

Assimilation of Sentinel-2 Remotely Sensed Data into a Wheat Model

Dr Yuxi Zhang



December 30, 2021

Monash University

# Executive summary

Crop models provide an insight into the crop development process and yield by forecasting the interaction among the plant, atmosphere and soil. With data assimilation techniques, remotely sensed observations about wheat and soil status can be assimilated into crop models and is expected to provide a more accurate yield estimation.

This study aimed to explore the assimilation of remotely sensed leaf area index (LAI) into a crop simulation model APSIM-Wheat to improve yield estimation in a field scale. The methodology consists of two steps. First, the Sentinel-2 remotely sensed spectral reflectance (3-4 days, 10 meters) was mapped to the leaf area index through the linear regression relationship between vegetation indices and the ground-measured leaf area index data. Second, the field average of LAI over the study area was assimilated into the APSIM-Wheat model.

The result found that the simple ratio showed the best linear regression to fit the ground-based LAI data. The assimilation of remotely sensed LAI improved the yield estimation with an absolute relative difference of yield reduced from 38.3% (no assimilation) to 25.0% (LAI assimilation). This improvement was found when LAI was assimilated in any phenological stages between the end of juvenile to the end grain filling phenological phases. Moreover, when testing the method in uncalibrated models, the LAI assimilation still provided a better yield estimation than with no observations assimilated. Eventually, this study illustrated the potential of improving yield estimation with satellite remote sensing techniques and data assimilation techniques.

Contents

[Executive summary i](#_Toc91766350)

[1 Introduction 1](#_Toc91766351)

[2 Methodology 1](#_Toc91766352)

[2.1 The EnKF-APSIM data assimilation framework 1](#_Toc91766353)

[2.2 Dataset 2](#_Toc91766354)

[2.3 Experimental design 2](#_Toc91766355)

[2.3.1 LAI mapping 3](#_Toc91766356)

[2.3.2 Data assimilation scenario 4](#_Toc91766357)

[3 Results and discussion 5](#_Toc91766358)

[3.1 LAI mapping 5](#_Toc91766359)

[3.2 Data assimilation 7](#_Toc91766360)

[3.2.1 Scenario 1 7](#_Toc91766361)

[3.2.2 Scenario 2 9](#_Toc91766362)

[4 Conclusion and perspectives 9](#_Toc91766363)

[Funding 10](#_Toc91766364)

[Acknowledgement 10](#_Toc91766365)

[References 11](#_Toc91766366)

# Introduction

Precision agriculture aims to maximise the agricultural output of the field and helps protect the environment by using the wise use of agricultural inputs by accurately estimating crop needs at sub-paddock scales [1-4]. Understanding spatial and temporal yield variability is vital to suggest site-specific management regarding when and where to apply irrigation, fertilisation, and weed/pest control [5].

Crop simulation models generally have insights into crop development by the biophysical processes they describe and provide estimation and prediction of wheat and soil status on a daily basis. However, for accurate predictions, these models require various parameter data that are often laborious and costly to collect or calibrate using traditional techniques [6,7]. Remote sensing techniques are now providing abundant information for agricultural monitoring. Satellite-borne remote sensing information about crop and soil information has broad coverage and high spatial resolution. This information can be assimilated into physically-based crop simulation models to reduce model uncertainty and improve its accuracy on crop monitoring and yield estimation [5].

Leaf area index (LAI) is a crucial indicator for photosynthesis productivity and grain yield. This project tests the assimilation of satellite-borne remote sensing LAI into an APSIM-Wheat crop model to improve the model performance in yield estimation.

# Methodology

## The EnKF-APSIM data assimilation framework

The Agricultural Production Systems sIMulator (APSIM) is a model developed to simulate the biophysical process of over twenty crop types and their interaction with the atmosphere, soil, climate, and management. The APSIM-Wheat module simulates wheat development from sowing to harvest, accounting for the wheat phenology, utilising solar radiation, biomass accumulation, leaf area development, rainfall-runoff process, plant root water uptake, etc. The model simulations soil and vegetation states (e.g., biomass, LAI, soil moisture, etc.) at a daily timestep and estimates grain yield at harvest.

