Advancing soil spectroscopy from laboratory to handheld applications to support site-specific sustainable phosphorus management.

By

Sifan Yang, MSc.

A thesis submitted for the award of PhD

School of Chemical Sciences

Dublin City University

Supervisors:

Prof. Blánaid White

Dr. Karen Daly (Teagasc, Crops, Environment and Land-Use Programme)

November 2025

DECLARATION

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Doctor of Philosophy is entirely my own work, and that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed: Sifan Yang Sifan Yang

Student ID No: 21268996

Date: 17/11/2025

ACKNOWLEDGEMENTS

I would like to acknowledge the Department of Agriculture, Food and the Marine for funding this research through the SENSUS project: Sensing Nutrients for Agronomic Advice and Sustainability Measures. I wish to express my deepest gratitude to my supervisors, Dr. Karen Daly and Prof. Blánaid White, to whom I will always be grateful for offering me the opportunity to pursue this PhD. Thank you for your expertise, patience, and encouragement. Your support has been fundamental throughout my four-year PhD journey and in the development of this thesis, particularly in enhancing my academic writing, strengthening my research skills, and shaping my growth as an early-career researcher.

I am sincerely grateful to the technical staff and postdoctoral researchers at both the School of Chemical Sciences, Dublin City University, and Teagasc (Johnstown Castle, Co. Wexford) for their invaluable assistance and support over the years. Special thanks to Felipe and Rebecca for generously sharing their knowledge, guiding me through research methods, and encouraging me to develop a more confident mindset. Their support has inspired me both in my scientific work and in shaping the standards I now set for myself as I strive to become as professional as they are in my future career.

I am profoundly thankful to my family for their unwavering support during my studies. Being away from home for many years, particularly while living alone in a foreign country, has been more difficult than I imagined. Your being always reachable, even in the middle of the night, gave me a deep sense of security and reminded me that no matter where I am, I always have a home and family behind me.

This journey has been arduous, but all the guidance, love, and encouragement I received from each of you have been my pillars through every moment of self-doubt and every time I felt like giving up. Thank you to everyone who has supported me throughout my challenging and often stressful PhD journey. I do not know what the world will become, but I hope it holds more peace and love. I do not know where I will be or what I will find, but I hope to get a good job that is worthy of the time and effort I have dedicated. I wish everyone good health, happiness, and the strength to face whatever lies ahead.

LIST OF PUBLICATIONS & PRESENTATIONS

Publications

Yang, S. *et al.* Integrating soil phosphorus sorption capacity with agronomic indices to improve sustainable P use in agriculture. *J. Soils Sediments* (2024) doi:10.1007/s11368-024-03900-z.

Yang, S., White, B., de Santana, F. B., Hall, R. L. & Daly, K. Comparing the potential of benchtop and handheld mid-infrared spectrometers for predicting soil phosphorus (P) sorption capacity and evaluating the influence of sample preparation. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **322**, 124856 (2024) doi.org/10.1016/j.saa.2024.124856.

Yang, S., White, B., de Santana, F.B., Hall, R.L., Daly, K. Mitigating moisture effects in spectral models for determining soil phosphorus sorption capacity using handheld MIR and NIR. *Geoderma* 463, 117579 (2025) doi.org/10.1016/j.geoderma.2025.117579.

Poster Presentations

From bench-top to handheld, how efficient by using Mid-infrared spectroscopy to predict soil phosphorous sorption? Analytical Research Forum 2022 (ARF2022), an event organised by the Royal Society of Chemistry (London - June 13th - 14th, 2022).

Coupling of P dynamics with P status for soil P management in precision agriculture. 74th Irish Universities Chemistry Research Colloquium, an event organised by the Institute of Chemistry of Ireland (University of Galway - June 14th - 15th, 2023).

Comparing the potential of bench-top and handheld mid-infrared (MIR) spectrometers for predicting soil phosphorus (P) sorption capacity and evaluating the influence of sample preparation. 4th International Workshop on Spectroscopy and Chemometrics, an event organised by the Vistamilk SFI Research Centre (Teagasc Moorepark - April 29th - 30th, 2024).

Oral Presentations

Prediction of soil phosphorus sorption capacity in agricultural soils using mid-infrared spectroscopy. EGU General Assembly 2024 (Vienna, April 14th – 19th, 2024).

