



HAL
open science

Using remote sensing, process-based crop models, and machine learning to evaluate crop rotations across 20 million hectares in Western Australia

Roger Lawes, Gonzalo Mata, Jonathan Richetti, Andrew Fletcher, Chris Herrmann

► **To cite this version:**

Roger Lawes, Gonzalo Mata, Jonathan Richetti, Andrew Fletcher, Chris Herrmann. Using remote sensing, process-based crop models, and machine learning to evaluate crop rotations across 20 million hectares in Western Australia. *Agronomy for Sustainable Development*, 2022, 42 (6), pp.120. 10.1007/s13593-022-00851-y . hal-04349554

HAL Id: hal-04349554

<https://hal.science/hal-04349554>

Submitted on 18 Dec 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Using remote sensing, process-based crop models, and machine learning to evaluate crop rotations across 20 million hectares in Western Australia

Roger Lawes¹ · Gonzalo Mata¹ · Jonathan Richetti¹ · Andrew Fletcher¹ · Chris Herrmann¹

Accepted: 17 November 2022 / Published online: 12 December 2022
© Crown 2022

Abstract

Remote sensing has been widely employed to identify crop types and monitor crop yields on farms. Here, we combine successive seasons of these products to identify crop rotations in each field across 20 million hectares of the Western Australian Wheatbelt. We used the APSIM crop model to define the starting soil water, temperature stresses, biomass, and crop yield to characterize the prevailing agro-environment of that field. These remote sensing data and APSIM crop modeling outputs were then combined, with machine learning, to predict the effect of the complex interaction between agro-environment and crop rotation on wheat yield. Predictions from machine learning are employed to evaluate the benefits or otherwise of crop rotation across Western Australia for every field in the study region. In general, if fields subjected to a wheat-cereal rotation were instead subjected to a wheat-canola rotation, then 68% of these fields were predicted to experience a yield increase of between 0 and 1850 kg ha⁻¹. However, only 28% of fields planted to canola were predicted to have a yield benefit of 200 kg ha⁻¹ or more on the following wheat crops. On average, annual pastures generated a slight yield penalty of 47 kg ha⁻¹ to the following wheat crop. The findings from this study, using crop models, remote sensing, and machine learning, indicate that the benefits of break crops and pastures to farmers is less than the 400 to 600 kg ha⁻¹ benefit commonly reported from field experiments. These management insights could underpin the development of future decision aids or agricultural digital twins for crop management decisions such as crop rotation planning. The approach provides farmers with tangible insights about their production using outputs from crop-based remote sensing and crop modeling.

Keywords Crop modeling · Remote sensing · APSIM · Crop rotation · Machine learning · Digital twin

1 Introduction

Crop rotation is widely used by farmers all around the globe. Break crops are grown in between main crops to break weed, disease, and pest cycles and provide a yield boost to the subsequent main crop. In the context of our study, cereals are considered the main crop and broadleaf crops or pastures are the break crops. The additional yield benefit to a cereal crop following a break crop is often referred to as the ‘break crop effect’ (Kirkegaard et al. 2008). The biological mechanisms that influence this break crop effect vary from one farming system to the next. For example, in low rainfall

North American wheat/legume systems, break crops may increase wheat yields through a reduction in pests (Lensen et al. 2013) or an increase in the nitrogen supply (Chen et al. 2012). In water-limited cropping regions, break crops may decrease wheat yields if the break crop uses additional soil moisture and reduces the amount of water available for the cereal crop (Chen et al. 2012). Reviews of break crop effects and the impact of crop rotation on cereal production by Kirkegaard et al. (2008) and Angus et al. (2015) reveal that the size of the break crop effect varies considerably and negative impacts can occur. In Australian conditions, break crop benefits to cereals can equate to 0.6 t ha⁻¹ yield benefit for cereal crops in Western Australia (Seymour et al. 2012) and 0.7 t ha⁻¹ elsewhere (Angus et al. 2015). The size of these benefits varies with season, soil type, and the particular break crop grown.

The decision to grow a break crop by the farmer is complex and depends on the size of the break crop effect, the

✉ Roger Lawes
roger.lawes@csiro.au

¹ CSIRO, 147 Underwood Avenue, Floreat, WA 6014, Australia

relative yield, and profitability of the break crop (Fletcher 2019). Information extracted from agronomic experiments may not reflect farmer management practices and may not provide farmers with the necessary insights that apply to their farm (Lacoste et al. 2022). Researchers may not know if their trials and findings are representative of farmer practice, while differences between survey findings about the benefits of break crops and experiments conducted by researchers have been noted in Western Australia (Harries et al. 2015, 2022). Therefore, insights about the agronomic benefit of a break crop to the following cereal crop can vary with the season, soil type, water availability, extent of underlying biotic stress, and soil nutrient status. It can be difficult for an individual farmer to determine what impact a break crop might have on the yield of a subsequent cereal crop. Previous surveys of farm practice about break crops found that farmers grew considerably fewer break crops than was economically optimal (Robertson et al. 2010). One possible reason for this difference is because farmers may not achieve the yield gains commonly reported in the agronomic literature about the benefits of break crops on cereal production. However, recent developments in crop monitoring and crop modeling may provide farmers with the tools needed to quantify these benefits. Crop monitoring and crop modeling have evolved to a point where it is possible to conduct landscape scale evaluations of crop rotations.

Crop species, or Crop ID technology, derived from satellite classifications of crops, have been developed to track the global food supply and monitor annual plantings across most agricultural areas. Typically, people use Sentinel-1 and Sentinel-2 satellites from the European Space Agency or Landsat 8 and Landsat 9 satellites from NASA to monitor the area sown to the significant crops (Adjemian 2012; Fritz et al. 2019; Velde et al. 2019). Local organizations finesse these technologies to cope with cloud (Shendryk et al. 2019; Magno et al. 2021), unbalanced training data (Waldner et al. 2019), small fields and irregular shapes (Burke and Lobell 2017; Lebourgeois et al. 2017), or double cropping systems (Paludo et al. 2020).

Like Crop ID, crop yield can now be monitored at scale. Methods, driven by Gross Primary Productivity (GPP) and pioneered by Reeves et al. (2005), are now available globally (Jaafar and Mourad 2021) through Google Earth Engine. Adaptations of these approaches to estimate crop yield with GPP have been developed for Australia (Chen et al. 2020; Donohue et al. 2018), Africa (Wellington et al. 2022), and China (Yan et al. 2022). Each method to estimate crop yield via satellite incorporates the local effects of climate and terrain to varying degrees. Local models may be calibrated with experimental data, data from farmers' fields or from government surveys, and typically predict crop yields with a root square mean error (RMSE) in the order of 0.6 t ha^{-1} for cereal crops.

Finally, crop modeling platforms, like The Agricultural Production Systems sIMulator (APSIM), have been used to create continental scale simulations and draw on local climate grids and soil type grids to generate estimates of crop yield potential (Hochman et al. 2016). Often researchers estimate the management to drive continental scale simulations. Local studies that collect farm data have been able to simulate farmer yields with some precision, but the data collection process is often time-consuming and is rarely repeated (Lawes et al. 2021; Lobell et al. 2005; Mourtzinis et al. 2018). Traditionally, process-driven crop models would be used to evaluate rotational components of the cropping system, following extensive, site-specific parameterization. For example, Araya et al. (2017) used the Decision Support System for Agrotechnology Transfer (DSSAT) to evaluate the benefit of rotation on long-term sorghum production in Kansas. Similarly, APSIM was used to model corn yield responses to crop rotation in Iowa (Puntel et al. 2016). However, crop rotations are not usually modeled at scale, in part because the permutations become intractable, and detailed site-specific data are not available.

Outputs from three modeling paradigms, satellite crop yield models, process crop models, and land use mapping with remote sensing may be combined to provide insight into the rotational benefits on cereal crops, at scale, and across seasons. The objective is to create crop intelligence and agronomic insight without exhaustive and expensive field surveys, the need to collect data directly from farmers, or conduct local and expensive small plot experiments. We describe how we combine multiple data products, created with earth observation technologies, earth observation crop models, process-based crop models, and big data analytics to evaluate the importance of crop rotations across 20 million hectares in Western Australia. We combine these approaches to test the hypothesis that farmers do achieve break crop benefits that are comparable to those achieved in research trials and research surveys. Earth observation is firstly used to predict the outcome of different management decision and then secondly to evaluate the value of crop rotations for every field at a regional scale. We argue that outputs from this suite of information products and big data analytics could provide the precursor to inputs for agricultural decision support systems or agricultural digital twins.

