Estimating crop emergence dates from satellite imagery fusion

Hamilton D and Berryman N

CSBP Fertilisers, Kwinana Beach Road, Kwinana, WA, doug.hamilton@csbp.com.au
CSBP Fertilisers, Kwinana Beach Road, Kwinana, WA, nicholas.berryman@csbp.com.au

Abstract

This project used publicly available satellite imagery datasets from the MODIS sensors (high temporal resolution, low spatial resolution) and Sentinel-2 sensors (moderate temporal resolution, moderate spatial resolution). A common smoothing method was used to fuse the MODIS and Sentinel-2 images based on relative differences to create a daily revisit moderate spatial resolution image timeseries dataset. On a per pixel basis, a double logistical curve was fitted to describe both the growth and senescence phase of the crop. A machine learning model was trained to calculate the baseline level of reflectance per pixel. The baseline and summer weeds were removed from the crop growth curve to determine an emergence date for each using the curve's first derivative. From a dataset of 154 point specific emergence date observations, the image fusion product was able to estimate the emergence date with a mean absolute error of 6 days (±2 weeks at a 97% rate). Scalable and accurate emergence dates can help build more accurate models to predict crop phenology and yield potential as well as inform growers and agronomists of delayed or failed crop emergence.

Keywords

Satellite image fusion, crop emergence, time-series, MODIS, Sentinel-2, smoothing, machine learning

Introduction

The usefulness of satellite images for observing timing critical agronomic parameters, such as crop emergence, is impacted by the satellite sensors' spatial resolution (pixel size in meters) and temporal resolution (revisit - days between images). Higher temporal resolution is important to enable a greater chance of cloud free images, where cloud obscure crops from the satellite sensors. Higher spatial and temporal resolution images are commercially available but can be cost prohibitive for many agricultural applications. In crop modelling, accurate plant emergence data are useful for modelling crop phenology, which is an important parameter for predicting crop yields and identifying potential climate risks to food security (Seidel et al., 2018; Jägermeyr et al., 2021; Liu et al., 2018). Examples of agronomic use cases include establishing spatial extents for re-seeding areas with no crop emergence, or where crop emergence has been staggered across multiple weeks, to treat with zone-based applications of soil wetting agents, herbicides or pesticides.

A common method in literature for emergence-date estimation involves fitting a curve to biomass imagery and performing a threshold operation on its derivative over time (Liu et al., 2023; Zhang & Li, 2021). Confounding factors can limit this approach's use on smaller (paddock/sub-paddock) scales, in particular, clouds, weeds, non-wetting soils and noise in the satellite imagery (Fischer, 1994; Zhao et al., 2021). The algorithm can be divided into a smoothing stage, a curve-fitting stage and an estimate stage. The smoothing stage aims to reduce noise and fill in cloud-affected areas of each image with plausible estimates of the actual biomass index at that time, while the curve-fitting stage aims to remove the effects of weeds from the smoothed imagery. From there a standard start-of-season estimation is performed using the derivative of the fit curve.

This paper will focus on the use of satellite-based imagery products that passively capture optical bands of light in the visible, near infrared and short-wave infrared regions (VNIR/SWIR), with its application on estimating emergence dates of cropping systems in agriculture. It proposes an algorithm to fuse coarse spatial resolution (250 m^2 pixel) daily revisit MODIS (https://modis.gsfc.nasa.gov/) images and moderate resolution (10 m^2 pixel) 5-day revisit Sentinel-2 (S2) (https://sentiwiki.copernicus.eu/web/sentinel-2) images to create a daily revisit 10m^2 pixel data set and test its ability to predict a crops emergence date.

Methods

Experimental design

A dataset of 154 point specific observations per paddock were collected through interviews with farmers across the 2022 and 2023 winter cropping seasons in Western Australia, with data spread across all

AgZones, as defined by the Western Australian Crop Variety Testing's AgZones. Weed germination and growth prior to emergence was not visually assessed on these sites. Based on the longitude and latitude coordinates of these specific points, and the surrounding paddock polygons, a data set of all available MODIS and S2 images from May to November were downloaded for 2022 and 2023.

