



What standard error metrics don't tell us about model performance – a fractional cover use case

Watson, F¹; Denham, R²

¹ Queensland Department of Environment, Science and Innovation; ² Joint Remote Sensing Research Program.

Key words: spatial prediction models; remote sensing; fractional cover

Abstract

Environmental monitoring by land managers and land administrators using remote sensing has entered the era of big data, machine learning and open data. Remotely sensed data are commonly used in models as predictors for machine learning algorithms trained on field observations, to develop new products and applications as more data and greater computing power become available.

Despite the wealth of remote sensing data, it is often difficult to obtain representative reference data for these applications, and complete coverage of the predictor space is rare. Users are left to assume that the model outputs will apply to regions where reference data are unavailable. This can result in situations where stated accuracy and error metrics show good model performance, however at a local and possibly regional scales, the map products may not be representative of the true state and therefore not serve its intended purpose. Users of new remote sensing products and applications need to be aware of the uncertainty inherent in the products they use.

In this paper, we present an approach to communicate uncertainty by adding a spatial component to performance metrics by applying the 'Area of Applicability' (AOA) of spatial prediction models, to the Joint Remote Sensing Research Program's Fractional Cover 3 (FC3) product.

The FC3 product is widely used in research and applied settings to monitor vegetation cover and bare ground and inform other models such as pasture biomass or land condition across Australia's rangelands. It is imperative that we continue to improve our understanding of fractional cover models, and their strengths and limitations, to provide appropriate advice and direction to users who are reliant on these data. It can also help to inform future investments in field data collection or other methods of training data collection.

Introduction

The geographic coverage and repeatability of remotely sensed monitoring approaches are useful characteristics to overcome the challenge of monitoring Australia's rangelands, which cover more than 80% of the continent (ACRIS, 2008). The Joint Remote Sensing Research Program Fractional Cover 3 (FC3)

product is one such approach which models the percentage of bare ground, photosynthetic vegetation and non-photosynthetic vegetation for each pixel. The model is built upon a collaboratively sourced database of field observations gathered by stakeholders across Australia, mostly within the rangelands.

As for many spatial prediction models, these training data may not meet the ideal standard of representativeness, due to practical constraints over its collection. This results in increased uncertainty in environments that are unsampled or under sampled, which is difficult to communicate to users of the product. The Area of Applicability (AOA) approach of Meyer and Pebesma (2021) is a useful tool that can enable us to better understand the limitations of the FC3 product, providing a method to assess the area to which the model can be reliably applied, and conversely providing insight into those locations which are not well represented. This also extends our ability to communicate uncertainty by adding a spatial component to performance metrics.

Methods

For this example, the methodology of Meyer and Pebesma (2021) was applied, with minor variations. Rather than assessing the AOA in-situ during model fitting we have performed a post-hoc assessment of an existing model by replicating as closely as possible the original model fitting procedure.

Two datasets are used in this example. The first is the training data for the FC3 model, consisting of approximately 4000 field observations of fractional cover collected across Australia between 1997 to 2018 following the method described in Muir et al. (2011). These have been matched to the nearest date, clear Landsat 5 TM, 7 ETM+ or 8 OLI overpass, and the surface reflectance values calculated using the procedures described in Flood et al. (2013), for a 3x3 pixel window. The second is the remotely sensed imagery for which fractional cover predictions can be made, which were Landsat seasonal composite reflectance mosaics for Australia, produced using the methods described in Flood et al. (2013), and Flood (2013), between winter 2014 and winter 2024 (DES, 2021-2024).

The AOA concept attempts to define the area in predictive space for which the model's performance is relevant. That is, the published model performance metrics are only appropriate when a location's characteristics are similar to the locations used to train the model. For each new location the Euclidean distance in predictor space to each of the training data points is calculated. Further, the distances are weighted by the model variable importance scores, under the assumption that locations further away in an important variable should be considered more serious than locations further away in a less important variable. Here, weights were based on SHAP (SHapley Additive exPlanations) scores (Lundberg & Lee 2017). The dissimilarity index (DI) is defined as the minimum of these distances. Those locations in which the DI is greater than a threshold are defined to be beyond the AOA. The threshold is determined by examining the distribution of the DI calculated in the training data under cross-validation. We follow Meyer and Pebesma's (2021) suggestion of a threshold set at the 75-percentile plus 1.5 times the interquartile range (IQR) of the DI values of the cross-validated training data.

