Field Crops Research 319 (2024) 109619

Contents lists available at ScienceDirect

Field Crops Research

journal homepage: www.elsevier.com/locate/fcr

Combining production ecology principles with random forest to model potato yield in China

Qiuhong Huang [a,b,c], Gerard B.M. Heuvelink [b,c,*], Ping He [a,**], Johan G.B. Leenaars [c], Antonius G.T. Schut [d]

[a] State Key Laboratory of Efficient Utilisation of Arid and Semi-Arid Arable Land in Northern China, Key Laboratory of Plant Nutrition and Fertiliser, Ministry of Agriculture and Rural Affairs/Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences (CAAS), Beijing 100081, PR China
[b] Soil Geography and Landscape Group, Wageningen University, PO Box 47, Wageningen, 6700 AA, the Netherlands
[c] ISRIC - World Soil Information, PO Box 353, Wageningen, 6700 AJ, the Netherlands
[d] Plant Production Systems Group, Wageningen University, PO Box 430, Wageningen, 6700 AK, the Netherlands

ARTICLE INFO

Keywords:
Cross-validation
Machine learning
Measurement error
Uptake
Uncertainty

ABSTRACT

Context: The random forest model (RF) has been widely applied for crop yield prediction. However, extrapolation, measurement errors, and uncertainty arising from limited predictive power of covariates may affect the model performance.
Objective: This study aimed to interpret and assess the accuracy of RF for potato yield prediction in China and quantify the main sources of uncertainty using the C.T. de Wit's three-quadrant diagram.
Methods: A dataset including 2182 plot-year combinations was derived from 63 potato field experiments covering nine Chinese provinces and three years. Model performance was evaluated by 10-fold cross-validation (CV), leave-block-out (LBOCV), leave-site-out (LSOCV), and leave-year-out cross-validation (LYOCV).
Results: The root mean square error (RMSE) was 3.5, 8.3, 9.9 and 10.3 t ha^{-1}, while the model efficiency coefficient (MEC) was 0.92, 0.64, 0.52 and 0.43 for 10-fold CV, LBOCV, LSOCV and LYOCV, respectively. Cumulated sunshine duration and topography position index were the most important covariates, while fertiliser variables were identified as least important for yield modelling. The standard deviation of the yield replicate variability estimated by a linear model accounted for 32 % of the RMSE for LSOCV. Introducing measured uptake of nutrient omission treatments, uptake of all treatments, and yields of nutrient omission treatments as additional covariates decreased the LSOCV RMSE by 2.3 t ha^{-1} on average.
Conclusions: The fitted models could explain up to 92 % of potato yield variability in China, although there was a considerable residual error when extrapolating to other areas or years. Yield replicate variability accounted for one-third of the residual error. Information about physiological efficiency was the main source of uncertainty, followed by available soil nutrients. Fertiliser recovery was least important because most of the experiments were conducted in fertile fields.
Implications: Combining a RF model with the three-quadrant diagram allows to better explain yield prediction uncertainty. The methodology used in this study can be applied to other crops, countries and data-driven models.

1. Introduction

Potato (Solanum tuberosum L.), a crop with a high yield potential, provides food security for many countries in the world. China is the world's largest potato producer with 5.7 million hectares and a total production of 95 million tons in 2022, accounting for 32 % of the harvested area and 25 % of the production in the world (FAO Food and Agriculture Organization, 2024). Over the past two decades, there has been an upward trend in potato production in China, increasing from 70 million tons in 2012 to 96 million tons in 2022 (FAO Food and Agriculture Organization, 2024).

Reliable yield prediction can provide valuable information for agricultural decisions related to management optimisation and resource allocation (Malone et al., 2007; Morell et al., 2016). There are two main

* Corresponding author at: Soil Geography and Landscape Group, Wageningen University, PO Box 47, Wageningen, 6700 AA, the Netherlands.
** Corresponding author.
E-mail addresses: gerard.heuvelink@wur.nl (G.B.M. Heuvelink), heping02@caas.cn (P. He).

https://doi.org/10.1016/j.fcr.2024.109619
Received 21 June 2024; Received in revised form 16 September 2024; Accepted 16 October 2024
Available online 4 November 2024
0378-4290/

approaches for yield prediction, process-based crop growth models (PBM) and data-driven statistical models, in particular machine learning models (MLM). PBMs consider logical principles of soil-plant-environment systems and simulate specific processes and their interactions. But PBMs need parameters, and a wide range of high-quality input variables, such as daily weather data and detailed soil and management information (Fleisher et al., 2017; Jones et al., 2017; Schut and Giller, 2020). Unlike PBM, machine learning models have flexible frameworks. MLMs can adaptably incorporate a wide set of data, such as local crop management (e.g., land preparation, cultivars, irrigation, and fertiliser history), natural phenomena (e.g., extreme weather effects, plant disease epidemics, and pest outbreaks), and socioeconomic changes, which provides valuable information to support decision-making (Lee et al., 2022; Ebrahimy et al., 2023). But MLMs also have limitations. For instance, extrapolation to other areas or years may produce unreliable results when predictions are made for conditions very different from those of the training data (Meyer and Pebesma, 2021). Another drawback is that MLMs ignore the spatial and temporal structure of the observations, and hence suitable cross-validation schemes are needed when evaluating model performance (de Bruin et al., 2022; Milà et al., 2022). Additionally, MLMs are not well suited to improve our understanding of mechanistic processes as they are purely data-driven (Scowen et al., 2021).

Among MLMs, the widespread use of the random forest (RF) model is attributed to its high accuracy, flexibility, cost-effectiveness, and minimal parameter requirements. RF allows for ranking the relative importance of covariates for model prediction. This ability is often used to unravel the key biophysical or crop management drivers of yield variability under varying growing conditions and management practices (Devkota et al., 2023; Raharimanana et al., 2023). RF has been applied in many important agriculture-related subjects, such as modelling yield response to fertiliser or climate change (Deng et al., 2023; Falconnier et al., 2023), explaining yield variability at various spatial and temporal scales (Paudel et al., 2021; Nayak et al., 2022), providing recommendations for future sowing and harvest plans (Krell et al., 2022) and developing fertilisation strategies (Wang et al., 2021). The RF model has widely been used for yield predictions of rice (Eugenio et al., 2023), wheat (Devkota et al., 2023), maize (Wang et al., 2021), but has rarely been tested for potato yields in China.

No model is perfect because any model is a simplified representation of reality and uncertainty and error always exist in model prediction. The uncertainty sources of MLMs mainly arise from measurement errors in the training data, insufficient training data, and the limited ability of covariates to explain the variability of the target variable. Measurement error refers to the inaccuracy introduced during data collection and measurement caused by varying measurement conditions, methods and instruments (van Leeuwen et al., 2021). Despite being a primary source of uncertainty in model predictions, measurement errors are rarely quantified. But even if measurement errors were negligibly small and a large calibration set is available for model training, the output remains imperfect because available covariates cannot explain all variability of the target variable. The covariates might not adequately represent key processes because they are merely proxies of the underlying mechanisms.

Explaining the variability in observed yields and yield prediction for a new field is a challenging task because it requires domain knowledge to understand the underlying relationships between nutrient supply (from soil and fertiliser), uptake and yield. In the three-quadrant diagram of de Wit (1953), the fertiliser supply-yield relationship is decomposed into NPK fertiliser supply-uptake relationships and NPK uptake-yield relationships. Only a portion of the nutrients supplied are available for crop uptake, with a typical recovery value of 0.1–0.6 for N, P and K. Nutrient recovery is influenced by the root morphology, soil nutrient contents before planting, amounts of N, P and K applied and crop demand. Fertiliser nutrients can fill the gap between soil supply and crop demand (Gondwe et al., 2020). The amount of nutrients ultimately absorbed and utilized by the crop depends on the congruence between plant demand and nutrient release from fertiliser and soil (de Wit, 1992). Finally, a proportion of the nutrient uptake is assigned to reproductive and storage organs, ultimately contributing to tuber yields. This process is largely influenced by genotype and environmental stress (nutrient deficiencies, drought stress, salt stress and heat stress) (Radanielson et al., 2018; Gupta et al., 2020). The relationships described above are complex and subject to considerable uncertainty, which can also impact yield modelling. To the best of our knowledge, the contributions of uncertainty in soil nutrients supply, fertiliser recovery, and physiological efficiency on yield prediction uncertainties have not been clearly quantified in previous research.

To fill the research gaps mentioned above, this study used potato cultivation in China as an example to: (1) calibrate an RF model and assess its performance to model potato yield in China; (2) interpret this RF model using variable importance plots; (3) estimate the potato yield replicate variability; (4) based on the three-quadrant diagram, introduce measured uptake in fertiliser nutrient omission treatments (i.e. uptake from soil nutrients), uptake of all treatments (i.e. uptake from soil nutrients and fertiliser nutrients), and yields in nutrient omission treatments (i.e. yield produced from soil nutrients) as extra covariates to understand and quantify the sources of potato yield prediction uncertainty.

2. Materials and methods

2.1. Study area

Considering the different cropping seasons, meteorological conditions and geographical locations, China can be classified into four different potato cultivation zones with distinctive characteristics (Teng et al., 1989) (Fig. 1). The most important zones for potato production in China are the northern single-crop zone and southwestern mixed single and multi-crop zone, accounting for 49 % and 39 % of the planting area, respectively, while the central double-crop zone and southern winter crop zone account for only 7 % and 5 % of the planting area.

The northern zone has a cold temperate climate, with average annual rainfall varying from 250 to 1000 mm. In this zone potatoes are typically planted in April and harvested in September. In the southwestern zone, potato production is primarily in mountainous areas at altitudes ranging from 700 to 3000 m, with an average annual air temperature of 6–22 ℃, and average annual precipitation 500–1500 mm. In this region, potatoes are planted in river valleys in winter and in high altitude mountainous areas in spring. The central zone experiences long summers and have an average annual temperature of 10–18 ℃ and average annual precipitation of 500–1750 mm. Potatoes are planted as the first crop in March or as the second crop in August. The southern winter crop zone has a maritime climate and experiences more than 300 frost-free days every year, with an annual average temperature of 18–24 °C. Potatoes are typically planted in autumn or winter in this region. This study included the northern, central and southwestern zones (Table 1).