Ensemble Kalman filter is an advanced data assimilation algorithm to update model state variables when observations are obtained from external sources. While observations of soil and vegetation states can be obtained from remote sensing, this information can be assimilated into the APSIM model to reduce uncertainty in the model and thus improve model performance for more accurate yield estimation. In this study, the APSIM-EnKF data assimilation frame developed by Zhang [5] (source code available on <https://github.com/yuxi-research/APSIM-EnKF>) was used to assimilate remotely sensed LAI data into the APSIM-Wheat model for improved performance of yield estimation.

## Dataset

The ground-based data used in this study is a dataset collected in Victoria over a 75 by 75 meters wheat field in the 2018/19 season. This dataset includes 1) daily weather forcings as model inputs: rainfall, radiation, maximum and minimum temperature, vapor pressure and wind speed; and 2) ground-based wheat and soil data for model calibration and validation: LAI, soil moisture, biomass, grain yield and phenology.

Sentinel-2 MultiSpectral Instrument Level-1C (MSI-1C) spectral reflectance product (3-4 days, 10 meters) over the growing period at the study area was downloaded from European Space Agency (ESA). The product contains thirteen spectral reflectance bands throughout visible (VIS) to short wave infrared (SWIR) with a cloud mask band indicating the presence of opaque and cirrus clouds. A shapefile 10-meters inside the study area boundary was used to extract the pixels from Sentinel-2 images to remove impure pixels mixed with roads, trees, bare soil and other crops close to the field. For each extracted pixel, multiple vegetation indices (VIs) were calculated in the absence of cloud cover, and a time series of each VI was produced by linear interpolation. Then, a daily time series of each VI at the field was obtained by averaging all pixels over the field.

## Experimental design

The experiment contains two steps: 1) mapping LAI from remote sensing spectral reflectance and validating the accuracy against observed LAI at the study area; and 2) assimilating LAI into the APSIM-Wheat model to estimate yield and validate the performance with ground-measured yield data. Two scenarios were explored in data assimilation experiments: 1) assimilating LAI into a calibrated APSIM-Wheat model in different wheat phenological stages, and 2) assimilating LAI into a group of uncalibrated APSIM-Wheat models.

### LAI mapping

Thirteen VIs (Table 1) were calculated to map LAI with linear regression models. Data pairs of each of the VIs and the ground-based measurements of LAI measured in the same day was taken to train the linear regression equation

$$VI=b\_{0}+b\_{1}×LAI,$$

where $b\_{0}$ and $b\_{1}$ are the slope and the intercept of the linear equation, respectively.

Table : Vegetation indices for LAI mapping (table reproduced from Sadeh*, et al.* [8]).

|  |  |  |
| --- | --- | --- |
| Vegetation index | Equation | Reference |
| Simple Ratio (SR) | $$\frac{NIR}{R}$$ | [9] |
| Enhanced Vegetation Index 2 (EVI2) | $$\frac{2.5(NIR-R)}{(NIR+2.4R+1)}$$ | [10,11] |
| Green Chlorophyll Vegetation Index (GCVI) | $$\left(NIR-G\right)-1$$ | [12,13] |
| Normalized Difference Vegetation Index (NDVI) | $$\frac{NIR-R}{NIR+R}$$ | [14] |
| Modified Triangular Vegetation Index 2 (MTVI2) | $$\frac{1.5 [1.2\left(NIR-G\right)-2.5\left(R-G\right)]}{\sqrt{(2NIR+1)^{2}-(6NIR-5\sqrt{R)}-0.5}}$$ | [15] |
| Modified Soil-Adjusted Vegetation Index (MSAVI) | $$0.5\left[2NIR+1-\sqrt{(2NIR+1)^{2}-8(NIR-R)}\right]$$ | [15,16] |
| Wide Dynamic Range Vegetation Index (WDRVI) | $$\frac{α ∙NIR-R}{α ∙NIR+R}+\frac{1-α}{1+α}$$ | [17,18] |
| Green Wide Dynamic Range Vegetation Index (Green-WDRVI) | $$\frac{α ∙NIR-G}{α ∙NIR+G}+\frac{1-α}{1+α}$$ | [18,19] |
| Optimized Soil-Adjusted Vegetation Index (OSAVI) | $$\frac{NIR-R}{NIR+R+0.16}$$ | [20] |
| Green Simple Ratio (GSR) | $$\frac{NIR}{Green}$$ | [21] |
| Green NDVI (GNDVI) | $$\frac{NIR-G}{NIR+G}$$ | [22] |
| Renormalized Difference Vegetation Index (RDVI) | $$\frac{NIR-R}{\sqrt{NIR+R}}$$ | [23] |
| Transformed Vegetative Index (TVI) | $$\sqrt{\frac{NIR-R}{NIR+R}+0.5}$$ | [14,24] |

