Precision Agriculture (2025) 26:85
https://doi.org/10.1007/s11119-025-10280-y

What makes on-farm experimental data suitable for data-driven decision-making? implications of trial design and spatial distribution of field data for machine learning models

André F Colaço[1,2] · Robert GV Bramley[1,3] · Jonathan Richetti[4] · Roger A Lawes[4]

Received: 13 May 2025 / Accepted: 28 August 2025

Abstract

Purpose On-farm experimentation (OFE) plays a key role in underpinning data-driven agronomic decision-making. For example, machine learning (ML) models can predict optimal nitrogen (N) fertiliser rates using trial crop responses alongside soil, plant, and climatic data. However, it is unclear how different OFE strategies, trial designs, and their consequences for the spatial distribution of field datasets, impact the development of such models. This work sought to investigate how trial design influences spatial autocorrelation in OFE data and the impact of this on model training. It was also of interest to explore whether tailoring OFE programs and ML models to specific regions might improve their performance compared to those generated for large geographic areas.

Methods Using 21 N strip trials across Australia, ML models were developed to predict optimal N rates under different scenarios of data autocorrelation and geographic coverage.

Results Spatial autocorrelation in OFE data had negligible impact on model performance. At the same time, models trained with fewer non-correlated observations showed similar performance to models trained with thousands of autocorrelated observations. This suggests that less replicated field trials providing more independent observations might be preferable – for their simplicity and pragmatism – to highly replicated, whole-field trials which generate highly autocorrelated field data. The results also indicate that regional models may perform better than global models.

Conclusion Overall, to improve both the quantity and quality of OFE data for ML models used to underpin a mid-season N fertiliser decision, prioritising a greater number of simpler experiments (less replicated strip or plot trials) across a region is likely to be more effective than increasing field coverage of individual experiments using highly replicated whole-field designs focused on field-specific models.

Keywords On-farm experimentation · Data-driven decision-making · Spatial autocorrelation · Regional models · Nitrogen management

Extended author information available on the last page of the article

Published online: 20 September 2025

Introduction

On-farm experimentation (OFE) may be characterised by large-scale field trials, often monitored and implemented with the aid of precision agriculture (PA) technology such as crop sensors, on-board yield monitors, guidance systems and variable rate technology (Cook & Bramley, 1998, Lacoste et al. 2022). The information obtained from OFE helps farmers understand how management practices affect system responses under specific field conditions, determined by the site and season in which the experiments are conducted. Consequently, this enables farmers to move away from generic agronomic recommendations and tailor their decision-making to the unique conditions of their farms. To achieve this, we believe that OFE should be 'farmer-centric'; that is, the experiments are typically planned and conducted with close farmer engagement, often with some 'specialist' support (Lacoste et al. 2022). This approach contrasts with trials primarily designed for 'discovery science,' driven by the interests of scientists rather than practical outcomes aimed at on-farm decision-making and farm business improvement. Typical examples of the use of OFE for decision-making have been large-scale seed rate (e.g. Anselmi et al., 2021) or N fertilisation trials (see references below).

Aside from promoting practical knowledge for farmers, irrespective of the degree of actual farmer engagement, OFE has been identified as a key element in the development of protocols for data-driven agronomic decision-making (Colaço & Bramley, 2018; Lawes et al., 2019; Bullock et al., 2019; Colaço et al., 2021; Hegedus et al., 2023). When observations of crop responses to treatments or management practices are combined with measures of other field variables – such as vegetation indices from sensors, soil characteristics, and climatic data – machine learning (ML) models can be trained to predict an optimal management strategy based on a set of observed factors (e.g. Colaço et al., 2021, 2024). In this case, aside from providing data to support immediate decision-making, the OFE trials are also critically used to build databases for model learning. Thus, a primary goal of OFE is to produce empirical evidence of the optimal management through field experimentation, which can then inform farmers and or be used to train AI models.

Recent research has explored various experimental designs used in OFE to achieve these goals (Fig. 1). Strip trials are arguably the most common and easiest to implement (Fig. 1b). In this design, adjacent parallel strips — one for each treatment — are implemented with a width equivalent to one or multiple passes of field machinery. Such trials can either cover the entire field with multiple replicates of each treatment (Panten & Bramley, 2012; Colaço et al., 2020) or, more commonly, focus on a portion of the field using one or a few replicates per treatment (Fig. 1b; Lawes & Bramley, 2012; Colaço et al., 2024). In the latter case, the location of the strips is determined using prior knowledge of the field's spatial variability, ensuring observations are collected across different zones within the same field (Fig. 1b). For the analysis of the trial, the field data (yield monitor, remote sensing data, etc.) is aggregated along the strip length, using multiple 'experimental units' (or areas of observation), to allow pair-wise treatment comparison and other analysis along the strips (Fig. 1b; Lawes & Bramley, 2012; Colaço et al., 2024).

Another approach involves the use of large-scale plot designs (Fig. 1a and c). The size of each plot is again determined by the width of the field machinery used for implementing and harvesting the trial, with a typical plot length of several tens of meters. Although not widely adopted by farmers, precision agriculture (PA)-based OFE research has increasingly

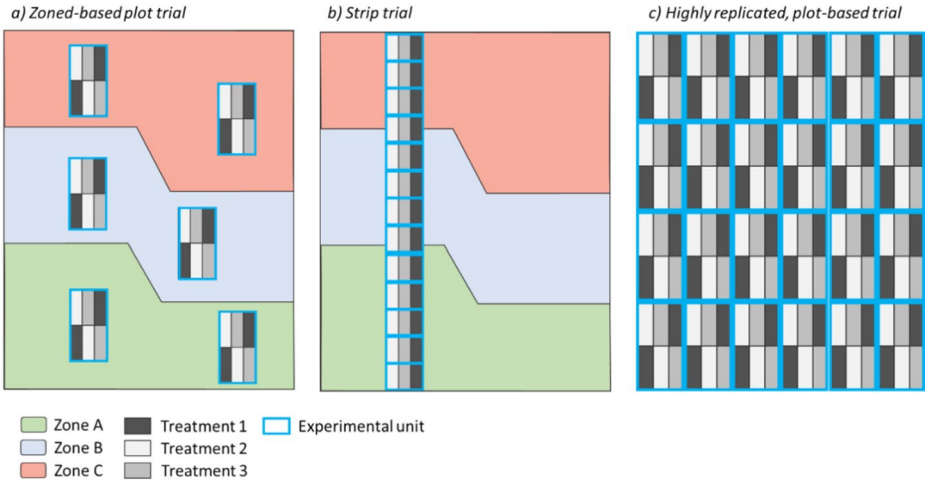

Fig. 1 Examples of different on-farm trial designs. (**a**) A large-scale plot trial in which experimental units are strategically distributed covering different zones within the field. (**b**) A strip trial crossing different zones of the field with multiple experimental analytical units to aggregate data enabling analysis along the length of strip. (**c**) A large-scale, highly replicated, whole-field plot trial covering the entire area of the field (**c**). From left to right more data are generated, but with increasing levels of spatial autocorrelation amongst these data

focused on highly replicated plot designs that cover the entire field area (Fig. 1c); a typical example is the whole-field 'checkerboard' fertiliser or seed rate trial to produce a map of optimal input rate across the entire field either with or without randomisation (Cook & Bramley, 1998; Kindred et al., 2015; Bullock et al., 2019; Trevisan et al. 2021). However, in contrast to the whole-field designs, farmers often tend to prefer less replicated trials, where treatment plots are strategically distributed across different zones of a field (Fig. 1a; Whelan et al., 2012). Much like simple strip trials, these designs are favoured for their simplicity, ease of analysis and implementation. Moreover, since the trial does not occupy the entire field, the farmer can implement the trial while also executing a desired management decision in the rest of the field. This is critically important as the information about treatment response obtained from the trial is specific to both the site and the season and so can inform on-going management decisions such as those related to mid-season N application.

Clearly, farmers, advisors and researchers have a range of trial designs and analytical approaches to choose from when it comes to planning an OFE. To make this choice, users must consider the main purpose of the trial, implications for data analysis, operational factors, availability of resources and other practical implications. When trials are designed to generate field data for developing ML models to aid management decisions, an important question arises: Do different trial strategies significantly impact the quality and utility of field data for model training? Generally, larger OFE areas favour the collection of a large number of observations for model training. However, quantity does not necessarily equate to quality. Data collected from the same field and from locations close to each other – for example, across the area of a checkerboard trial (Fig. 1c) or along the length of a strip trial (Fig. 1b) – may exhibit high levels of autocorrelation. This means that neighbouring data points could be redundant, contributing limited new information to the training of non-spatial ML models. In fact, some studies report that such autocorrelation can hinder the

training process of such models (Liu et al., 2022; Meyer et al., 2019). However, statistical and machine learning theory generally acknowledges that spatial autocorrelation in the input or outcome variables does not necessarily lead to autocorrelated residuals – and it is the spatial structure in the residuals that is most relevant for evaluating model adequacy (Cressie, 1993). In other words, if a model is able to explain the spatial variability using appropriate covariates and algorithms, then residual spatial autocorrelation may be minimal or absent, and thus not a concern. This can be formally assessed by analysing residuals after model fitting. Alternatively, if spatial autocorrelation is indeed regarded as problematic, one possible strategy is to train models using fewer but more spatially independent observations (scattered data as in Fig. 1a). However, such datasets may lack the volume needed to effectively train modern ML models.