2.2. Datasets

2.2.1. Field experimental data

A dataset including 2182 plot-year combinations was collected from a total of 63 field experiments covering nine Chinese provinces from 2017 to 2019 (Fig. 1) (Xu et al., 2019, 2022; Jiang et al., 2023). Each experiment was laid out as a randomized complete block design and included nutrient omission treatments with PK, NK, NP fertiliser, local farmer practices, as well as optimised NPK fertilisation treatments. The fertiliser application rates of experimental treatments are shown in Fig. S1. Each treatment had three replicate plots. These plots were randomly arranged in a block and separated by 1-meter pathways, with individual plot sizes ranging from 30 to 60 m^2. We had a total of 79 blocks. Potato tuber fresh yield response to fertiliser application is

Fig. 1. Geographic locations of potato experimental plots in different agroecological zones in China. The green dots with different sizes represent the number of experimental plots.

Table 1
Growing periods, cultivars and planting densities used in the selected field experiments from 2017 to 2019. The numbers in brackets represent the number of plot-year combinations in each province.

Zone	Province	Planting date (mm/dd)	Harvest date (mm/dd)	Potato cultivars	Planting density (Plants m^{-2})
Northern single-crop	Inner Mongolia (240)	05/07–05/12	09/13–09/22	Kennebec	4.0–5.0
	Gansu (158)	04/26–05/04	10/10–10/25	Longshu10	5.3
	Heilongjiang (286)	04/30–05/18	09/08–09/19	Yanshu4, Kexin13, Innovator	6.0–7.5
	Jilin (107)	04/24–04/26	09/05–09/07	Favorita	6.0
	Shanxi (706)	04/27–05/18	09/24–10/15	Jinshu6, Tongshu29, Datongliwaihuang	4.5
Southwestern mixed single and multi-crop	Guizhou (141)	02/15–03/28	06/23–07/06	Weiyu5	5.7
	Yunnan (162)	03/09–03/16	09/19–09/25	Kaihuashu1	4.8
	Sichuan (266)	09/13–09/25	12/14–12/21	Xingjia2, Favorita	8.5
Central double-crop	Jiangxi (116)	01/22–03/12	05/06–06/05	Favorita	6.8

shown in Fig. S2. The growing periods, potato cultivars, and planting densities are listed for each province in Table 1. The potato cultivars varied among different provinces. The planting density varied from 4.0 to 8.5 plants m^{-2}. Weeds, pests, and diseases were well controlled by spraying herbicides and pesticides. Potato experiments in Inner Mongolia used irrigation, while no irrigation was used in other provinces.

Before planting, five soil samples from 0 to 30 cm depth within each block were randomly collected, mixed, and placed in a cool and ventilated area for air-drying. These samples were then taken into a laboratory to determine soil pH, organic matter (g kg^{-1}), alkali-hydrolyzable nitrogen (g kg^{-1}), Olsen-P (mg kg^{-1}) and exchangeable K (mg kg^{-1}). Soil pH was determined by a pH meter (Seven Excellence S470-K, Mettler Toledo, Switzerland) with a soil-to-water ratio of 1:2.5. The soil organic matter was analysed using the Walkley-Black method (Walkley and Black, 1934). Alkali-hydrolyzable nitrogen was measured by the alkali hydrolyzation-diffusion method (Mulvaney and Khan, 2001). A total of 412 soil samples were analysed for total N instead of alkali-hydrolyzable N. Soil alkali-hydrolyzable N and total N were strongly correlated in this study, so a linear regression model based on 356 paired observations at a given total N was fitted to estimate the 412 missing soil alkali-hydrolyzable nitrogen values ($R^2 = 0.90$) (Fig. S3). Plant available P was determined by the Olsen method (Olsen-P) and exchangeable K was measured with the ammonium acetate method (Chinese Society of Soil Science, 2000). Table S1 shows the proportion of the soil samples at five different nutrient levels, i.e. very abundant, abundant, moderate, deficient, very deficient, and extremely deficient nutrient level (National Soil Survey Office. Soils of China, 1998). At harvest, two rows of plants (15–30 m^2) were harvested from the middle of each plot to measure tuber fresh yield. Three to five representative plants were sampled from each plot and separated into tubers and straw, and oven-dried (60 ℃ for 72 h) to determine dry matter contents and calculate dry matter weights. Plant samples were digested with H_2SO_4-H_2O_2 to determine total N by Kjeldahl, total P with the vanado-molybdate yellow color method, and total K with flame spectrophotometry, respectively (Chinese Society of Soil Science, 2000).

2.2.2. Covariates

The final dataset contained a total of 38 covariates (Table S2). Nineteen covariates were obtained from the potato field experiments. The other 19 covariates were collected from secondary data sources, as described below. A statistical summary of these variables is given in Table S3.

Daily average temperature, maximum temperature, minimum temperature, evaporation, precipitation, and sunshine duration were collected from meteorological stations near the experimental location. Soil temperatures above 25 °C can delay or even impede emergence, reduce plant survival, and reduce the number of main stems per plant (Sale, 1979; Midmore, 1984). Exposure to freezing temperatures can damage the plant tissue, particularly the foliage (Burke and Hatfield, 1987). This damage hinders the plant's ability to photosynthesize and produce energy. We calculated the number of days during which the minimum daily temperature is below 0 ℃, the days during which maximum daily temperature is higher than 25℃, the number of days to reach the effective accumulated temperature required by potato maturity (950 ℃ for early-maturity, 1115 ℃ for medium-maturity cultivars and 1205 ℃ for late-maturity cultivars), the cumulative difference in precipitation and evaporation 30 days, 60 days and 90 days after planting, and the cumulated sunshine duration during growing months. The soil cation exchange capacity, soil bulk density and soil texture were derived from the SoilGrids dataset version 2.0 (Poggio et al., 2021). Other datasets used included the EarthEnv-DEM90 dataset (Robinson et al., 2014) for elevation, MODIS for enhanced vegetation index, and the WorldGrids (Reuter and Hengl, 2012) dataset for topographic position index, downslope curvature, upslope curvature and SAGA wetness index.

2.3. Modelling tuber yield

2.3.1. Model calibration and prediction

As an ensemble machine learning algorithm, RF constructs a collection of decision trees using bootstrap sampling (Breiman, 2001) and then averages the outputs of all decision trees to make the final prediction. A decision tree is a tree where each branch node represents a choice among multiple alternatives, and each leaf represents a decision (Myles et al., 2004).

In this study, RF models were used for modelling potato fresh yield using 38 covariates (Table S2) and 2182 observations. To provide more comprehensive information and improve model interpretation, we retained all 38 covariates without applying a variable reduction technique, such as recursive feature elimination, despite high correlations among some covariates (Fig. S4). The Pearson correlations between organic N, organic P, and organic K fertiliser application and between Alkali N and soil organic matter were larger than 0.90, while between other variables it was below 0.85. We used default hyper-parameters in the *ranger* function of R, setting mtry to 6 (the square root of the number of covariates), ntree to 500, and node size to 5.

2.3.2. Model evaluation

Cross-validation schemes are commonly applied to assess model performance. Compared to a single split into a training and test set, they offer the advantage of utilizing all data for testing. In k-fold cross-validation, the original dataset is randomly divided into k equally sized subsets. One subset is used for model testing, while the remaining k-1 subsets are used as training data. This process is repeated k times, with each of the subsets used once as a test set. However, for datasets that are spatially clustered, k-fold cross-validation will likely generate overoptimistic results (Roberts et al., 2017; Wadoux et al., 2021; de Bruin et al., 2022), because the test data will often be from locations close to training data locations. Similarly, using data from the same year for training as well as testing may also lead to overly optimistic model performance metrics. Leave-block-out (LBOCV), leave-site-out (LSOCV), and leave-year-out (LYOCV) cross-validation are designed to evaluate the performance of a model when extrapolating the model to new blocks, sites, or years (Takoutsing and Heuvelink, 2022; Silva et al., 2023). Which cross-validation scheme best reflects the actual model performance depends on the model purpose that a user has in mind. For instance, if a user intends to use a model to predict the yield of a future year, then LYOCV would be most appropriate. In addition to 10-fold cross-validation, the model performance was therefore also evaluated by LBOCV, LSOCV, and LYOCV (Takoutsing and Heuvelink, 2022; Silva et al., 2023). Here, a 'site' was defined as a group of blocks that are within 200 m distance from each other. We had 46 sites that contained one block, 15 sites that contained two blocks and one site that included three blocks.

We employed three model evaluation metrics: mean error (ME), root mean squared error (RMSE), and the model efficiency coefficient (MEC). ME was used to assess the model's systematic error (bias), while the RMSE includes the effect of both systematic and random errors and provides a measure of how well a model's predictions align with observed data. In hydrology, MEC is known as the Nash Sutcliffe Model Efficiency (Nash and Sutcliffe, 1970). The MEC equals 1 in case of a perfect model, while it is 0 for a model that is as good as taking the mean of all observations as a prediction. A negative MEC indicates a very poor model that performs worse than taking the mean of all observations for prediction (McCuen et al., 2006; Criss and Winston, 2008).

$$\mathrm{ME} = \frac{1}{n}\sum_{i=1}^{n}(O_i - P_i) \tag{1}$$

$$\mathrm{RMSE} = \frac{1}{n}\sqrt{\sum\nolimits_{i=1}^{n}(P_i - O_i)^2} \tag{2}$$

$$\mathrm{MEC} = 1 - \frac{\sum_{i=1}^{n}(P_i - O_i)^2}{\sum_{i=1}^{n}(O_i - \overline{O})^2} \tag{3}$$

where n is the number of observations, O_i and P_i are the i-th observation and prediction, respectively; and $\overline{O}$ is the mean of the observations.

2.3.3. Model interpretation

Variable importance plots were used to rank the importance of the 38 covariates. Variable importance is evaluated based on changes in the model evaluation metrics upon random permutation of the data (Strobl et al., 2008). A variable is considered "important" if random shuffling its values increases the model error considerably because in this case the model relied on that variable for the prediction (Fisher et al., 2019). The estimation of variable importance involves four steps: (1) calculating the model error in the original model, (2) re-estimating the model error by shuffling the values of the selected covariate (permuted model), (3) computing variable importance values, i.e., the ratio of the model error in the permuted model to the model error in the original model, and (4) ranking the variables based on the descending order of variable importance values. Variable importance was estimated with the *importance* function of the *iml R* package (Molnar, 2018). One must be careful to interpret the variable importance of highly correlated covariates. When covariates are strongly correlated, the values obtained by permutation might be misleading and result in incorrect ranking of variable importance (Hooker et al., 2021). In this case, one alternative method is to permute a group of correlated covariates instead of each individual covariate (Wadoux and Molnar, 2022).