Prediction of soil phosphorus sorption capacity in agricultural soils using mid-infrared spectroscopy. Teagasc Seminar Series, an event organised by the Teagasc Research Centre (Wexford, April 9th, 2024).

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Agilent	Agilent 4300 FTIR (Agilent Technologies, Inc., Santa Clara, CA, USA)
Al	Aluminium
Al^{3+}	Aluminium ion
Al_{ox}	Oxalate-extractable aluminium
ATR	Attenuated total reflectance
BD	Bulk density
Bruker	Benchtop FTIR Spectrometer INVENIO S® (Bruker, Massachusetts - USA)
Ca^{2+}	Calcium
$CaCl_2$	calcium chloride
$CaCO_3$	Calcium carbonate
CEC	Cation-exchange capacity
CO_2	Carbon dioxide
CO_3^{2-}	Carbonate ion
CWT	Continuous wavelet transform
DG-NIR	Dispersive near-infrared
DPS	Degree of soil phosphorus saturation
DRIFTS	Diffuse reflectance infrared Fourier-transform spectroscopy
DRS	Diffuse reflectance spectroscopy
DS	Direct standardisation
EC	Electrical conductivity
EMR	Electromagnetic radiation
EPO	External parameter orthogonalisation
FAO	Food and Agriculture Organization
Fe	Iron
Fe^{3+}	Iron ion
Fe_{ox}	Oxalate-extractable iron
FIR	Far-infrared
FTIR	Fourier-transform infrared
FT-NIR	Fourier-transform near-infrared
GMC	Gravimetric water content
H_2O	Water
Inno	IR-S-G1 portable NIR spectrometer (900 - 1700 nm)
IR	Infrared
KH_2PO_4	Monopotassium orthophosphate
LR	Lime requirement
LUCAS	Land Use/Land Cover Area Frame Survey
LV	Latent variable
LVF	Linear variable filter
M3-Al	Mehlich-3-aluminium
M3-Ca	Mehlich-3-extractable calcium
M3-Fe	Mehlich-3-extractable iron
MBC	Maximum buffer capacity ($mg \cdot kg^{-1}$)
MDKS	Mahalanobis distance
MEMS	Micro-electro-mechanical systems
Mg^{2+}	Magnesium
mid-DRIFTS	Mid-infrared diffuse reflectance infrared Fourier-transform spectroscopy

MIR	Mid-infrared
ML	Machine learning
MSC	Multiplicative scatter correction
N	Nitrogen
NAP	Nitrated Action Programme
Neo	NeoSpectra scanner (Si-Ware Systems, 1350-2500 nm)
NIR	Near-infrared
OSC	Orthogonal signal correction
P	Phosphorus
P_2O_5	Phosphorus pentoxide
PAS	Photoacoustic spectroscopy
PBC	Phosphorus buffering capacity
PBI	Phosphorus buffer index
PCA	Principal component analysis
PCR	Principal component regression
PCs	Principal components
PDS	Piece-wise direct standardization
Pi	Inorganic phosphorus
PLSR	Partial least squares regression
Po	Organic phosphorus
PR	Phosphate rock
PSC	Phosphorus sorption capacity
PSI	Phosphorus sorption index
RF	Random forest
SG	Savitzky–Golay
Smax	P sorption maximum ($mg \cdot kg^{-1}$), derived from the Langmuir model
SNR	Signal-to-noise ratio
SO_4^{2-}	Sulphate ion
SOC	Soil organic carbon
SOM	Soil organic matter
STP	Soil test phosphorus
SVM	Support vector machines
TC	Total carbon
TN	Total nitrogen
TP	Total phosphorus
Vis-NIR	Visible to near-infrared

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Advancing soil spectroscopy from laboratory to handheld applications to support site-specific sustainable phosphorus management.

Sifan Yang, MSc.

Soil phosphorus (P) dynamics are of agronomic and environmental importance. Near-infrared (NIR) and mid-infrared (MIR) soil spectroscopy offer rapid and cost-effective alternatives to traditional wet chemical tests. Testing soil P sorption capacity using traditional isothermal sorption experiments allows for the establishment of reference values and the definition of behavioural classes for integration into agronomic guidelines. The aim of this study was to develop infrared spectroscopy as a practical alternative, advancing from laboratory-based to handheld on-site testing of P sorption capacity, with reference to the refined Soil Index System that integrates soil P sorption characteristics to support agronomic decision-making for long-term, site-specific sustainable P management. Mineral topsoil samples (loss-on-ignition ≤ 20%, depth 0.05 - 0.20 m), covering 35,716 km^2 of Ireland, were taken from the archive of the Tellus national programme.