R: red, G: green, B: blue, NIR: near-infrared. L is a soil adjustment factor (usually 0.5).

The coefficient of correlation (R-Squared, $R^{2}$) and the root mean squared error (RMSE) of the mapped LAI were calculated to evaluate the goodness-of-fit of these linear regression equations. The linear regression that best fits the ground-observed LAI will be chosen as observations in the data assimilation experiments.

### Data assimilation scenario

A stochastic model run will be first performed as an 'open-loop' run. The term 'open-loop' is opposite to 'closed-loop' where the data assimilation module is applied and thus means no observation is assimilated to update the model states. The open-loop is the control run of the experiment used to evaluate the performance of a data assimilation experiment.

In the data assimilation experiments, a time series LAI with the lowest RMSE estimated by the VIs were chosen to be assimilated as remotely sensed observation into the APSIM-Wheat model. The entire assimilation window was set to be between the beginning of the phenological stage 4 (end of juvenile) and the middle of stage 7 (grain filling). The assimilation of the LAI in the entire assimilation windows into a calibrated APSIM-Wheat model was taken as a baseline scenario in this study.

To initialise the ensemble, the weather forcings, parameters and initial conditional were perturbed to create 50 simulations as the member of the ensemble. Each ensemble member was perturbed by adding Gaussian noise with a mean of zero and standard deviation of their uncertainty estimated by observations or reasonably assumed to create an appropriate ensemble size. The observational uncertainty of the remotely sensed LAI was the RMSE of the best-fit LAI estimated in the LAI mapping stage. Two data assimilation scenarios were performed as described in the remaining of this section.

Scenario 1 aimed to explore the assimilation performance in different phenological stages and to suggest which stage the observation of LAI measurement is valuable in improving the model performance for yield estimation. In this scenario, the assimilation of these stages in different phenological stages as stage 4 (end of juvenile to floral initiation), stage 5 (floral initiation to flowering), and stage 6-7 (flowering to end of grain filling) were assimilated, respectively. The performance of yield estimation was compared with the open-loop and within all assimilation runs.

Scenario 2 aims to explore whether the assimilation of LAI improves the performance of yield estimation in an area when local calibration or site investigation is not available. In this scenario, a group of uncalibrated models with different parameters in the soil module was used to replace the calibrated model. The uncalibrated models use 14 sets of different soil characteristics in the APSIM soil library, taken from a dataset measured in 14 locations close to the study area in Victoria, Australia.

In both scenarios, the performance of yield estimation is evaluated by the yield relative difference (RD) calculated as:

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

where $yield\_{est}$ and $yield\_{obs}$ are the estimated and ground-observed yield measured at harvest, respectively. A RD of yield value equals zero means a perfect yield estimation, with positive and negative values meaning that the yield was over-estimated and under-estimated, respectively. If the RD of yield value from data assimilation scenarios showed a closer-to-zero value than the open-loop, the assimilation of the state is considered to have a better yield performance than no observation is assimilated.