2 Methods

2.1 Study area

The study area comprises the wheatbelt region of Western Australia (WA), which occupies 20 million hectares in the Southwest of WA (Fig. 1). Mixed crop and livestock systems, as well as continuous cropping systems, are widely

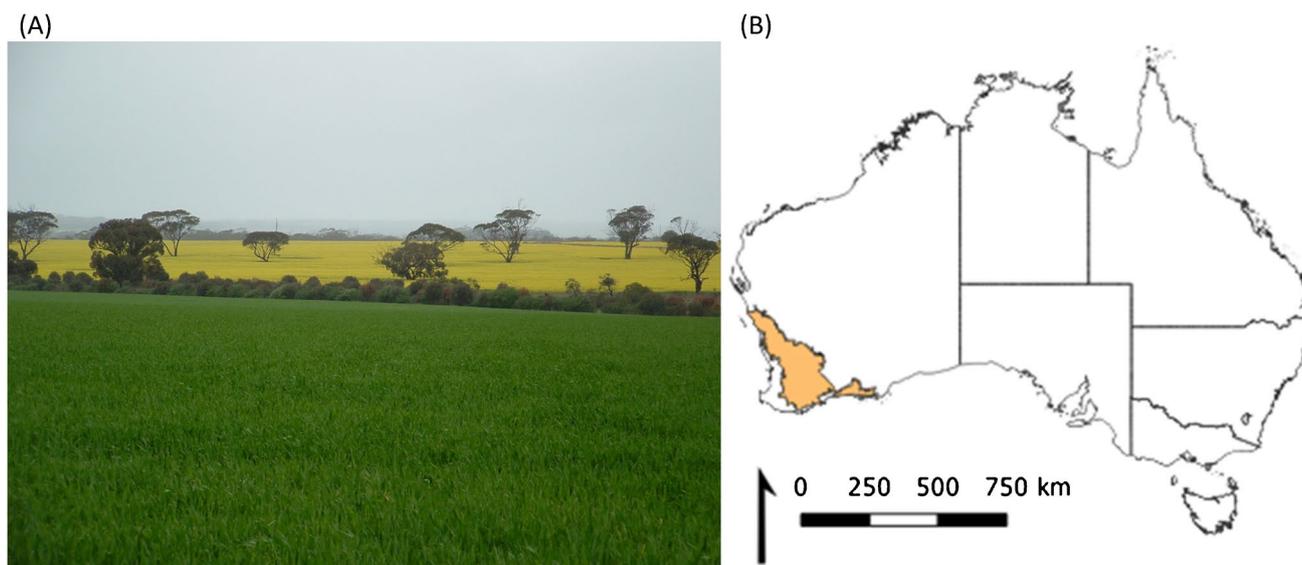


Fig. 1 **a** A typical Western Australian cropping landscape, with wheat in the foreground, and canola planted behind the wheat crop. **b** The location of the Western Australian Wheatbelt.

implemented across this region (Harries et al. 2022). Soils are old, weathered, and of light texture, while the climate is typically Mediterranean. That is, summers are typically hot and dry, and winters are cold and wet. Crops and annual pastures are sown in the Autumn (April/May) and harvested in late Spring or Summer (November/December). Soils are typically the lighter, sandy Tenosols, or the sandy loam Kandosols. Clays may also be classified as Vertosols or Dermosols, although these soil types are less common (Isbell 2016). The plant available water holding capacity across Western Australia ranges from 30 mm to 150 mm (Oliver and Robertson 2009). Across the Western Australian Wheatbelt, winter dominant rainfall drives grain production (Ludwig et al. 2009; Turner and Asseng 2005). Growing season rainfall generally defines an upper limit to cereal grain production (French and Schultz 1984) and averaged 260 mm from 1900 to 2016 across the entire region (Fletcher et al. 2020). The spatial distribution of growing season rainfall and thermal time is presented in Fig. 2.

2.2 Data description: field boundaries, crop identification, and crop rotation

Field polygons, to create field boundaries, were defined for the Western Australian Wheatbelt using a process of semantic segmentation and image processing of Sentinel-2 imagery. Specific details of this process were defined by Waldner and Diakogiannis (2020). Field polygons for the entire wheatbelt, known as ePaddocks, are available at the following website (<https://agdatashop.csiro.au/epaddock-australian-paddock-boundaries>). In total, 301,209 fields were

identified, equating to an arable farmed area of 13,980,729 ha. The total area of 20 million hectares relates to the entire land mass and includes all land uses.

The land use relating to the crop and pasture identification, for each field, was defined annually from 2017 to 2020. Each season, training data were collected across the Western Australian Wheatbelt, and Landsat 8 and Sentinel-1 imagery were used to create a crop identification (Crop ID) for every field in the Western Australian Wheatbelt (Table 1). Details about the acquisition of training data are provided in Lawes et al. (2022). A Random Forest Classifier (Breiman, 2001) was used to create the Crop ID classes that included the crop types of wheat, barley, oats, legume crops, or pasture. Classifications were generated on a per pixel basis of 25 m². Each field was classified as a particular crop class, based on the most numerous pixel present within the field boundary. Classification accuracies for each year ranged from 74% to 77% across all land use types. Further details about the crop classification processes are described by Fowler et al. (2020). From the CROP ID classifications, both the current crop and the previous crop could be defined, and a two-year crop rotation deduced.

2.3 Aggregating the classification and visualizing classification output

To assist with the visualization of the geographic spread of various crop choices, crop classifications were aggregated onto a 20 km grid to illustrate how species choice varied across the Western Australian Wheatbelt. For each crop, the percentage of crop area occupied for that grid square

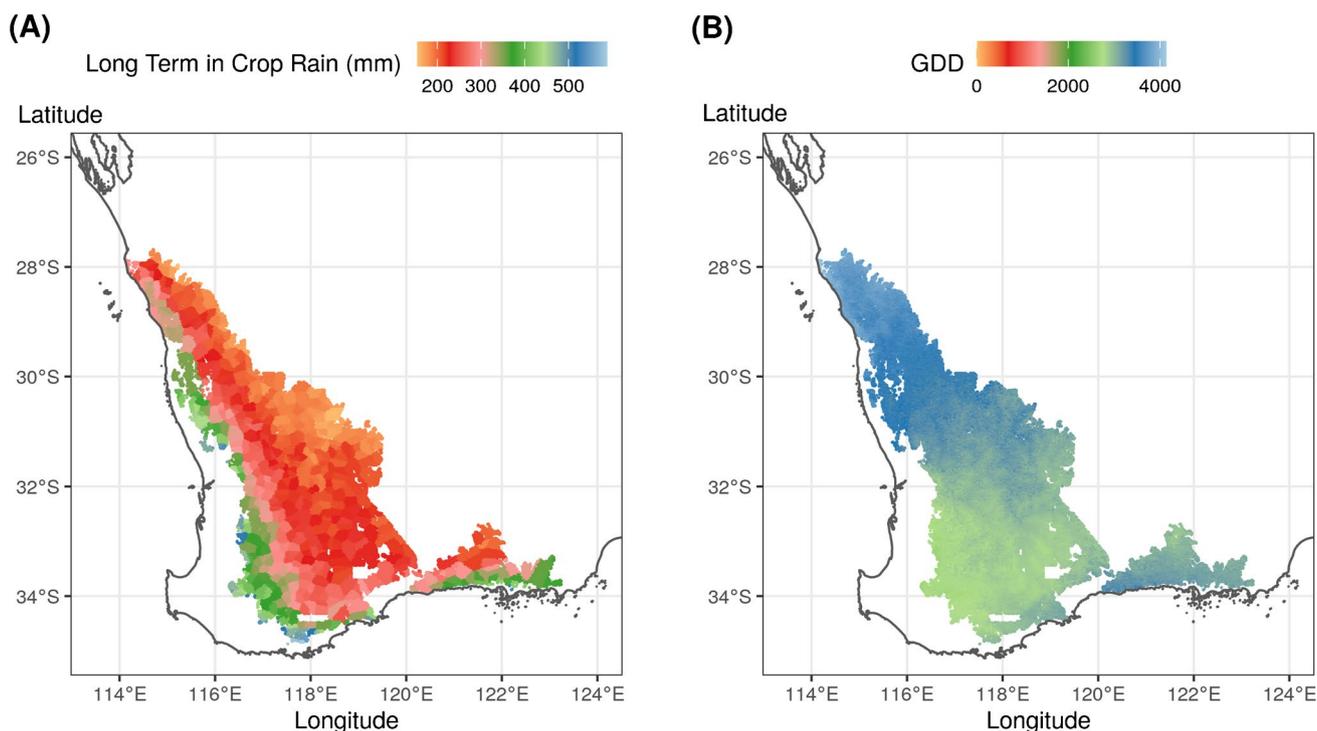


Fig. 2 The spatial distribution of growing season rainfall and cumulative thermal time across the wheatbelt of Western Australia.

Table 1 Landsat 8 paths and rows and Sentinel-1 orbits for the Western Australian Wheatbelt used for crop classification.

