophys. Res. Atmos.* **2003**, *108*. Available online: <https://agupubs.onlinelibrary.wiley.com> (accessed on 12 November 2021). [[CrossRef](#)]
45. Barker, D.; Huang, X.-Y.; Liu, Z.; Auligné, T.; Zhang, X.; Rugg, S.; Ajjaji, R.; Bourgeois, A.; Bray, J.; Chen, Y. The weather research and forecasting model's community variational/ensemble data assimilation system: WRFDA. *Bull. Am. Meteorol. Soc.* **2012**, *93*, 831–843. [[CrossRef](#)]
46. Buehner, M.; Morneau, J.; Charette, C. Four-dimensional ensemble-variational data assimilation for global deterministic weather prediction. *Nonlinear Proc. Geoph.* **2013**, *20*, 669–682. [[CrossRef](#)]

47. Yue, J.; Feng, H.; Yang, G.; Li, Z. A comparison of regression techniques for estimation of above-ground winter wheat biomass using near-surface spectroscopy. *Remote Sens.* **2018**, *10*, 66. [[CrossRef](#)]
48. Zhou, X.; Zhu, X.; Dong, Z.; Guo, W. Estimation of biomass in wheat using random forest regression algorithm and remote sensing data. *Crop J.* **2016**, *4*, 212–219.
49. Yue, J.; Yang, G.; Li, C.; Li, Z.; Wang, Y.; Feng, H.; Xu, B. Estimation of winter wheat above-ground biomass using unmanned aerial vehicle-based snapshot hyperspectral sensor and crop height improved models. *Remote Sens.* **2017**, *9*, 708. [[CrossRef](#)]
50. Bendig, J.; Bolten, A.; Bennertz, S.; Broscheit, J.; Eichfuss, S.; Bareth, G. Estimating biomass of barley using crop surface models (CSMs) derived from UAV-based RGB imaging. *Remote Sens.* **2014**, *6*, 10395–10412. [[CrossRef](#)]