# Results and discussion

## LAI mapping

Table 2 shows the parameters and evaluation criteria of the linear regression equations between the ground-measured LAI and the VIs. The simple ratio (SR) was found to bet-fit the data pairs with R2 of 0.91 and RMSE of 0.2 m2/m2. Thus, a best-fit time series of LAI was produced by mapping the SR with the linear equation of slope b1=2.15 and intercept b0=1.69. The best-fitted LAI time series is used as the assimilated observations in the subsequent assimilation experiments. A scatter of the fitted and ground-measured LAI is presented in Figure 1.

Table : Linear regression parameters between LAI and 13 vegetation indices (sort in ascending order according to the value of RMSE fitted data).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| VI | intercept (b0) | slope (b1) | Fitted data | All data |
| R2 | RMSE | R2 | RMSE |
| SR | 1.69 | 2.15 | 0.91 | 0.20 | 0.82 | 0.45 |
| WDRVI | 0.14 | 0.21 | 0.90 | 0.20 | 0.79 | 0.50 |
| GSR | 1.92 | 1.01 | 0.87 | 0.24 | 0.77 | 0.53 |
| EVI2 | 0.61 | 0.44 | 0.86 | 0.25 | 0.72 | 0.60 |
| GWDRVI | 0.15 | 0.12 | 0.85 | 0.26 | 0.73 | 0.58 |
| RDVI | 18.4 | 15.7 | 0.83 | 0.28 | 0.68 | 0.67 |
| NDVI | 0.36 | 0.18 | 0.82 | 0.29 | 0.67 | 0.67 |
| OSAVI | 0.36 | 0.18 | 0.82 | 0.29 | 0.67 | 0.67 |
| GCVI | 880 | 1087 | 0.82 | 0.29 | 0.67 | 0.68 |
| TVI | 0.93 | 0.09 | 0.80 | 0.31 | 0.65 | 0.70 |
| GNDVI | 0.34 | 0.13 | 0.77 | 0.34 | 0.64 | 0.72 |
| MSAVI | 0.54 | 0.16 | 0.76 | 0.35 | 0.61 | 0.77 |
| MTVI2 | 0.52 | 0.17 | 0.75 | 0.36 | 0.59 | 0.80 |



Figure : Scatter of LAI estimation and observations. Data pairs with LAI observation less than 3 m2/m2 (fitted data) were used to fit the linear regression.

## Data assimilation

### Scenario 1

The data assimilation scenario 1 shows that the assimilation of wheat LAI improved yield estimation when assimilated. The RD of yield was improved from -38.3% in the open-loop to -25% (stage all), -15.2% (stage 4), -24.9% (stage 5), -30.2% (stage 6–7) in assimilating LAI in different phenological stages. Particularly, stage 4 improved yield estimation more effectively than other stages, with an RD of yield value closest to zero.

The improvement in wheat monitoring was reflected by the root mean squared error (RMSE) of the estimated wheat states. Generally, reduced RMSE values in the LAI and biomass were found when assimilated in all stages and each separate stage (Table 3) compared to the open-loop. Different from the wheat states, the impact of LAI assimilation on surface soil moisture (SM1) and root-zone soil moisture (SM RZ) is not distinct, with minor changes found in the RMSE of SM1 and SM RZ (Table 3).

Table : Relative difference (RD) of yield and root mean square root (RMSE) of wheat states when assimilating LAI in different phenological stages.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Assimilated phenological stage | RD | RMSE |  |  |  |
| yield | LAI | Biomass | SM1 | SM RZ |
| % | m2/m2 | kg/ha | m3/m3 | m3/m3 |
| Open-loop | -38.3 | 0.91 | 150.4 | 0.041 | 0.022 |
| All (end of juvenile to to grain filling) | -25.0 | 0.48 | 102.7 | 0.046 | 0.024 |
| Stage 4 (end of juvenile to floral initiation) | -15.2 | 0.49 | 91.2 | 0.046 | 0.024 |
| Stage 5 (floral initiation to flowering) | -24.9 | 0.87 | 110.8 | 0.042 | 0.023 |
| Stage 6–7 (flowering to end of grain filling) | -30.2 | 0.91 | 147.3 | 0.041 | 0.022 |

SM1: soil moisture in the first layer (0-5 cm). SM RZ: soil moisture in the root zone (5-55 cm).