It is important to emphasise that this dilemma applies to non-spatial ML approaches. However, the majority of ML applications explored for data-driven decision making in agriculture have been non-spatial; with clustering and regression analysis being typical examples (Chilingaryan et al., 2018; Liakos et al., 2018). While these non-spatial models can be used site-specifically for PA purposes, they generally make no direct use of the spatial information inherent in the data.

Another important factor that influences data quantity and variability is the combination of data from multiple fields. On the one hand, some studies focus on field-specific ML models (Hegedus et al., 2023), despite the need for multiple trials over multiple years to capture the field-specific effects of seasonal variability. Conversely, OFE networks spanning larger geographical areas (district or region) may increase data variability to a point where a single 'global' model may not be effective (Colaço et al., 2024). A solution may be an intermediate approach; that is, regional models produced by combining data from similar fields or from within readily constrained geographical regions to limit data variability. Achieving the right balance between these factors by matching OFE programs and trial designs to the data requirements for modelling, are critical when it comes to developing ML-based frameworks for agronomic decision-making informed by OFE data.

Recent research on OFE seems to have given little focus to these questions. Many OFE studies focus on developing appropriate statistical tools to analyse the experimental results per se, that is, the treatment response across the landscape (Lawes & Bramley, 2012; Córdoba et al., 2025); some even use geostatistical approaches for this (Bishop & Lark, 2006; Panten & Bramley, 2011; Jin et al., 2021). Other studies that do employ OFE to build databases to support the development of ML-based decision tools often do not address how the OFE strategy influences the data generated and how this, in turn, affects the quality and reliability of model training. Colaço et al. (2024), for example, developed a data-driven model for optimal N rate prediction using OFE data, and found that it could not outperform the farmer prescription unless unlimited historical field data – a simulated 'perfect data' scenario – were available. While these authors did not mention the potential negative effect of autocorrelation in their strip trial data for model learning and generalization, they hypothesized that model regionalization could be investigated to improve model performance. In other cases, authors have mentioned and recognized data autocorrelation in their analysis (de Lara et al., 2023; Evans et al., 2020) but without evaluating its effect on ML model performance.

In light of the above, the present research sought to promote understanding on how OFE strategies and the spatial distribution of field data may affect the training and applicability of

an ML model designed to predict optimal application rates of N fertiliser for wheat and barley crops. More specifically, the present study builds upon the work of Colaço et al. (2024) and aimed to enhance the results achieved using their data-driven N decision model by investigating two specific issues related to the spatial characteristics of OFE data. The first and main goal of this research was to investigate the spatial autocorrelation of field data and its influence (if any) on model learning and performance. To achieve this, autocorrelation was assessed both in the original field data and in the residuals produced by the models. The hypothesis to be tested was whether reduced data autocorrelation, using less replicated trials, favours model learning and generalization. A second interest of this work was to explore the role of the geographic extent of a model by comparing 'global' models (combining OFE data within a large geographical area) and 'regional' models (with a subset of OFE trial data from the same region). It was hypothesised that reducing the variability of useful covariates – rainfall, temperature, etc. – through regionalisation may favour the model's performance when compared to a 'global' model. In other words, the benefits of restricting the region of a model on a sensible biophysical basis was explored.

Materials and methods

This study used a structured database from OFE trials used in the work of Colaço et al. (2024). The data were obtained from N fertilisation strip trials in commercial wheat and barley fields conducted between 2018 and 2021 in different graingrowing regions of Australia. The database contained target variables – 'optimal N fertiliser rate' and 'grain yield' – and predictor variables, including on-farm sensor data, and off-farm data from public sources such as satellite images and weather observations. These on-farm target and predictor variables were obtained along the length of the strip trials – every ten meters – and combined with the off-farm data to train a random forest regression model as described by Colaço et al. (2024). Their main goal was to make predictions of optimal mid-season N application rates based on the observed field conditions. For the present analysis of the effect of data autocorrelation, an iterative data thinning process was performed through successive sampling followed by re-modelling (see below). Thus, with each iteration, the distance between each observation increased – simulating less replicated field trials – consequently reducing data autocorrelation. Finally, regional models based on different data subgroups, obtained by combining OFE trials through cluster analysis, were compared to the initial model, that is, the one generated from the entire database. All data preparation, analysis, and visualization steps were performed in Python using the Google-Colab platform.

On farm experiments and crop response database

Twenty-one on-farm N fertilisation experiments, thirteen in wheat and eight in barley, were conducted between 2018 and 2021 in different grain growing areas in Australia (Table 1). Each experiment consisted of N application strips similar to those shown in Fig. 2. Trials were planned with close engagement with the farmer and designed to minimize the burden of adopting OFE practices. In fact, in most cases, farmers were already using N-rich strips to fine-tune their fertiliser management and so the most important research modification was the inclusion of 'zero' strips (see below). Therefore, there was some variation in the specific

Table 1 Site characteristics of the on-farm experimental program

Year	Location [a]	Field size (ha)	*n*° of observations	Clay content (%) [b]	Growing season rainfall (mm) [c]	Average yield (t/ha) [d]	Average EONR (kg/ha) [e]
2018	Tarlee - SA[1]	64	254	38.8	181	5.0	87
2018	Woorak - VIC[1]	119	151	45.6	232	5.5	45
2018	Kalannie - WA[1]	227	97	15.2	232	3.5	56
2019	Tarlee - SA[1]	64	70	38.8	218	3.5	43
2019	Woorak - VIC[1]	119	152	45.6	237	4.7	130
2019	Kalannie - WA[2]	357	80	17.2	194	2.2	85
2020	Narrabri - NSW[1]	183	159	43.4	261	5.4	39
2020	Tarlee - SA[1]	64	65	38.8	352	7.5	93
2020	Woorak - VIC[2]	232	167	36.2	290	4.2	87
2020	Kalannie - WA[2]	357	80	17.2	162	2.5	21
2019	Booleroo Centre - SA[1]	100	169	25.1	145	0.8	4
2019	Urania - SA[1]	53	66	24.8	176	4.2	117
2019	Wharminda - SA[1]	31	60	11.4	179	4.0	65
2019	Tumby Bay - SA[1]	77	102	30.3	186	2.9	77
2019	Loxton - SA[1]	153	108	3.8	92	0.9	13
2020	Urania - SA[1]	53	66	24.8	330	5.1	38
2020	Booleroo Centre - SA[2]	86	104	24.2	343	4.8	29
2020	Tumby Bay - SA[2]	41	79	25.4	195	2.3	44
2020	Loxton - SA[1]	153	108	3.8	194	2.1	24
2021	Booleroo Centre - SA[1]	100	134	25.1	217	2.3	9
2021	Loxton - SA[2]	218	81	6.4	141	0.8	4

[a] The same numerical index for each location indicates that the same field was used in different years

[b] Average soil clay content at the 0.3 m depth layer obtained from soil sampling

[c] Rainfall between April and October sourced from the Australian Bureau of Meteorology (BOM, 2024) using the nearest weather station from the field

[d] Average yield measured from the N 'rich' treatment

[e] Average EONR (economically optimal N rate) obtained from the N strip trial data

SA – South Australia; VIC – Victoria; WA – Western Australia; NSW – New South Wales

design of each trial depending on the farmer equipment and farmer preference at each site. In the majority of trials, the N strips were 24–36 m wide, equivalent to two or three header widths. Typically, strips were positioned to run across 'management zones' determined through analysis of available spatial data – previous yield maps, remotely sensed imagery and electromagnetic soil survey data when available – coupled with the farmer knowledge of the field. At, or soon after sowing, an N 'rich' strip was established with N applied at approximately twice the normal farmer rate. Also, an N 'zero' strip was established with no additional application of N beyond that applied at sowing; the remainder of the field was fertilised according to normal farmer practice with the area adjacent to the strips used as the 'normal field' treatment. The experiments were harvested using a machine equipped with a yield monitor, which recorded the yield harvested along each experimental strip, logged at 1 Hz, equivalent to every 1–2 m, approximately.

