



A machine learning approach to predicting plant available phosphorus that accounts for soil heterogeneity and regional variability

Rebecca L. Hall¹ · Felipe Bachion de Santana¹ · Eric C. Grunsky² · Margaret A. Browne² · Victoria Lowe² · Mairéad Fitzsimons² · Suzanne Higgins³ · Vincent Gallagher² · Karen Daly¹

Received: 18 January 2023 / Accepted: 1 September 2023 / Published online: 20 September 2023
© The Author(s) 2023

Abstract

Purpose Mehlich-3 extractable P, Al, Ca, and Fe combined with pH can be used to help explain soil chemical processes which regulate P retention, such as the role of Al, Ca, Fe, and pH levels in P fixation and buffering capacity. However, Mehlich-3 is not always the standard test used in agriculture. The objective of this study is to assess the most reliable conversion of Mehlich-3 Al, Ca, Fe, and P and pH into a commonly used soil P test, Morgan's P, and specifically to predict values into decision support for fertiliser recommendations.

Methods A geochemical database of 5631 mineral soil samples which covered the northern area of Ireland was used to model soil test P and P indices using Mehlich-3 data.

Results A random forest machine learning algorithm produced an R^2 of 0.96 and accurately predicted soil P index from external validation in 90% of samples (with an error range of $\pm 1 \text{ mg L}^{-1}$). The model accuracy was reduced when predicted Morgan's P concentration was outside of the sampled area.

Conclusions It is recommended that random forest is used to produce Mehlich-3 conversions, especially when data covers large spatial scales with large heterogeneity in soil types and regional variations. To implement conversion models into P testing regimes, it is recommended that representative soil types/geochemical attributes are present in the dataset. Furthermore, completion of a national scale geochemical survey is needed. This will enable accurate predictions of Morgan's P concentration for a wider range of soils and geographical scale.

Keywords Random forest · Soil test phosphorus · Soil index · Nutrient availability · Mehlich-3 · Morgan's phosphorus

1 Introduction

A central objective of Europe's agri-food policy is to ensure food security in the face of climate change and biodiversity loss, as well as reduce the environmental and

climate footprint of the food system. Nutrient loss from agriculture is one of the main pressures on European water quality (European Commission 2019). Fertiliser critical concentrations of soil test P are widely used in agriculture to provide P concentrations which cause the lowest possible environmental impact. Whilst ensuring optimal agricultural output and provide soil P concentration 'build up' or 'draw down' advice (Dunne et al. 2021a, b). However, soil P dynamics in agroecosystems are multifaceted and single point soil test P concentrations underrepresent the complex mechanisms of P behaviour. For instance, buffering capacity, sorption capacity, and soil biogeochemical processes control bioavailability and P loss potential.

Soil P testing for agronomic recommendations relies on chemical extraction of available P using reagents, such as Morgan's P and Mehlich-3. The Mehlich-3 soil test is widely used for P analysis to derive agronomic indicators, which has the added benefit of including extractable

Responsible editor: Jun Zhou

Highlights

- Machine learning can be used to convert between soil P tests.
- Conversion algorithms/equations can reduce the dependency on agronomic field trials.
- Morgan's P can be predicted from Mehlich-3 Al, Ca, Fe, and P and pH with R^2 0.94.
- Random forest correctly predicted 90% of external validation samples within $\pm 1 \text{ mg L}^{-1}$ error.
- To increase model prediction, calibration samples should contain representative soil types.

Extended author information available on the last page of the article

elements relevant to P behaviour in soils (Iatrou et al. 2014). For instance, Mehlich-3 extractable aluminium (Al), calcium (Ca), and iron (Fe), combined with pH, have been used to help explain soil chemical processes which regulate P retention (Bhatta et al. 2021). Consequentially, providing P nutrient advice with a substantial evaluation of the soil's biogeochemical processes in European states, the USA, and Canada. This is also the case in Ireland where Mehlich-3 has successfully been used to compare P sorption, supply potential, and plant availability (Daly et al. 2015; Dunne et al. 2021a, b).

Mehlich-3 has also been used on Irish soils in developing P bioavailability indicators (Daly et al. 2015; Dunne et al. 2020; Schulte and Herlihy 2007). For instance, field trials that have quantified fertiliser response to P found that both Morgan's P and Mehlich-3 P explained 95% of potential yield and herbage-P concentration in large-scale field trials (Schulte and Herlihy 2007), which confirmed the potential for Mehlich-3 P as an agronomic soil test P. However, Mehlich-3 concentrations when compared to Morgan's P values have been recorded as 3–30 times higher (Herlihy and McCarthy 2006; Song and Ketterings 2010), and have had a poor correlation for some soil types (Daly et al. 2015; Dunne et al. 2021a, b), therefore making it problematic to substitute soil test P methodology directly into agronomic advice without using large-scale, expensive, time-consuming field trials.

Using multiple linear regression (MLR) conversion equations on archived samples is a time-efficient and cost-effective way to introduce new methods in agronomic soil testing regimes (Ketterings et al. 2001; Vero et al. 2022). For instance, MLR was used to derive accurate conversion models for Morgan's P/Mehlich-3 in soils from New York State (Ketterings et al. 2001). The best model fit was obtained by using pH and Mehlich-3 P, Ca, and Al as independent variables, which gave an R^2 value of 0.88. Conversion models developed by Ketterings et al. (2001) have been approved by Natural Resources Conservation Service standards. However, these methods can potentially generate more error than conducting large-scale field trials.

One example of statistical error is multicollinearity, which is intrinsic in mineral soils due to mineral stoichiometry (Daoud, 2018). When more than two variables correlate, there is a risk of coefficient standard errors increasing, causing overfitting of the standard error, and moreover making variables statistically insignificant when they should be significant (Daoud, 2018). Consequently, alternative statistical analysis may be needed to convert Mehlich-3 P into Morgan's P (Ketterings et al. 2001).

An alternative form of analysis that can be used to avoid multicollinearity is random forest. Random forest is a machine learning and rule-induced algorithm that orders data by inferring relationships between the dependent variable and a set of predictors (Hounkpatin et al. 2018). Random

forest uses a decision tree method to build prediction rules by recursive binary partitioning of the dataset (Cutler et al. 2007). In tree-based modelling, nodes and leaves are used, i.e. each node uses an 'if then' statement to split the data into its most homogeneous subgroups (Heung et al. 2014). As an ensemble method of decision trees, random forest fits many decision trees to a dataset and combines the predictions of all of these trees (Harrison et al. 2021; Heung et al. 2014). The benefit of using a random forest machine learning model is that large environmental heterogeneous datasets can be analysed. For example, data which do not follow the normal distribution, contain many outliers, and may not have linearity or co-linearity can be analysed.

Conversion equations can be used in Irish soil testing regimes to predict Morgan's P bioavailability and fertiliser application rates, whilst also providing P sorption dynamics from Mehlich-3 analysis. However, the complex nature of soils and the contrast of soil parent material and/or type on large-scale datasets may need more complex analysis than previously reported by MLR analysis. The objective of this study was to find a modelling approach that accounted for soil type and regional variation when predicting available P from soil geochemical parameters. Therefore, we hypothesised that random forest would be a more reliable model than MLR, particularly when supplementing soil test conversions into soil testing programmes that cover landscapes with regional geological variation. This study used an intensively sampled soil geochemical survey of Ireland as a case study to apply the machine learning algorithm random forest to convert Mehlich-3 and pH to Morgan's P soil test.