Calibration samples were characterised by P status and Langmuir sorption parameters, including P sorption maximum capacity (Smax, $mg \cdot kg^{-1}$) and binding affinity from the wet chemical extractions. Segmented regression identified a significant change-point in P behaviour at Smax = 450.03 $mg \cdot kg^{-1}$, classifying soils into 'low' (SL) and 'high' (SH) P sorption capacities.

Chemometric models for Smax prediction were developed using samples in a dry state. Benchtop (Bruker, 4000–400 cm^{-1}) and handheld (Agilent, 4000–650 cm^{-1}) MIR spectrometers were used to scan <0.100 mm (ball-milled) and <2 mm soil samples. In comparison, Agilent obtained spectral libraries contained more noise and less resolution. Unlike Bruker, ball milling preparation step significantly influenced the performance of Agilent. Handheld NIR instruments (InnoSpectra, 900–1700 nm, and NeoSpectra, 1350–2550 nm) were also evaluated for <2 mm samples. Only Agilent and NeoSpectra handheld instruments demonstrated potential for discriminating SL/SH soils.

A systematic re-wetting experiment assessed moisture effects on <2 mm samples scanned with handheld NIR and MIR spectrometers. The External Parameter Orthogonalisation (EPO) algorithm successfully removed moisture effects. However, predictions from EPO-transformed wet spectra remained less accurate than those from dry samples. In practice, to simplify and improve the reliability of predictions from MIR spectroscopy, maintaining relatively dry field conditions is highly recommended.

The findings in this thesis contribute to site-specific P management, water quality protection, and policy development for future studies.

Chapter 1

Introduction & Literature Review

1.1 Soil Phosphorus

1.1.1 Background: Soil P for achieving food security and ecosystem sustainability

The global population reached eight billion in mid-November 2022 and is projected to reach 9.7 billion by 2050[1]. To meet the rising demand driven by population growth and dietary changes, particularly exacerbated by increased food insecurity after the COVID-19 pandemic, the Food and Agriculture Organization (FAO) estimates that global food production will have to increase by 60% by 2050[2,3]. Phosphorus (P) is an essential plant nutrient, the second most important after nitrogen (N), and is critical for achieving optimum crop yields.

Soil P is stored in various formations: (i) plant-available P, soluble in soil solution; (ii) labile-P, P held by soils that are likely to be transferred to plant-available P; and (iii) non-labile P. Less than 1% of the total P in soils is plant-available, i.e., in a form readily accessible for plant uptake[4–6]. However, a large amount of plant-unavailable P is stored in (i) organic matter (e.g., dead plant and animal residues, soil microorganisms, and humic substances), (ii) primary P minerals (e.g., apatite, strengite, and variscite), and (iii) P-bearing secondary minerals (e.g. clay minerals and iron-rich clay aggregates); in addition to (iv) the labile pool, where inorganic P is likely bound to soil clay minerals, iron (Fe), aluminium (Al) oxides and calcium carbonate ($CaCO_3$)[4,7,8]. Over the past century, human existence has depended on producing P fertilisers and their application to agricultural soils to sustain food production[9].

The application of P fertilisers, including mineral fertilisers and organic amendments (e.g., manure and compost), is a conventional approach to quickly addressing P deficiency and maintaining P levels in natural soil pools[10]. P fertilisation contributes to increased crop yields by stimulating root growth, increasing drought tolerance, improving disease resistance, and facilitating nutrient and water absorption in the seedling[5,11]. In Europe, Ireland ranks among the highest in fertiliser use intensity (i.e., consumption per hectare)[12]. In 2015, Ireland was the ninth-highest fertiliser consumer among EU member states[13], with an average application rate of 146.4 kg of artificial fertilisers per hectare of agricultural land, including 4.2 kg of P[14].