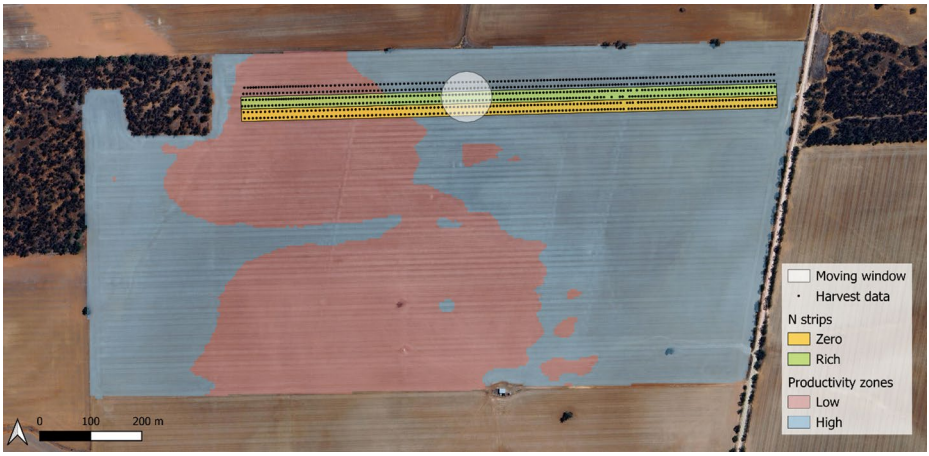

Fig. 2 Example of a field experiment with 'rich' and 'zero' rates of nitrogen fertiliser application along strips crossing different productivity zones in a commercial barley field in South Australia. For illustration purposes, the yield monitor data shown as dark dots were thinned to allow better visualization. Also highlighted is a 'moving window' exemplifying the data grouping along the length of the experimental strip

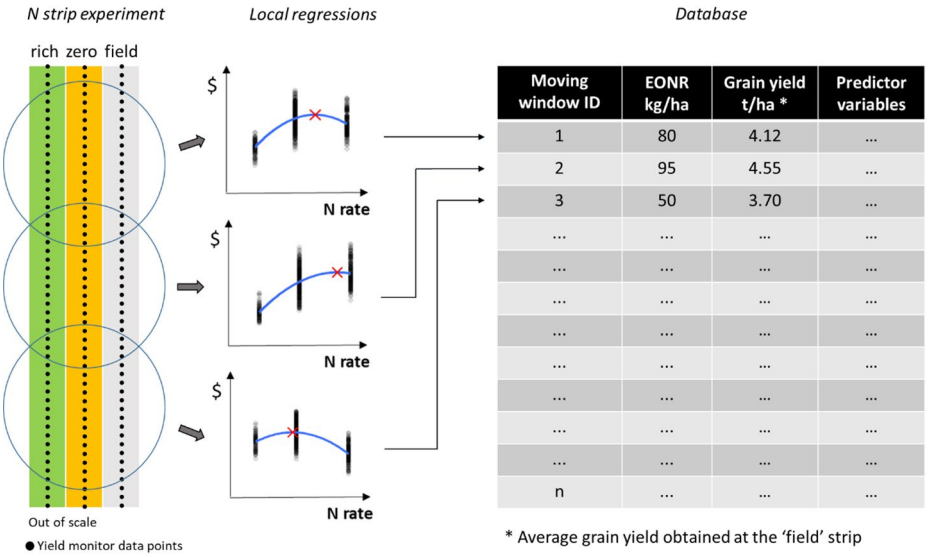

Fig. 3 The approach for determining economically optimal nitrogen rates (EONR) and building the database, where each row represents observations collected in a moving window along the length of the experimental strips. Adapted from Colaço et al. (2024).

The database was constructed from observations grouped every ten meters along the length of each strip trial using a circular moving window of 50 m radius (Figs. 2 and 3); note that the windows overlapped, and some data points were shared between neighbouring windows. To obtain the main response variable – 'optimal N rate (in kg/ha) – in each moving window, a simple quadratic regression was generated between the applied N rate

(*x*-axis) and the partial profitability obtained in each strip (*y*-axis). The 50 m radius of the moving window was chosen to ensure sufficient data points were available for each regression analysis. Partial profitability was calculated by subtracting the cost of fertiliser from the gross income obtained from production, with the optimal rate being the one that maximised profitability (Fig. 3).

Some studies have taken a different approach to predicting optimal N rates using ML and OFE. To avoid using a modelled value for the optimal N rate – derived from a simple quadratic regression, as described above – as the target variable for model training, these studies instead use ML to generate the 'actual' response function, with applied N as one of the predictor variables. They then retrieve the optimal N rate using an *argmax* optimization of a profitability function derived from the ML response function (de Lara et al., 2023; Trevisan et al., 2021; Tanaka et al., 2024). In contrast, the present study draws inspiration from the approach of Lawes et al. (2019), which was developed using a simulated dataset, but applies it to a real-field dataset under the constrains imposed by the pragmatic nature of the trials. It assumes that the simple regression used to derive optimal N rates, but conducted many times along the length of the strips, closely approximates 'actual' crop responses within each field. This estimated response, generated using readily available on-farm equipment, is then used as a reference against which an ML model can be trained. While this approach was used in the previous study by Colaço et al. (2024), it is acknowledged that the uncertainty associated with the simple regression – particularly when based on limited variation in N rates – can affect the final modelling outputs when such estimates are used as the response variable. Therefore, additional analyses were conducted using grain yield from the 'normal field' N application strip as the target variable, thereby avoiding the use of estimated values as the response.

Predictor variables

In addition to the target variables described above, predictor variables were collected and matched to the resolution of the regression analysis for optimal N rate. Therefore, observations within the 50 m radius of the moving window (Figs. 2 and 3) were averaged, also every ten meters along each experiment. These predictor variables were collected from public databases (such as satellite images and climate data) and using on farm sensors (such as proximal reflectance sensors). These variables characterise the field history, in-season crop status, soil and landscape features and weather (Table 2; further description is available in Colaço et al., 2024). Only variables that could be obtained at, or prior to the mid-stage of the early vegetative development phase (crop stage GS-31 according to the classification of Zadoks et al. 1974), when the decision on the amount of additional mid-season N fertiliser to be applied should be made, were used.

Baseline modelling

Initial random forest models for prediction of the optimal N rate and grain yield were developed to serve as a baseline for comparison with other variations to be tested (described below). To train and validate these models, an iterative process for spatial cross validation was conducted following a 21-fold cross validation approach with each fold defined by one experiment (Richetti et al., 2023). In each iteration, the data were divided into training and

Table 2 Variables used for optimal N rate and grain yield prediction

Variable	Description	Source	Used for optimal *N* rate prediction	Used for grain yield prediction
NDVI response function recommendation	N rate that maximised NDVI at GS-31 crop stage using a quadratic model between NDVI and trial N rates	On farm trial and Sentinel 2 satellites	yes	no
Historic yield	Average yield from previous 3 to 4 years	On board yield monitoring	yes	yes
NDVI (field)	NDVI from the adjacent field area ('field' strip) at GS-31 crop stage	Sentinel 2 satellites	yes	yes
NDVI (zero)	NDVI from the 'zero' strip at GS-31 crop stage	Sentinel 2 satellites	yes	no
NDVI (rich)	NDVI from the 'rich' strip at GS-31 crop stage	Sentinel 2 satellites	yes	no
Historic NDRE (95th percentile)	95th NDRE percentile from historic imagery	Landsat 8 satellite	yes	yes
Historic NDRE (5th percentile)	5th NDRE percentile from historic imagery	Landsat 8 satellite	yes	yes
Rainfall	Accumulated rainfall from sowing until GS-31	BOM[a]	yes	yes
Air temperature	Average daily max temperature from sowing until GS-31	BOM[a]	yes	yes
Land surface temperature (phase)	Model parameter of a sinusoid function fitted to a land surface temperature dataset	MODIS[b]	yes	yes
Land surface temperature (amplitude)	Model parameter of a sinusoid function fitted to a land surface temperature dataset	MODIS[b]	yes	yes
Gamma radiometric (k)	K^{40} radiometry from airborne gamma-ray spectrometric survey	Radiometric Grid of Australia[c]	yes	yes
Gamma radiometric (th)	Th^{232} radiometry from airborne gamma-ray spectrometric survey	Radiometric Grid of Australia[c]	yes	yes
Gamma radiometric (u)	U^{238} radiometry from airborne gamma-ray spectrometric survey	Radiometric Grid of Australia[c]	yes	yes
Soil clay content	Soil clay content at the top 0.3 m layer	ASRIS[d]	yes	yes
Aspect	Landscape attribute from digital elevation model	Digital Elevation Model of Australia[e]	yes	yes
Slope	Landscape attribute from digital elevation model	Digital Elevation Model of Australia[e]	yes	yes
Hill shade	Landscape attribute from digital elevation model	Digital Elevation Model of Australia[e]	yes	yes

[a] Bureau of Meteorology (BOM, 2024)

[b] Retrieved using the approach of Jakubauskas and Legates (2000)

[c] Australian wide Gama radiometric survey (Poudjom & Minty, 2019)

[d] Australian Soil Resource Information System (ASRIS, 2024)

[e] Wilson et al. (2011)

testing sets, followed by modelling. In other words, to ensure independence between training and testing data, in each iteration, data from one experiment were reserved for testing, and the remaining 20 experiments were used for training. Subsequently, a random forest regression model was trained. Since the focus was on comparing models across a range of scenarios, no fine-tuning was conducted. Therefore, the configuration (hyperparameters) to train the models – number of trees, maximum depth of each tree, and node splitting rules – was kept at the default values from the 'sklearn' package in Python. Finally, the model was validated using the root mean squared error (RMSE). Twenty-one iterations were performed – using all combinations of data splitting between the field trials.