2 Methods

2.1 Tellus geochemical survey sample collection

Soil samples were collected by the Tellus programme of Geological Survey Ireland from 2011 to 2018, from a total of 9921 sites. Samples were collected at a typical density of one bulked representative sample per 4 km². In peri-urban areas, samples were collected at a density of one sample per 1 km².

Samples were taken from undisturbed and un-forested land where possible. Sample sites were located greater than 200 m from major infrastructure and water bodies/rivers, and greater than 100 m from mapped and unmapped infrastructure and small water bodies/streams. Where possible, sites were upslope and away from all observed contaminants. A field sheet was completed for each site to record observations regarding the site, including the type, use, soil type, vegetation, and weather conditions. Further details of the survey methods and data downloads are contained in Geological Survey Ireland (2020) and can be accessed at GSI (2023).

2.2 Soil preparation and analysis

Sample preparation was undertaken as part of the Tellus programme on behalf of Geological Survey Ireland on Tellus 'A' soils. Samples were dried at 30°, disaggregated, and sieved to 2 mm.

Organic matter (%) was measured by loss-on-ignition analysis. Milled sample splits were prepared loss-on-ignition at 450 °C for organic matter. A pre-dried, 1-g weighed sample was combusted in a pre-ignited crucible in a temperature-controlled benchtop muffle furnace at 450 °C for 4 h. It was then cooled in a controlled (moisture-free) atmosphere and re-weighed. Loss-on-ignition was calculated as the proportionate mass difference before and after combustion.

Sample splits of coarse < 2 mm fraction were used for pH by CaCl₂ slurry method. A 5-g weighed sample was mixed with 12.5 mL of 0.01 M CaCl₂ and placed on a reciprocal shaker for ca. 5 min to form a slurry. The soil suspension was allowed to settle for ca. 1 h. A pH electrode and TITRANDO titrators were used to measure the solution pH potentiometrically.

Morgan's P (Morgan 1941) and Mehlich-3 (Mehlich 1984) extractants were used to calculate plant available P fractions. Soil samples were dried (40 °C) for 24 h and sieved to < 2 mm. Morgan's P was extracted using 0.5 M sodium acetate acetic acid, in a soil solution ratio of 1:5 v/w, and analysed colorimetrically with molybdate blue method by a Lachat autosampler (Daly et al. 2015). Mehlich-3 was measured with the modified Mehlich test to extract Al, Ca, Fe, and P. Soil samples were extracted at a 1:10 soil:solution ratio of 2 g and shaken with 20 mL Mehlich-3 reagent (0.2 m CH₃COOH + 0.25 m NH₄NO₃ + 0.015 m NH₄F + 0.13 m HNO₃ + 0.001 m EDTA) for 5 min on a reciprocating shaker (Daly et al. 2015). Extracts were then filtered and analysed by inductively coupled plasma–optical emission spectroscopy.

2.3 Statistical analysis

To predict Morgan's P independent variables, Mehlich-3 Al, Ca, Fe, and P and pH were selected along with categorical variables 'Season' and 'Land use' from survey field notes. Land use categories which were used to build the MLR and random forest models are 'Agricultural grassland', 'Rough land and grazing', and 'Arable land'. Only agricultural or natural lands were used to build the MLR and random forest models. For example, forestry, extraction sites, or industrial sites were removed from the dataset as these do not apply to an agricultural index system.

Samples which had > 20% organic matter as determined by loss-on-ignition were not included in the model as under Irish guidance these soils are not measured for Morgan's P as fertiliser application and management differs. The original Tellus dataset consists of 9911 samples, but with the

removal of these samples, there were 5800 samples (before outlier removal). The Terra Soil dataset of Mehlich-3 Al, Ca, Fe, and P and pH was then tested for outliers using robust regression and outlier removal analysis in Sigma Plot 14.0.

Open access geological data such as bedrock, parent material, and great soil group were not included in model development but have been discussed in terms of context. The aim of this study was to provide a simple conversion between soil test P methods (and pH) which are already used in practice and have current guidance. As soil type and geological categories are not currently used in soil testing, it was not built into MLR or random forest models.

2.4 Model development

Independent variables and Morgan's P were log-transformed prior to correlation matrix/linear regression analysis. The dataset was then split into a machine learning training dataset (70% of samples) and model internal validation dataset (30% of samples). This split was made using the Kenstone function in R studio. This function selected validation samples within the range of calibration samples that had a uniform distribution within the predictor space (R Documentation, 2023).

MLR (lrm package) and random forest (randomForest package) models were both built in R studio version 4.2.0. Calibration- and internal validation–predicted results were tested by R², root mean square error calibration (RMSEC), and prediction (RMSEP).

2.5 External validation and model testing

MLR and random forest models were tested using external validation archived data. The external validation datasets are published in Graça et al. (2021), Mellander et al. (2016), Vero et al. (2022), and Wall et al. (2013). The external validation set combined archived datasets which covered the island of Ireland (including parts of Rep. Ireland and Northern Ireland). For instance, 72% of the external dataset was collected within the Terra Soil sample area, whereas 28% of the dataset was sampled from Northern Ireland or the southern regions of Ireland. All external data sets were collected and analysed in a comparable way to the calibration and internal validation sets. Furthermore, the same sample treatment was applied, i.e. log transformation and samples > 20% organic matter removed.

2.6 Bland–Altman agreement statistics

Bland–Altman agreement statistics were used to identify bias in predicted Morgan's P values (in SigmaPlot 14.0). For instance, 'agreement' was calculated as the difference between the actual and predicted means (0 bias being perfect agreement). 'Bias' was then plotted in a scatter plot to identify

Table 1 Soil index system

Soil index	Index description	Response to fertilisers	Soil P ranges (mg L ⁻¹)	
			Grassland crops	Other crops
1	Very low	Definite	0.0–3.0	0.0–3.0
2	Low	Likely	3.1–5.0	3.1–6.0
3	Medium	Unlikely/tenuous	5.1–8.0	6.1–10
4	Sufficient/excess	None	Above 8.0	Above 10.0

whether the dataset had a ‘random relative error’ or ‘systematic error’. The limits of agreement (90% confidence interval) were plotted as the 90% overall error in model prediction. This was calculated as $2 \times$ the standard error of agreement. The standard error of bias was also used to give a tighter threshold of ‘acceptable prediction error’ in the dataset.

2.7 Soil index systems

Fertiliser recommendations are often provided in the form of a soil P index which takes into account P concentration and land use (Wall and Plunkett 2021). The soil P index system is based on large field scale experiments to calibrate P concentrations against crop yield and fertiliser response to provide critical concentrations (Schulte and Herlihy 2007). Therefore, values of Morgan’s P were placed into four bands or indices based on their fertiliser response for grassland production (Table 1).

Results from the internal and external validation sets were converted into a soil index. This was to assess whether the MLR and random forest models correctly predicted Morgan’s P in the correct agronomic management index category (Table 1).

3 Results

3.1 Soil geochemical and survey data

Mineral soils (% organic matter ≤ 20) collected from agricultural land were used to analyse the relationship between Morgan’s P and Mehlich-3 (5800 samples, 5631 after outlier analysis was performed). Morgan’s P/Mehlich-3 P values ranged from 0.27 to 209 mg kg⁻¹ and from 0.12 to 449 mg kg⁻¹, respectively. In addition, Pearson’s rank correlation analysis of Morgan’s P/Mehlich-3 P resulted in a significant positive relationship (p value = ≥ 0.000) (Pearson’s correlation coefficient: 0.57). However, when analysed by linear regression, a weak relationship was observed (R^2 0.32), which was slightly improved when analysed by fourth-order polynomial non-linear regression (R^2 0.42).