The development of synthetic P fertilisers has contributed significantly to increased crop production per unit area; however, the associated environmental impacts have often gone unmeasured. For example, phosphate rock (PR), the primary raw material used in the production of nearly all commercial phosphate fertilisers, is a non-renewable resource. In Europe, PR is additionally considered as a strategic resource. Due to limited domestic PR reserves, Europe is predominantly a P importer and is therefore vulnerable to geopolitical tensions[15–18]. Since 2022, Irish farmers have faced a significant cost increases, experiencing a doubling to tripling of the cost of phosphate fertilisers[19]. Not only are PR reserves dwindling, but the quality of PR used globally is also steadily declining as high-grade reserves become increasingly scarce. The use of low-grade PR increases the contamination of phosphate fertilisers with heavy metals (e.g., cadmium, lead); these toxic pollutants are subsequently released into water and soils[20,21].

On the other hand, P surpluses have been evidenced by studies comparing fertiliser inputs to P offtakes[22–26]. A unique characteristic of soil P is its low availability due to slow diffusion and strong fixation in mineral soils[27]. In many agricultural systems, the recovery of applied P by crops within a single growing season is very low, with more than 80% of applied P becoming immobile and unavailable for plant uptake due to adsorption, precipitation, or conversion into organic forms[26–29]. In a global terrestrial P fertilisation study, Luo et al. (2024)[26] reported that, on average, only 12.6% of applied P is taken up by plants, while 67.2% of inorganic P fertilisers remain stored in soils. Legacy-P, defined as accumulated P from long-term fertiliser applications, is sorbed as labile-P or stored in secondary P minerals[30,31]. Agricultural soils with high legacy-P/labile-P exhibit reduced buffering capacity, increasing the susceptibility of newly applied P to environmental losses[32]. Numerous studies have shown that agricultural P losses to water have intensified significantly in recent years[33–36]. Globally, farmers apply approximately 25 million tonnes of P annually, of which about 14 million tonnes remain unused by crops, leading to a build-up of P in excess of crop requirements and increasing the risk of soil P loss, e.g., via surface runoff during periods of heavy rainfall[37,38]. One major

consequence is the significant contribution of P to eutrophication (e.g., the 2007 Algal Bloom Crisis), which is considered an endemic global problem[26,39–42].

Balancing food security and environmental sustainability while minimising the negative impacts of P fertilisation remains a significant challenge for sustainable agricultural management.

1.1.2 Managing soil P: Legislation and agronomic use of soil testing

Policies and advice are used to regulate fertiliser application, encouraging farmers to apply P fertilisers more responsibly with the objective of reducing losses to water. For example, the 'Zero Pollution Action Plan', launched within the European Green Deal that aims to reduce nutrient losses by 50% by 2030, emphasises the need for improved management of N and P throughout their life cycle[43]. At the European level, although there is currently no comprehensive legal instrument specifically addressing soil protection[44], the Nitrates Directive (Directive 91/676/EEC, OJ 1991 L 375, p. 1 ff.), the Water Framework Directive (WFD) (Directive 2000/60/EC, OJ 2000 L 327, p. 1 ff.), and the Habitats Directive (Directive 92/43/EEC, OJ 1992, L 206, p. 7 ff.) seek to protect water quality through reducing diffuse P loss from agricultural areas. In 2023, the European Commission proposed a Soil Monitoring Law, which aims to monitor and improve soil health in all EU Member States. As of 10th April 2025, a provisional political agreement has been reached between the European Parliament and the Council[45].

Ireland currently regulates N and P use through the Nitrates Action Programme (NAP), as set out in Statutory Instrument S.I. No. 113 of 2022[46,47]. The agronomic Soil Index System for P measures soil test P (STP) based on Morgan's extractable P (sodium acetate and acetic acid reagent with a pH of 4.8), providing a static assessment of plant-available soil P relative to crop/grass demand[48]. Tested Morgan's P values are grouped into bands, with concentrations of 0 - 3 mg·L^{-1} and 3.1 – 5.0 mg·L^{-1} corresponding to grassland soils at Index 1 and Index 2, respectively. Croplands (e.g., tillage and vegetable crops) require higher P levels for optimum production, with Index 2 corresponding to STP values of 3.1 – 6.0 mg·L^{-1}. Soils at Index 1 or Index 2 are considered to have low levels of plant-available P and are expected to show a definite or likely response to P fertilisation. Index 3 is the target level for

the optimum P range for crop (6.1 - 10.0 mg·L^{-1}) and grass growth (5.1 - 8.0 mg·L^{-1}). When STP exceeds 10 mg·L^{-1} for croplands or 8.0 mg·L^{-1} for grasslands, soils are classified as Index 4, indicating sufficient/excess plant-available P, where additional P application is not allowed[49].