Autocorrelation analysis

It is axiomatic that the characteristics of the field experiment and the data grouping along the strip length lead to autocorrelation in the database, which may negatively affect the training of non-spatial ML models; note that each observation in the database represents data grouped within a 50 m radius every 10 m along the length of the strip trial. Thus, an autocorrelation analysis, using a correlogram, was conducted. The correlogram shows the correlation between sequential data and the same data with a progressive lag, making it possible to observe the lag level – or, in this case, the distance between points – that ensures 'independence' (i.e., non-significant correlation) between the data. This analysis was performed for grain yield and optimal N rate data. In each case, data from all strips were concatenated into a single sequence ordered by spatial location along each strip. Although spatial continuity across experiments was not explicitly modelled, this aggregation allowed for a general assessment of autocorrelation structure in the dataset. To produce the correlograms, the 'plot_acf' function from the 'statsmodels' library was used, and spatial autocorrelation was visualized up to 80 lags. The function computes the sample autocorrelation at each lag and plots confidence intervals (by default, 95%). These intervals are based on an approximate normal distribution, with bounds set at $\pm 1.96/\sqrt{n}$, where n is the number of observations.

In addition to the analysis of autocorrelation in the response variables, we also investigated whether the autocorrelation present in the field data was retained in the residuals after model fitting. Therefore, correlograms were also produced based on the residuals from the optimal N rate and grain yield models.

Next, a data thinning strategy was used by sampling the training data to progressively increase the distance between observations in the database and consequently decrease the autocorrelation between them. It is important to note that the goal of this analysis was not to test data filtering as a potential strategy to reduce data autocorrelation. Rather, it sought to simulate data generated from simpler, less replicated trials, such as small strips or spatially distributed plot-based trials (Fig. 1a), as opposed to highly replicated designs (Fig. 1c), for example. Therefore, values of q were defined as 1, 2, 10, 20, 30, 40, 50, and 60, with q being the 'step' size of the data sampling; that is, one observation was retained every q rows of the database. An iterative process (for each q) was defined, where sampling of the database was performed with step q, followed by a new iteration for spatial cross-validation. The same approach as before was used where data from one experiment was used for validation and the remaining twenty for training, with twenty-one iterations testing all combinations of training and testing among the experiments. It is important to note that for all data sampling scenarios, model validation was conducted using the original, un-thinned data – meaning

the data from the validation trial was maintained at its original spatial resolution (every 10 m along the strips). This approach allowed the models to be evaluated in a scenario compatible with site-specific management (e.g., continuous variable rate fertiliser application), even though they were developed using sparse, thinned data. For each iteration, a random forest model was trained, again using the default hyperparameters of the 'sklearn' package in Python. The RMSE was used as the evaluation metric, and Tukey's test (at a 5% significance level) was conducted to compare model performance across the different q scenarios. Additionally, residual autocorrelation was calculated at lag 1 for each q scenario to assess whether reducing spatial autocorrelation in the input data could also reduce residual autocorrelation after model fitting.

Regionalisation analysis

In Table 1, it is noted that an OFE trial network was implemented in nine different locations spread across four states in Australia, namely Kalannie (WA), Booleroo Centre (SA), Loxton (SA), Tarlee (SA), Tumby Bay (SA), Urania (SA), Wharminda (SA), Woorak (Vic), and Narrabri (NSW), with the number of experiments per location varying between one and three. In the initial modelling (baseline), the data were combined, and a 'global' model – with data from all experiments – was generated and tested. This approach sought to maximize the amount of data used for training and validation. However, given the agroclimatic differences between regions, data from Narrabri (NSW), for example, may have little to contribute to a model applied about 3,500 km away in Kalannie (WA). Therefore, an attempt was made to create sub-regions from the initial database and subsequently train and validate models locally.

To do so, each location was characterised by the average observed yield and optimal N rate across the experiments. Then, a cluster analysis using the k-means algorithm was performed. The ideal number of clusters was defined using an 'elbow graph' that depicts the inertia value (the sum of squared distances between each point and the cluster centroid) for a varied number of clusters. Once the nine locations were grouped, cluster information was added to the initial database based on the location of each observation. Each subset of data for each cluster was then used for training and validating 'regional' models. In each case, the data were subdivided again between training and testing by separating an independent experiment for testing. For example, if a cluster grouped data from Booleroo Centre - SA and Loxton - SA, this dataset would comprise six experiments, three from each location (Table 1). Thus, data from one experiment would be set aside for testing, and data from the other five experiments would be used to train the model. In this example, six iterations would have been performed to test all combinations of training and testing among the available experiments. In each scenario, random forest models were trained again using the default hyperparameters of the 'sklearn' package in Python, and the RMSE across the different iterations was used as the evaluation metric. Finally, the performance of the models at each cluster were compared to the error obtained by the global model when applied to the respective locations of each cluster. A t-test (at a 5% significance level) was used to statistically compare the local and global models within each cluster.

It is acknowledged that this clustering strategy, combined with the constraints of a limited number of available trials, may not produce geographically cohesive 'regions'. However, it demonstrates a data-driven approach to defining 'regions' and allows this study

to leverage the varying degrees of geographical cohesiveness in the resulting clusters to interpret the findings.

Results

Baseline model

The results obtained for the initial modelling – a random forest model with all available data – showed a prediction error (RMSE) on the test set of 1.60 t/ha for grain yield and 48.6 kg/ha for the optimal N rate which can be regarded as relatively high considering the average yields and optimal N rates obtained in each trial (Table 1). The first hypothesis of this study was that the presence of autocorrelation in the training data was limiting the model's learning and generalization capacity. A second hypothesis was that the significant variability observed among locations and experimental years (Table 1) made it challenging to generate a single 'global' model capable of performing well in the different study regions.

Autocorrelation

The correlograms show the strong presence of autocorrelation in the field data, especially for grain yield (Fig. 4). This result was expected since the data were collected sequentially and in close proximity, every 10 m along the length of each experiment. In the case of grain yield data, such autocorrelation is especially expected due to the operational principles and dynamics involved in yield monitor data collection during harvest (Whelan & McBratney, 2002).The correlograms also indicate that non-significant autocorrelation is only achieved when there is a shift in the data of around 50 to 70 lag for the optimal N rate and yield, representing a distance between observations of 500 m to 700 m. In practice, this means that for an experiment of 1500 m in length, as shown in Fig. 2 (representing a typical experiment in this project), only around two to three observations would be independent, possibly coinciding with one in each of the productivity zones of the field; noting that each observation here represents data aggregated across a 50 m radius (Fig. 3).

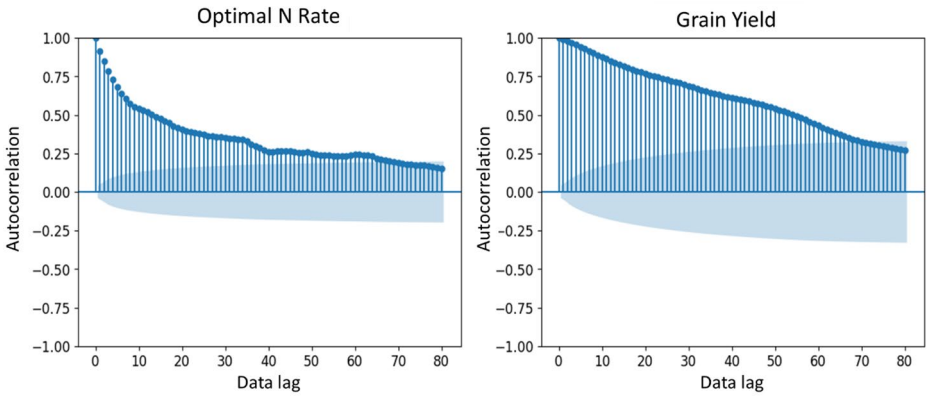

Fig. 4 Correlograms for optimal N rate and grain yield data. The blue background area represents 95% confidence intervals

The residuals obtained after fitting random forest models for optimal N rate prediction exhibited low spatial autocorrelation (Fig. 5). This suggests that, despite the spatial structure present in the original data, the trained models were generally effective in capturing the variability in this response variable. However, for grain yield – where the original data showed higher spatial autocorrelation compared to optimal N rate (Fig. 4) – the residuals still displayed some autocorrelation at short lags. This indicates that the stronger spatial dependence in grain yield data was not fully accounted for by the model, suggesting some degree of underperformance.

Figure 6 shows the effect of thinning the data on the prediction performance of grain yield and optimal N rate using a random forest model. Overall, models performed similarly across the various scenarios of data availability and autocorrelation. No statistically significant differences were observed among the filtering scenarios for either model. These results indicate that training a model using data with reduced autocorrelation – but with fewer data – did not affect prediction accuracy, as originally hypothesised. It is also seen that, having less data also did not lead to a loss of performance likely due to the variation in the dataset being maintained despite the thinning process. In addition, these results reject a possible alternative hypothesis that a larger dataset could improve model performance. As expected, abundant data do not contribute to the model performance when strong autocorrelation is present, as will inevitably be the case when highly replicated OFE designs are used.