Strong correlations ensued from all of the independent variables in the dataset (Fig. 1). Al, Ca, Fe, and P from Mehlich-3 analysis, as well as pH, correlated with Morgan’s P (Fig. 1; Table 2). The strongest relationship between the independent variables was Mehlich-3 Ca and pH which had a correlation coefficient of 0.70.

Fig. 1 Correlation and regression matrix of Terra Soil dataset (Log₁₀). ***Pearson correlation as highly significant

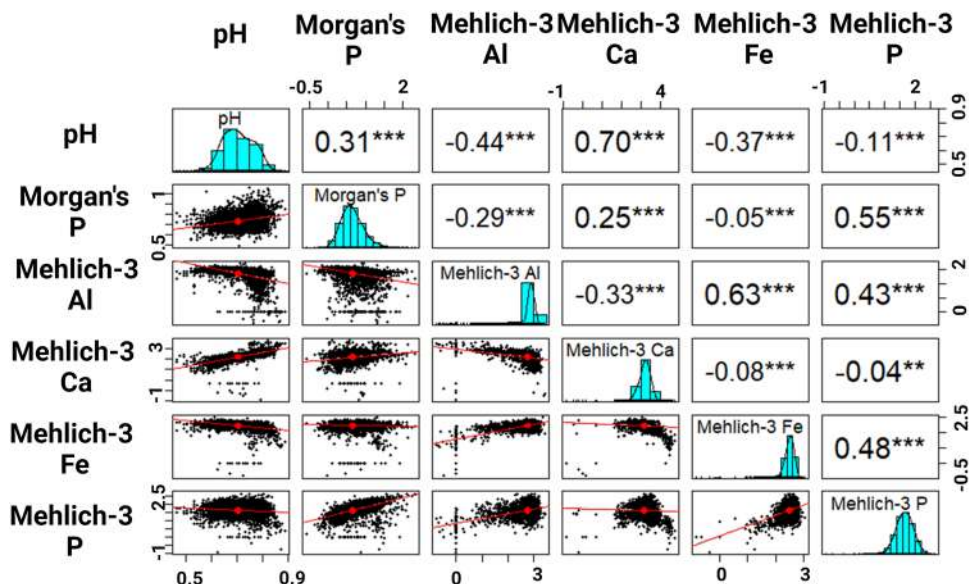


Table 2 Summary of soil geochemical and phosphorus data used in the modelling

	All soils (n = 9911)			Mineral soils (n = 5631)		
	Mean	Median	Range	Mean	Median	Range
pH	4.7	4.7	2.5–7.8	5.13	5.02	2.90–7.74
Organic matter (%)	33	15	0.63–99.9	-	-	-
	Mehlich-3 (mg kg ⁻¹)			Mehlich-3 (mg kg ⁻¹)		
Al	639	676	0.02–2671	733	747	0.17–2661
Ca	2264	1398	0.10–56970	2219	1653	0.10–56,970
Fe	314	301	0.08–1446	320	313	0.15–1064
P	48	38	0.06–449	55	45	0.12–449
	Morgan's (mg L ⁻¹)			Morgan's (mg L ⁻¹)		
P	7.63	4.95	0.20–209	6.38	4.20	0.27–209

3.2 Model development

The dataset was divided into model training (calibration) and testing (validation) datasets (Table 3; Fig. 2a). The calibration and validation datasets contained 3930 (70%) and 1690 (30%) samples, respectively (see Table 3; Fig. 2a). The mean and range of the validation datasets were within the limits of the calibration set. Additionally, the median values were all representative of the calibration dataset.

3.3 Multiple linear regression

Initial attempts to model Morgan's P produced a model which had a less than satisfactory output (R^2 value of

0.67 and an RMSEP of 0.20; Fig. 3). For this reason, outlier analysis was performed on residuals to reduce data variability. After residual outliers were removed from the calibration dataset, the R^2 value increased to a more reliable value of 0.78. The validation dataset was then used to test the MLR model. As a result, the MLR model produced $R^2 = 0.61$ and RMSEP = 0.14 (Eqs. 1–3; Fig. 3).

Agricultural grassland:

$$\text{Morgan's P} = -0.0661 - 0.33397^\circ \text{Al} - 0.0138^\circ \text{Ca} - 0.0952^\circ \text{Fe} + 0.8171^\circ \text{P} + 0.8234^\circ \text{pH} \quad (1)$$

Arable land:

Table 3 Summary of dependent, independent, and categorical variables used in model development

	Calibration (n = 3390)			Independent validation (n = 1690)			External validation (n = 369)		
	Mean	Median	Range	Mean	Median	Range	Mean	Median	Range
	—mg L ⁻¹ —								
Morgan's P	7.15	4.44	0.27–209	4.59	3.86	0.98–30	8.5	6.9	1.2–44
	—mg kg ⁻¹ —								
Mehlich-3 Al	705	724	0.17–2661	799	779	1–1601	883	886	69–1739
Mehlich-3 Ca	2442	1699	0.10–56,970	1700	1586	5–5494	1821	1602	191–8660
Mehlich-3 Fe	315	307	0.15–1064	331	322	1–806	428	444	694
Mehlich-3 P	57	44	0.12–449	52	46	1–189	85	66	7–302
pH	5.16	5.10	2.90–7.74	5.04	4.98	3.00–7.00	5.75	5.70	4.60–7.31
Categorical variables	Category	Count	Categorical code	Count			Count		
Land use	Agricultural grassland	2833	2	1593			468		
	Arable land	423	1	45			15		
	Rough land and grazing	685	3	52			-		
Season	Autumn	1997	1	1162			480		
	Winter	384	4	47			1		
	Spring	36	2	1			1		
	Summer	1524	3	480			1		