Landsat 8 pathrow	Sentinel-1 orbit
Source: USGS	Source: European Space Agency
Resolution: 25 m	Resolution: 10 m
Purpose: crop classification	Purpose: crop classification
108/83–108/84	017
109/83–108/84	046
110/82–110/84	061
111/81–111/84	090
112/80–112/84	119
113/80–113/82	163
114/79–114/80	

was calculated. This value was calculated in each year and then averaged across the four seasons. The Simpsons diversity index was applied to these data to illustrate where crop diversity was greatest and where it was least, across the Western Australian Wheatbelt (Eq. 1). The Simpsons diversity index operates between 0 and 1, and grid squares approaching 1 have a more diverse suite of crops than those where the diversity index approaches zero. A zero index is a monoculture, and diversity indices have been used to evaluate crop diversity in the USA (Larsen and Noack 2017). These outputs on the 20 km grid were used to highlight regions of the wheatbelt with high and low diversity that

may relate to spatial variation in break crop effects. These aggregated data were not part of the machine learning process. The machine learning process was conducted at the field level.

$$D = 1 - \left(\frac{\sum n(n-1)}{N(N+1)} \right) \tag{1}$$

D is the Simpsons Diversity Index, which ranges between 0 and 1 for a 20 km grid square. *n* is the total area, in hectares of a particular crop or pasture enterprise for a 20 km grid square. *N* is the total area, in hectares, of all enterprises for a 20 km grid square.

2.4 Crop yield estimation with the C-Crop model

Crop yield for each wheat field was estimated with the C-Crop model from 2017 to 2020 (Donohue et al. 2018). This satellite-driven crop model, which uses GPP to estimate carbon and then wheat yield, has been developed for Australian conditions. It estimates wheat yield at the field scale with an *r*² of 0.72 (Donohue et al. 2018) and provides a means of generating vast quantities of yield data, suitable for machine learning applications. Output from the C-Crop model is used as the dependent variable to explore the effect of crop rotation on wheat yield. Actual yields from farmers’ fields were not available, and this remotely sensed estimate of crop yield

was the only variable that could be accessed across 20 million hectares easily to provide an estimate of the yield for a particular field. However, the method employed is repeatable and provides consistent insight across years and locations.

2.5 Crop simulation modeling with APSIM, to define the agro-environment

The APSIM crop model was used to define the agro-environment and wheat yield potential for each field. The agro-environment relates to the plant available water holding capacity of the soil (PAWC), plant available water at sowing, plant available water at harvest, and the hot and cold extremes in temperature. Variables relating to climate, such as growing season rainfall (April to October), growing degree days (season length), soil water status, and the yield potential and total above ground biomass, were output from APSIM. Here, yield potential is defined as the water limited yield potential of the crop. The nitrogen supply is unlimited and is comparable to other definitions of water limited yield potential as used by Hochman et al. (2016) and Lawes et al. (2021) for the Australian continent. To create these variables with APSIM, the centroid of each field (from ePaddocks) was used to locate the soil type from the national ASRIS soil grid and the nearest climate file from the Australian Bureau of Meteorology available from the SILO gridded dataset on a 0.05-degree grid. The ASRIS grid estimates soil type for the Australian continent at a 90 m × 90 m resolution (Grundy et al. 2015). An APSIM wheat simulation was generated for every field for every season, following the methods of Hochman et al. (2016).

Output from the APSIM model, crop rotation, and climate information are defined as the independent variables

to predict wheat crop yield, derived from the C-Crop model. The details of each attribute are described in Table 2.

A summary of these variables and the analytical workflow used to determine the importance of crop rotation in the Western Australian Wheatbelt is provided in Fig. 3.

Wheat crop yield, from C-Crop, was then predicted from the variables in Table 2 using a Random Forest. A recursive feature elimination (RFE) was then used to optimize feature selection and identify the most important predictors (Kuhn 2008). The RFE is a backward selection method where the predictors of any given model are ranked. The least important predictors are sequentially eliminated as part of the RFE process that is applied to the base Random Forest algorithm. The Random Forest builds many decision trees via bagging, in which each tree samples the dataset randomly with replacement. The algorithm's final prediction is produced by averaging the estimates of each decision tree.

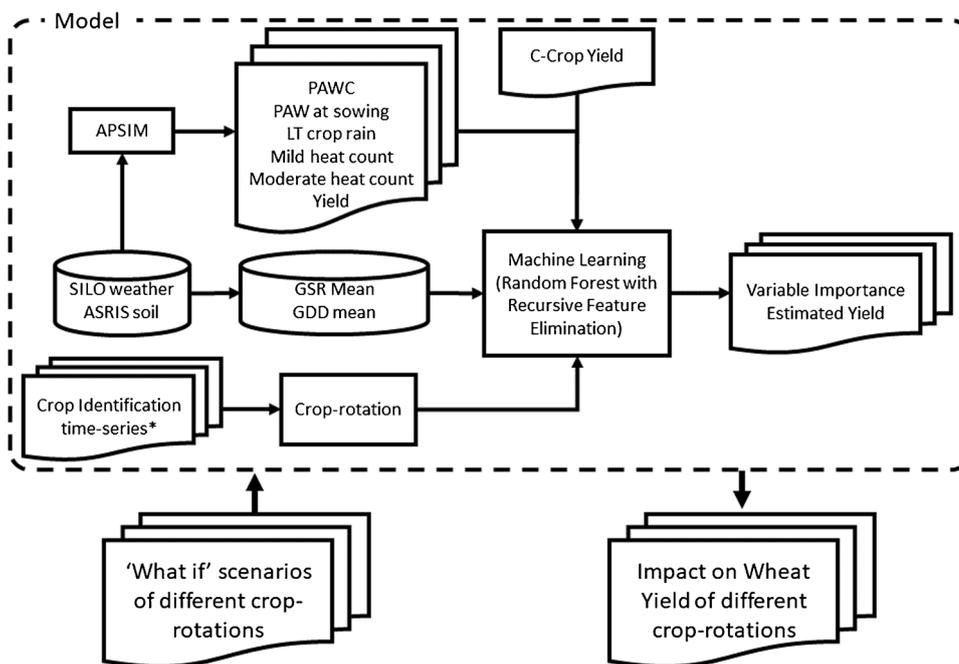
Two prediction models were created. The first is a simple analysis with only four variables: year, growing season rainfall (GSR), crop rotation, and thermal time expressed as growing degree days (GDD), that predict C-Crop yield. The objective was to discover what predictive capacity a limited data set could provide the industry about crop rotation and discover how important crop rotation was to wheat productivity across the broader Western Australian Wheatbelt.

The second analysis was again designed to assess the importance of crop rotation, where additional variables provide local and seasonal context to the prediction. The objective of this analysis was to improve the prediction of crop yield at the field level. Here, the RFE and Random Forest were used to build the model of crop yield, where additional variables that are output from the APSIM crop yield model were used in the prediction. In all, 11 variables were used and included PAW at sowing, PAW at harvest, long-term average growing season rainfall, heat count, soil

Table 2 Description of variables derived from Crop ID and the APSIM model, captured from Agri-Yieldz output.

Variable label	Description
Year	The year when the crop wheat yield was monitored
GSR mean	Growing Season Rainfall, extracted from SILO weather data, from April 1 to October 31 for the current year
Group rotation	Crop rotation, which refers to the previous crop grown prior to wheat and was identified in for each field
GDD mean	Thermal time calculated from April 1 to October 31 each season
AP PAW SOW	Plant available water (mm), at sowing, as estimated by APSIM
AP PAW harvest	Plant available water (mm) at harvest, as estimated by APSIM
AP LT in crop rain	Long term mean growing season rainfall for the particular field. Extracted from SILO weather data
AP mild heat count	Number of days temperatures fell between 32°C and 34 °C during the heat shock window (growth stage 70 to growth stage 80). Extracted from the APSIM crop model
AP soil ID	The soil type identification from the ASRIS soil grid
AP moderate heat count	Number of days temperatures fell between 34°C and 46°C during the heat shock window (growth stage 70 to growth stage 80). Extracted from the APSIM crop model
AP yield	The APSIM predicted yield for the field for the current year

Fig. 3 A graphical summary of the analytical approach to predict the effect of crop rotation on wheat cereal yields in every field.



type, and long-term mean yield from APSIM. These data are available from the Agri-Yieldz website <https://agdatashop.csiro.au/agriyieldz-the-wheat-crop-production-estimator>. Variables provide an insight into the agronomic state of the field, as derived from the APSIM crop model.

Both analyses used a 10-fold cross validation scheme. For each fold, two splits were used, where 90% of the dataset (164,175 observations) were used in training and 10% (18,242 observations) for testing (i.e., accuracy assessment). A graphical summary of the data, analytical workflow, and creation of output about crop rotations to underpin an agricultural digital twin are presented in Fig. 3.

3 Results

The most dominant crop rotations were wheat following a cereal (38%), wheat following an annual pasture (30%), wheat following canola (15%), wheat following a legume crop (9%), and wheat following a long fallow (6%). The area available for analysis, over 4 years, ranged from 6.4 million ha for wheat following a cereal to 1 million ha for wheat following fallow (Table 3). Crop rotation and the crop diversity varied across the Western Australian Wheatbelt, where crop diversity was greater in the southern regions and higher rainfall zones (Figs. 4 and 5).