To reduce the disturbing effect of strong correlations among covariates on variable importance statistics and evaluate the relative importance of different variable groups, we also computed the variable importance of each of five variable groups (VG): soil (S), fertiliser (F), management (M), weather (W) and topography (T) (Table S2). RF models were separately trained for these five different variable groups and their performance assessed with 10-fold CV, LBOCV, LSOCV, and LYOCV. Eq. (4) was used to quantify the relative importance (RI) of each variable group (Torres-Matallana et al., 2021). A higher RI value indicates a more important variable group.

$$\mathrm{RI}(VG_i) = \frac{\mathrm{MEC}(VG_i)}{\sum_{j=1}^{5}\mathrm{MEC}(VG_j)} \tag{4}$$

2.3.4. Analysis of yield prediction uncertainty

We analysed the model prediction uncertainty and attributed it to yield replicate variability, soil nutrient supply uncertainty, fertiliser recovery uncertainty, and physiological efficiency uncertainty. Yield replicate variability is the combined effect of yield measurement error and yield differences between plots that have the same block, treatment and year. We estimated the yield replicate variability by fitting a linear model for fresh potato yield using treatment, block, year and their interactions as covariates in the *lm R* function (Table S4). The standard deviation of the residual error was computed and compared with the RMSE of the RF model, thus providing insight into the effect of yield replicate variability on model performance. If the standard deviation of yield replicate variability and the RMSE have a similar magnitude then yield replicate variability is a major source of model uncertainty, while it would play only a minor role if the RMSE is much bigger than the standard deviation of yield replicate variability. Note that in this study the value of all covariates, including soil properties and climate, were the same for all plots that had the same block, treatment and year. This means that the performance of the RF model cannot be better than a linear model that uses block, treatment and year and all their interactions as explanatory variables. Therefore, the RMSE of the RF model cannot be smaller than the standard deviation of the linear model residual error (i.e., the standard deviation of the yield replicate variability).

The three-quadrant diagram (Fig. 2) (de Wit, 1953) was introduced to further separate the uncertainty sources existing in the fertiliser supply-uptake relationship (quadrant IV) and uptake-yield relationship (quadrant I). We first added the intercept of the fertiliser supply-uptake relationship (yellow dot of quadrant IV in Fig. 2) as an extra covariate to evaluate how much this decreases the RMSE and increases the MEC. The yellow dot refers to the amount of a soil nutrient that is available for crop uptake without fertilizer application. We calculated it by taking the mean of the three replicates of the tuber nutrient uptake in the respective N, P or K omission treatment plots. Next, to eliminate all uncertainty in the relationship shown in quadrant IV, we added the measured nutrient uptake in all treatments (i.e., the right x-axis in Fig. 2) as extra covariates. Finally, to unravel the uncertainty introduced by the uptake-yield relationship (i.e. physiological efficiency) in quadrant I, we further added yield of fertiliser nutrient omission treatments (i.e., the blue dot in quadrant I, which has the same information as the green dot) as an extra covariate.

We designed a scheme to evaluate the model accuracy for all above

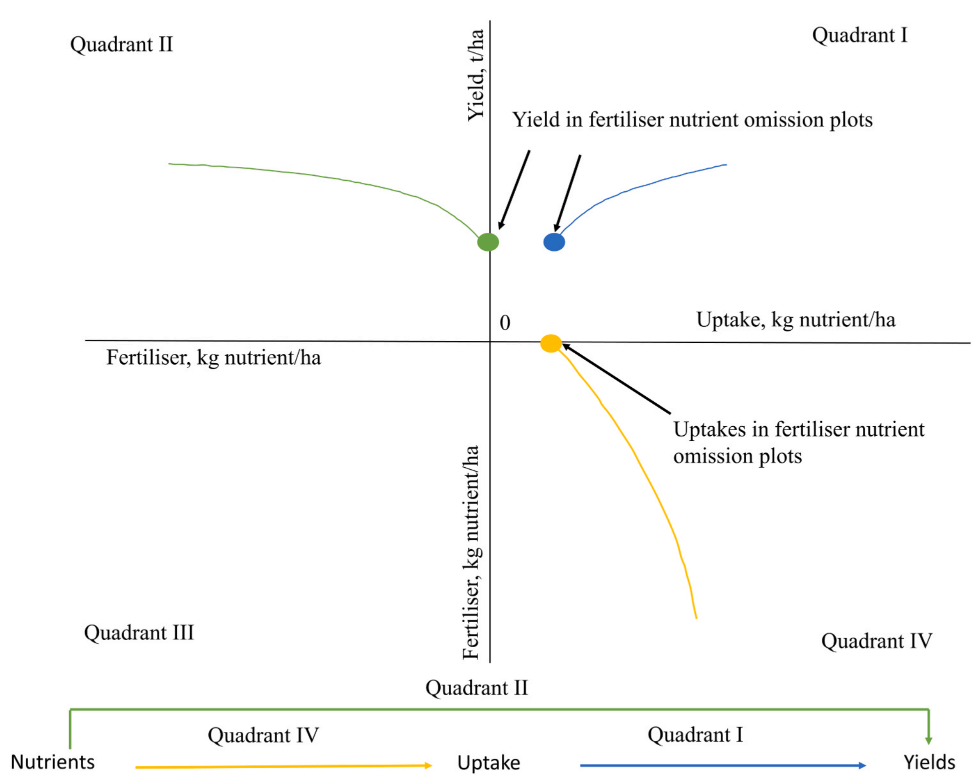

Fig. 2. Three-quadrant diagram of the relationship between fertiliser, uptake and yield (de Wit, 1953). Relationship between fertiliser application and yield (quadrant II); between uptake and yield (quadrant I); between fertiliser application and uptake (quadrant IV).

cases, as shown in Table 2. The benchmark model only considered the 38 basic covariates (no additional covariates). Models M1N, M1P and M1K referred to cases where the N uptake in N fertiliser omission treatments, P uptake in P fertiliser omission treatments, and K uptake in K fertiliser omission treatments (i.e., yellow dot of quadrant IV in Fig. 2) were respectively included as additional covariates, while model M1 included all three nutrient omission uptakes. These M1 models were used to evaluate the contribution of available soil nutrients on yield prediction uncertainty. Similarly, M2N, M2P and M2K included measured N uptake, P uptake and K uptake (i.e., the right x-axis in Fig. 2) as additional covariates, respectively, whereas M2 was the case where all three uptakes were included as additional variables. On top of the available soil nutrients, these M2 models were used to evaluate the contribution of fertiliser recovery. Cases M3N, M3P and M3K considered measured uptake plus tuber yield in fertiliser omission treatments, for N, P, and K fertiliser (i.e., the right x-axis plus blue dot in Fig. 2) as additional covariates, respectively, while M3 included all three nutrient uptakes plus the three nutrient omission yields. On top of available soil nutrients and fertiliser recovery, these M3 models further assessed the uncertainty introduced by physiological efficiency, where physiological efficiency is defined as the amount of yield (e.g. grains, tubers) produced per unit of nutrient uptake (Sattari et al., 2014), represented by the blue line in quadrant I of Fig. 2. We compared the ME, RMSE and MEC of the benchmark model with those of the 12 extended models in Table 2. To ensure a fair comparison, all models used the same dataset and cross-validation scheme. We used LSOCV in combination with a subset of the dataset with only observations where full NPK fertilisers were applied (n=1309).

3. Results

3.1. Yield modelling

The performance statistics of the RF model that uses the 38 covariates is shown in Table 3. The model accuracy as assessed with 10-fold CV had an RMSE of 3.45 t ha^{-1} and a MEC of 0.92. These metrics suggested that a large part of the yield variability could be predicted by the model, because the MEC is close to 1 and the RMSE is much smaller than the fresh tuber yield standard deviation of 12.56 t ha^{-1} (Table S3). But 10-fold CV benefits from the spatial and temporal clustering and always had training data from the same sites and seasons as the test data. Model accuracy significantly dropped when 10-fold CV was replaced by LBOCV, LSOCV, and LYOCV. When moving from 10-fold CV to LBOCV, the RMSE increased by 4.84 t ha^{-1} and the MEC decreased by 0.28. From LBOCV to LSOCV, where the RF model was cross-validated at a larger spatial scale, and test data were never from the same site as the training data, the model accuracy further declined. The LYOCV showed the poorest model accuracy, with the largest RMSE of 10.31 t ha^{-1} and the smallest MEC of 0.43. Negative ME values were observed in LBOCV and LSOCV, indicating that the RF model overpredicted the yield. But the ME of 10-fold CV was close to 0 and the ME of LYOCV was positive.

Table 3
Root mean squared error (RMSE), model efficiency coefficient (MEC) and mean error (ME) of random forest model for potato yield prediction.

	RMSE (t ha^{-1})	MEC	ME (t ha^{-1})
10-fold CV	3.5	0.92	−0.02
LBOCV	8.3	0.64	−1.8
LSOCV	9.9	0.52	−2.8
LYOCV	10.3	0.43	0.98

Predicted values obtained with 10-fold CV exhibited closer alignment with the observed values compared to the other cross-validation schemes (Fig. 3). This corroborates with the smallest RMSE and largest MEC of 10-fold CV. The deviation from the 1:1 line is larger for LSOCV than for LBOCV. Fig. 3 also shows that the RF model tends to overpredict lower yield values and underpredict higher yield values, which is common for regression models. For LYOCV, fresh tuber yields between 50 and 80 t ha^{-1} were severely underpredicted, which contributed to the ME value being positive.

3.2. Model interpretation

3.2.1. Individual variable importance

The relative importance of each covariate is shown in Fig. 4. Cumulated sunshine duration (Cum_SSD) during the growing season was identified as the most important in predicting potato yield variability, followed by topography position index (TPI). Three of the top 15 important covariates (enhanced vegetation index, upslope curvature, SAGA wetness index) were from the topography group, six (bulk density, sand, clay, and Olsen P, organic matter, available K and Alkali-hydrolyzable N) from the soil group, and two (planting date and cultivar) from the management group. Fertiliser-related covariates were identified as least important, with organic fertiliser ranked lowest.