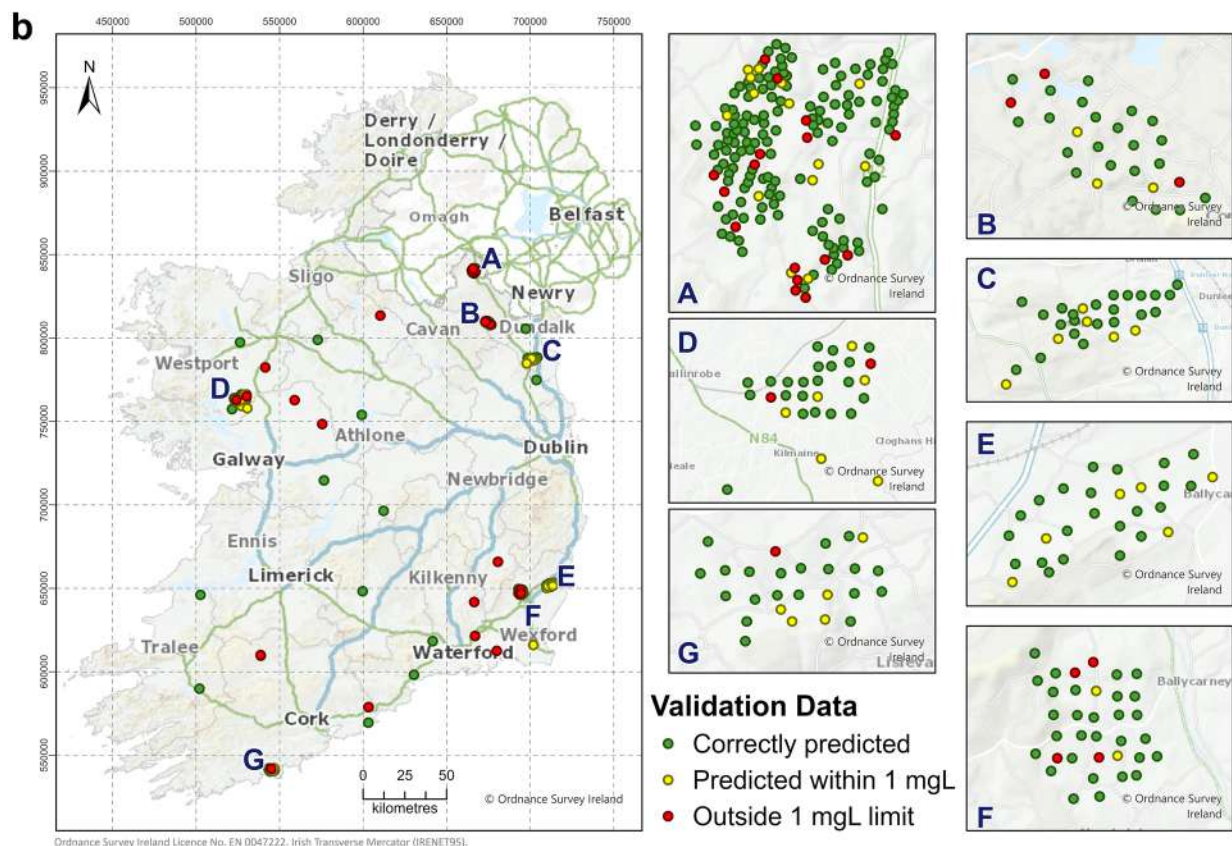
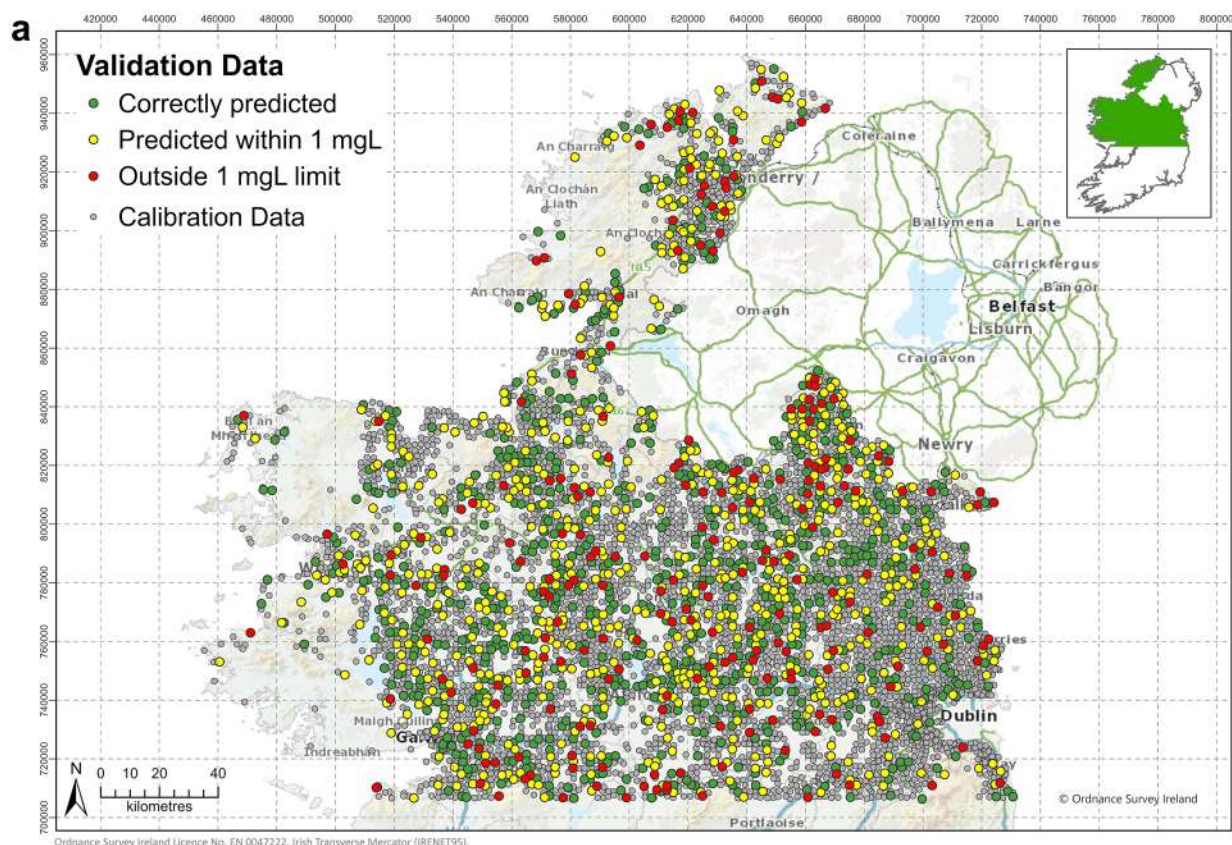


Fig. 2 **a** Map of sample points from calibration and validation datasets. **a** Calibration and validation datasets with external validation–predicted Morgan's P soil index. **b** Map of sample points from calibration and validation datasets. External validation dataset with predicted soil index. A, B, C, D, E, F, G High resolution of sample areas

$$\text{Morgan's P} = -0.0171 - 0.33397^{\circ}\text{AI} - 0.0138^{\circ}\text{Ca} - 0.0952^{\circ}\text{Fe} + 0.8171^{\circ}\text{P} + 0.8234^{\circ}\text{pH} \quad (2)$$

Rough land and grazing:

$$\text{Morgan's P} = -0.0520 - 0.33397^{\circ}\text{AI} - 0.0138^{\circ}\text{Ca} - 0.0952^{\circ}\text{Fe} + 0.8171^{\circ}\text{P} + 0.8234^{\circ}\text{pH} \quad (3)$$

ANOVA and regression coefficients were analysed to test for independent variable significance and multicollinearity. Mehlich-3 P, Al, and Fe, pH, and land use were found to be highly significant ($p=0.000$), whereas Mehlich-3 Ca and Season were not ($p=0.198$ and 0.237 , respectively). As a result, Eqs. 1–3 do not show 'Season' as a categorical variable.