Figure 2 shows the evolution of grain weight (GrainWt), LAI, biomass and surface soil moisture (SM1) from the assimilation of remotely sensed LAI in all stages. With the LAI observations assimilated, the LAI estimates were updated to a value closer to its observation (Figure 2-b). Meanwhile, the biomass and surface soil moisture estimates were updated to approach their observations, respectively (Figure 2-c, d). Eventually, a more accurate yield estimation was provided by the model (Figure 2-a).



Figure : Evolution of (a) grain weight (GrainWt), (b) LAI, (c) biomass, and (d) surface soil moisture (SM1) with the assimilation of remotely sensed LAI in all stages.

### Scenario 2

In scenario 2, the open-loop from 8 out of the 14 uncalibrated models showed a substantial underestimated grain yield. With the assimilation of remotely sensed LAI (Figure 3) into these uncalibrated models, the underestimates in the yield were reduced in all 14 cases, although one simulation was slightly over-corrected. This result showed that the assimilation of both remotely sensed LAI improved APSIM-Wheat yield estimation, even in a wheat field where soil properties cannot be accurately measured or calibrated. Therefore, the assimilation of remotely sensed LAI showed a potential to provide a more accurate yield estimation under the condition that local calibration and site investigation of soil properties is not available.



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

# Conclusion and perspectives

This report presented a data assimilation case study for a 75 by 75 meters wheat field using the dataset collected in the 2018/19 season with only remotely sensed observation assimilated. This study successfully mapped LAI over the study site from multiple vegetation indices from Sentinel-2 spectral reflectance images with R2 of 0.91 and RMSE of 0.2 m2/m2. The data assimilation scenario 1 showed that wheat yield monitoring and prediction could be improved by assimilating remotely sensed leaf area index into the APSIM-Wheat model at various phenological stages. Moreover, scenario 2 revealed the potential that a crop simulation model could be applied to uncalibrated wheat fields to provide more accurate yield estimates if leaf area index observations can be obtained from remote sensing.

Notably, the method presented in this study is not only limited to the assimilation of the leaf area index. With more observations of wheat and soil states (biomass, soil moisture, soil nitrogen, etc.) conveniently available from the rapid development of remote sensing techniques, these state variables can also be assimilated into the wheat model. Moreover, the methodology is also applicable to more than twenty crop species (maise, barley, etc.) integrated into the APSIM system with a minor modification to the source code of the APSIM-EnKF framework.

# Funding

This research is supported by the 2020 Grants for Accelerating Precision Agriculture, Bayer CropScience LP.

# Acknowledgement

The author thanks Bayer CropScience LP for the grant it provides to complete this study. Acknowledgment 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> for details).

# References

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

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

3. Shaw, R.; Lark, R.; Williams, A.; Chadwick, D.; Jones, D. 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* **2016**, *230*, 294-306.

4. Paustian, M.; Theuvsen, L. Adoption of precision agriculture technologies by German crop farmers. *Precis Agric* **2017**, *18*, 701-716.

5. Zhang, Y. Towards improved crop growth and yield estimation: observation constrained wheat modelling. Doctoral thesis, Monash University, Victoria, Australia. **2020**, doi:10.26180/13151318.v1.

6. Batchelor, W.D.; Basso, B.; Paz, J.O. Examples of strategies to analyse spatial and temporal yield variability using crop models. *European Journal of Agronomy* **2002**, *18*, 141-158.

7. Mosleh, M.K.; Hassan, Q.K.; Chowdhury, E.H. Application of remote sensors in mapping rice area and forecasting its production: a review. *Sensors (Basel)* **2015**, *15*, 769-791, doi:10.3390/s150100769.