3.2.2. Relative importance of variable groups

Fig. 5 shows the relative importance of variable groups for each of the four cross-validation schemes as computed using Eq. (4). Generally, the relative importance values of management variables were higher than those of other variable groups, while weather group and soil group ranked second-important position, and just slightly lower than management group. Management group ranked second-highest in relative importance values for 10-fold CV, and highest for the other four cross-validation schemes. Weather group ranked third in relative importance for LBOCV, and second for the other four cross-validation methods. The soil group was the highest in relative importance value

Table 2
Benchmark RF model included 38 basic covariates and extended RF models including extra covariates for measured uptake in nutrient omission treatments (M1), measured uptake in all treatments (M2) or yield in the nutrient omission treatments (M3).

RF model with LSOCV	Basic covariates	Uptake in N, P or K nutrient omission treatments			Uptake in all treatments			Yield in N, P, K nutrient omission treatments		
		N	P	K	N	P	K	N	P	K
Benchmark	√	×	×	×	×	×	×	×	×	×
M1N	√	√	×	×	×	×	×	×	×	×
M1P	√	×	√	×	×	×	×	×	×	×
M1K	√	×	×	√	×	×	×	×	×	×
M1	√	√	√	√	×	×	×	×	×	×
M2N	√	×	×	×	√	×	×	×	×	×
M2P	√	×	×	×	×	√	×	×	×	×
M2K	√	×	×	×	×	×	√	×	×	×
M2	√	×	×	×	√	√	√	×	×	×
M3N	√	×	×	×	√	×	×	√	×	×
M3P	√	×	×	×	×	√	×	×	√	×
M3K	√	×	×	×	×	×	√	×	×	√
M3	√	×	×	×	√	√	√	√	√	√

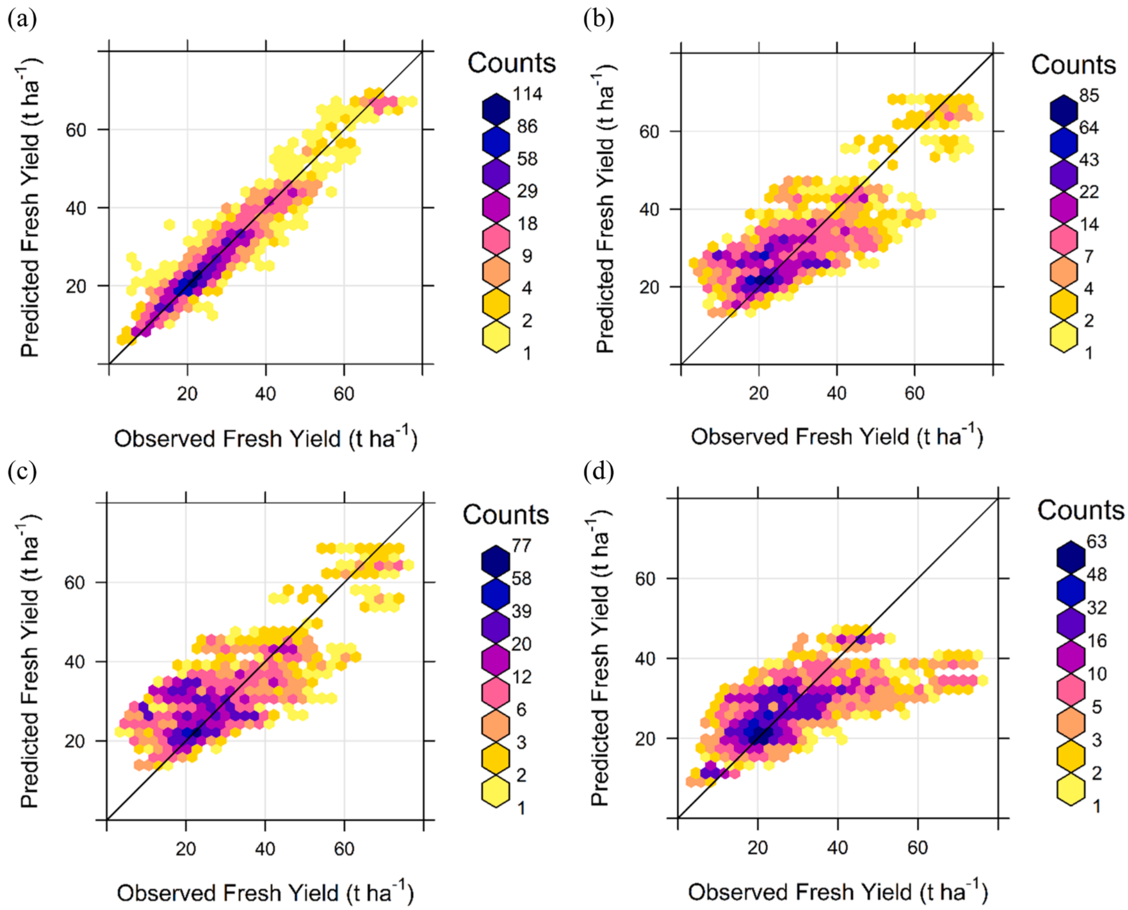

Fig. 3. Scatter density plots of RF predictions against observations: (a) 10-fold CV, (b) Leave block out cross-validation (LBOCV), (c) Leave site out cross-validation (LSOCV), (d) Leave year out cross-validation (LYOCV). The solid lines represent the 1:1 line.

for 10-fold CV, second highest for LBOCV and LSOCV, but fourth for LYOCV. The lowest relative importance was observed for the fertiliser group for all cross-validation methods, which is in line with the results of individual variable importance (Fig. 4). Topography group had a low relative importance in all cross-validation schemes, with values that were only slightly higher than those of the fertiliser group.

3.3. Prediction uncertainty analysis

The standard deviation of the residual of the linear model that predicts fresh yield from treatment, block, year and their interactions was 3.2 t ha^{-1} (purple line in the Fig. 6**b**). This estimate of the yield replicate variability standard deviation was substantially smaller than the LSOCV RMSE of 9.9 t ha^{-1}, but far from negligible. Thus, yield replicate variability had a significant contribution to the prediction uncertainty of the ML model. Note that here we used LSOCV for comparison, which is reasonable since 10-fold CV produces too optimistic results.

Adding uptake in nutrient omission treatments, uptake of all treatments and yields from nutrient omission treatments significantly decrease model prediction error. As shown in Fig. 6, compared to the benchmark model, the performance of models that incorporate both uptake of all treatments and nutrient omission yield (M3N, M3P, M3K, M3) was best, followed by models only incorporating uptake (M2N, M2P, M2K, M2). The nutrient omission uptake models (M1N, M1P, M1K, M1) had the least improvement compared to the benchmark model, where available soil N supply had more effect than available soil P or K supply, as can be seen by comparing the performance metrics of M1N, M1P and M1K.

Adding uptake in all treatments (i.e., information about fertiliser recovery) had only a small effect, reducing RMSE values by only 0.7 t ha^{-1} when comparing the M2 model with the M1 model. Adding nutrient omission yield (i.e., information about physiological efficiency) had a much stronger effect, reducing the RMSE by 2.5 t ha^{-1} when comparing M3 with M2. This information had the strongest effect for P and K, as can be seen by comparing M2P with M3P and by comparing M2K with M3K.

The RMSE of M3 was only 4.6 t ha^{-1} and could explain 90 % of yield variability, which showed a marked improvement over the benchmark model, although the RMSE of the M3 model was still 42 % larger than the standard deviation of yield replicate variability.

4. Discussion

4.1. Model performance and extrapolation potential

The fitted RF model could explain potato yield in China well using 38 covariates (Table 3 and Fig. 3), although prediction accuracy decreased for an unknown site or an unknown year with MEC of 0.52 and 0.43, respectively. When evaluated by 10-fold CV, the RF model explained 92 % of the yield variation and had a RMSE of 3.5 t ha^{-1}, explaining more variation than reported in previous studies (Li et al., 2020; Prodhan et al., 2022; Ruan et al., 2022). However, k-fold CV tends to provide an overoptimistic assessment for spatially or temporally clustered data (Ploton et al., 2020; Wadoux et al., 2021), because k-fold CV uses test and training data from the same cluster (i.e., the same site or year). As expected, when evaluated with leave block, site and year out cross-validation, the MEC of the RF model significantly decreased to 0.43–0.64 and the RMSE considerably increased to 8.3–10.3 t ha^{-1}. The

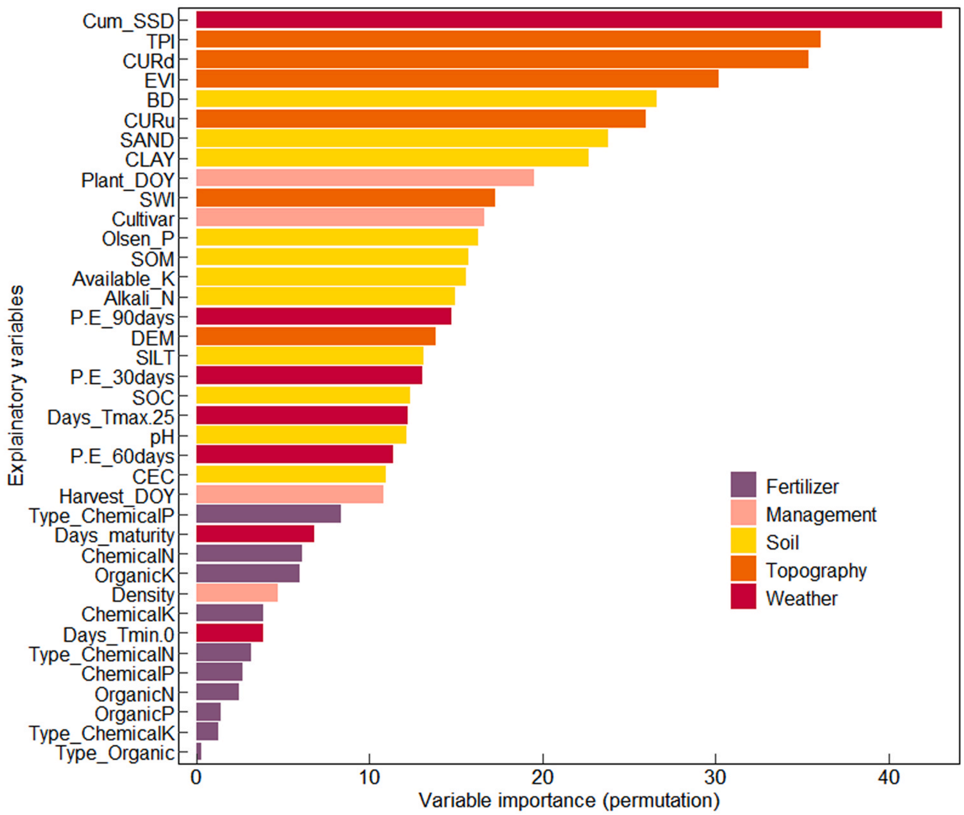

Fig. 4. Relative importance of covariates for the RF model. See Table S2 Abbreviations for explanation of the variable code names.