Morgan's solution has been effective for assessing P status in neutral to acidic Irish soils (e.g., by maintaining a stable reaction below pH 7.5); however, its inability to account for system dynamics, such as plant-driven P mobilisation or long-term P accumulation, limits its agronomic and ecological relevance. For example, in Ireland, a national soil survey showed that 55% of dairy farms, 61% of drystock farms, and 57% of tillage farms had soils classified as Index 1 and Index 2 in 2021. A decline in soil P levels was also observed in both grassland and tillage farms[50]. Although soil testing indicates that P fertilisation is needed to rebuild soil P levels and prevent the reversal of previous fertility gains, gaps remain in understanding the accumulation of legacy-P and its potential role in increasing the risk of P loss to water systems in the future.

1.1.3 Soil P cycling: Sorption and desorption processes

Soil P can be lost to water from both point sources (e.g. sewage treatment works, industrial discharges, septic tank systems or farmyard runoff entering watercourses via pipes or drains) and diffuse sources (e.g. surface runoff, leaching, or subsurface drainage from agricultural lands)[38,51]. With the identification and management of point-source water pollution, research focus has shifted to diffuse pollution from agricultural lands, which arises from multiple interacting factors[52]. Climate change[53,54], land-use change[55,56], vegetation coverage[56], increasing population density[57] and dietary changes[58] *etc*. all contribute to soil P losses. For example, increased rainfall intensity and severe storms caused by climate change have led to a greater topsoil loss from agricultural land[59].

The movement, transformation, and availability of P within the soil-plant-environment continuum are governed by various processes within the soil P cycle (Figure 1-1), including plant uptake, desorption–sorption, mineralisation–immobilisation, dissolution–precipitation, soil erosion and runoff to surrounding waterbodies[60–62]. For example, the plant-available P originates from (i) natural inorganic

P (Pi) released through primary mineral weathering; (ii) desorption of P from clay, Fe/Al oxides; (iii) dissolution of secondary minerals that bind P; (iv) mineralisation of organic P (Po) compounds derived from plant residues, microbial biomass; inputs from (v) mineral P fertilisers; and (vi) organic fertilisers. It is removed from the agricultural system through harvesting or through diffuse losses to surrounding water bodies, e.g., soil erosion and surface runoff. The mobilisation of Po is dependent on variables such as plant root systems, the activity of mycorrhizal fungi, bacterial, redox potential, soil type and pH[63]. Plant roots can exude phosphatases that mineralise Po, and organic acids (e.g., malic and citric acids) that solubilise mineral-bound P. Symbiotic arbuscular mycorrhizal fungi and ectomycorrhizal fungi also contribute to mobilising mineral-bound P by producing chelating compounds[64,65].