Since the distribution of DI are based on cross-validation, the method is sensitive to the type of cross-validation used. There is some evidence that, when training points are spatially correlated, the common k -fold cross-validation leads to overly optimistic results. Conversely, spatial cross-validation, such as block cross-validation, has been found to be overly pessimistic. To account for the spatial structure in the sampling data but avoid under-estimating performance, we chose to use k -fold nearest-neighbour distance matching cross-validation (k NNDM) (Linnenbrink et al 2023).

A sequence of seasonal (four per year) surface reflectance composites between 2014 and 2024 were used as a proxy for representative conditions for which we would likely require predictions. A corresponding DI and AOA map was derived from each one and the full set of 40 seasonal AOA maps then summarised as a frequency map representing the proportion of seasons that a location was outside the AOA.

Results

The results of the kNNNDM cross validation are shown in Table 1, for the three cover fractions which the FC3 model predicts. The DI threshold to define AOA was calculated using the same cross-validation folds and found to be 0.128. This threshold was applied to the DI images computed from seasonal composites between winter 2014 and winter 2024, and the AOA summary map shown in Figure 1 produced.

Table 1: kNNNDM Cross-validation results.

	Bare (% Cover)	Photosynthetic Vegetation (% Cover)	Non-photosynthetic Vegetation (% Cover)
Root Mean Square Error (RMSE)	13.1	10.0	15.4
Mean Absolute Error (MAE)	9.6	7	11.6

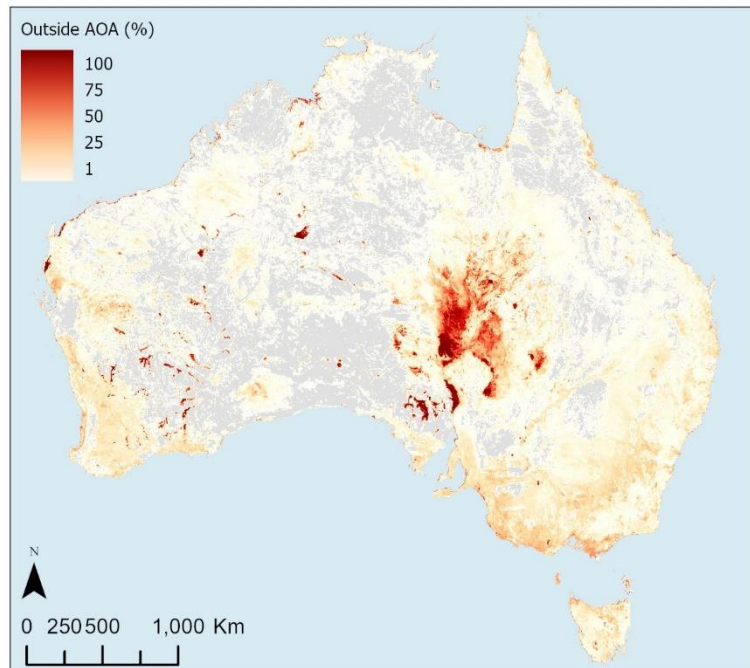


Figure 1 Percentage of seasons in which the surface reflectance was found to be outside the AOA of the FC3 model, 2014 to 2024. Locations that are always within the AOA are grey.

On average less than 10% of Australia is outside the AOA, and less than 5% is frequently outside the AOA (i.e. >50% of the time). Based on visual inspection of the summary map, the areas frequently outside the AOA often appear to be locations where either bare ground or high green vegetation cover dominate, or where cover is highly dynamic, such as the central deserts and salt lake regions, coastal areas and cropping or modified pastures across south-eastern Australia and south-western Western Australia. The bare ground category represents a variety of features including beaches, mudflats, claypans, saltpans, rocky areas and soil of diverse geologies, presumably spectrally distinct from other 'bare ground' locations sampled within the model training data. The high green vegetation cover category is unsurprising, as field site selection has mostly targeted rangeland environments dominated by open native vegetation as per the original intended use of this product. While Australian rangelands can have brief periods of high green cover, these are poorly sampled due to access constraints.

Additionally, all the major cropping regions of Australia show a moderate percentage of seasons outside AOA. This is also unsurprising given that cropping areas may cycle through bare and high green cover states, and in some locations soil moisture from irrigation may also modify the spectral response. Additionally, calibration sites within cropped locations represent a small component of the training data. Further investigation could explore any relationship between specific stages of the crop cycle and DI.