Data pre-processing

Biomass images were generated using the Enhanced Vegetation Index (EVI2) (Miura et al., 2008) applied to L2A S2 images, and L2G MODIS images (Terra satellite). EVI2 was selected as it could easily be applied to both MODIS and S2. S2Cloudless (Skakun et al., 2022) and the MODIS cloud mask (*Cloud Mask (35_L2) | Atmosphere Discipline Team Imager Products*, n.d.) were applied, respectively, to each image set.

Data smoothing

The smoothing stage first involves assigning each pixel of a S2 biomass image to a MODIS pixel as a "context feature", where coarse local estimates such as S2 paddock average biomass can also be used. A time series of the context feature is smoothed using a common smoothing algorithm. Relative differences between S2 pixels and their assigned context feature is calculated for each day of the timeseries. These relative differences are then smoothed in the same way. In the final phase, the smoothed relative differences are recombined with the smoothed context features to provide the daily timeseries of $10m^2$ paddock images. This method improves on the simple smoothing common in literature (Fischer, 1994; Liu et al., 2023; Zhang & Li, 2021) by ensuring that local trends (as signalled by the context features) are followed, reducing the possibility of over-smoothing in cases where clouds are severe. Smoothing parameters were selected in both cases by visual inspection of a set of image timeseries, not including those presented in the results section. We also present a comparison using paddock averages of S2 imagery as the sole context feature rather than MODIS pixels.

Data curve-fitting

This stage begins by identifying local-minima inflection points within the smoothed image timeseries, where all values prior to that in the timeseries are set to the minimum value of the entire timeseries. This aims to eliminate the confounding effect of weed growth on the emergence estimates. A double-logistic curve (Fischer, 1994) is fitted to the timeseries, where the first date that the derivative of this curve passes a given threshold is taken as the emergence date estimate. Initial parameters for the curve fit were selected using the same method as the smoothing parameters above. The curve is fit to guarantee a consistently shaped rise (logistic rise), as the smoothed timeseries alone may not always follow a perfectly consistent shape, which could confound estimation.

Evaluation

Quantitative evaluation of paddock-wide emergence date accuracy was performed by 'training' an implementation of the above algorithm on a set of 43 paddocks for the year 2023 (March-November), before being used to produce estimates on a test set of 111 paddocks and then recording the error. The training/testing years were then reversed, and the process was repeated. The dataset available did not supply sub-paddock emergence estimates, and so quantitative evaluation could only be performed at a paddock-wide level. Qualitative evaluation was thus utilised for sub-paddock estimates. This involved visual inspection of the image timeseries, with an eye for verisimilitude, limited artefacting, and limited apparent confounder influence (cloud, weed, noise).

Results

Paddock-level accuracy

Table 1: Accuracy of paddock-wide estimation methods (MAE = Mean Absolute Error)

Year (test set)	MAE (days) S2 Only	MAE (days) MODIS Fused	Within 14 days (%) S2 Only	Within 14 days (%) MODIS Fused
2022	5.83	6.18	98%	98%
2023	9.43	9.02	77%	78%

Paddock-level estimates using fused and non-fused imagery appear to perform similarly across both tests (Table 1). The fusion imagery achieved a slightly lower MAE on the 2023 test set, while the non-fused imagery achieved a slightly lower MAE on the 2022 test set. In both cases, accuracy was reasonably high (MAE <1 week in 2022 test, <10 days in 2023 test).