The analysis of residual autocorrelation across the subsampling scenarios showed that, for the optimal N rate models, it remained largely unaffected by changes in the subsampling scenarios (Fig. 7). In contrast, grain yield prediction models appeared to have benefited from reduced spatial autocorrelation in the field data – residual autocorrelation decreased under sparser and less autocorrelated field data (Fig. 7) – leading to a slight improvement (although not statistically significant) in predictive performance, as shown in Fig. 6.

Regionalisation

For the analysis of model regionalisation, the ideal number of clusters for grouping the nine experimental sites based on yield and average optimal N rate data was first evaluated. The elbow graph indicated that little improvement in the clustering quality occurred beyond three clusters (Fig. 8). Thus, three clusters were used (Fig. 9). Overall, a relatively even

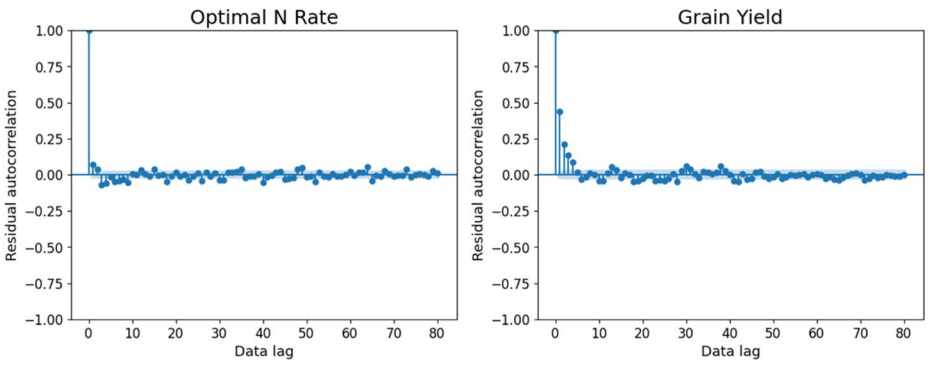

Fig. 5 Correlograms of the residuals after model fitting for optimal N rate and grain yield data. The blue background area represents 95% confidence intervals

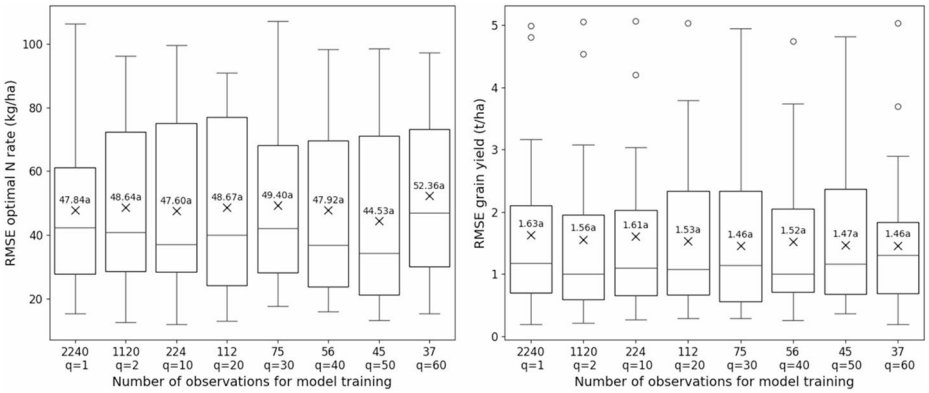

Fig. 6 Prediction error (in the test set) of models for optimal N rate and grain yield generated with progressively smaller databases and lower autocorrelation in the data. In each sampling scenario (x axis), one data point was retained every *q* observations. The boxplots indicate the error distribution across the different test sites during the cross-validation. In each box, the lower edge indicates the 25th percentile (Q1), the upper edge indicates the 75th percentile (Q3), the line inside the box indicates the median, and the cross indicates the mean. Whiskers indicate the lowest and highest data points that fall within 1.5 × (Q3 - Q1) below Q1 and above Q3. Data points outside the whiskers are considered outliers. Mean values are accompanied by a letter indicating the Tukey test result at 5% significance

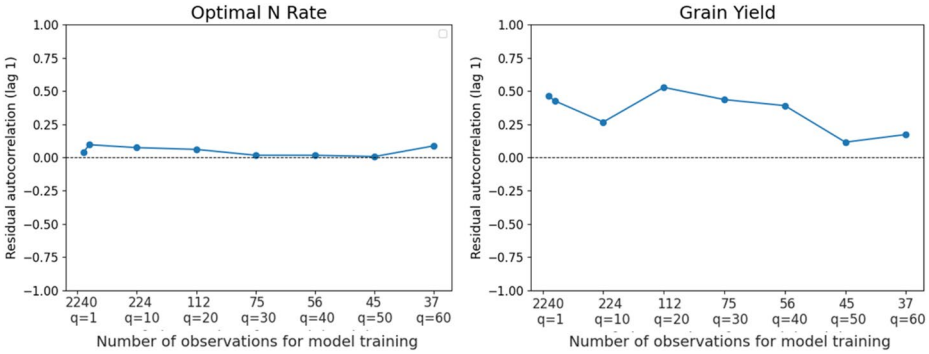

Fig. 7 Spatial autocorrelation (using a data lag of 1) of the residuals of models for optimal N rate and grain yield generated with progressively smaller databases and lower autocorrelation in the data. In each sampling scenario, one data point was retained every *q* observations

division in terms of data quantity was observed among the clusters; each cluster contained data from six to eight experiments. The first cluster (A) consisted of the Booleroo Centre and Loxton sites, totalling six experiments, three at each site. These sites are located in arid regions of South Australia, albeit ~ 300 km apart, typically characterized by low crop yield and low demand for N fertiliser and in particular, low annual and growing season rainfall; indeed the site at Booleroo Centre is north of Goyder's line (Nidumolu et al. 2012). Cluster B comprised a mix of sites with intermediate values of optimal N rate and yield, including seven experiments spread across four locations states: Kalannie - WA, Tumby Bay and Wharminda - SA, and Narrabri - NSW. The remaining three sites (Tarlee and Urania - SA, and Woorak - VIC), with higher crop yield and optimal N rate, encompassing eight experi-

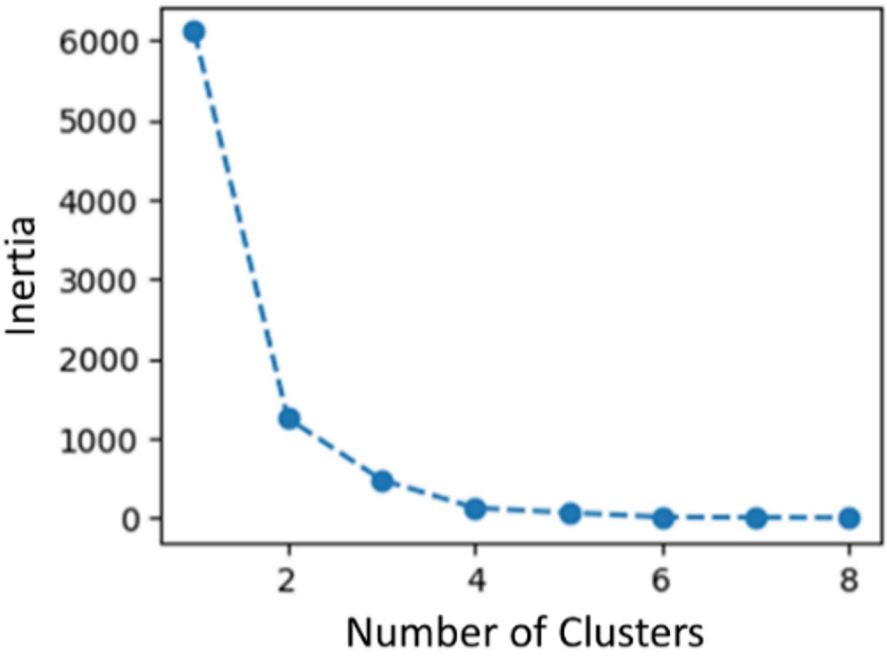

Fig. 8 Inertia (sum of squared distances between each point and the cluster centroid) for a varying number of clusters during the clustering analysis of average yield and optimal N rate data from each experimental site

ments, were grouped into cluster C. Generally, cluster A was the only one that showed some cohesiveness, even though it could still be further refined with the availability of more OFE data (region marked in red on the map, Fig. 9). Cluster C, in a more southern region of South Australia and Victoria (region marked in blue on the map, Fig. 9), and especially Cluster B, scattered around the continent, were much less sensible from an agronomic perspective.