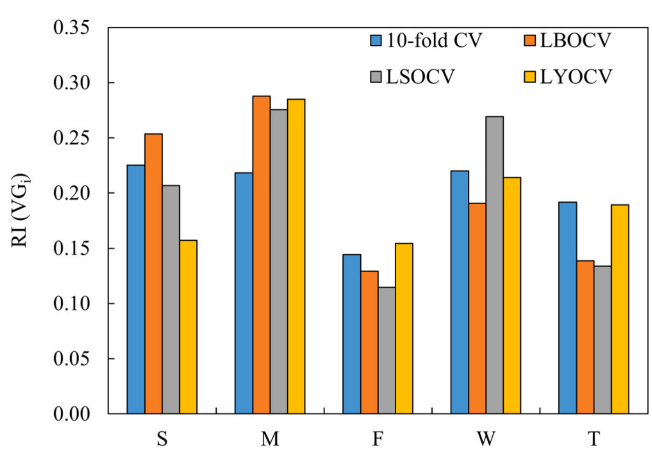

Fig. 5. Relative importance of five variable groups for 10-fold CV (blue bar), leave block out CV (orange bar), leave site out CV (grey bar), and leave year out CV (yellow bar). S = soil group, M = management group, F = fertiliser group, W = weather group, T = topographic group.

considerable differences between the various CV methods show that it is important to address the clustering effect in model evaluation using cross-validation. de Bruin et al. (2022) recommended weighted cross-validation for moderately clustered samples, while Milà et al. (2022) applied nearest neighbour distance matching to avoid biased estimates of model performance metrics. The significant drop in model performance from 10-fold CV to LBOCV, LSOCV and LYOCV is not only due to a clustering effect but also because of extrapolation in geographic space or time. Extrapolation in geographical space or time will lead to extrapolation in covariate space, and the model performance will get worse when predictions are made for cases that are spatially or temporally distant from the training data (Meyer and Pebesma, 2021).

The RF model performed better when extrapolating over space (i.e. LBOCV and LSOCV) than when extrapolating over time (i.e. LYOCV). A possible reason is that the time series used for model training was insufficient to capture important temporal yield variability. The total dataset comprised yield observations from three years, of which only two years were used for training in LYOCV. A model trained on data from a longer time period can more effectively capture extreme weather events and climate patterns, and thereby better model yield temporal variation.

4.2. Variable importance analysis

Cumulative sunshine duration was identified as the most important variable for potato yield modelling. Sunshine hours are causally related to intercepted radiation, thereby increasing the accumulation of tuber dry matter (Haverkort, 2007). Topographic position index (TPI) was the second-most important variable to explain yield variation (Fig. 4). The effect of TPI depends on the biophysical context. Leuthold et al. (2022) found that the relationship between yield and topography varied with precipitation. Crop yield response to TPI is driven by underlying variation in soil moisture (Rabia et al., 2022), with lower yields in low landscape positions during the wet season due to poor drainage (Kravchenko and Bullock, 2000) and higher yields in the same landscape position in dry years.

Interestingly, fertiliser-related variables were the least important for explaining potato yield variability in our RF models (Fig. 4). A likely reason is the generally adequate soil nutrient content in our experimental fields, reflected by especially Olsen P and exchangeable K. About 52 % and 70 % of the soil samples were abundant or very abundant in P and K, respectively, and about 48 % of the soil samples contained abundant Alkali-hydrolyzable N (Table S1). Crops could absorb adequate nutrients from the soil in these fields so that fertiliser application had little effect on yield production, reflected by the lower yield

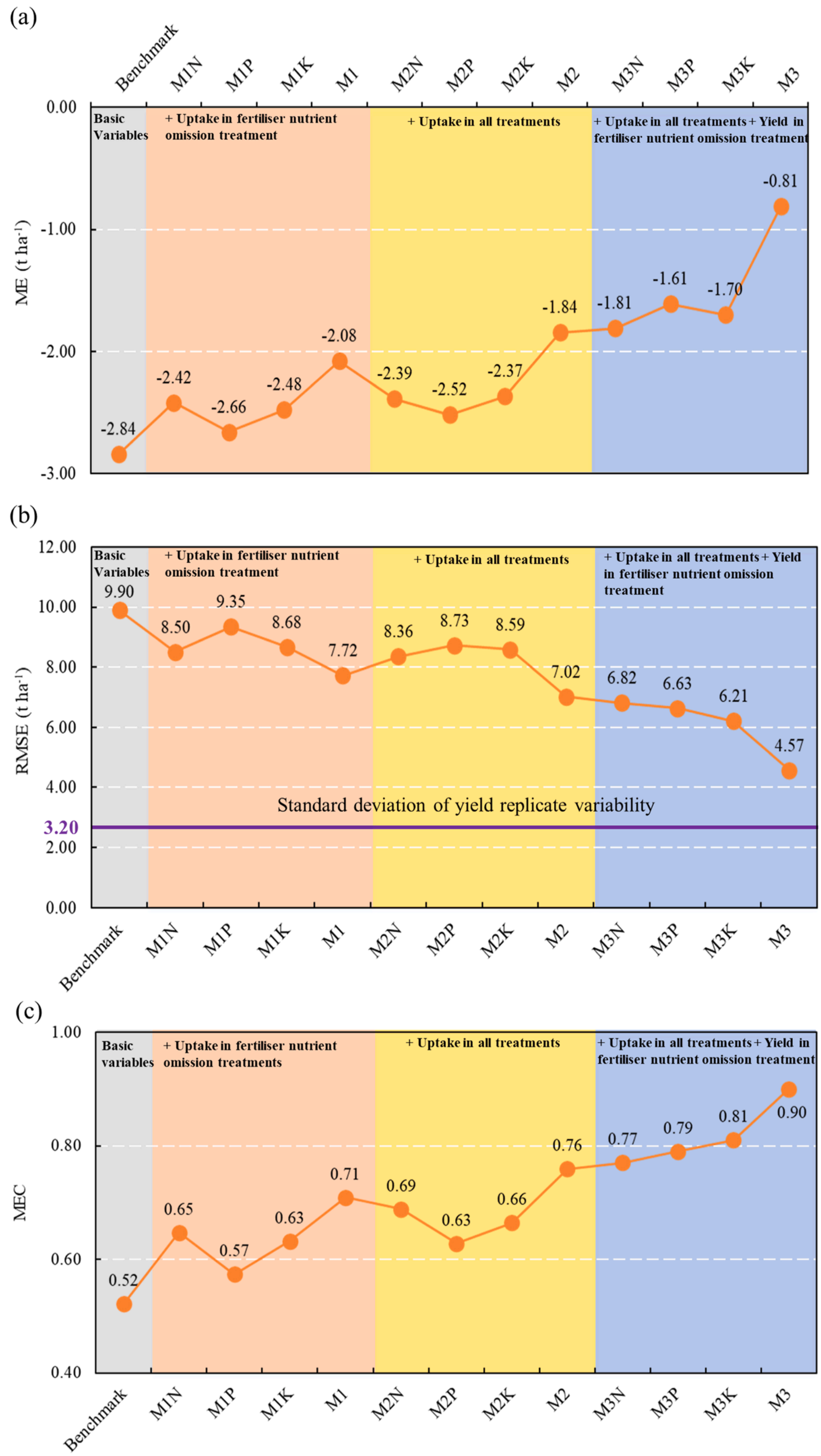

Fig. 6. Evaluating the performance of RF models with LSOCV by stepwise adding nutrient omission uptake, uptake of all treatments and nutrient omission yields as additional covariates. ME (a), RMSE (b), MEC (c). The explanation of the model acronyms is shown in Table 3. Note that the straight lines connecting dots do not represent time series or continuous functions, and are only used as a visual aid to clarify how the metrics change when additional covariates are gradually included.

response to fertiliser applications in our study compared to findings in previous studies (Fig. S2) (Bélanger et al., 2000; Shillito et al., 2009; Wang et al., 2020; Jiang et al., 2021).

For group variable important analysis, generally, management covariates were the most important for potato yield prediction (Fig. 5), even though only slightly surpassing the second most important covariates (i.e., weather group and soil group). Coulibali et al. (2020) reported that tuber planting density is the most important factor determining medium-size tuber yield, followed by N fertiliser amount, soil P, soil Al, and soil type. On the global scale, climate variation explains a third of crop yield variability (Ray et al., 2015). Same with individual variable importance analysis, fertiliser-related covariates were still identified as the least important for explaining potato yield variability. The importance of topography-related covariates was different in individual variable importance analysis and group variable importance analysis. In group variable importance analysis, topography-related covariates were the second-least important for yield modelling, but in individual variable importance analysis, they were the second-most important variable. This is because the computation methods of individual variable importance analysis and group variable importance analysis were different. Individual variable importance analysis is evaluated by shuffling the selected individual covariate while keeping the others fixed. Group variable importance analysis is calculated by using the MEC of an entire group of covariates (Eq. 4).

The relative importance ranking of grouping variables was different among the four cross-validation schemes (Fig. 5). It is difficult to explain these differences, but the overall results summarized from four cross-validation methods seem more reliable than those from the single cross-validations.