Crop yield, as estimated by the C-Crop model also, varied across the Western Australian Wheatbelt with the season. C-Crop predicted wheat yields averaged 1.84 t ha^{-1} , with a standard deviation of 0.8 t ha^{-1} . Mean wheat yield also varied with season and ranged from an average of 1.4 t ha^{-1}

Table 3 Percentage of land area occupied by each land use in the Western Australian Wheatbelt, averaged over 4 years of the study.

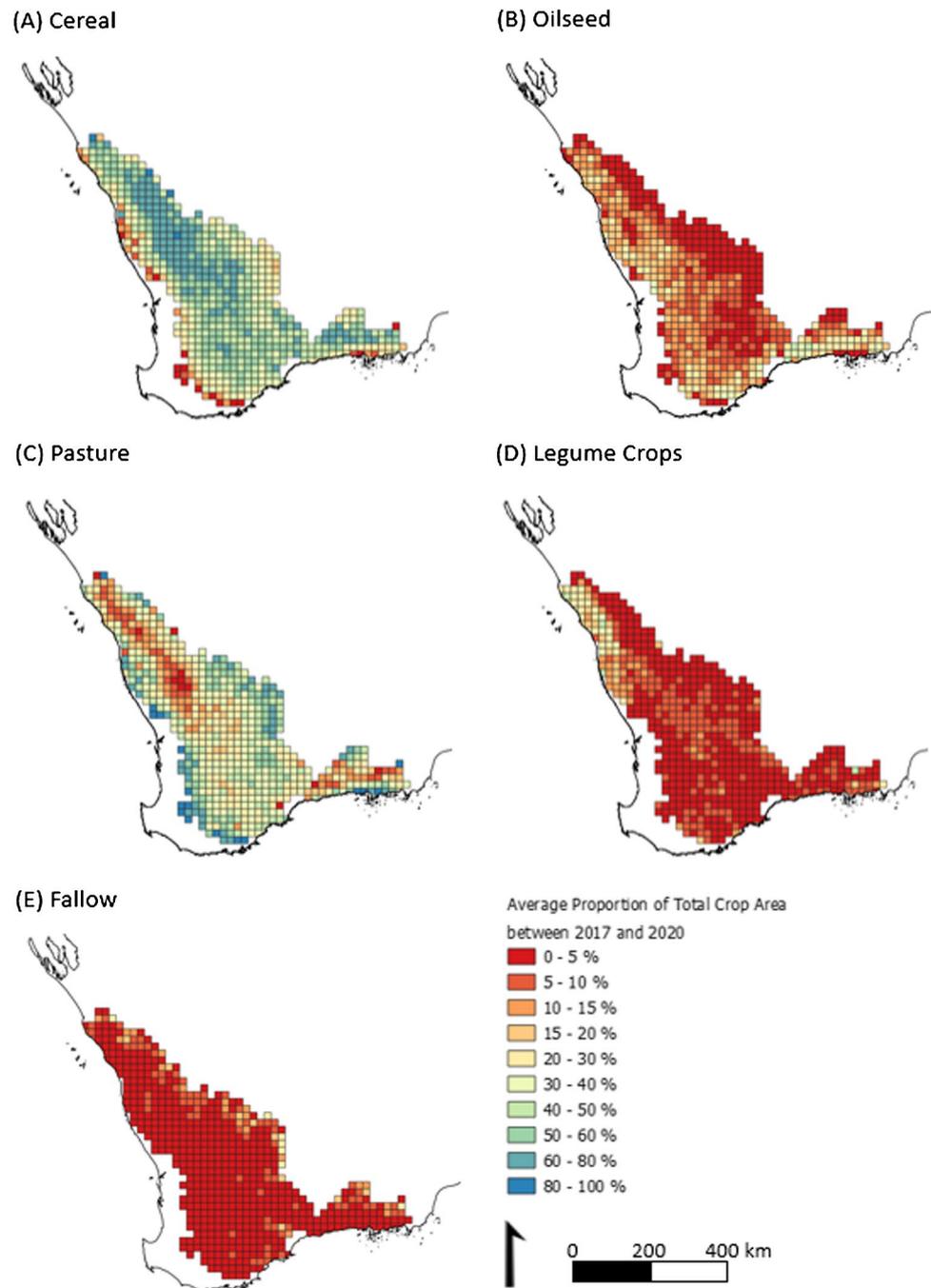
Previous land use prior to wheat	Percent of wheat crops with that prior land use	Area (millions of ha), surveyed, with that land use over 4 years
Cereal	38%	6.4
Pasture	30%	5.2
Canola	15%	2.6
Legume crop	9%	1.5
Fallow	6%	1.0

in 2019 to an average of 2.4 t ha^{-1} in 2018. Estimates from the Australian Bureau of Statistics were available for 2017, 2018, and 2019 production seasons, and trends detected by C-Crop for WA wheat production are in line with government figures (Table 4).

The first, simple Random Forest, defined as model 1, with just 4 variables, was able to predict C-Crop wheat yield from the test data with an r^2 of 0.72 (std dev 0.004) and an RMSE of 0.56 t ha^{-1} (std dev 0.004 t ha^{-1}). Year was the most important variable, followed by growing season rainfall, crop rotation, and thermal time. Crop rotation was important in this analysis, but the effects of crop rotation with model 1 were not conclusive (Fig. 6), and output from this model was not interrogated further.

The second, more complex model, defined as model 2, with numerous variables that explain the complex interaction between crop rotation and the agro-environment was able to predict C-Crop wheat yield from the test data with an r^2 of 0.84 (std dev 0.003) and an RMSE of 0.40 t ha^{-1} (std

Fig. 4 Variation in crop rotations across the Western Australian Wheatbelt for **a** cereal, **b** oilseed, **c** pasture, **d** legume crops, and **e** fallow, where the land use refers to the management carried out prior to growing a wheat crop.



dev 0.003 t ha^{-1}). Variable importance scores changed markedly from model 1 to model 2. The most important variable was crop rotation, followed by thermal time and then year. Soil water status and temperature stress events were also important predictors (Fig. 7).

Output from model 2 was interrogated to understand why crop rotation became the most important variable to predict wheat yield, as defined by the C-Crop model. From this suite of crop rotations in Table 1, partial deviance output from model 2 revealed a 87 kg ha^{-1} yield advantage, on average,

for wheat on legumes compared to wheat on cereal. This advantage increased to 97 kg ha^{-1} for wheat on canola vs wheat on cereal. In contrast, wheat following pasture and wheat following fallow had small yield disadvantages compared to wheat on wheat. These yield differences were -47 kg ha^{-1} and -60 kg ha^{-1} , respectively. These average differences are smaller than the overall model RMSE. This is to be expected with a Random Forest, where crop rotation accounts for part of the variation explained by the model. Whilst useful, these partial differences do not completely

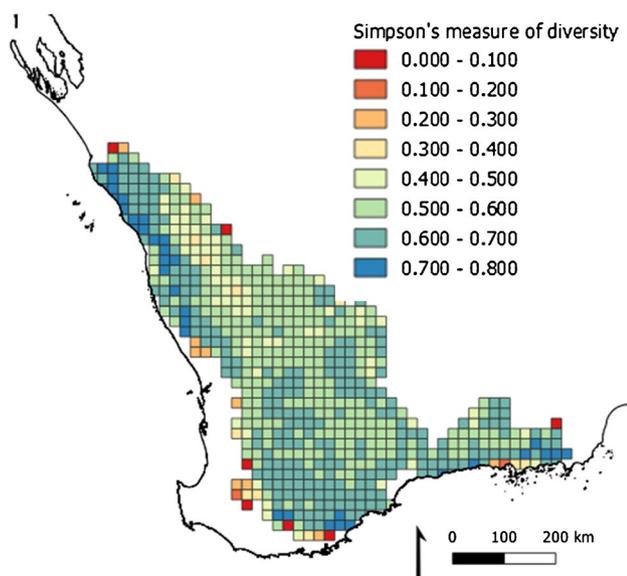


Fig. 5 Geographic variation on crop diversity, as measured with a Simpsons Diversity Index across the Western Australian Wheatbelt.

Table 4 Mean C-Crop estimate of wheat yield and Australian Bureau of Statistics (ABS) wheat yield.

	C-Crop wheat yield (t ha ⁻¹)	ABS wheat yield (t ha ⁻¹)
2017	1.81	1.89
2018	2.4	2.28
2019	1.38	1.41
2020	1.80	2.0

explain the value of crop rotation in this farming system. For wheat grown after canola and wheat grown after grain legumes the median, positive skew of the distribution demonstrate that crop rotation can occasionally deliver large yield benefits (Table 5, Fig. 8). The consequence of the skewed distribution meant that there was a wheat yield increase of 200 kg ha⁻¹ or more for 27% of wheat crops grown after canola and 26% for crops grown after a legume relative to wheat grown after wheat. For pasture and fallow, the number of fields with a wheat yield increase of 200 kg ha⁻¹ or more declined to just 8% and 7% of fields, respectively.