Clustering the data reduced the model error for optimal N rate prediction only for cluster A, from 23.9 kg/ha (RMSE obtained by the global model) to 18.4 kg/ha (RMSE obtained by the regional model) (Fig. 10). Although this represents a reduction of approximately 5.5 kg/ha or 23%, the difference was not statistically significant according to Tukey's test at 5% level. Overall, the difference in performance between the global model and the regional model was very small for clusters B and C, showing a slight deterioration of performance with clustering (Fig. 10). The results seem to make sense considering the characteristics of each group. Cluster A showed greater coherence from a biophysical and agronomic perspective, representing locations in regions with consistently low yield potential and similar limitations related to climate and soil characteristics. Cluster C, although including sites with similar average yield potential and N demand (Fig. 9), had highly contrasting soil and climate conditions in different locations. For instance, Urania - SA, located on Yorke Peninsula (SA), experiences a strong influence of an oceanic climate with regular rainfall, contrasting with Tarlee (SA) and Woorak (VIC) where the weather is much less locally influenced by proximity to the sea. The soil in Urania (SA) also differs markedly from the others, being more sandy, rocky, shallow, and calcareous, often with calcrete occurring at

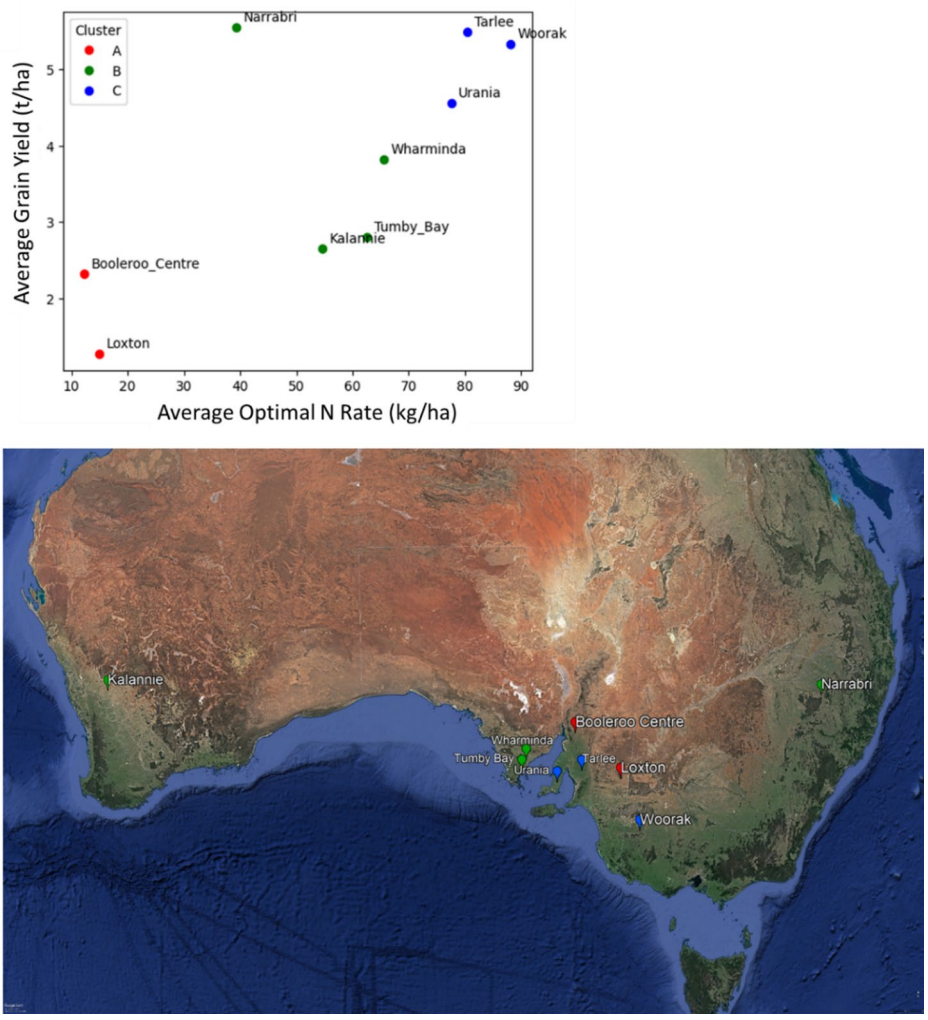

Fig. 9 Clustering of nine experimental sites based on grain yield and optimal N rate

shallow depths. Therefore, the worse result for the regional model compared to the 'global' model is not surprising. Cluster B also did not show any regional or agronomic coherence as it included distinct and geographically scattered locations; thus, it did not achieve positive effects for modelling either.

Finally, to speculate on the potential benefit of further restricting the coverage of a model, single location models were also tested. Models generated for Kalannie (WA), Tarlee (SA), and Woorak (VIC), each with three experiments (2018, 2019, and 2020), did not result in a reduction of error compared to the global model (data not shown). This indicates that, despite the specificity of the models, in this case for a single location or farm, modelling was not effective when based on limited data (only three seasons).

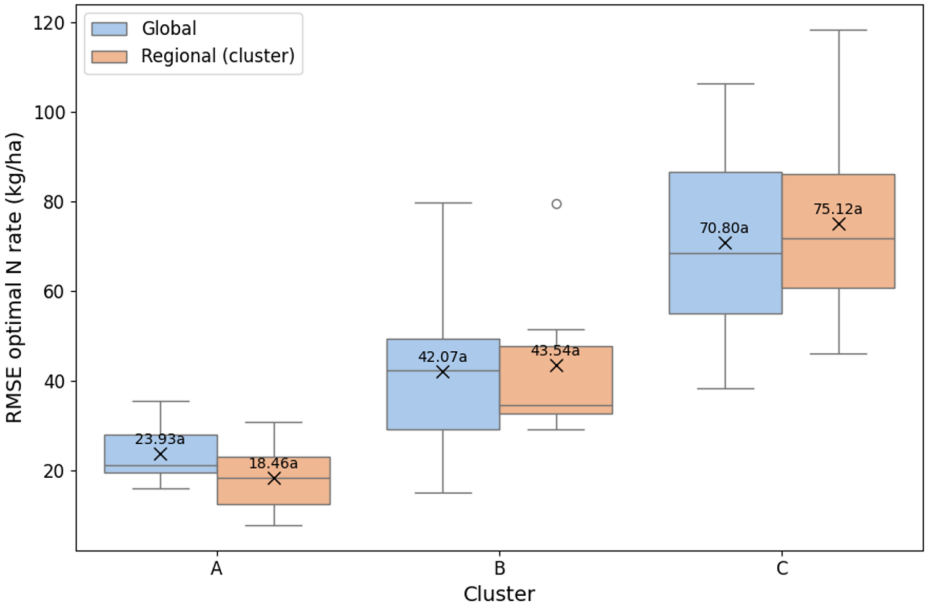

Fig. 10 Predictive performance of optimal N rate from regional models and a global model. The boxplots indicate the error distribution across the different test sites during the cross validation. In each box, the lower edge indicates the 25th percentile (Q1), the upper edge indicates the 75th percentile (Q3), the line inside the box indicates the median, and the cross indicates the mean. Whiskers indicate the lowest and highest data points that fall within 1.5 × (Q3 - Q1) below Q1 and above Q3. Data points outside the whiskers are considered outliers. Mean values are accompanied by a letter indicating the t-test (5% significance) results between global and regional models at each cluster

Discussion

Autocorrelation is a common characteristic of any OFE database, and for that matter, of almost any agronomic dataset (Mcbratney and Pringle, 1999). It is also obvious that without spatial autocorrelation (i.e. discernible spatial structure in within-field agronomic datasets), there would be no case for the targeted management that is implicit in PA. While it is accepted that autocorrelation needs to be avoided in inferential statistics – because it violates the assumption of independence among observations (Lovell, 2013), leading to artificially narrow confidence intervals and lower p-values – this study aimed to assess whether caution is also warranted when using autocorrelated databases to train ML models intended to underpin agronomic decision-making. After all, autocorrelation may imply greater redundancy in observations, potentially making extensive databases with autocorrelation less effective than a small amount of non-correlated data. While some studies paid attention to this factor in regard to model validation using spatial data splitting between training and test sets (e.g. de Lara et al., 2023), they overlooked the autocorrelation still present in the training set which may affect model training.

In this study, models remained essentially unaffected by spatial autocorrelation in field data; although a slight improvement was observed for grain yield prediction models using thinned (ie less autocorrelated) data, the practical benefit is likely negligible. Overall, results indicate that spatial autocorrelation present in OFE data may not be a major concern for

ML training. Another important lesson was that models generated from a few independent observations had a similar performance to those generated from thousands of autocorrelated observations. Similar results were obtained by Richetti et al. (2023). In that case, ML yield prediction models were created from yield monitor data, and vegetation index and climatic data as predictor variables, collected across nine fields in Australia. They found that ML models produced using 5% of the data were as accurate as those based on the entire database. It is acknowledged that, at first glance, these results may appear to contradict the findings of Colaço et al. (2024), who emphasised the importance of abundant data for model performance. However, it is crucial to clarify that the data-abundant scenario in that study was designed to represent a situation in which historical field data used to train the model perfectly reflects the current field conditions where the model is applied. Therefore, in that context, 'abundance' refers to the representativeness of the data, not simply its quantity.