Care is needed when interpreting the importance of covariates, because it could yield misleading results if there is a strong dependence among covariates (Hooker et al., 2021). It is challenging to define the exact cause of observed yield variability explicitly because we can only infer a causal effect from well-balanced experimental designs (van Es et al., 2007). In this research, we used a balanced experimental design for fertiliser application but not for other covariates.

Analysing the importance of covariates for model prediction in a future year would be very interesting for future work. Using a LYOCV approach, one could compare the mean squared error of the original model with the mean squared error of a model with one selected covariate randomly shuffled. The decrease in the mean squared error would then signify the importance of that covariate for a model that is extrapolated in time.

4.3. Analysis of yield prediction uncertainty

The standard deviation of yield replicate variability was estimated as 3.2 t ha^{-1}, which was 32 % of the RMSE of the RF model when using LSOCV. As noted before, the yield replicate variability is the combined effect of yield measurement error and within-block-treatment-year yield variation. We could not separate these two sources because the estimation of yield measurement error would require repeated independent yield measurements of the same plot, which were not available in our dataset. We analysed the main causes of the remaining RMSE by gradually offering more useful information from production ecology principles. Soil nutrient uptake efficiency, fertiliser recovery efficiency and physiological use efficiency were valuable covariates for yield modelling. When these three covariates are known and included as covariates, users can more accurately predict yield. This study incrementally incorporated nutrient omission uptake (i.e., soil nutrients available for uptake), uptake of all treatments (i.e., information about fertiliser recovery), and nutrient omission yield (i.e., information about physiological use efficiency) to gradually evaluate prediction uncertainty.

The increased performance of the M1 model (adding nutrient omission uptake) represented the contribution of available soil nutrients on yield modelling, while the increased performance of the M2 model (adding uptake of all treatments) represented the contribution of both available soil nutrients and fertilisation (Fig. 6). The performance of the M2 model was only slightly better than that of M1, which revealed that fertiliser recovery information had a small contribution to yield prediction uncertainty. This finding was supported by the variable importance analysis, i.e. fertiliser-related variables were least important for yield modelling (Fig. 4 and Fig. 5). The RMSE of M2 (7.0 t ha^{-1}) was much bigger than the yield replicate variability standard deviation (3.2 t ha^{-1}). One plausible reason for that is uncertainty about the uptake-yield relationship (quadrant I of Fig. 2), while another likely reason is the measurement error in nutrient uptake. The measurement error in uptake is a combined effect of laboratory measurement error in tuber nutrient concentration and tuber dry weight. If uptake measurement error is substantial then this will affect the performance of M2, since this model uses the measured uptake instead of the 'true' uptake as a covariate.

The performance of M3 (adding both nutrient omission yield and uptake of all treatments) showed a marked improvement over the M2 model. This indicates that nutrient omission yield provided valuable reference values for predicting yield. Yields in nutrient omission treatment can be used to explain yield differences between sites and between seasons. The yield variation within a block was not very large because potato was cultivated in the same climatic and environmental conditions and potato yield had a relatively weak response to fertiliser. However, the RMSE of M3 (4.5 t ha^{-1}) was still somewhat greater than the yield replicate variability standard deviation (3.2 t ha ^{1}). This is likely caused by measurement error of potato uptake and uncertainty in uptake-yield relationship under fertilized treatments.

The ME results shown in Fig. 6a indicate that in absolute value terms, the ME is substantially smaller than the RMSE, but also show that the ME is not negligible. The ME of all extended models was negative (Fig. 6a), which indicated a systematic overprediction (Eq. 3). We have no obvious explanation for the prediction bias, since RF models tend to produce unbiased predictions (Hengl et al., 2018), but it might be caused by the data splitting by sites in LSOCV, since bias did not occur for 10-fold CV (Table 3). Note also that an explanation for all ME values being negative and having the same sign might be because all models used the same training and test data and the same splitting procedure. Taking a closer look revealed that the observed yields were significantly lower than the predicted yields in Jiangxi province during 2018 and 2019 and at some sites in Shanxi province during 2017 and 2018. This is likely the main reason for the negative ME values. Excluding these cases in LSOCV for all 13 models evaluated in Fig. 6 produced nine negative and four positive ME values.

Comparing models in the same group but with different nutrient covariates shows that the models that include N uptake (M1N and M2N cases) generally performed best, next followed by adding K uptake (M1K and M2K cases), and lastly followed by adding P uptake (M1P and M2P cases). In this study, potato yield was more strongly correlated to N uptake than to P uptake and K uptake, indicating that N was more yield limiting than P or K. Information about N availability for the crop was therefore more important than P or K availability.

In this study, all replicate plots that had the same treatment, block and year had identical covariates. This suggests that model prediction uncertainty could be decreased by averaging the observed yields of replicate plots prior to model training (as in Sarijaloo et al., 2021). This 'change of support' would eliminate part of the short-distance spatial variability and hence improve prediction accuracy (Heuvelink, 2018; Szatmári et al., 2021). But it would also change the scale of modelling from plot level to block level (i.e., the average of three plots within one block). Since we aimed to quantify the model prediction accuracy and the main sources of prediction uncertainty at plot level, we decided not to average the yields before modelling.

4.4. Strengths and weaknesses

This study was based on an extensive experimental dataset from China, covering multiple seasons and a wide range of geographic locations. This allowed for a comprehensive analysis of yield variability across various environmental conditions, which enhanced the reliability and generalizability of the fitted RF models. The observed yield data, soil nutrients, fertiliser application, management information and uptake were collected from field experiments and laboratory measurements, which is a more reliable data source compared to farm surveys or prediction maps. Another strength is that to the best of our knowledge, this study is the first to combine production ecology principles (i.e. the three-quadrant diagram) and RF to analyse yield prediction uncertainty. Previous studies have revealed the contribution of soil nutrients, fertiliser application to yield modelling (Jones et al., 2022; Carneiro et al., 2023; Silva et al., 2023), but none of these used nutrient omission uptake, uptake of all treatments, and nutrient omission yield to explain prediction uncertainty. It is also worthwhile to note that while this study analysed potato yield in China, the methodology used here can easily be applied to other crops and countries, or even other machine learning methods.

Our research has some limitations that should be considered by further studies. One topic for future research is to quantify the laboratory measurement error of tuber dry weight and tuber nutrient concentration. This would allow isolating the contribution of uptake measurement uncertainty on the performances of the M2 and M3 models. Finally, it is important to note that the models developed in this study cannot easily be used for making prediction maps, since many of the covariates (including nutrient omission uptake, uptake of all treatments and nutrient omission yield) used here were measured at the sites and are not available for the whole geography of China, which would be required for mapping.

5. Conclusion

This study aimed to explain and evaluate the accuracy of a random forest model for potato yield prediction in China and to quantify the main sources of prediction uncertainty. The fitted models could explain up to 92 % of the yield variability, although the model efficiency coefficient dropped to 0.52 and 0.43 when extrapolating to unknown sites or years, respectively. These results highlight that care should be taken when using RF models for yield prediction in case of strongly clustered training data and assessing their performance with standard cross-validation methods. The importance of covariates was assessed both individually and in groups. Cumulated sunshine duration and topography position index were the most important covariates, while the weather group and management group were most important for yield prediction. Fertiliser variables were identified as the least important variables for yield modelling. Yield replicate variability had a significant contribution to the prediction uncertainty of the RF model. The production ecology principle allowed us to better understand the sources of yield prediction uncertainty. We conclude that including information about physiological efficiency leads to a large increase of yield prediction accuracy, followed by information about available soil nutrients. In fertile fields such as those used in this study, information about fertiliser recovery was less important and did not significantly improve the model predictions.

CRediT authorship contribution statement

Johan Leenaars: Writing – review & editing, Validation, Supervision, Methodology, Formal analysis, Conceptualization. **Antonius Schut:** Writing – review & editing, Validation, Supervision, Methodology, Formal analysis, Conceptualization. **Qiuhong Huang:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Conceptualization. **Ping He:** Writing – review & editing, Supervision, Resources, Project administration, Investigation, Funding acquisition. **Gerard Heuvelink:** Writing – review & editing, Validation, Supervision, Project administration, Methodology, Formal analysis, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was financed by Smart Fertilisation Project (05) and China Agricultural Research Systems (CARS-09-P31). We thank the Chinese Academy of Agricultural Sciences for their support in experimental design, implementation, and laboratory analysis. We thank Zhuo Xu who contributed to data collection and analysis. We acknowledge all those who provided local assistance or technical help for the field sampling. Huang Q. thanks the China Scholarship Council for awarding her a PhD fellowship (No.202203250061) to study at Wageningen University.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.fcr.2024.109619.

Data Availability

Data will be made available on request.

References

Bélanger, G., Walsh, J.R., Richards, J.E., Milburn, P.H., Ziadi, N., 2000. Comparison of three statistical models describing potato yield response to nitrogen fertiliser. J. Agron. 92 (5), 902–908. https://doi.org/10.2134/agronj2000.925902x.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.

de Bruin, S., Brus, D.J., Heuvelink, G.B.M., van Ebbenhorst Tengbergen, T., Wadoux, A. M.J.C., 2022. Dealing with clustered samples for assessing map accuracy by cross-validation. Ecol. Inform. 69, 101665. https://doi.org/10.1016/j.ecoinf.2022.101665.

Burke, J.J., Hatfield, J.L., 1987. Plant morphological and biochemical responses to field water deficits: III. effect of foliage temperature on the potential activity of glutathione reductase. Plant Physiol. 85 (1), 100–103. https://doi.org/10.1104/pp.85.1.100.

Carneiro, F.M., Filho, A.L. d B., Ferreira, F.M., Junior, G. d F.S., Brandão, Z.N., da Silva, R.P., Shiratsuchi, L.S., 2023. Soil and satellite remote sensing variables importance using machine learning to predict cotton yield. Smart Agric. Technol. 5, 100292. https://doi.org/10.1016/j.atech.2023.100292.

Chinese Society of Soil Science, 2000. Methods of Soil and Plant Analysis. China Agriculture Scientech Press, Beijing.

Coulibali, Z., Cambouris, A.N., Parent, S.-É., 2020. Site-specific machine learning predictive fertilisation models for potato crops in Eastern Canada. PLOS ONE 15 (8), e0230888. https://doi.org/10.1371/journal.pone.0230888.