3.4 Random forest

The random forest model predicted Morgan's P concentration with an $R^2=0.96$ and $\text{RMSEP}=0.07$ (Fig. 3). The validation dataset also provided a very accurate prediction accuracy ($R^2=0.87$, $\text{RMSEP}=0.08$), albeit, slightly less than the calibration dataset. Mehlich-3 P was the most important variable in the model, followed by Mehlich-3 Al, Ca, Fe, and Fe, pH, Land use, and Season (Table 4). The categorical variables 'Land use' and 'Season' had the lowest variable importance for both mean squared error and Gini importance (Table 4). 'Land use' and 'Season' only accounted for ca. 20% of the model error (mean squared error); however,

Table 4 Random forest regression model variable importance output for increase in mean squared error and random forest Gini importance

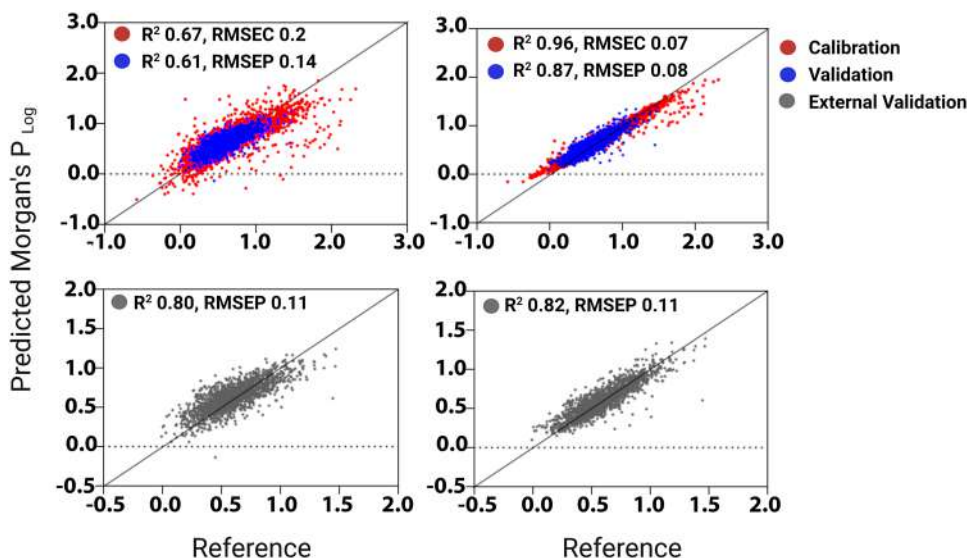
	% Mean square error	Gini importance
M3-P	212.98	238.29
M3-Al	83.89	77.46
M3-Ca	44.75	52.09
M3-Fe	42.21	44.82
pH	37.15	50.66
Land use	20.19	12.07
Season	18.50	9.54

when removed from the model the R^2 reduced from 0.96 to 0.94. Thus, they have been kept in the model to test external samples.

3.5 External validation

Independent variables and Morgan's P values in the external validation dataset presented a representative range of the calibration dataset (Table 3). Similarly, the independent variables correlated with Morgan's P and one another, with high significance ($p=0.000$). However, when outlier analysis was performed on the external validation dataset, Mehlich-3 P/pH and Mehlich-3 Al/Fe were not significantly correlated. Morgan's P and Mehlich-3 P results were significantly correlated ($p=0.000$, Pearson's correlation coefficient = 0.851), with a moderate linear relationship ($R^2=0.60$). Nonetheless, they showed poor prediction certainty ($\text{RMSEP}=3.59$). Yet, when Morgan's P values were predicted from the external validation set, both MLR and random forest performed well ($R^2=0.82$ and 0.80 , $\text{RMSEP}=0.11$ and 0.11 , respectively; Fig. 3).

Fig. 3 Regression results. **a** Calibration and validation results for MLR. **b** Calibration and validation results for random forest. **c** Independent validation results for MLR. **d** Independent validation results for random forest



3.6 Bland–Altman agreement statistics

The Bland–Altman agreement method was used to identify predicted bias and to calculate acceptable predicted error in the model output (Fig. 4). Bland–Altman agreement plots identified very low levels of random bias in the validation dataset for both MLR and random forest models (0.05 and 0.03 mgL^{-1} , respectively). The predicted bias slightly increased in the external validation datasets for MLR and random forest (-0.30 and 0.07 mgL^{-1} , respectively), yet showed no indication of fixed bias. However, Morgan's P predicted values from MLR and random forest analysis both showed discrepancies when values were above 12 mg L^{-1} (Fig. 4a–d). This was more pronounced in the internal validation set (Fig. 4c, d).

The accepted predicted error computed by Bland–Altman agreement statistics was 2 mgL^{-1} for both MLR and random forest models (with a 95% confidence interval; Fig. 4a–d). When Morgan's P predicted values of $\geq 12 \text{ mg L}^{-1}$ were removed from Bland–Altman plots, the acceptable prediction error reduced significantly to $\leq 1 \text{ mg L}^{-1}$.

3.7 External validation predictions of soil P index

Predicted values from the external validation dataset were placed into agronomic indices based on their projected response to fertiliser (Table 1). The MLR model accurately predicted soil index in 73% of the independent validation samples. Nonetheless, random forest accurately predicted soil index in 79% of the independent validation samples

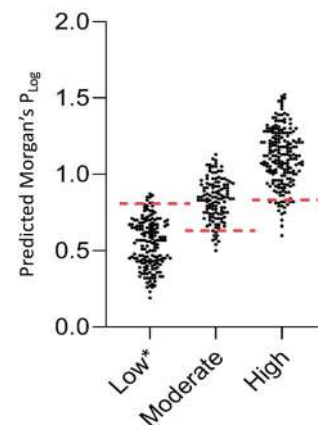
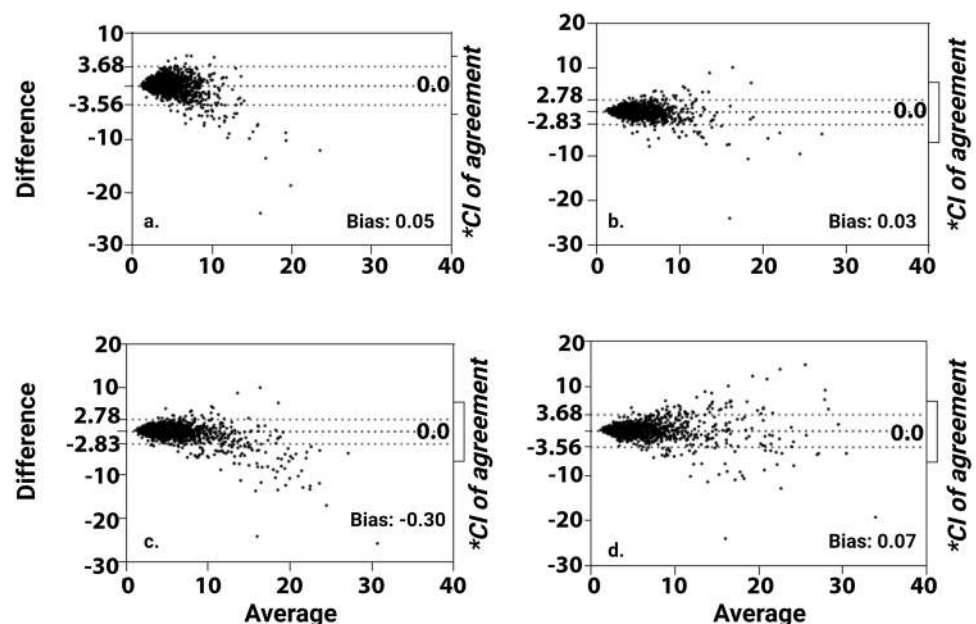


Fig. 5 Soil index predicted concentrations in actual soil index classes

(Fig. 5). Furthermore, random forest predicted 79 samples inaccurately in total. Although, 43 of these samples were within the 1 mg L^{-1} threshold, meaning 36 samples were incorrectly predicted.