In summary, the analysis of autocorrelation effects indicates that, since model performance was generally maintained across the various scenarios of data availability and autocorrelation, it can be concluded that the choice of an OFE trial design – which impacts on both data availability and data autocorrelation – can primarily follow more pragmatic considerations regarding practicality and utility of the OFE. Here, the results suggest that users such as farmers, advisors or service providers developing data driven decision tools, can opt for simpler trials such as strips (Fig. 1a) or less replicated plot-based trials (Fig. 1b) for training ML models like the one developed here. Likewise, in choosing an approach to process and analyse strip trial data, observations can be aggregated in steps greater than 10 m, which seems sensible considering the typical width of a combine header and the effects of so-called 'convolution' in yield monitor data (Whelan & McBratney, 2002). Although these designs may generate fewer data than highly replicated whole-field trials (Fig. 1c), they can still provide sufficient data for ML training – especially when strategically distributed in the field based on knowledge of management zone location. While the value of continuous representation of crop response obtained from highly replicated whole-field trials is recognised – especially for the learning potential it offers to farmers (Bullock et al., 2019; Cook & Bramley, 1998; Hegedus et al., 2023; Lacoste et al., 2022) – its added value for developing data driven decision tools based on ML may be limited. It is therefore also worth noting that an important limitation of whole-field designs is that a management decision arising from the trial – e.g. how much top dress N fertiliser to apply – cannot be implemented in the same season without it destroying the experiment, because the trial occupies the entire field area. As a consequence, the farmer must choose between implementing a trial or executing a management decision (Hegedus et al., 2023). If the additional data provided by these trials do not add value to the modelling, as the results of this study demonstrate, such a dilemma may not arise. A farmer can opt for smaller trials without compromising either the development of a decision model or their desire for management optimization in a particular season.

With that said, the discussion above – and the potential concern regarding autocorrelation – only applies to non-spatial ML approaches, which are predominant in PA applications (Chilingaryan et al., 2018; Liakos et al., 2018). On the contrary, spatial autocorrelation from whole-field trial data – and more precisely, the spatial variability it captures – is not a concern at all and is actually key, when using convolutional neural networks (CNNs) or other spatial ML approaches (Barbosa et al., 2020). However, the disadvantage of these approaches is that the spatial patterns of a given field are only useful information for that particular field. In other words, the limitation of CNNs, and other spatial approaches, lies

in the fact that models will likely be field specific. The first problem that arises from such an approach is the need to build many models, one for each field. Second, is the fact that each field-specific model might require many years of data collection to capture seasonal variability (Bullock et al., 2020). In contrast, models generated from databases that mix data from different trials – using OFE networks – may capture 'seasonal variability' by the fact that variables representing climatic conditions (for example water availability) can vary between different fields in the same season. Further, field-specific databases collected over several years tend to be less balanced in terms of the variation captured in each of the variables; that is, while on-farm sensor data varies continuously within a field, weather-related variables often change only between seasons for any given field at the spatial resolution at which they are available. For these reasons, a strategy that integrates less replicated trials (i.e., fewer observations per field) with data from multiple fields can help create more balanced datasets, which may be beneficial for model training. This aspect links to the question of regionality in model development.

While some studies have tested models covering large geographical regions with limited success (Colaço et al., 2024), others focused on field specific approaches (Bullock et al., 2020; Hegedus et al., 2023), which, as mentioned, may require several years of experimentation until an effective model can be generated such that it is useful across a range of season types. Here, an intermediate approach was tested in which OFE data from a few different fields were grouped through cluster analysis before modelling. Results show that a model for optimal N rate prediction generated and validated from two agronomically similar locations in the arid region of South Australia showed potential to outperform a more global model built using data from all regions of Australia combined. Therefore, regionalisation offers a pathway for improvement in model performance by limiting the variability in the conditions under which the models are implemented. However, regionalisation must take into account extensive agronomic and climatic knowledge for each location to group different areas in the most coherent and relevant manner and here, we stress that our clustering of sites (Fig. 9) was illustrative of an approach rather than intended to be definitive. In this connection, it is accepted that our available dataset was probably inadequate for a more robust regionalisation study, but we suggest that as a means of illustrating an approach, the present work has been useful. Future studies with access to more OFE data could not only enhance the approach used here but also explore alternative classification strategies that account for both geographical and seasonal variability.

In summary, the combined analyses of data autocorrelation and model regionalization suggest that to enhance both the quantity and quality of OFE data for ML models in the development of data-driven approaches to more robust fertiliser decision-making, prioritising a larger number of less complex experiments (e.g., less replicated strip or plot trials) across reasonably similar locations is likely more effective than increasing spatial coverage within individual experiments (e.g., highly replicated whole-field designs) focussed on field specific models. This approach is also more likely to be accepted by farmers due to the simpler trial designs, potential for more 'regional coherence' (farms, groups of nearby farms, districts, regions, etc.) and possibly the smaller amount of data required from a single location to build effective models.

Conclusion

This study investigated the implications of OFE strategies for the spatial characteristics of field data and how this, in turn, affects the development of ML decision models for N fertiliser management. The analysis focused specifically on the effect of trial design on spatial autocorrelation in field data and its impact on model training and whether regionalisation of OFE data could improve the performance of prediction models. Twenty-one N strip trials spread across Australia were used to develop ML models for optimal N rate and grain yield prediction using a range of scenarios of data autocorrelation and model geographical coverage.

In this study, spatial autocorrelation in OFE data had negligible impact on model performance. Likewise, abundant yet autocorrelated data did not add extra value for the model either. Overall, in obtaining data for constructing ML models, OFE designs can follow more pragmatic considerations. For example, less replicated strip or plot-based trials providing fewer but strategically distributed observations in contrasting regions of a field might be preferable – for their simplicity and pragmatism – to highly replicated designs that aim to cover the entire extent of a field. Similarly, in order to obtain diverse and independent data, increasing the number of experiments in different locations and years might be preferable to increasing the spatial coverage of observations within a single experiment.

In finding the right combination of data from different trials, limiting the geographical extent of an OFE trial network may improve model performance compared to a global model encompassing areas with distinct agronomic and climatic characteristics. This strategy reduces the variability of conditions under which the models are implemented. However, it will be essential to define the geographical coverage of such OFE network based on relevant field characteristics and expert agronomic knowledge so that regions are sensible from an agronomic perspective with further work in this area highly desirable.

Acknowledgements This research was cofunded by the Commonwealth Scientific and Industrial Research Organisation (CSIRO), the University of Sydney, Agriculture Victoria, the University of Southern Queensland, the Queensland University of Technology and the Grains Research and Development Corporation (GRDC). We are also most grateful to the various farmers who collaborated with fields and trial data — Ashley Wakefield, Ben Pratt, Bob Nixon, Ed Hunt, Jessica and Joe Koch, Mark and Sam Branson, Mark Swaffer, Rob Cole, Robin Schaeffer, Stuart Modra, Kieran Shephard and Peter Bell. We also thank Prof José Molin (USP) for a critical review of this manuscript.

Funding Open access funding provided by CSIRO Library Services.

Declarations

Competing interests The authors declare that they have no conflict of interest.

References

Anselmi, A. A., Molin, J. P., Bazame, H. C., & Corrêdo, L. P. (2021). Definition of optimal maize seeding rates based on the potential yield of management zones. *Agriculture*, *11*, 911. https://doi.org/10.3390/agriculture11100911

ASRIS (2024). Australian Soil Resource Information System. Available online: https://www.asris.csiro.au (last accessed, 12/08/2024).

Barbosa, A., Trevisan, R., Hovakimyan, N., & Martin, N. F. (2020). Modeling yield response to crop management using convolutional neural networks. *Computers and Electronics in Agriculture*. https://doi.org/10.1016/j.compag.2019.105197

Bishop, T. F. A., & Lark, R. M. (2006). The geostatistical analysis of experiments at the landscape-scale. *Geoderma, 133*(1–2), 87–106. https://doi.org/10.1016/j.geoderma.2006.03.039

BOM (2024). Bureau of Meteorology - Climate Data Online. Available online: http://www.bom.gov.au/climate/data (last accessed, 12/08/2024).