Criss, R.E., Winston, W.E., 2008. Do Nash values have value? Discussion and alternate proposals. Hydrol. Process. 22 (14), 2723–2725. https://doi.org/10.1002/hyp.7072.

Deng, X., Huang, Y., Yuan, W.P., Zhang, W., Ciais, P., Dong, W.J., Smith, P., Qin, Z.C., 2023. Building soil to reduce climate change impacts on global crop yield. Sci. Total Environ. 903. https://doi.org/10.1016/j.scitotenv.2023.166711.

Devkota, M., Devkota, K.P., Karrou, M., Nangia, V., 2023. Genotype and agronomic management interaction to enhance wheat yield and water use efficiency in the Mediterranean rainfed environment of Morocco: I. Field data analysis. Eur. J. Agron. 151, 126972. https://doi.org/10.1016/j.eja.2023.126972.

Ebrahimy, H., Wang, Y., Zhang, Z., 2023. Utilisation of synthetic minority oversampling technique for improving potato yield prediction using remote sensing data and machine learning algorithms with small sample size of yield data. ISPRS J. Photogramm. Remote Sens. 201, 12–25. https://doi.org/10.1016/j.isprsjprs.2023.05.015.

van Es, H.M., Gomes, C.P., Sellmann, M., van Es, C.L., 2007. Spatially-balanced complete block designs for field experiments. Geoderma 140 (4), 346–352. https://doi.org/10.1016/j.geoderma.2007.04.017.

Eugenio, F.C., Grohs, M., Schuh, M., Venancio, L.P., Schons, C., Badin, T.L., Mallmann, C. L., Fernandes, P., Pereira da Silva, S.D., Fantinel, R.A., 2023. Estimated flooded rice

grain yield and nitrogen content in leaves based on RPAS images and machine learning. Field Crops Res. 292, 108823. https://doi.org/10.1016/j.fcr.2023.108823.
Falconnier, G.N., Leroux, L., Beillouin, D., Corbeels, M., Hijmans, R.J., Bonilla-Cedrez, C., van Wijk, M., Descheemaeker, K., Zingore, S., Affholder, F., Lopez-Ridaura, S., Malézieux, E., Makowski, D., Rurinda, J., van Ittersum, M.K., Vanlauwe, B., Giller, K. E., Lammoglia, S.K., Waha, K., 2023. Increased mineral fertiliser use on maize can improve both household food security and regional food production in East Africa. Agric. Syst. 205. https://doi.org/10.1016/j.agsy.2022.103588.
FAO (Food and Agriculture Organization), 2024. FAOSTAT Database collections. FAO, Rome. ⟨https://www.fao.org/faostat/en/#data⟩.
Fisher, A., Rudin, C., Dominici, F., 2019. All models are wrong, but many are useful: learning a variable's importance by studying an entire class of prediction models simultaneously. J. Mach. Learn. Res 20 (177), 1–81.
Fleisher, D.H., Condori, B., Quiroz, R., Alva, A., Asseng, S., Barreda, C., Bindi, M., Boote, K.J., Ferrise, R., Franke, A.C., Govindakrishnan, P.M., Harahagazwe, D., Hoogenboom, G., Naresh Kumar, S., Merante, P., Nendel, C., Olesen, J.E., Parker, P. S., Raes, D., Raymundo, R., Ruane, A.C., Stockle, C., Supit, I., Vanuytrecht, E., Wolf, J., Woli, P., 2017. A potato model intercomparison across varying climates and productivity levels. Glob. Chang. Biol. 23 (3), 1258–1281. https://doi.org/10.1111/gcb.13411.
Gondwe, R.L., Kinoshita, R., Suminoe, T., Aiuchi, D., Palta, J.P., Tani, M., 2020. Available soil nutrients and NPK application impacts on yield, quality, and nutrient composition of potatoes growing during the main season in Japan. Am. J. Potato Res 97 (3), 234–245. https://doi.org/10.1007/s12230-020-09776-2.
Gupta, A., Rico-Medina, A., Caño-Delgado, A.I., 2020. The physiology of plant responses to drought. Science 368 (6488), 266–269. https://doi.org/10.1126/science.aaz7614.
Haverkort, A.J., 2007. Potato crop response to radiation and daylength. In: Vreugdenhil, D., Bradshaw, J., Gebhardt, C., Govers, F., Mackerron, D.K.L., Taylor, M.A. (Eds.), Potato Biology and Biotechnology. E-Publishing Inc, Amsterdam, pp. 353–365. https://doi.org/10.1016/B978-044451018-1/50059-2.
Heuvelink, G.B.M., 2018. Uncertainty and uncertainty propagation in soil mapping and modelling. In: McBratney, A.B., Minasny, B., Stockmann, U. (Eds.), Pedometrics. Springer International Publishing, Cham, pp. 439–461. https://doi.org/10.1007/978-3-319-63439-5_14.
Hooker, G., Mentch, L., Zhou, S., 2021. Unrestricted permutation forces extrapolation: variable importance requires at least one more model, or there is no free variable importance. Stat. Comput. 31 (6).
Jiang, L., Jin, G., Zhang, G., Zhang, C., 2021. Tuber yield and french fry processing quality response of potatoes to nitrogen rate. Potato Res 65 (2), 255–271. https://doi.org/10.1007/s11540-021-09525-w.
Jiang, L., Qiu, S., Ding, W., Xu, X., He, P., 2023. Synergizing potato productivity and environmental performance with Nutrient Expert recommendation approach in northern China. J. Clean. Prod. 382. https://doi.org/10.1016/j.jclepro.2022.135258.
Jones, J.W., Antle, J.M., Basso, B., Boote, K.J., Conant, R.T., Foster, I., Godfray, H.C.J., Herrero, M., Howitt, R.E., Janssen, S., Keating, B.A., Munoz-Carpena, R., Porter, C. H., Rosenzweig, C., Wheeler, T.R., 2017. Brief history of agricultural systems modeling. Agric. Syst. 155, 240–254. https://doi.org/10.1016/j.agsy.2016.05.014.
Jones, E.J., Bishop, T.F.A., Malone, B.P., Hulme, P.J., Whelan, B.M., Filippi, P., 2022. Identifying causes of crop yield variability with interpretive machine learning. Comput. Electron. Agric. 192. https://doi.org/10.1016/j.compag.2021.106632.
Kravchenko, A.N., Bullock, D.G., 2000. Correlation of corn and soybean grain yield with topography and soil properties. J. Agron. 92 (1), 75–83. https://doi.org/10.2134/agronj2000.92175x.
Krell, N., Davenport, F., Harrison, L., Turner, W., Peterson, S., Shukla, S., Marter-Kenyon, J., Husak, G., Evans, T., Caylor, K., 2022. Using real-time mobile phone data to characterize the relationships between small-scale farmers' planting dates and socio-environmental factors. Clim. Risk Manag. 35. https://doi.org/10.1016/j.crm.2022.100396.
Lee, D., Davenport, F., Shukla, S., Husak, G., Funk, C., Harrison, L., McNally, A., Rowland, J., Budde, M., Verdin, J., 2022. Maize yield forecasts for Sub-Saharan Africa using Earth Observation data and machine learning. Glob. Food Secur. 33, 100643. https://doi.org/10.1016/j.gfs.2022.100643.
van Leeuwen, C.C.E., Mulder, V.L., Batjes, N.H., Heuvelink, G.B.M., 2021. Statistical modelling of measurement error in wet chemistry soil data. Eur. J. Soil Sci. 73 (1). https://doi.org/10.1111/ejss.13137.
Leuthold, S.J., Wendroth, O., Salmerón, M., Poffenbarger, H., 2022. Weather-dependent relationships between topographic variables and yield of maize and soybean. Field Crops Res. 276, 108368. https://doi.org/10.1016/j.fcr.2021.108368.
Li, B., Xu, X., Zhang, L., Han, J., Bian, C., Li, G., Liu, J., Jin, L., 2020. Above-ground biomass estimation and yield prediction in potato by using UAV-based RGB and hyperspectral imaging. ISPRS J. Photogramm. Remote Sens 162, 161–172. https://doi.org/10.1016/j.isprsjprs.2020.02.013.
Malone, R.W., Ma, L., Heilman, P., Karlen, D.L., Kanwar, R.S., Hatfield, J.L., 2007. Simulated N management effects on corn yield and tile-drainage nitrate loss. Geoderma 140 (3), 272–283. https://doi.org/10.1016/j.geoderma.2007.04.011.
McCuen, R.H., Knight, Z., Cutter, A.G., 2006. Evaluation of the Nash–Sutcliffe efficiency index. Journal of hydrologic engineering, 11 (6), 597–602. https://doi.org/10.1061/(ASCE)1084-0699(2006)11:6(597).
Meyer, H., Pebesma, E., 2021. Predicting into unknown space? Estimating the area of applicability of spatial prediction models. Methods Ecol. Evol. 12 (9), 1620–1633. https://doi.org/10.1111/2041-210X.13650.
Midmore, D.J., 1984. Potato (Solanum spp.) in the hot tropics I. Soil temperature effects on emergence, plant development and yield. Field Crops Res. 8, 255–271. https://doi.org/10.1016/0378-4290(84)90073-X.
Milà, C., Mateu, J., Pebesma, E., Meyer, H., 2022. Nearest neighbour distance matching Leave-One-Out Cross-Validation for map validation. Met. Ecol. Evol. 13 (6), 1304–1316. https://doi.org/10.1111/2041-210x.13851.
Molnar, C., 2018. iml: An R package for Interpretable Machine Learning. J. Open Source Softw. 3 (26). https://doi.org/10.21105/joss.00786.
Morell, F.J., Yang, H.S., Cassman, K.G., Wart, J.V., Elmore, R.W., Licht, M., Coulter, J.A., Ciampitti, I.A., Pittelkow, C.M., Brouder, S.M., Thomison, P., Lauer, J., Graham, C., Massey, R., Grassini, P., 2016. Can crop simulation models be used to predict local to regional maize yields and total production in the U.S. Corn Belt? Field Crops Res 192, 1–12. https://doi.org/10.1016/j.fcr.2016.04.004.
Mulvaney, R.L., Khan, S.A., 2001. Diffusion methods to determine different forms of nitrogen in soil hydrolysates. Soil Sci. Soc. Am. J. 65 (4), 1284–1292. https://doi.org/10.2136/sssaj2001.6541284x.
Myles, A.J., Feudale, R.N., Liu, Y., Woody, N.A., Brown, S.D., 2004. An introduction to decision tree modeling. J. Chemom. 18 (6), 275–285. https://doi.org/10.1002/cem.873.
Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I — A discussion of principles. J. Hydrol. 10 (3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6.
National Soil Survey Office. Soils of China, 1998. China Agricultural Press, Bejing, pp. 106 (in Chinese).
Nayak, H.S., Silva, J.V., Parihar, C.M., Krupnik, T.J., Sena, D.R., Kakraliya, S.K., Jat, H.S., Sidhu, H.S., Sharma, P.C., Jat, M.L., Sapkota, T.B., 2022. Interpretable machine learning methods to explain on-farm yield variability of high productivity wheat in Northwest India. Field Crops Res. 287, 108640. https://doi.org/10.1016/j.fcr.2022.108640.
Paudel, D., Boogaard, H., de Wit, A., Janssen, S., Osinga, S., Pylianidis, C., Athanasiadis, I.N., 2021. Machine learning for large-scale crop yield forecasting. Agric. Syst. 187, 103016. https://doi.org/10.1016/j.agsy.2020.103016.
Ploton, P., Mortier, F., Rejou-Mechain, M., Barbier, N., Picard, N., Rossi, V., Dormann, C., Cornu, G., Viennois, G., Bayol, N., Lyapustin, A., Gourlet-Fleury, S., Pelissier, R., 2020. Spatial validation reveals poor predictive performance of large-scale ecological mapping models. Nat. Commun. 11 (1), 4540. https://doi.org/10.1038/s41467-020-18321-y.
Poggio, L., De Sousa, L.M., Batjes, N.H., Heuvelink, G., Kempen, B., Ribeiro, E., Rossiter, D., 2021. SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. Soil 7 (1), 217–240. https://doi.org/10.5194/soil-7-217-2021.
Prodhan, F.A., Zhang, J., Pangali Sharma, T.P., Nanzad, L., Zhang, D., Seka, A.M., Ahmed, N., Hasan, S.S., Hoque, M.Z., Mohana, H.P., 2022. Projection of future drought and its impact on simulated crop yield over South Asia using ensemble machine learning approach. Sci. Total Environ. 807, 151029. https://doi.org/10.1016/j.scitotenv.2021.151029.
Rabia, A.H., Neupane, J., Lin, Z., Lewis, K., Cao, G., Guo, W., 2022. Chapter Four - Principles and applications of topography in precision agriculture. In: Sparks, D.L. (Ed.), Advances in Agronomy, 171. Academic Press, pp. 143–189.
Radanielson, A.M., Angeles, O., Li, T., Ismail, A.M., Gaydon, D.S., 2018. Describing the physiological responses of different rice genotypes to salt stress using sigmoid and piecewise linear functions. Field Crops Res. 220, 46–56. https://doi.org/10.1016/j.fcr.2017.05.001.
Raharimanana, V., Yamaguchi, T., Tsujimoto, Y., Oo, A.Z., Nishigaki, T., Rakotonindrina, H., Katsura, K., 2023. A machine learning approach is effective to elucidate yield-limiting factors of irrigated lowland rice under heterogeneous growing conditions and management practices. Field Crops Res. 304. https://doi.org/10.1016/j.fcr.2023.109170.
Ray, D.K., Gerber, J.S., MacDonald, G.K., West, P.C., 2015. Climate variation explains a third of global crop yield variability. Nat. Commun. 6, 5989. https://doi.org/10.1038/ncomms6989.
Reuter, H.I., Hengl, T., 2012. Worldgrids—a public repository of global soil covariates. Digit. Soil Assess. Beyond 6.
Roberts, D.R., Bahn, V., Ciuti, S., Boyce, M.S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J., Schröder, B., Thuiller, W., Warton, D.I., Wintle, B. A., Hartig, F., Dormann, C.F., 2017. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography 40 (8), 913–929. https://doi.org/10.1111/ecog.02881.
Robinson, N., Regetz, J., Guralnick, R.P., 2014. EarthEnv-DEM90: a nearly-global, void-free, multi-scale smoothed, 90 m digital elevation model from fused ASTER and SRTM data. ISPRS J. Photogramm. Remote Sens. 87, 57–67. https://doi.org/10.1016/j.isprsjprs.2013.11.002.
Ruan, G., Li, X., Yuan, F., Cammarano, D., Ata-Ui-Karim, S.T., Liu, X., Tian, Y., Zhu, Y., Cao, W., Cao, Q., 2022. Improving wheat yield prediction integrating proximal sensing and weather data with machine learning. Comput. Electron. Agric. 195, 106852. https://doi.org/10.1016/j.compag.2022.106852.
Sale, P., 1979. Growth of potatoes (Solanum tuberosum L.) to the small tuber stage as related to soil temperature. Aust. J. Agric. Res. 30 (4), 667–675. https://doi.org/10.1071/AR9790667.
Sarijaloo, F.B., Porta, M., Taslimi, B., Pardalos, P.M., 2021. Yield performance estimation of corn hybrids using machine learning algorithms. Artif. Intell. Agric. 5, 82–89. https://doi.org/10.1016/j.aiia.2021.05.001.
Sattari, S.Z., van Ittersum, M.K., Bouwman, A.F., Smit, A.L., Janssen, B.H., 2014. Crop yield response to soil fertility and N, P, K inputs in different environments: testing and improving the QUEFTS model. Field Crops Res. 157, 35–46. https://doi.org/10.1016/j.fcr.2013.12.005.
Schut, A.G.T., Giller, K.E., 2020. Soil-based, field-specific fertiliser recommendations are a pipe-dream. Geoderma 380, 114680. https://doi.org/10.1016/j.geoderma.2020.114680.