In addition, it was of importance to note the location of the inaccurately predicted samples (Fig. 2b). For instance, from the 36 samples which were not predicted into the correct soil P index (with accepted 1 mgL^{-1} variation), 30% of these samples were outside the sampled area (Fig. 2a, b). Furthermore, ca. 90% of samples within the sampled area were correctly predicted into the correct index category with an error range of $\pm 1 \text{ mg L}^{-1}$.

Fig. 4 Bland–Altman plots. **a** MLR internal validation. **b** MLR external validation. **c** Random forest internal validation. **d** Random forest external validation



4 Discussion

4.1 Soil geochemical survey data

Morgan's P and Mehlich-3 P were significantly positively correlated (Fig. 1), which was in line with other published studies (Dunne et al. 2020, 2021a, b). However, the relationship between both tests in regression analysis was weak. This was largely due to the geochemical variability in the soils analysed. Furthermore, the observed contrast in ranges in Morgan's P and Mehlich-3 in this study originated from the design of the different extraction methods. For instance, Mehlich-3 was designed to enhance the release of Al and Fe phosphate and suppress adsorption from P colloids, and reagent acidity aids the dissolution of available P from Ca and Fe bound P (Song et al. 2010), whereas Morgan's P was exclusively designed to dissolve plant available P compounds (Daly and Casey 2005).

Correlation coefficient analysis for Mehlich-3 Al, Ca, Fe, and P all showed strong significant relationships with Morgan's P and Mehlich-3 P, albeit noisy, in regression analysis (Fig. 1). This was unsurprising as the soil archive and data used in this study were unique in providing a wide range of element concentrations and pH. As this study used a large heterogeneous landscape with varying geochemical attributes, representative of Irish mineral soils, a new perspective previous studies have not captured was provided (Daly et al. 2015; Dunne et al. 2020; Schulte and Helrihy 2007). For instance, the Northern region is dominated by Schists and Dalradian metasediments, the Eastern regions by Grey Wacke, the Southern region by igneous rocks, the West by Peatlands, and the Midland areas by Limestone (Geological Survey Ireland 2020). For purposes relating geochemistry to P bioavailability, this translates to high Al and Fe hydrous oxide concentrations in the North/Eastern regions, and higher Ca concentrations and lower Fe/Al content in the midlands/southern region. And, mineral soils in western regions typically have higher levels of Fe (Geological Survey Ireland 2020; Fay et al. 2007).

4.2 Multiple linear regression and random forest model output

Random forest outperformed MLR in predicting Morgan's P concentration and was overall the better model for Morgan's P prediction from pH and Mehlich-3-Al, Ca, Fe, and P (Fig. 2). In general, it is expected that predicted variation from field trials or survey data is within a threshold of 75% (Song et al. 2010). Both models in this study exceeded the 75% threshold once residual outliers had been removed from MLR ($R^2 = 0.78$). Yet, Random forest had a higher

prediction accuracy (both R^2 and RMSEP) in the calibration and validation models. However, it was clear that multicollinearity was present in the dataset from coefficient and variable importance analysis.

Analysis of variance (ANOVA) and coefficient analysis were included in MLR analysis due to the multicollinearity of independent variables (Fig. 1). ANOVA and coefficient analysis of MLR showed independent variables Mehlich-3 P, Al, and Fe and pH to be highly significant in the MLR model ($p = 0.000$). However, Mehlich-3 Ca was not significant in the calibration dataset. As Ireland consists of higher Ca concentrations and lower Fe/Al content in the midlands/southern region, when applying predicted results in practice, there may be inconsistencies in results in these regions. Furthermore, there was a large range of Ca concentration in the dataset (Tables 2 and 3), and Mehlich-3 Ca and Mehlich-3 P were positively correlated (Fig. 1), and it was expected that Ca would have provided a significant interaction.

Multicollinearity can be overcome by using a logratio approach in the application of MLR or random forest (Aitchison 1986). However, analysis from this dataset showed there was no effect on the predicted data and therefore was not included. Random forest provided a robust way to analyse large datasets as it did not require the data to be normally distributed. Furthermore, it did not depend on parameters being linear, or co-correlated. Therefore, not only did the random forest model yield better predictions, but it was also more statistically coherent. Besides, random forest was better suited to this soil archive as many of the variables were co-correlated.

Variable importance in random forest analysis was valuable when interoperating the underlying mechanisms of soil test P bioavailability (Table 4). Variable importance (mean square error) from random forest calculated Mehlich-3 P was the most important variable in the random forest model, followed by Mehlich-3 Al, Ca, and Fe, pH, Land use, and Season. Mehlich-3 Al, Ca, and Fe metals represented labile forms of metals. These metals are associated with sorption sites on clay particle surfaces and amorphous forms of metals that can be involved in the binding and release of P into soluble forms (Hall et al. 2020), therefore understandably provided the highest importance when predicting Morgan's P. Another important mechanistic property in P bioavailability is pH, as sorption of P to amorphous forms of metals can depend on pH (Penn and Camberato 2019).

Soil pH played an important role in Morgan's P concentration prediction (error increased by 37% once removed; Table 4). The optimum pH for P binding to Al oxides is 4.5–6.3 (Penn and Camberato 2019), whereas P fixation to Ca oxides in soils is most prominent in neutral to alkaline soils, which were present in the southern regions of the survey area (Penn and Cambareto 2019; GSI 2020). Furthermore, the dataset contained ca. 75% of low pH soils, which

may explain why Al and then Ca followed Mehlich-3 P in variable importance. Additionally, some regions in Ireland have areas of low pH and high Fe accumulation (Cunningham et al. 2001), which resulted in Fe having a lower, yet significant, contribution.

Land use and Season had the lowest importance when analysed by random forest (Table 4). Land use and Season only accounted for 20% and 19% mean squared error, respectively. However, when Land use and Season were removed from the random forest model, the R^2 reduced from 0.96 to 0.92. Land use and Season were initially included in the model as each of the classifications may have had different P management strategies. For example, P fertiliser management processes in agricultural systems vary depending on production needs (Wall and Plunkett 2021), with more P being applied to sustain arable land (Table 1). As noted in Table 3, the majority of samples in this study were collected from agricultural grassland (Table 3). However, in the southern regions of Ireland, arable farming is more prevalent (Corine 2019). Therefore, the random forest model may be unrepresentative in regions outside of the dataset, and require a wider distribution of arable land use categories.

4.3 External validation

When using the external validation data, the random forest method was only 2% more accurate than MLR at predicting Morgan's P ($R^2=0.82$ and 0.80 ; Fig. 3). When Morgan's P and Mehlich-3 P were analysed by linear regression, there was a significant linear relationship, albeit with poor prediction certainty ($R^2=0.60$). Yet, this was stronger than the calibration dataset ($R^2=0.60$). Therefore, the success of Morgan's P prediction from MLR in the external dataset can be explained by the similar association of Morgan's/Mehlich-3 P. However, as random forest was the most statistically accurate model, only random forest was further discussed.

4.4 Prediction error limits (Bland–Altman agreement statistics)

When using predicted results, it is good practice to provide a comprehensive evaluation of the prediction error. The average bias of the validation and external validation sets had near to perfect agreement (≤ 0.03 ; Fig. 4) for the random forest model. The 90% limits of agreement calculated by the Bland–Altman method were ca. $2\text{--}3\text{ mg L}^{-1} \pm$. However, it was not practicable to have a 2 mg L^{-1} threshold limit on the concentration in this study. For instance, the Low Soil Index category shown in Table 1 increased by an increment of 2 mg L^{-1} . As a result, predicted Morgan P values in this category could be placed in the incorrect index, consequently providing incorrect fertiliser management advice. However, when samples $> 12\text{ mg L}^{-1}$

were removed from the dataset, the lesser value of $1\text{ mg L}^{-1} \pm$ was calculated.