Model 2 accounts for the complex agro-climatic interactions with crop rotation. For the individual field, the influence of the other variables related to climate and soil type can either increase or decrease the impact of a crop rotation choice for that particular field. Simple “rules of thumb” could not be extracted by directly comparing fields with positive break crop effects and those with negative break crop effects. The benefit of the rotation effect on wheat yield varied (Fig. 8).

The complexities of individual seasons, soil types, starting soil moisture, and crop stress do complicate where and when particular rotation choices benefit cereal yields. In Fig. 9, the effect of canola on wheat yields is depicted spatially. In 2018 and in 2020, yield benefits to cereals are most evident in the north-eastern section of the Western Australian Wheatbelt. The benefits to cereals from canola were noticeably lower in 2019 and 2017. The north-eastern section of the Western Australian Wheatbelt is characterized as drier, with a shorter growing season than the southern regions and central regions (Figure 2).

4 Discussion

Crop rotation is an important component of most cropping systems, as crop rotation can increase the yields of the dominant crop by reducing the extent of biotic threats and, in the case of legume breaks, increase the nutrient supply. However, the value, in terms of the actual yield benefit, that a break crop provides to a subsequent crop can vary. This variability may complicate the decision to grow a break crop. Therefore, in this research, we utilized remote sensing information about crop identification and crop yield, with process-based crop modeling and machine learning to quantify the yield benefit of crop rotation to the yield of the subsequent cereal crop. We performed this assessment, at scale, for every field in the entire 20 million hectares of Western Australia over four growing seasons. While our study used the Western Australian Wheatbelt as a test case, the approach could be applied to any agricultural region in the world provided that appropriate methodologies to estimate crop type and yield were available.

The specific questions tested in this research were 1. What are the rotations in WA and 2 What is the break crop effect on wheat production in WA? The study demonstrated that the most common rotation in WA was wheat following a cereal (Table 3 and Fig. 4). This was followed by wheat grown after a pasture and then wheat after canola. Broadly, the findings in this study relating to the relative popularity of particular farming practices and crop rotations agree with insights from conventional ground surveys (Harries et al. 2015, 2022). This was to be expected, given that Australian crop classification approaches have been developed with extensive training data and predict with accuracies in the order of 80% (Lawes et al. 2022).

The agronomic impact of crop choice and rotation on the subsequent yield of the wheat crop did vary from insights derived from farmer surveys and from individual agronomic trials, conducted in Western Australia. Here, we identified moderate benefits and predict that in at least 70% of wheat fields, the break crop benefit is less than 200 kg ha⁻¹. This contrasts with field surveys, where wheat yields following

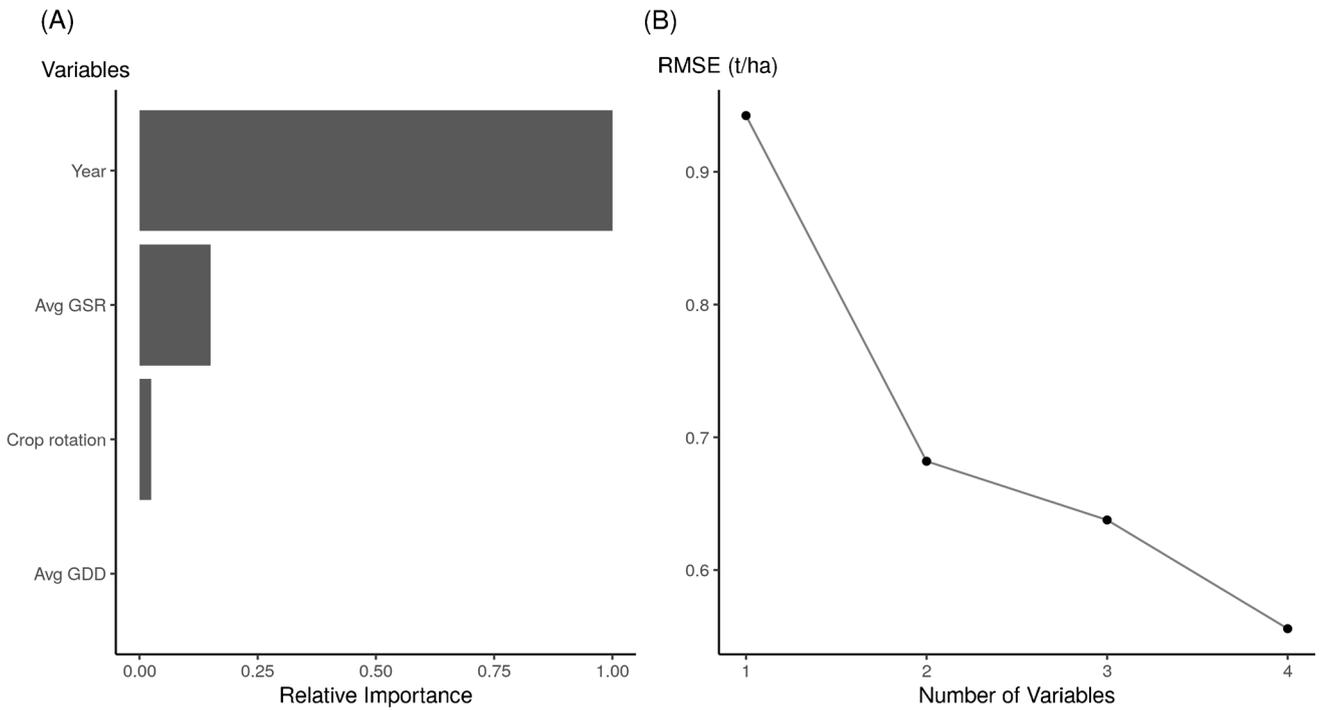


Fig. 6 Variable importance (A) and normalized root mean square error (B) of the model 1, from the 10-fold cross-validation recursive feature elimination with random forest.

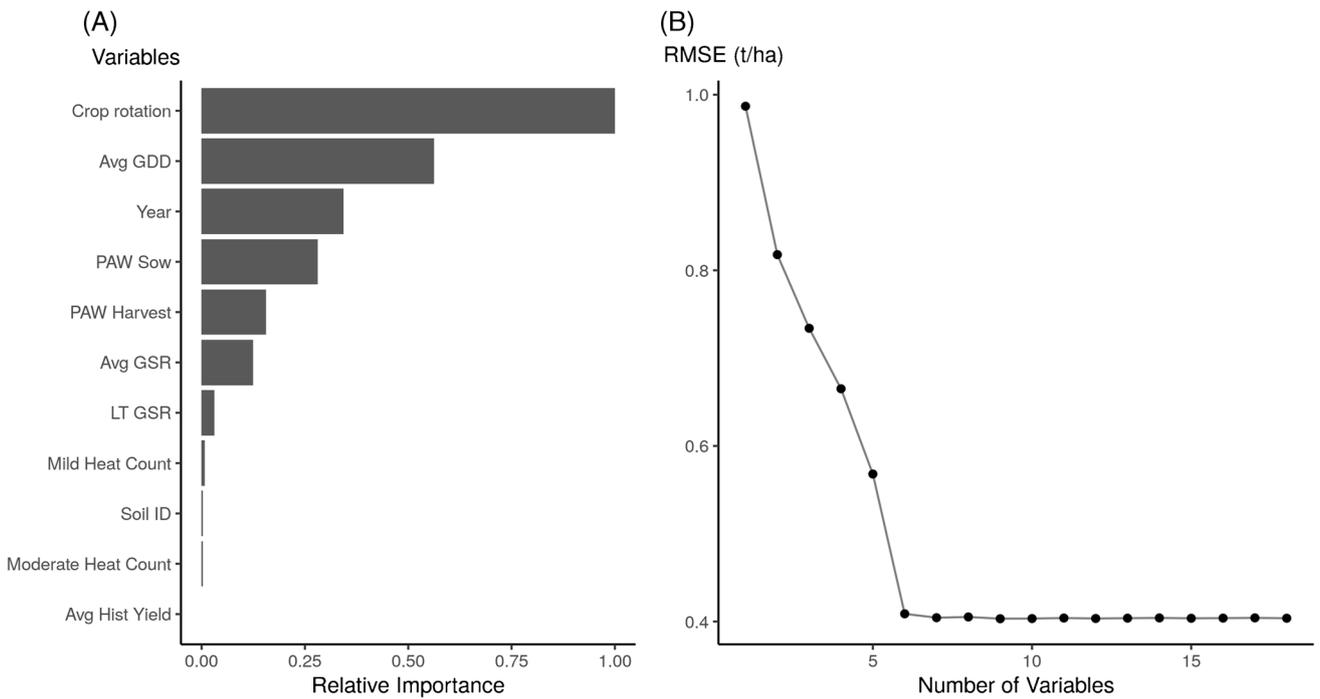


Fig. 7 Variable importance (A) and normalized root mean square error (B) of the model 2, from the 10-fold cross-validation recursive feature elimination with random forest.