8. 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. *International Journal of Applied Earth Observations and Geoinformation* **2020**.

9. Jordan, C.F. Derivation of leaf‐area index from quality of light on the forest floor. *Ecology* **1969**, *50*, 663-666.

10. Jiang, Z.; Huete, A.R.; Didan, K.; Miura, T. Development of a two-band enhanced vegetation index without a blue band. *Remote Sensing of Environment* **2008**, *112*, 3833-3845, doi:10.1016/j.rse.2008.06.006.

11. Nguy-Robertson, A.; Gitelson, A.; Peng, Y.; Viña, A.; Arkebauer, T.; Rundquist, D. Green leaf area index estimation in maise and soybean: Combining vegetation indices to achieve maximal sensitivity. *Agronomy Journal* **2012**, *104*, 1336-1347.

12. Gitelson, A.A.; Viña, A.; Arkebauer, T.J.; Rundquist, D.C.; Keydan, G.; Leavitt, B. Remote estimation of leaf area index and green leaf biomass in maise canopies. *Geophysical Research Letters* **2003**, *30*, n/a-n/a, doi:10.1029/2002gl016450.

13. Gitelson, A.A.; Vina, A.; Ciganda, V.; Rundquist, D.C.; Arkebauer, T.J. Remote estimation of canopy chlorophyll content in crops. *Geophysical Research Letters* **2005**, *32*.

14. Rouse, J.; Haas, R.; Schell, J.; Deering, D. Monitoring vegetation systems in the Great Plains with ERTS. In Proceedings of the Washington, DC, 1974; pp. 309–317.

15. Haboudane, D.; Miller, J.R.; Pattey, E.; Zarco-Tejada, P.J.; Strachan, I.B. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote sensing of environment* **2004**, *90*, 337-352.

16. Qi, J.; Chehbouni, A.; Huete, A.; Kerr, Y.; Sorooshian, S. A modified soil adjusted vegetation index. *Remote sensing of environment* **1994**, *48*, 119-126.

17. Gitelson, A.A. Wide dynamic range vegetation index for remote quantification of biophysical characteristics of vegetation. *Journal of plant physiology* **2004**, *161*, 165-173.

18. Nguy-Robertson, A.L.; Peng, Y.; Gitelson, A.A.; Arkebauer, T.J.; Pimstein, A.; Herrmann, I.; Karnieli, A.; Rundquist, D.C.; Bonfil, D.J. Estimating green LAI in four crops: Potential of determining optimal spectral bands for a universal algorithm. *Agricultural and forest meteorology* **2014**, *192*, 140-148.

19. Peng, Y.; Gitelson, A.A. Application of chlorophyll-related vegetation indices for remote estimation of maise productivity. *Agricultural and Forest Meteorology* **2011**, *151*, 1267-1276.

20. Rondeaux, G.; Steven, M.; Baret, F. Optimization of soil-adjusted vegetation indices. *Remote sensing of environment* **1996**, *55*, 95-107.

21. Sripada, R.P.; Heiniger, R.W.; White, J.G.; Meijer, A.D. Aerial color infrared photography for determining early in-season nitrogen requirements in corn. *Agronomy Journal* **2006**, *98*, 968-977.

22. Gitelson, A.; Merzlyak, M.N. Spectral reflectance changes associated with autumn senescence of Aesculus hippocastanum L. and Acer platanoides L. leaves. Spectral features and relation to chlorophyll estimation. *Journal of Plant Physiology* **1994**, *143*, 286-292.

23. Roujean, J.-L.; Breon, F.-M. Estimating PAR absorbed by vegetation from bidirectional reflectance measurements. *Remote sensing of Environment* **1995**, *51*, 375-384.

24. Haas, R.; Deering, D.; Rouse Jr, J.; Schell, J. Monitoring vegetation conditions from LANDSAT for use in range management. In Proceedings of the NASA Earth Resources Survey Symposium, NASA. Lyndon B. Johnson Space Center, June 01, 1975, 1975; pp. 43-52.