Bullock, D. S., Boerngen, M., Tao, H., Maxwell, B., Luck, J. D., Shiratsuchi, L., Puntel, L., & Martin, N. F. (2019). The data-intensive farm management project: Changing agronomic research through on-farm precision experimentation. *Agronomy Journal, 111*(6), 2736–2746. https://doi.org/10.2134/agronj2019.03.0165

Bullock, D. S., Mieno, T., & Hwang, J. (2020). The value of conducting on-farm field trials using precision agriculture technology: A theory and simulations. *Precision Agriculture*, *21*(5), 1027–1044. https://doi.org/10.1007/s11119-019-09706-1

Chlingaryan, A., Sukkarieh, S., Whelan, B., Computers and Electronics in Agriculture. (2018). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *Computers and Electronics in Agriculture, 151*, 61–69. https://doi.org/10.1016/j.compag.2018.05.012

Colaço, A. F., & Bramley, R. G. V. (2018). Do crop sensors promote improved nitrogen management in grain crops? *Field Crops Research, 218*, 126–140. https://doi.org/10.1016/j.fcr.2018.01.007

Colaço, A. F., Pagliuca, L. G., Romanelli, T. L., & Molin, J. P. (2020). Economic viability, energy and nutrient balances of site-specific fertilisation for citrus. *Biosystems Engineering*, *200*, 138–156. https://doi.org/10.1016/j.biosystemseng.2020.09.007

Colaço, A. F., Richetti, J., Bramley, R. G. V., & Lawes, R. A. (2021). How will the next-generation of sensor-based decision systems look in the context of intelligent agriculture? A case-study. *Field Crops Research*. https://doi.org/10.1016/j.fcr.2021.108205

Colaço, A. F., Whelan, B. M., Bramley, R. G. V., Richetti, J., Fajardo, M., McCarthy, A. C., Perry, E. M., Bender, A., Leo, S., Fitzgerald, G. J., & Lawes, R. A. (2024). Digital strategies for nitrogen management in grain production systems: Lessons from multi-method assessment using on-farm experimentation. *Precision Agriculture*, *25*(2), 983–1013. https://doi.org/10.1007/s11119-023-10102-z

Cook, S. E., & Bramley, R. G. V. (1998). Precision agriculture - Opportunities, benefits and pitfalls of site-specific crop management in Australia. *Australian Journal of Experimental Agriculture*, *38*(7), 753–763. https://doi.org/10.1071/EA97156

Córdoba, M., Paccioretti, P., & Balzarini, M. (2025). A new method to compare treatments in unreplicated on-farm experimentation. *Precision Agriculture*. https://doi.org/10.1007/s11119-024-10206-0

Cressie, N. A. C. (1993). Statistics for Spatial Data. Revised edition. New York: Wiley. https://doi.org/10.1002/9781119115151

de Lara, A., Mieno, T., Luck, J. D., & Puntel, L. A. (2023). Predicting site-specific economic optimal nitrogen rate using machine learning methods and on-farm precision experimentation. *Precision Agriculture*, *24*(5), 1792–1812. https://doi.org/10.1007/s11119-023-10018-8

Evans, F. H., Salas, A. R., Rakshit, S., Scanlan, C. A., & Cook, S. E. (2020). Assessment of the use of geographically weighted regression for analysis of large on-farm experiments and implications for practical application. *Agronomy*. https://doi.org/10.3390/agronomy10111720

Hegedus, P. B., Maxwell, B., Sheppard, J., Loewen, S., Duff, H., Morales-Luna, G., & Peerlinck, A. (2023). Towards a Low-Cost comprehensive process for On-Farm precision experimentation and analysis †. *Agriculture (Switzerland)*, *13*(3). https://doi.org/10.3390/agriculture13030524

Jakubauskas, M., & Legates, D. R. (2000). Harmonic analysis of time-series AVHRR NDVI data for characterizing U.S. Great plains land use/land cover. *International Archives of for Photogrammetry and Remote Sensing*, *32*, 384–389.

Jin, H., Shuvo Bakar, K., Henderson, B. L., Bramley, R. G. V., & Gobbett, D. L. (2021). An efficient geostatistical analysis tool for on-farm experiments targeted at localised treatment. *Biosystems Engineering, 205*, 121–136. https://doi.org/10.1016/j.biosystemseng.2021.02.009

Kindred, D. R., Milne, A. E., Webster, R., Marchant, B. P., & Sylvester-Bradley, R. (2015). Exploring the spatial variation in the fertilizer-nitrogen requirement of wheat within fields. *Journal of Agricultural Science, 153*(1), 25–41. https://doi.org/10.1017/S0021859613000919

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., García, F. O., Maxwell, B., Griffin, T., Oberthür, T., Huyghe, C., & Hall, A. (2022). On-Farm Experimentation to transform global agriculture. In Nature Food (Vol. 3, Issue 1, pp. 11–18). Springer Nature. https://doi.org/10.1038/s43016-021-00424-4

Lawes, R. A., & Bramley, R. G. V. (2012). A simple method for the analysis of on-farm strip trials. *Agronomy Journal, 104*(2), 371–377. https://doi.org/10.2134/agronj2011.0155

Lawes, R. A., Oliver, Y. M., & Huth, N. I. (2019). Optimal nitrogen rate can be predicted using average yield and estimates of soil water and leaf nitrogen with infield experimentation. *Agronomy Journal*, *111*(3), 1155–1164. https://doi.org/10.2134/agronj2018.09.0607

Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. In Sensors (Switzerland) (Vol. 18, Issue 8). MDPI AG. https://doi.org/10.3390/s18082674

Liu, X., Kounadi, O., & Zurita-Milla, R. (2022). Geo-information incorporating spatial autocorrelation in machine learning models using spatial lag and eigenvector spatial filtering features. *International Journal of Geo-Information, 11*, 242. https://doi.org/10.3390/ijgi

Lovell, D., P (2013). Biological importance and statistical significance. *Journal of Agricultural and Food Chemistry*, *61*, 8340–8348. https://doi.org/10.1021/jf401124y

Mcbratney, A. B., & Pringle, M. J. (1999). Estimating average and proportional variograms of soil properties and their potential use in precision agriculture. *Precision Agriculture, 1*, 125–152.

Meyer, H., Reudenbach, C., Wöllauer, S., & Nauss, T. (2019). Importance of spatial predictor variable selection in machine learning applications – Moving from data reproduction to spatial prediction. *Ecological Modelling*. https://doi.org/10.1016/j.ecolmodel.2019.108815

Nidumolu, U. B., Hayman, P. T., Howden, S. M., & Alexander, B. M. (2012). Re-evaluating the margin of the South Australian grain belt in a changing climate. *Climate Research*, *51*(3), 249–260. https://doi.org/10.3354/cr01075

Panten, K., & Bramley, R. G. V. (2011). Viticultural experimentation using whole blocks: Evaluation of three floor management options. *Australian Journal of Grape and Wine Research, 17*, 136–146. https://doi.org/10.1111/j.1755-0238.2011.00129.x

Panten, K., & Bramley, R. G. V. (2012). Whole-of-block experimentation for evaluating a change to canopy management intended to enhance wine quality. *Australian Journal of Grape and Wine Research, 18*(2), 147–157. https://doi.org/10.1111/j.1755-0238.2012.00183.x

Poudjom, D. Y., & Minty, B. R. S. (2019). *Radiometric grid of Australia (Radmap) v4 2019 filtered Ppm thorium*. Geoscience Australia. https://doi.org/10.26186/5dd48e3eb6367

Richetti, J., Diakogianis, F. I., Bender, A., Colaço, A. F., & Lawes, R. A. (2023). A methods guideline for deep learning for tabular data in agriculture with a case study to forecast cereal yield. *Computers and Electronics in Agriculture*. https://doi.org/10.1016/j.compag.2023.107642

Tanaka, T. S. T., Heuvelink, G. B. M., Mieno, T., & Bullock, D. S. (2024). Can machine learning models provide accurate fertilizer recommendations? *Precision Agriculture*, *25*(4), 1839–1856. https://doi.org/10.1007/s11119-024-10136-x

Trevisan, R. G., Bullock, D. S., & Martin, N. F. (2021). Spatial variability of crop responses to agronomic inputs in on-farm precision experimentation. *Precision Agriculture*, *22*(2), 342–363. https://doi.org/10.1007/s11119-020-09720-8

Whelan, B. M., & McBratney, A. B. (2002). A parametric transfer function for grain-flow within a conventional combine harvester. *Precision Agriculture*, *3*, 123–134.

Whelan, B. M., Taylor, J. A., & McBratney, A. B. (2012). A small strip approach to empirically determining management class yield response functions and calculating the potential financial net wastage associated with whole-field uniform-rate fertiliser application. *Field Crops Research, 139*, 47–56. https://doi.org/10.1016/j.fcr.2012.10.012

Wilson, N., Tickle, P. K., Gallant, J., Dowling, T., & Read, A. (2011). *1 second SRTM derived hydrological digital elevation model (DEM-H) version 1.0. Record 1.0.4*. Geoscience Australia.

Zadoks, J. C., Chang, T. T., & Konzak, C. F. (1974). A decimal growth code for the growth stages of cereals. *Weed Research*, *14*(14), 415–421.

Authors and Affiliations

André F Colaço[1,2] · Robert GV Bramley[1,3] · Jonathan Richetti[4] · Roger A Lawes[4]

✉ André F Colaço
andre.colaco@usp.br

[1] CSIRO, Waite Campus, Glen Osmond, SA 5064, Australia

[2] Present address: University of São Paulo, Luiz de Queiroz College of Agriculture, Piracicaba 13418-900, SP, Brazil

[3] Present address: 'robbramley.ag', Myrtle Bank, Glen Osmond, SA 5064, Australia

[4] CSIRO, Floreat, WA 6014, Australia