Table 5 Overall differences from model 2 outputs comparing the effects of each crop rotation choice on wheat yields.

Yield (ton ha ⁻¹)	Mean	Std Dev	Median	Skew	Kurtosis
(A) Cereal after cereal	-0.02	0.19	0.00	-0.19	5.74
(B) Cereal after fallow	-0.06	0.22	-0.02	-1.04	3.42
(C) Cereal after oilseed	0.10	0.24	0.05	0.69	2.69
(D) Cereal after pasture	-0.05	0.21	0.00	-1.02	5.08
(E) Cereal after pulse	0.09	0.23	0.06	0.43	2.18

canola crops generated yield gains of 220 kg ha⁻¹ in a 200 mm rainfall environment and 330 kg ha⁻¹ in a 300 mm rainfall environment (Harries et al. 2022). Harries et al. (2022) analysis suggest break crop effects increase with increasing rainfall. Furthermore, Harries et al. (2022) also identified that wheat yields after pasture would generate 400 kg ha⁻¹ more yield than wheat after wheat in a 200 mm rainfall environment. Harries et al. (2022) finding contrasts with the earlier meta-analysis of Western Australian field trials by Seymour et al. (2012), who identified that the break crop effect was not dependent on rainfall and averaged 0.6 t ha⁻¹ for wheat following a lupin crop and 0.4 t ha⁻¹ for wheat following a canola crop. Therefore, farm surveys have provided different insights into crop rotation, relative to those

Fig 8 Histograms illustrating the number of fields that would have experienced either a yield benefit, or loss, if an alternative rotation choice compared to wheat on wheat was implemented by the farmer, for each year of the investigation. A red dashed line is a negative mean yield effect. A blue dashed line is a positive mean yield effect.

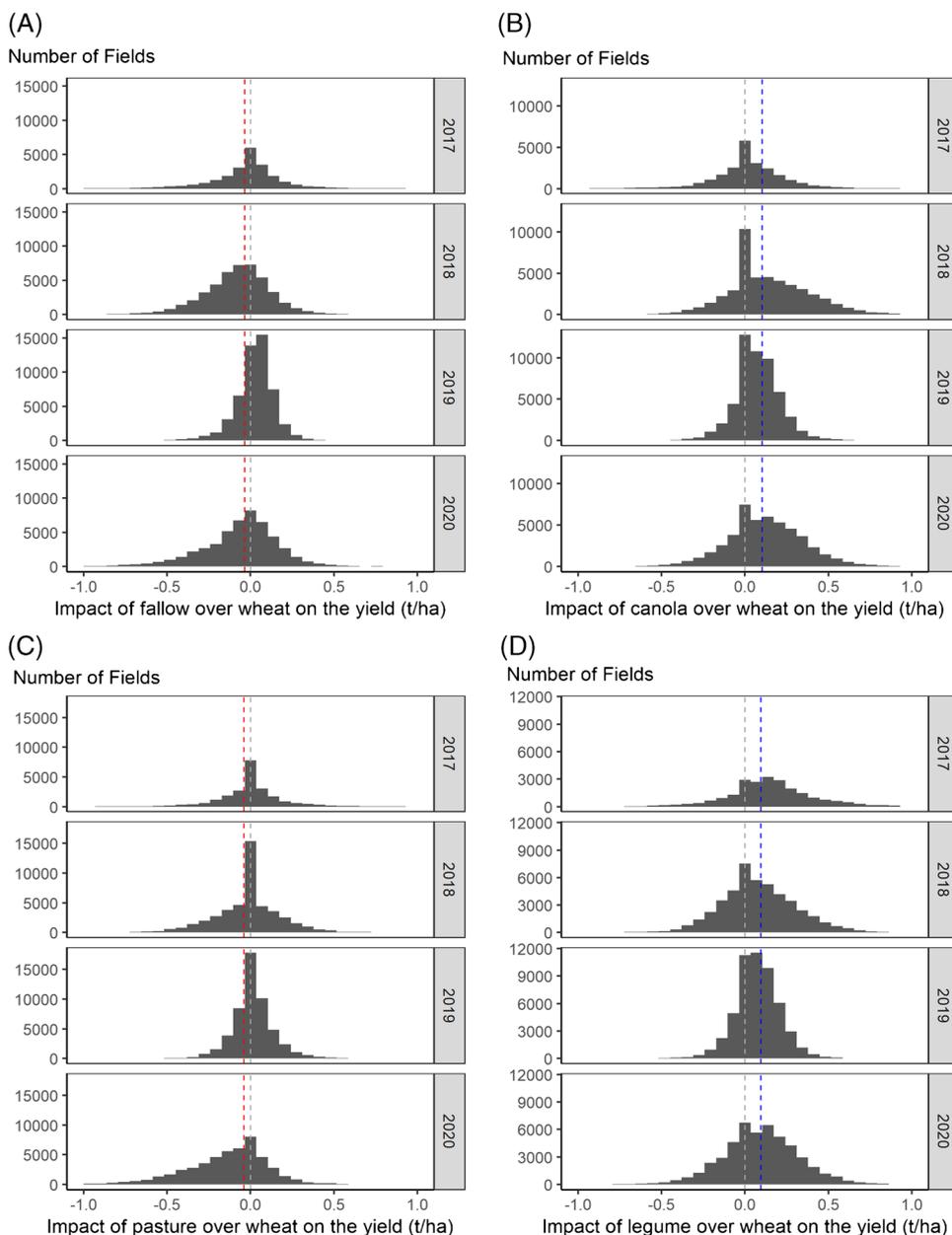
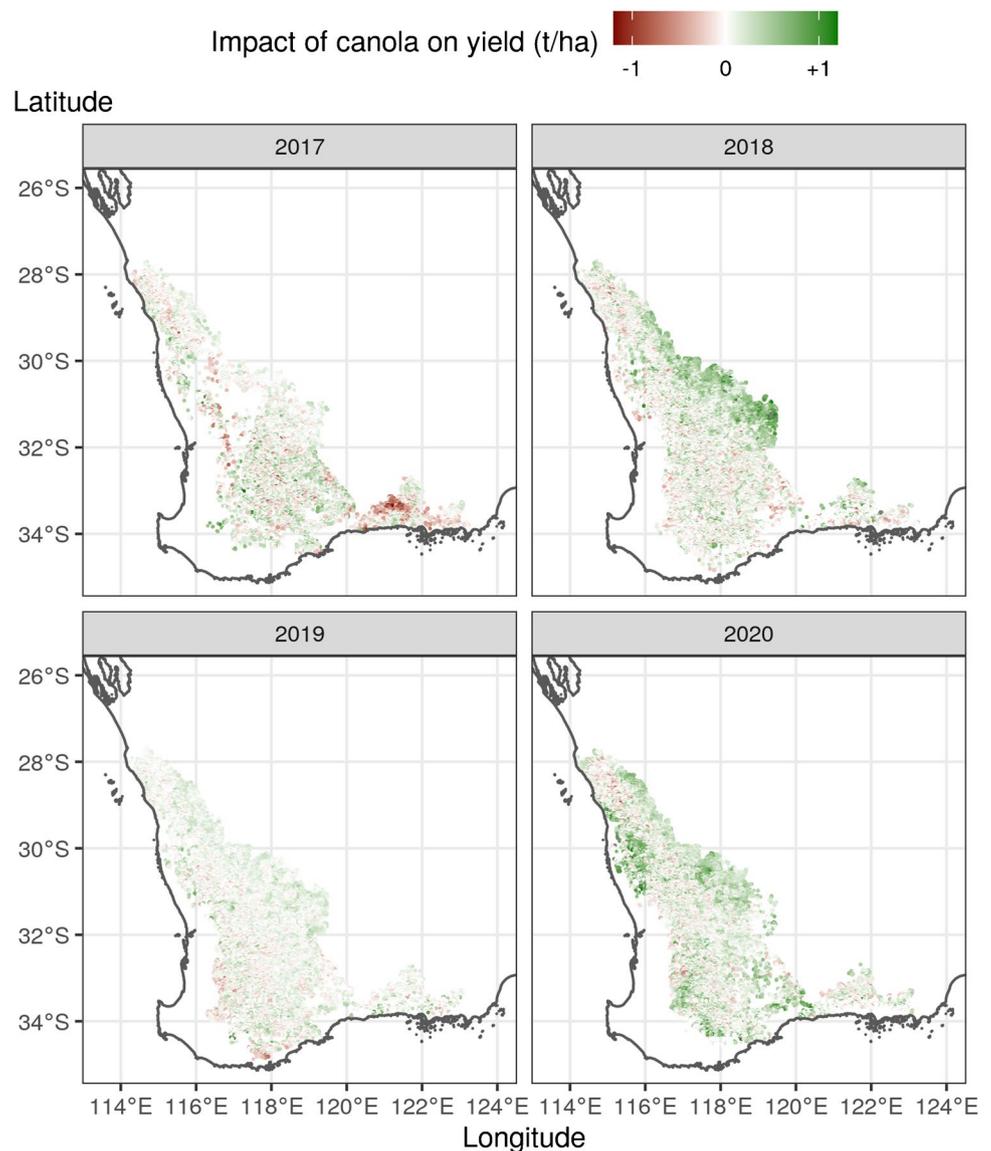


Fig. 9 Geographic spread of the positive and negative benefit of canola on wheat yield across the Western Australian Wheatbelt from 2017 to 2020, if a wheat-canola crop rotation was implemented instead of the wheat-cereal rotation.



derived from a meta-analysis of numerous field trials. These differences complicate creating simplified rules for farmers to follow. The differences add weight to Lacoste et al. (2022) view that on farm experimentation is often required to help farmers evaluate a technology or management concept.