Figure 2 shows exemplar locations outside the AOA. The upper images a) and c) show the DI images for a cropped field and an area of bare soil respectively. The corresponding plots b) and d) below illustrate the spectral space covered by the training data (shown in pink) and the exemplar yellow polygons (shown in blue), as pairwise relationships between each band used for model fitting. A density distribution plot for each band is also shown. The limited overlap between the two sets of points in each example shows that the training data lacks coverage for these features, and that the AOA approach has yielded sensible and interpretable results.

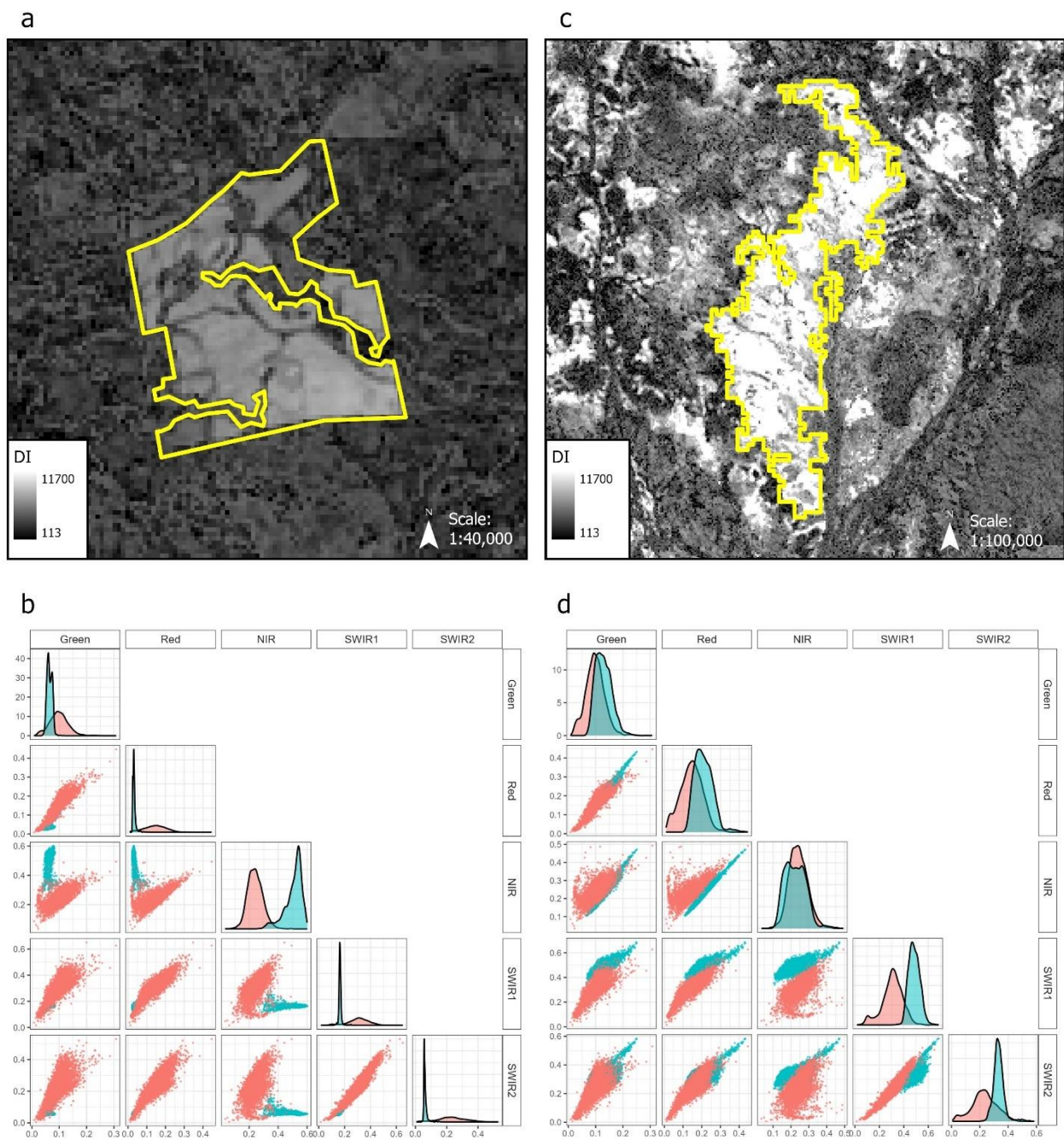


Figure 2 Examples of locations frequently (> 80%) outside AOA (yellow polygon), and comparison of their reflectance with the model training data: (a) DI image for a cropped field in north Queensland; (b) pairs plot of reflectance values for training data (pink shade) and pixels within the yellow polygon in (a) (blue shade); (c) DI image for bare soil in the Channel Country of far western Queensland; and (d) pairs plot of reflectance values for training data (pink shade) and pixels within the yellow polygon in (c) (blue shade).

Discussion

The AOA frequency product presented, provides an aid to users to assess the suitability of the FC3 product for their purposes. Areas which are frequently outside the AOA should be treated cautiously, as the error metrics provided for the product are not applicable here. It should be noted that this does not mean that the model is wrong in these locations, rather that we are unable to say anything about the likely performance in these areas.

The application of the AOA methodology to the FC3 model has proven to be a useful tool for understanding the representativeness of the underlying training data and the corresponding limitations imposed on model performance. It extends the information derived from typical performance or error metrics by giving a spatially explicit extent to which those metrics apply and identifying locations that are not represented by the training data and for which model performance is therefore unknown.

In this case we have shown that while the training data are relatively sparse for a model applied across such a broad geographic area, most of Australia is within the AOA. It is hoped that by making the AOA summary layer publicly available, the diverse users of the FC3 products can better understand how the limitations of the model may impact their intended uses of the product.