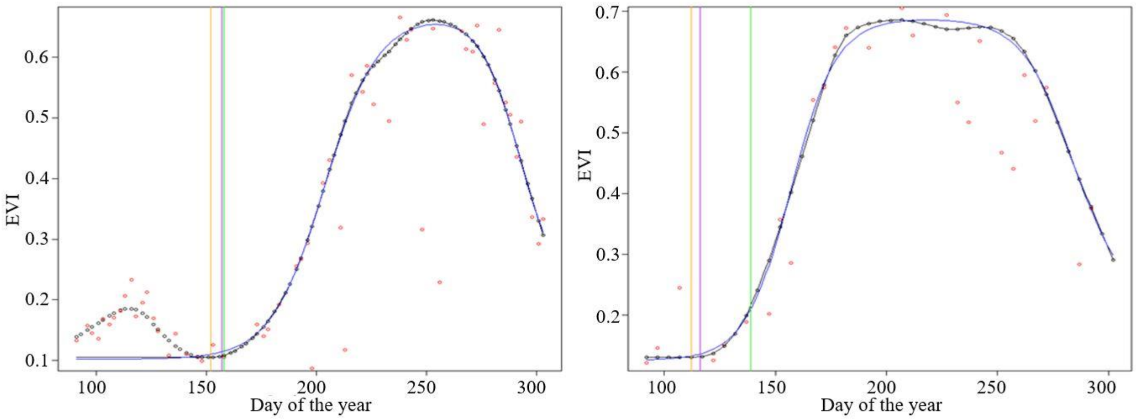

Figure 1: Plot of processed average EVI (Enhanced Vegetation Index) with MODIS fusion applied in smoothing stage across a typical example paddock (left) and an example of a poor estimate of emergence (right). Red dots – raw S2; black dots – smoothed S2; black line – smoothed S2 with weed filtering; blue trendline – double logistic fit curve; Orange vertical line – weed inflection point; blue vertical line – emergence estimate; green vertical line – emergence ground truth.

The smoothing and weed filtering components of the algorithm appear to perform reasonably well, with the large noise and early peaks apparent in the raw S2 imagery being largely eliminated after applying the smoother and weed inflection filter (Figure 1). In at least some circumstances, ground truth emergence dates appear to occur well into apparent biomass growth (Figure 1, right-hand image).

Sub-paddock level accuracy

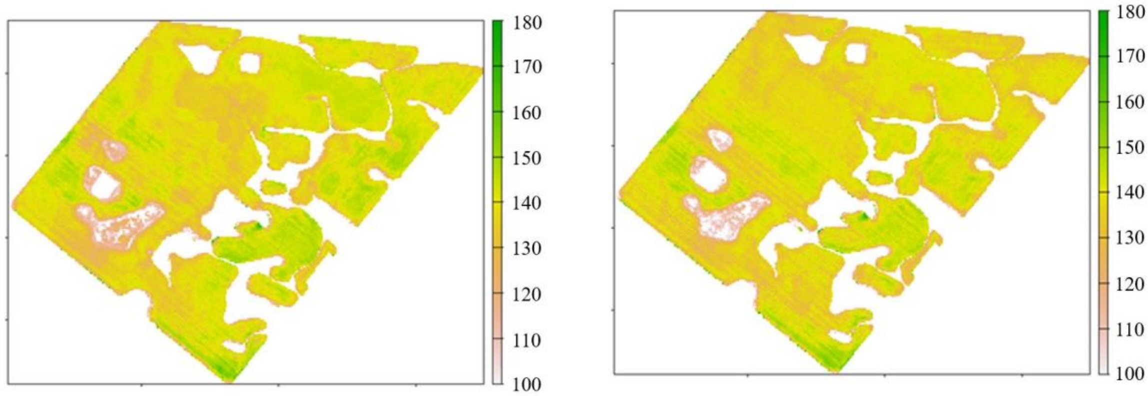

Figure 2: Example image of sub-paddock emergence day of the year estimates. S2 only (left) vs MODIS fused (right). Colour scale is day of year.

In the provided image (Figure 2), both methods identify most emergence occurring within two weeks, with some small areas having estimated emergence within a week earlier or later than that. When compared to each other, differences in emergence can be seen in places where artifacts (top centre of left image) are less prominent in the MODIS fused imagery, along with some difference in other areas such as bottom centre.

Discussion

In paddock-wide emergence estimates, both variations of the algorithm presented in this paper appear to perform well, and comparably so between them. While a small number of possible erroneous ground truth values were identified (based on large discrepancy vs visual analysis of biomass imagery timeseries), these were left in as they could not be undoubtedly confirmed as erroneous. Sub-paddock emergence estimates were unable to be quantified in this paper, though the combination of good results for paddock-wide estimates, and visual inspection of sub-paddock images (for both the smoothed imagery and emergence estimates) suggests the strong possibility of this algorithms usefulness in this task, with MODIS imagery fusion potentially providing an improvement. Further work in this area would likely find fertile ground in replicating this work with a larger dataset across more years, introducing additional information to the estimation model such as crop type or soil type, and by quantitatively analysing the accuracy of the sub-paddock estimates. It would also be of interest to analyse the accuracy of the sub-paddock smoothing and interpolation component of this algorithm behind fake cloud masks, as described by (Zhao et al., 2021).