In the present study, pastures generated a net penalty to wheat crops, and break crops such as canola and legume crops offered a moderate yield improvement to wheat crops. Both of these findings are at odds with the agronomic literature and with insights derived from farm surveys. This raises important questions about the value of specific agronomic trials for individual farmers, where the conditions of the trial vary from the farm. For example, break crop experiments may be located on sites that have a biotic stress present. This stress would bias the outcome in favor of cereal crops grown following a break crop that mitigated the biotic stress.

Researchers may deliberately choose sites with such a stress to demonstrate the management advantages of this practice to farmers. Such practices are normal, as researchers deliberately design and locate field trials to provide the insight required to address a problem. However, if farms experience lower levels of biotic stress than what is commonly seen in field trials, then the benefits of a break crop may be smaller than anticipated.

Biases, or inconsistencies between trialled data, and farm data may materialize for other reasons. For example, the pastures commonly grown by farmers throughout Western Australia may no longer be legume dominant. This would limit the amount of nitrogen fixed, and grass dominant pastures would not provide a break from pathogens. The benefits from these volunteer, weedy pastures would be considerably lower than the benefits that optimally managed, legume

dominant pastures provide to the following cereal crop. A recent study by Loi et al. (2022) found that cereal yields following a well-managed, modern legume pasture cultivar produced similar yields to wheat following a fallow, but required less applied nitrogen to achieve that yield. Again, though direct comparisons between wheat following wheat and wheat following a managed pasture was not conducted.

The variation between individual fields, particularly across the Western Australian Wheatbelt, will be greater than the variation that exists in field trials. Therefore, defining the environment of that field, in the context of starting soil water, soil type, growing season rainfall, and other stresses become important, when those factors interact with a particular management strategy. Here, this study predicted a wheat yield response to crop rotation suggesting that break crop benefits were most likely in the lower rainfall, and shorter growing season regions of the Western Australian Wheatbelt. This finding also conflicts with the findings of the earlier studies. Therefore, the ability to provide management insights, at scale, using local information as a component of a machine learning model is highly novel. We contend that the Random Forest predictions can cope with high-level interactions and cope with factors such as starting soil moisture and end of season temperature stress. The approach adopted here was able to quickly provide a prediction for every field in the Western Australian Wheatbelt, and each prediction could be evaluated on a case-by-case basis by a farmer. The discrepancies between the findings in this study and those of Harries et al. (2022) and Seymour et al. (2012) suggest that field-level predictions will require considerable development and engagement with farmers before these more abstract, machine-learned approaches could be deployed into an on-farm management decision support systems.

The analytical approach combined insights from remote sensing and APSIM using machine learning to estimate the rotational benefits for cereal crops. A necessary evolutionary step may involve updating or reconciling each of these estimates with actual data recorded on farm. Since data will not exist on every farm, analytical methods that allow the seamless integration and updating of model predictions for farms with data will need to be developed. This challenge is not trivial from an analytical perspective. Further model-based refinements may be required if farmers have detailed management information about fertilizer or weed control. Thus, in essence, the framework developed here provides the industry with the ability to benchmark production and explores possible benefits of alternative rotation choices. The approach here does combine multiple information sources, but in the future, the approach may need to combine even more disparate sources of information to create a believable and actionable management recommendation for a farmer. That is, this approach allows

more nuanced findings to be provided to an individual field and draws on a vast number of permutations relating to season, soil type, and climate to arrive at those predictions. The approach can provide predictions for an individual field and allow users to evaluate different scenarios with a prediction.

Further developments would need to actively consider who the user is and what the users' needs are. Outputs from the modeling would be valuable for individual farmers and their consultants and agronomists. Researchers may benefit, particularly, if machine learning approaches such as RFE and modeling based approaches deliver insights that are counter-factual to insights derived from field trials. The analytics and model-based information systems developed in this study that created insights about crop rotation at scale could become the precursor to the next generation of agricultural decision support systems or inform an agricultural digital twin.

5 Conclusion

We used machine learning to combine output from process-based crop models, crop identification systems and satellite driven crop models to obtain insights into the benefits of crop rotation across the vast Western Australian Wheatbelt. The outputs identified different crop yield responses to those commonly reported in the agronomy literature. The approach identified that the yield responses varied spatially across the Western Australian Wheatbelt, and climatic factors did influence the outcome. The machine learning approach was employed to produce scenarios for every field in the 20-million-hectare farming region. Outputs from this exercise could form the basis behind future digital decision support systems or agricultural digital twins that farmers, consultants, and researchers may use.

Authors' contributions R. Lawes conceptualised and drafted the manuscript. A. Fletcher contributed to the conceptualisation. J. Richetti conducted the statistical analysis. C. Herrmann and G. Mata created images and maps.

Funding This study was funded through the CSIRO Drought Mission and Digiscape initiative (OD-221548).

Data availability The outputs from the APSIM crop modelling are available for download via a subscription purchase: <https://agdatashop.csiro.au/agriyieldz-the-wheat-crop-production-estimator>. Remote Sensing Outputs have been licensed to Digital Agricultural Services and are available through that organization.

Code availability Code is unavailable.

Declarations

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Conflict of interest The authors declare no competing interests.