Given the FC3 model is applied in an ongoing operational context, an additional insight gained through this work is knowledge about where new calibration field data could be collected, to expand the AOA for future model iterations.

Acknowledgements

All Landsat imagery used in this study courtesy of NASA Goddard Space Flight Center and U.S. Geological Survey.

References

- Bastin, G. & National Land and Water & Australian Collaborative Rangeland Information System. (2008) Rangelands 2008: taking the pulse. Canberra: National Land & Water Resources Audit <http://lwa.gov.au/products/PN21387>
- Department of Environment and Science (2017-2023) Queensland Government (2021): Seasonal surface reflectance - Landsat, JRSRP algorithm, Australia Coverage. Version 1.0. Terrestrial Ecosystem Research Network. (Dataset). <https://portal.tern.org.au/metadata/TERN/5a31eed4-e43a-404d-b534-3f820305ed61>
- Flood, N. (2013) Seasonal Composite Landsat TM/ETM+ Images Using the Medoid (a Multi-dimensional Median). *Remote Sensing* 5(12), 6481-6500.
- Flood, N. Danaher, T. Gill, T. Gillingham, S. (2013) An Operational Scheme for Deriving Standardised Surface Reflectance from Landsat TM/ETM+ and SPOT HRG Imagery for Eastern Australia. *Remote Sensing* 5, 83-109. <https://doi.org/10.3390/rs5010083>
- Linnenbrink J, Milà, C, Ludwig, M, and Meyer, H (2024) kNNDM CV: k-fold nearest-neighbour distance matching cross-validation for map accuracy estimation. *Geoscientific Model Development* 17, 5897–5912, <https://doi.org/10.5194/gmd-17-5897-2024>.
- Meyer H, Ludwig L, Milà C, Linnenbrink J, Schumacher F. (2024) The CAST package for training and assessment of spatial prediction models in R. arXiv <https://doi.org/10.48550/arXiv.2404.06978>.
- Lundberg SM and Lee, S (2017) A Unified Approach to Interpreting Model Predictions. In ‘Advances in Neural Information Processing Systems 30 (NIPS 2017)’. (Curran Associated, Inc.)
- Meyer H, Pebesma E (2021) Predicting into unknown space? Estimating the area of applicability of spatial prediction models. *Methods in Ecology and Evolution* 12(9), 1620-1633.
- Muir J, Schmidt M, Tindall D, Trevithick R, Scarth P, Stewart J. (2011). Guidelines for field measurement of fractional ground cover: a technical handbook supporting the Australian collaborative land use and management program. Tech. rep., Queensland Department of Environment and Resource Management for the Australian Bureau of Agricultural and Resource Economics and Sciences, Canberra, ACT.