Dunne et al. (2021a, b) identified Morgan P values $> 12\text{ mg L}^{-1}$ as P saturated. P decline in these soils to the target agronomic P index takes longer than soils between 8 and 12 mg L^{-1} . In the Irish soil P index system, Morgan's P concentrations which are $> 8\text{ mg L}^{-1}$ (for grassland soils) or $> 10\text{ mg L}^{-1}$ (for arable soils) are classed as high P soils. When a soil index is in this threshold, management strategies to draw down P and reduce the risk of P loss to water are to be implemented. When predicted Morgan's P results were $> 12\text{ mg L}^{-1}$, it was understood that they are in a 'P saturated' category and the random forest model was not as accurate in analysing the results. Therefore, traditional wet chemistry methods should be used in these soils.

4.5 Predicting into a soil P index

Validation is important in predictive modelling as it tests the robustness of the model performance in practice. When predicted P concentration from the external validation dataset was placed into a soil P index, the majority of samples (90%) were predicted within the correct index (or within the 1 mg L^{-1} error threshold). The external validation samples which were not correctly predicted mainly came from County Monaghan (Fig. 2b; (A)), and counties Carlow, Cork, and Kilkenny (Fig. 2b; (E, F, G)). These samples were collected in the southern regions of Ireland, with differing soil types. However, most (ca. 90%) of the dataset was correctly predicted.

Soils are formed via weathering of parent material, which includes bedrock, Quaternary deposits, and organic matter. Hence, the mineralogy of soils is determined by the parent material type, which changes depending on location (Schulte et al. 2018). The spatial extent of the database focused on the Northern half of Ireland. As a result, some of the soils and parent materials which occurred in the external dataset (as identified in Fig. 2) were not represented in the calibration set. Of the 369 samples in the external validation sets, 36 samples were predicted outside the 2 mg L^{-1} Morgan P threshold, and 30% were derived from areas which were not represented in the model. Therefore, random forest accounted for soil type and regional variation when predicting available P from soil geochemical parameters. However, to use the random forest model in practice, a national geochemical survey is needed to ensure that the model is representative of all Irish soils.

Moreover, the Soil Health Monitoring Law proposal has an emphasis on management to reduce nutrient loss, increase carbon storage, and mitigate against CO_2 emissions. Generally, soil P/carbon stoichiometry is an important indicator relating to effective nutrient cycling and ecological interactions (Cui et al. 2022). For instance, soil carbon release via quotient CO_2 is reduced when there is microbial P limitation (Chen et al. 2023; Cui et al. 2022; Zhai et al. 2022). Furthermore, Cui et al.

(2022) noted soil CO₂ increased by 19–26% after a high rate of P fertiliser was applied to incubated soils. There was also an observed trade-off between P and microbial carbon use efficiency in managed systems (Chen et al. 2023; Zhai et al. 2022). Therefore, when supplementing field trials for conversion equations, it is vital that accurate tools are used to reduce the environmental and climate footprint of the agri-food system.

5 Conclusions

Mehlich-3 provided available P levels as well as a comprehensive overview of available elements which interact with P. Therefore, providing Mehlich-3 soil analysis alongside pH can help understand and manage P behaviour more effectively than only providing a single parameter P test result. When converting Mehlich-3 Al, Ca, Fe, and P and pH to Morgan's P, random forest was a superior multivariate model. Machine learning conversion models, such as random forest, benefit from frequent data updates to improve accuracy. For a universal model, and implementation into Irish P testing regimes, it is recommended that additional soil types are added to the calibration dataset. This will provide a more representative prediction model that reflects the wide range of parent materials in Irish soils.

Acknowledgements This work has been jointly supported by Geological Survey Ireland and Teagasc (Terra Soil Collaborative Agreement 2018). The authors acknowledge and are grateful to Ms. Courtney Doyle, Patricia Berry, and Dr. Anna Fenelon for technical support in sample processing and technical support at Teagasc Johnstown Castle laboratories. Geological Survey Ireland authors publish with the permission of the Director of Geological Survey Ireland.

Data availability Data will be made available upon request.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aitchison J (1986) The statistical analysis of compositional data. Chapman and Hall, New York, 416p
- Bhatta A, Prasad R, Chakraborty D, Shaw JN, Lamba J, Brantley E, Torbert HA (2021) Mehlich 3 as a generic soil test extractant for environmental phosphorus risk assessment across Alabama soil regions. *Agrosyst Geosci Environ* 4(3). <https://doi.org/10.1002/agg2.20187>
- Chen J, Cordero I, Moorhead DL, Rowntree JK, Simpson LT, Bardgett RD, Craig H (2023) Trade-off between microbial carbon use efficiency and specific nutrient-acquiring extracellular enzyme activities under reduced oxygen. *Soil Ecol Lett* 5(2). <https://doi.org/10.1007/s42832-022-0157-z>
- Corine (2019) Corine Land Cover 2018 (vector) - version 20, Jun. 2019. Online. Available at: <https://sdi.eea.europa.eu/catalogue/srv/api/records/53ef1493-e7a1-4216-b043-87a7c2a5a68d>. Accessed 11 Sept 2023
- Cui Y, Moorhead DL, Wang X, Xu M, Wang X, Wei X, Zhu Z, Ge T, Peng S, Zhu B, Zhang X, Fang L (2022) Decreasing microbial phosphorus limitation increases soil carbon release. *Geoderma* 419(February):115868. <https://doi.org/10.1016/j.geoderma.2022.115868>
- Cunningham DA, Collins JF, Cummins T (2001) Anthropogenically-triggered iron pan formation in some Irish soils over various time spans. *CATENA* 43(3):167–176. [https://doi.org/10.1016/S0341-8162\(00\)00161-2](https://doi.org/10.1016/S0341-8162(00)00161-2)
- Cutler DR, Edwards TC, Beard KH, Cutler A, Hess KT, Gibson J, Lawler JJ (2007) Random forests for classification in ecology. *Ecology* 88(11)
- Daly K, Casey A (2005) Environmental aspects of soil phosphorus testing. *Irish J Agric Food Res* 44(2):261–279
- Daly K, Styles D, Lalor S, Wall DP (2015) Phosphorus sorption, supply potential and availability in soils with contrasting parent material and soil chemical properties. *Eur J Soil Sci* 66(4):792–801. <https://doi.org/10.1111/ejss.12260>
- Daoud JI (2018) Multicollinearity and regression analysis. *J Phys Conf Ser* 949(1). <https://doi.org/10.1088/1742-6596/949/1/012009>
- Dunne KS, Holden NM, O'Rourke SM, Fenelon A, Daly K (2020) Prediction of phosphorus sorption indices and isotherm parameters in agricultural soils using mid-infrared spectroscopy. *Geoderma* 358(June 2019):113981. <https://doi.org/10.1016/j.geoderma.2019.113981>
- Dunne KS, Holden NM, Daly K (2021a) A management framework for phosphorus use on agricultural soils using sorption criteria and soil test P. *J Environ Manage* 299(July):113665. <https://doi.org/10.1016/j.jenvman.2021.113665>
- Dunne KS, Holden NM, Daly K (2021b) Predicting phosphorus sorption isotherm parameters in soil using data of routine laboratory tests. *Pedosphere* 31(5):694–704. [https://doi.org/10.1016/S1002-0160\(21\)60012-7](https://doi.org/10.1016/S1002-0160(21)60012-7)
- European Commission (2019) Communication on the European Green Deal. European Commission, Brussels
- Fay D, Kramer G, Zhang C, McGrath D, Grennan E (2007) Soil geochemical atlas of Ireland. *Environ Protect* 128
- Geological Survey Ireland (2020) Geochemical survey Ireland: Tellus geochemical survey: shallow topsoil data from the border and west of Ireland. Available at: https://gsi.geodata.gov.ie/downloads/Geochemistry/Reports/Tellus_A_geochemistry_data_report_2020_v1.0.pdf. Accessed 13 Apr 2023
- Graça J, Daly K, Bondi G, Ikoyi I, Crispie F, Cabrera-Rubio R, Cotter PD, Schmalenberger A (2021) Drainage class and soil phosphorus availability shape microbial communities in Irish grasslands. *Euro J Soil Biol* 104:103297. <https://doi.org/10.1016/j.ejsobi.2021.103297>
- GSI (2023) Geochemistry. Online. Available at: <https://www.gsi.ie/en-ie/data-and-maps/Pages/Geochemistry.aspx#ShallowTopsoilA>. Accessed 13 Apr 2023
- Hall RL, Boisen Staal L, Macintosh KA, McGrath JW, Bailey J, Black L, Gro Nielsen U, Reitzel K, Williams PN (2020) Phosphorus speciation and fertiliser performance characteristics: A comparison of waste recovered struvites from global sources. *Geoderma* 362(November 2019):114096. <https://doi.org/10.1016/j.geoderma.2019.114096>