References

- Adjemian MK (2012) Quantifying the WASDE announcement effect. *Am J Agr Econ* 94:238–256. <https://doi.org/10.1093/ajae/aar131>
- Angus JF, Kirkegaard JA, Hunt JR, Ryan MH, Ohlander L, Peoples MB (2015) Break crops and rotations for wheat. *Crop Pasture Sci* 66:523–552. <https://doi.org/10.1071/CP14252>
- Araya A, Kisekka I, Gowda PH, Prasad PVV (2017) Evaluation of water-limited cropping systems in a semi-arid climate using DSSAT-CSM. *Agr Syst* 150:86–98. <https://doi.org/10.1016/j.agry.2016.10.007>
- Burke M, Lobell DB (2017) Satellite-based assessment of yield variation and its determinants in smallholder African systems. *PNAS* 114:2189–2194. <https://doi.org/10.1073/pnas.1616919114>
- Chen C, Neill K, Burgess M, Bekkerman A (2012) Agronomic benefit and economic potential of introducing fall-seeded pea and lentil into conventional wheat-based crop rotations. *Agron J* 104:215–224. <https://doi.org/10.2134/agronj2011.0126>
- Chen Y, Donohue RJ, McVicar TR, Waldner F, Mata G, Ota N, Houshmandfar A, Dayal K, Lawes RA (2020) Nationwide crop yield estimation based on photosynthesis and meteorological stress indices. *Agr Forest Meteorol* 284:107872. <https://doi.org/10.1016/j.agrformet.2019.107872>
- Donohue RJ, Lawes RA, Mata G, Gobbett D, Ouzman J (2018) Towards a national, remote-sensing-based model for predicting field-scale crop yield. *Field Crop Res* 227:79–90. <https://doi.org/10.1016/j.fcr.2018.08.005>
- Fletcher A (2019) Benchmarking break-crops with wheat reveals higher risk may limit on farm adoption. *Eur J Agron* 109:125921. <https://doi.org/10.1016/j.eja.2019.125921>
- Fletcher AL, Chen C, Ota N, Lawes RA, Oliver YM (2020) Has historic climate change affected the spatial distribution of water-limited wheat yield across Western Australia? *Clim Change* 159:347–364. <https://doi.org/10.1007/s10584-020-02666-w>
- Fowler J, Waldner F, Hochman Z (2020) All pixels are useful, but some are more useful: efficient in situ data collection for crop-type mapping using sequential exploration methods. *Int J Appl Earth Obs* 91:102114. <https://doi.org/10.1016/j.jag.2020.102114>
- French R, Schultz J (1984) Water use efficiency of wheat in a Mediterranean-type environment. I. The relation between yield, water use and climate. *Aust J Agr Res* 35:743–764. <https://doi.org/10.1071/AR9840743>
- Fritz S, See L, Bayas JCL, Waldner F, Jacques D, Becker-Reshef I, Whitcraft A, Baruth B, Bonifacio R, Crutchfield J, Rembold F, Rojas O, Schucknecht A, Van der Velde M, Verdin J, Wu B, Yan N, You L, Gilliams S, Muecher S, Tetrault R, Moorthy I, McCallum I (2019) A comparison of global agricultural monitoring systems and current gaps. *Agr Syst* 168:258–272. <https://doi.org/10.1016/j.agry.2018.05.010>
- Grundy MJ, Rossel RAV, Searle RD, Wilson PL, Chen C, Gregory LJ (2015) Soil and landscape grid of Australia. *Soil Res* 53:835–844. <https://doi.org/10.1071/SR15191>
- Harries M, Anderson GC, Hüberli D (2015) Crop sequences in Western Australia: what are they and are they sustainable? Findings of a four-year survey. *Crop Pasture Sci* 66:634–647. <https://doi.org/10.1071/CP14221>
- Harries M, Flower KC, Renton M, Anderson GC (2022) Water use efficiency in Western Australian cropping systems. *Crop Pasture Sci* 73:1097–1117. <https://doi.org/10.1071/CP21745>
- Hochman Z, Gobbett D, Horan H, Garcia JN (2016) Data rich yield gap analysis of wheat in Australia. *Field Crop Res* 197:97–106. <https://doi.org/10.1016/j.fcr.2016.08.017>
- Isbell R (2016) The Australian soil classification. CSIRO publishing, Melbourne. <https://doi.org/10.1071/9781486304646>
- Jaafar H, Mourad R (2021) GYMEE: a global field-scale crop yield and ET mapper in Google Earth engine based on Landsat, weather, and soil data. *Remote Sens* 13:773. <https://doi.org/10.3390/rs13040773>
- Kirkegaard J, Christen O, Krupinsky J, Layzell D (2008) Break crop benefits in temperate wheat production. *Field Crop Res* 107:185–195. <https://doi.org/10.1016/j.fcr.2008.02.010>
- Kuhn M (2008) Building predictive models in R using the caret package. *J Stat Softw* 28:1–26. <https://doi.org/10.18637/jss.v028.i05>
- Lacoste M, Cook S, McNee M, Gale D, Ingram J, Bellon-Maurel V, MacMillan T, Sylvester-Bradley R, Kindred D, Bramley R, Tremblay N, Longchamps L, Thompson L, Ruiz J, Garcia FO, Maxwell B, Griffin T, Oberthür T, Huyghe C, Zhang W, McNamara J, Hall A (2022) On-farm experimentation to transform global agriculture. *NatFood* 3:11–18. <https://doi.org/10.1038/s43016-021-00424-4>
- Larsen AE, Noack F (2017) Identifying the landscape drivers of agricultural insecticide use leveraging evidence from 100,000 fields. *PNAS* 114:5473–5478. <https://doi.org/10.1073/pnas.1620674114>
- Lawes R, Chen C, Whish J, Meier E, Ouzman J, Gobbett D, Vadakattu G, Ota N, van Rees H (2021) Applying more nitrogen is not always sufficient to address dryland wheat yield gaps in Australia. *Field Crop Res* 262:108033. <https://doi.org/10.1016/j.fcr.2020.108033>
- Lawes R, Hochman Z, Jakku E, Butler R, Chai J, Chen Y, Waldner F, Mata G, Donohue R (2022) Graincast™: monitoring crop production across the Australian grainbelt. *Crop Pasture Sci* (early View). <https://doi.org/10.1071/CP21386>
- Lebourgeois V, Dupuy S, Vintrou É, Ameline M, Butler S, Bégue A (2017) A combined random forest and OBIA classification scheme for mapping smallholder agriculture at different nomenclature levels using multisource data (simulated Sentinel-2 time series, VHRS and DEM). *Remote Sens* 9:259. <https://doi.org/10.3390/rs9030259>
- Lenssen AW, Long DS, Grey WE, Blodgett SL, Goosey HB (2013) Spring wheat production and associated pests in conventional and diversified cropping systems in North Central Montana. *Crop Manage* 12:1–8. <https://doi.org/10.1094/CM-2013-0017-RS>
- Lobell DB, Ortiz-Monasterio JI, Asner GP, Naylor RL, Falcon WP (2005) Combining field surveys, remote sensing, and regression trees to understand yield variations in an irrigated wheat landscape. *Agron J* 97:241–249. <https://doi.org/10.2134/agronj2005.0241a>
- Loi A, Thomas DT, Yates RJ, Harrison RJ, D'Antuono M, Re GA, Norman HC, Howieson JG (2022) Cereal and oil seed crops response to organic nitrogen when grown in rotation with annual aerial-seeded pasture legumes. *J Agric Sci* 160:207–219. <https://doi.org/10.1017/S0021859622000326>
- Ludwig F, Milroy SP, Asseng S (2009) Impacts of recent climate change on wheat production systems in Western Australia. *Clim Change* 92:495–517. <https://doi.org/10.1007/s10584-008-9479-9>
- Magno R, Rocchi L, Dainelli R, Matese A, Di Gennaro SF, Chen C-F, Son N-T, Toscano P (2021) AgroShadow: a new Sentinel-2 cloud

- shadow detection tool for precision agriculture. *Remote Sens* 13:1219. <https://doi.org/10.3390/rs13061219>
- Mourtzinis S, Edreira JIR, Grassini P, Roth AC, Casteel SN, Ciampitti IA, Kandel HJ, Kyveryga PM, Licht MA, Lindsey LE, Mueller DS, Nafziger ED, Naeve SL, Stanley J, Staton MJ, Conley SP (2018) Sifting and winnowing: analysis of farmer field data for soybean in the US North-Central region. *Field Crop Res* 221:130–141. <https://doi.org/10.1016/j.fcr.2018.02.024>
- Oliver YM, Robertson MJ (2009) Quantifying the benefits of accounting for yield potential in spatially and seasonally responsive nutrient management in a Mediterranean climate. *Soil Res* 47:114–126. <https://doi.org/10.1071/SR08099>
- Paludo A, Becker WR, Richetti J, Silva LCDA, Johann JA (2020) Mapping summer soybean and corn with remote sensing on Google Earth Engine cloud computing in Parana state – Brazil. *Int J Digit Earth* 13:1624–1636. <https://doi.org/10.1080/17538947.2020.1772893>
- Puntel LA, Sawyer JE, Barker DW, Dietzel R, Poffenbarger H, Castellano MJ, Moore KJ, Thorburn P, Archontoulis SV (2016) modeling long-term corn yield response to nitrogen rate and crop rotation. *Front Plant Sci* 7:1630. <https://doi.org/10.3389/fpls.2016.01630>
- Reeves MC, Zhao M, Running SW (2005) Usefulness and limits of MODIS GPP for estimating wheat yield. *Int J Remote Sens* 26:1403–1421. <https://doi.org/10.1080/01431160512331326567>
- Robertson MJ, Lawes RA, Bathgate A, Byrne F, White P, Sands R (2010) Determinants of the proportion of break crops on Western Australian broadacre farms. *Crop Pasture Sci* 61:203–213. <https://doi.org/10.1071/CP09207>
- Seymour M, Kirkegaard JA, Peoples MB, White PF, French RJ (2012) Break-crop benefits to wheat in Western Australia – insights from over three decades of research. *Crop Pasture Sci* 63:1–16. <https://doi.org/10.1071/CP11320>
- Shendryk Y, Rist Y, Ticehurst C, Thorburn P (2019) Deep learning for multi-modal classification of cloud, shadow and land cover scenes in PlanetScope and Sentinel-2 imagery. *ISPRS J Photogram* 157:124–136. <https://doi.org/10.1016/j.isprsjprs.2019.08.018>
- Turner NC, Asseng S (2005) Productivity, sustainability, and rainfall-use efficiency in Australian rainfed Mediterranean agricultural systems. *Aust J Agric Res* 56:1123–1136. <https://doi.org/10.1071/AR05076>
- Velde Mvd, Diepen CAV, Baruth B (2019) The European crop monitoring and yield forecasting system : celebrating 25 years of JRC MARS bulletins. *Agric Sys* 168:56–57. <https://doi.org/10.1016/j.agry.2018.10.003>
- Waldner F, Diakogiannis FI (2020) Deep learning on edge: extracting field boundaries from satellite images with a convolutional neural network. *Remote Sens Env* 245:111741. <https://doi.org/10.1016/j.rse.2020.111741>
- Waldner F, Chen Y, Lawes R, Hochman Z (2019) Needle in a haystack: mapping rare and infrequent crops using satellite imagery and data balancing methods. *Remote Sens Env* 233:111375. <https://doi.org/10.1016/j.rse.2019.111375>
- Wellington MJ, Kuhnert P, Renzullo LJ, Lawes R (2022) Modelling within-season variation in light use efficiency enhances productivity estimates for cropland. *Remote Sens* 14:1495. <https://doi.org/10.3390/rs14061495>
- Yan H, Du W, Zhou Y, Luo L, Ze N (2022) Satellite-based evidences to improve cropland productivity on the high-standard farmland project regions in Henan Province, China. *Remote Sens* 14:1724. <https://doi.org/10.3390/rs14071724>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.