Scowen, M., Athanasiadis, I.N., Bullock, J.M., Eigenbrod, F., Willcock, S., 2021. The current and future uses of machine learning in ecosystem service research. Sci. Total Environ. 799, 149263. https://doi.org/10.1016/j.scitotenv.2021.149263.

Shillito, R.M., Timlin, D.J., Fleisher, D., Reddy, V.R., Quebedeaux, B., 2009. Yield response of potato to spatially patterned nitrogen application. Agric. Ecosyst. Environ. 129 (1), 107–116. https://doi.org/10.1016/j.agee.2008.07.010.

Silva, J.V., Heerwaarden, J.V., Reidsma, P., Laborte, A.G., Tesfaye, K., Ittersum, M.K.V., 2023. Big data, small explanatory and predictive power: lessons from random forest modeling of on-farm yield variability and implications for data-driven agronomy. Field Crops Res. 302, 109063. https://doi.org/10.1016/j.fcr.2023.109063.

Strobl, C., Boulesteix, A.L., Kneib, T., Augustin, T., Zeileis, A., 2008. Conditional variable importance for random forests. BMC Bioinforma. 9, 307. https://doi.org/10.1186/1471-2105-9-307.

Szatmári, G., Pásztor, l, Heuvelink, G.B.M., 2021. Estimating soil organic carbon stock change at multiple scales using machine learning and multivariate geostatistics. Geoderma 403, 115356. https://doi.org/10.1016/j.geoderma.2021.115356.

Takoutsing, B., Heuvelink, G.B.M., 2022. Comparing the prediction performance, uncertainty quantification and extrapolation potential of regression kriging and random forest while accounting for soil measurement errors. Geoderma 428. https://doi.org/10.1016/j.geoderma.2022.116192.

Teng, Z.F., Zhang, C., Wang, Y.Z., 1989. Study on China's potato-cultivation divisions. Sci. Agric. Sin. 22, 35–44.

Torres-Matallana, J.A., Leopold, U., Heuvelink, G.B.M., 2021. Multivariate autoregressive modelling and conditional simulation for temporal uncertainty analysis of an urban water system in Luxembourg. Hydrol. Earth Syst. Sci. 25 (1), 193–216. https://doi.org/10.5194/hess-25-193-2021.

Wadoux, A.M.J.C., Heuvelink, G.B.M., de Bruin, S., Brus, D.J., 2021. Spatial cross-validation is not the right way to evaluate map accuracy. Ecol. Modell. 457, 109692. https://doi.org/10.1016/j.ecolmodel.2021.109692.

Wadoux, A.M.-C., Molnar, C., 2022. Beyond prediction: methods for interpreting complex models of soil variation. Geoderma 422, 115953.

Walkley, A., Black, I.A., 1934. An examination of the degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. Soil Sci. 37 (1), 29–38. https://doi.org/10.1097/00010694-193401000-00003.

Wang, X., Miao, Y., Dong, R., Zha, H., Xia, T., Chen, Z., Kusnierek, K., Mi, G., Sun, H., Li, M., 2021. Machine learning-based in-season nitrogen status diagnosis and side-dress nitrogen recommendation for corn. Eur. J. Agron. 123, 126193. https://doi.org/10.1016/j.eja.2020.126193.

Wang, C., Zang, H., Liu, J., Shi, X., Li, S., Chen, F., Chu, Q., 2020. Optimum nitrogen rate to maintain sustainable potato production and improve nitrogen use efficiency at a regional scale in China. A meta-analysis. Agron. Sustain. Dev. 40 (5). https://doi.org/10.1007/s13593-020-00640-5.

de Wit, C.T., 1953. A physical theory on placement of fertilisers, Verslagen Landbouwkundig Onderzoek (Agricultural Research Reports) 594. Staatsdrukkerij's Gravenhage, The Netherlands, p. 71.

de Wit, C.T., 1992. Resource use efficiency in agriculture. Agric. Syst. 40 (1), 125–151. https://doi.org/10.1016/0308-521X(92)90018-J.

Xu, X., He, P., Qiu, S., Zhao, S., Ding, W., Zhou, W., 2022. Nutrient management increases potato productivity and reduces environmental risk: evidence from China. J. Clean. Prod. 369. https://doi.org/10.1016/j.jclepro.2022.133357.

Xu, Y., He, P., Xu, X., Qiu, S., Ullah, S., Gao, Q., Zhou, W., 2019. Estimating nutrient uptake requirements for potatoes based on QUEFTS analysis in China. J. Agron. 111 (5), 2387–2394. https://doi.org/10.2134/agronj2018.09.0572.