- Harrison JW, Lucius MA, Farrell JL, Eichler LW, Relyea RA (2021) Prediction of stream nitrogen and phosphorus concentrations from high-frequency sensors using random forests regression. *Sci Total Environ* 763. <https://doi.org/10.1016/j.scitotenv.2020.143005>
- Herlihy M, McCarthy J (2006) Association of soil-test phosphorus with phosphorus fractions and adsorption characteristics. *Nutr Cycl Agroecosyst* 75(1–3):79–90. <https://doi.org/10.1007/s10705-006-9013-2>
- Heung B, Bulmer CE, Schmidt MG (2014) Predictive soil parent material mapping at a regional-scale: a random forest approach. *Geoderma* 214–215:141–154. <https://doi.org/10.1016/j.geoderma.2013.09.016>
- Houknpatin KOL, Schmidt K, Stumpf F, Forkuor G, Behrens T, Scholten T, Amelung W, Welp G (2018) Predicting reference soil groups using legacy data: a data pruning and random forest approach for tropical environment (Dano catchment, Burkina Faso). *Sci Rep* 8(1). <https://doi.org/10.1038/s41598-018-28244-w>
- Iatrou M, Papadopoulos A, Papadopoulos F, Dichala O, Psoma P, Bountla A (2014) Determination of soil available phosphorus using the Olsen and Mehlich 3 methods for Greek soils having variable amounts of calcium carbonate. *Commun Soil Sci Plant Anal* 45(16):2207–2214. <https://doi.org/10.1080/00103624.2014.911304>
- Ketterings QM, Bellows B, Czymmek K, Reid W (2001) Do modified Morgan and Mehlich-III P have a Morgan P equivalent? Whole-farm nutrient mass balances for New York dairy farms View project On-farm Research: strip trial analysis, Precision Agriculture method View project. <https://www.researchgate.net/publication/241752390>
- Mehlich A (1984) Mehlich-3 soil test extractant: a modification of Mehlich-2 extractant. *Comm Soil Sci Plant Anal* 15:1409–1416. <https://doi.org/10.1080/00103628409367568>
- Mellander PE, Jordan P, Shore M, McDonald NT, Wall DP, Shortle G, Daly K (2016) Identifying contrasting influences and surface water signals for specific groundwater phosphorus vulnerability. *Sci Total Environ* 541:292–302. <https://doi.org/10.1016/j.scitotenv.2015.09.082>
- Morgan MF (1941) Chemical diagnosis by the universal soil testing system. Connecticut Agricultural Experiment Station (New Haven. Bulletin 450)
- Penn CJ, Camberato JJ (2019) A critical review on soil chemical processes that control how soil pH affects phosphorus availability to plants. *Agric (Switz)* 9(6). <https://doi.org/10.3390/agriculture9060120>
- R Documentation (2023) Kennard-Stone algorithm for calibration sampling. Online, available at: <https://search.r-project.org/CRAN/refmans/prospectr/html/kenStone.html>. Accessed 13 Apr 2023
- Schulte R, O'Sullivan L, Creamer R (2018) Soil functions—an introduction. https://doi.org/10.1007/978-3-319-71189-8_13
- Schulte RPO, Herlihy M (2007) Quantifying responses to phosphorus in Irish grasslands: interactions of soil and fertiliser with yield and P concentration. *Eur J Agron* 26(2):144–153. <https://doi.org/10.1016/j.eja.2006.09.003>
- Song C, Ketterings QM (2010) Impact of soil temperature and moisture on Mehlich-3 and Morgan soil test phosphorus. *Soil Sci* 175(10):511–518. <https://doi.org/10.1097/SS.0b013e3181f850d4>
- Vero SE, Doody D, Cassidy R, Higgins S, Nicholl G, Campbell J, Mellander PE, McDonald N, Burgess E, Daly K, Sherry E (2022) Comparison of soil phosphorus index systems for grassland in the cross-border region of Ireland#. *J Plant Nutr Soil Sci* 185(1):110–119. <https://doi.org/10.1002/jpln.202100194>
- Wall D, Plunkett M (2021) Major and micro nutrient advice for productive agricultural crops. 5:51–53. <http://hdl.handle.net/11019/2475>
- Wall DP, Jordan P, Melland AR, Mellander PE, Mechan S, Shortle G (2013) Forecasting the decline of excess soil phosphorus in agricultural catchments. *Soil Use Manag* 29(1):147–154. <https://doi.org/10.1111/j.1475-2743.2012.00413.x>
- Zhai Z, Luo M, Yang Y, Liu Y, Chen X, Zhang C, Huang J, Chen J (2022) Trade-off between microbial carbon use efficiency and microbial phosphorus limitation under salinization in a tidal wetland. *Catena* 209(P1):105809. <https://doi.org/10.1016/j.catena.2021.105809>

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

Authors and Affiliations

Rebecca L. Hall¹ · Felipe Bachion de Santana¹ · Eric C. Grunsky² · Margaret A. Browne² · Victoria Lowe² · Mairéad Fitzsimons² · Suzanne Higgins³ · Vincent Gallagher² · Karen Daly¹

✉ Rebecca L. Hall
rebecca.hall@teagasc.ie

¹ Environment, Soils and Land Use Department, Teagasc Johnstown Castle Research Centre, Wexford Y35 TC97, Ireland

² Geological Survey Ireland, Block 1 Booterstown Hall, Booterstown Co Dublin, Blackrock A94 N2R6, Ireland

³ Agri-Food and Biosciences Institute, 18a Newforge Lane, Belfast BT9 5PX, UK