Furthermore, while this paper made use of seasonal retrospective biomass imagery, it could also be applied to in-season emergence estimate.

Conclusion

This paper presents a method to combine imagery products from two different satellite sensors with different spatial and temporal resolution, with the purpose of being able to create an earth observation image product more suitable for crop monitoring purposes of extensive and low margin cropping systems. The results of using this method to estimate crop emergence dates was shown to provide reasonably accurate estimates, with MODIS fusion potentially improving sub-paddock accuracy. Being able to more accurately estimate the scale, spatial extent and fine resolution differences in crop emergence dates can help improve sub-paddock modelling of crop phenology and subsequently yield potential of the crop, which is an important driver of demand for crop inputs such as nitrogen. Being able to provide this information back to farmers within an appropriate time frame will help to improve farm profitability and sustainability.

References

Seidel SJ, Palosuo T, Thorburn P and Wallach D (2018). Towards Improved Calibration of Crop Models-Where Are We Now and Where Should We Go? European Journal of Agronomy 94, 25–35.

Jägermeyr J, Müller C, Ruane AC, Elliott J, Balkovic J, Castillo O, Faye B, Foster I, Folberth C and Franke JA (2021). Climate Impacts on Global Agriculture Emerge Earlier in New Generation of Climate and Crop Models. Nature Food *2*, 873–885.

Liu L, Wallach D, Li J, Liu B, Zhang L, Tang L, Zhang Y, Qiu X, Cao W and Zhu Y (2018). Uncertainty in Wheat Phenology Simulation Induced by Cultivar Parameterization under Climate Warming. European Journal of Agronomy 94, 46–53.

Liu, X., Chen, Y., Li, Z., & Li, Y. (2023). Evaluating the Consistency of Vegetation Phenological Parameters in the Northern Hemisphere from 1982 to 2015. Remote Sensing, 15(10), Article 10. https://doi.org/10.3390/rs15102559

Zhang, Y., & Li, M. (2021). A new method for monitoring start of season (SOS) of forest based on multisource remote sensing. International Journal of Applied Earth Observation and Geoinformation, 104, 102556. https://doi.org/10.1016/j.jag.2021.102556

Fischer, A. (1994). A model for the seasonal variations of vegetation indices in coarse resolution data and its inversion to extract crop parameters. Remote Sensing of Environment, 48(2), 220–230. https://doi.org/10.1016/0034-4257(94)90143-0

Zhao, M., Olsen, P. A., & Chandra, R. (2021). Seeing Through Clouds in Satellite Images (arXiv:2106.08408). arXiv. https://doi.org/10.48550/arXiv.2106.08408

Miura, T., Yoshioka, H., Fujiwara, K., & Yamamoto, H. (2008). Inter-Comparison of ASTER and MODIS Surface Reflectance and Vegetation Index Products for Synergistic Applications to Natural Resource Monitoring. Sensors (Basel, Switzerland), 8(4), 2480–2499. https://doi.org/10.3390/s8042480

Skakun, S., Wevers, J., Brockmann, C., Doxani, G., Aleksandrov, M., Batič, M., Frantz, D., Gascon, F., Gómez-Chova, L., Hagolle, O., López-Puigdollers, D., Louis, J., Lubej, M., Mateo-García, G., Osman, J., Peressutti, D., Pflug, B., Puc, J., Richter, R., … Žust, L. (2022). Cloud Mask Intercomparison eXercise (CMIX): An evaluation of cloud masking algorithms for Landsat 8 and Sentinel-2. Remote Sensing of Environment, 274, 112990. https://doi.org/10.1016/j.rse.2022.112990

Cloud Mask (35_L2) | Atmosphere Discipline Team Imager Products. (n.d.). Retrieved July 2, 2024, from https://atmosphere-imager.gsfc.nasa.gov/products/cloud-mask

Fischer J, Smith R and Jones CR (2004). Old directions for a new planet. In: Proceedings of the 10th World Fishing Congress, 20 September - 1 October 2004. Timbuktu, Mali. (https://www.google.com.au/). http://dx.doi.org/10.1016/j.agee.2009.08.014).