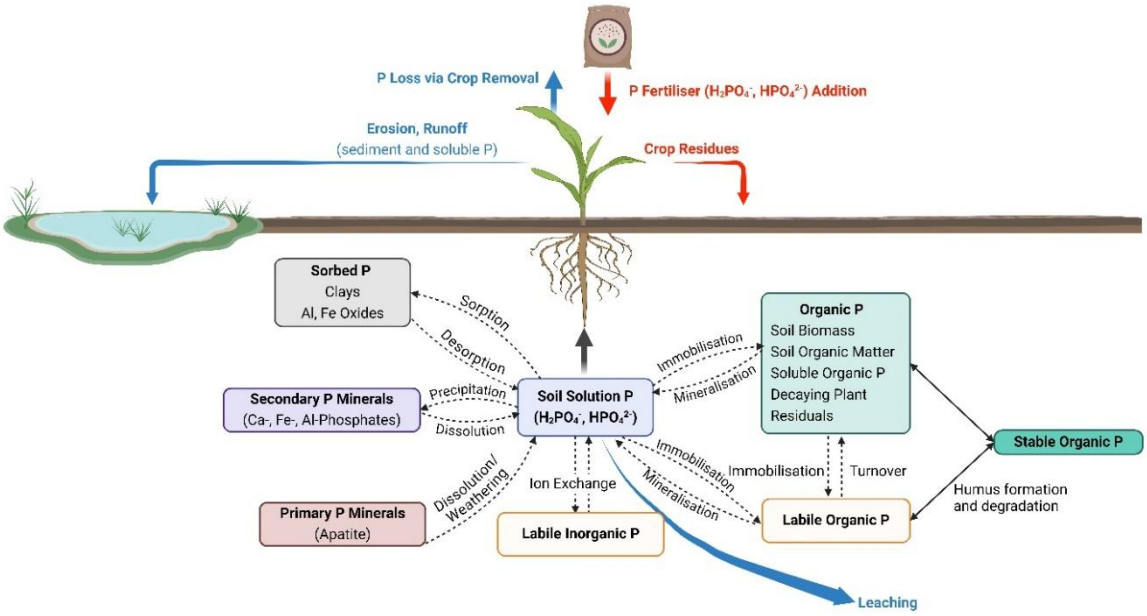


Figure 1-1. Schematic of the soil phosphorus cycle. Redrawn from Sims and Sharpley (2005)[66].

In this study we focused on Pi dynamics, e.g., sorption – desorption mechanisms.

Sorption broadly describes the transfer of P from the soil solution to the solid phase soil sorbents, and the inverse process is desorption. The main sorption mechanisms include inner-/outer-sphere adsorption, absorption, and precipitation on mineral surfaces[67,68]. Soil sorbents include clays, metal-oxides/hydroxides (e.g., Fe and Al) and soil organic matter (SOM)[68]. For example, in acidic soils rich in Al or Fe oxides (e.g., gibbsite and goethite), outer-sphere (anion exchange) and inner-sphere (ligand exchange) sorption mechanisms facilitate the surface adsorption of phosphate. In alkaline calcareous soils, phosphate is often adsorbed and precipitated with calcium (Ca^{2+}) and magnesium (Mg^{2+})[27,68–70]. Clay soils, with their small particle size and high specific surface area, often contain high levels of Fe/Al oxide minerals and typically exhibit high P sorption capacity[71,72]. However, the effects of SOM on P sorption in mineral soils are more complex. SOM can indirectly inhibit the crystallisation of Fe/Al oxides, potentially increasing P adsorption; while also directly competing with phosphate for adsorption sites, thereby reducing P adsorption[27]. For example, in highly weathered soils (e.g., lateritic soils), SOM has been found to be negatively correlated with P adsorption capacity[73,74].

Soil P sorption capacity is closely related to soil properties (e.g., soil mineralogy, SOM)[75–79]. In particular, soil pH is one of the most important factors influencing P dynamics[28,69]. This is because soil pH plays a key role in determining surface charge. Increasing soil pH (e.g., from 4 to 7) leads to a decrease in surface positive charge, resulting in reduced P outer-sphere adsorption. Ligand exchange of phosphate onto minerals also tends to decrease with increasing pH (e.g., pH < 7.5), while extremely acidic pH values (e.g., pH < 4.5) can also potentially limit surface adsorption (adsorption envelope phenomenon). In contrast, in calcareous soils, precipitation of Ca-phosphate minerals is favoured by increasing pH (e.g., pH from 6.5 to 8.5)[69].

In practice, understanding soil P sorption mechanisms has informed strategies for soil management and nutrient dynamics. For example, in highly P-fixing and poorly P-responsive soils, organic amendments can reduce P sorption by temporarily raising soil pH during decomposition (the rapid proton exchange between the soil and the organic amendments[80–82]) and by complexing Al^{3+} and Fe^{3+} ions with organic compounds, thereby decreasing the activity of P-fixing cations. In Ireland, liming

plays an important role in correcting soil acidity in agricultural soils. Panagos et al. (2022)[83] analysed the Land Use/Land Cover Area Frame Survey (LUCAS) topsoil dataset (0–20 cm), where plant-available P was determined using the Olsen method. The study reported that in Ireland, the ratio of total P to plant-available P in agricultural topsoil was 18.3, compared to a European average of 17, suggesting lower P bioavailability and a greater proportion of total P retained as labile-P/legacy-P in Irish soils. Raising soil pH through liming improves the efficiency of applied mineral/organic P and enhancing the availability of labile-P/legacy-P by breaking ligand exchange bonds[84]. Maximum P availability is typically achieved at a soil pH of around 6.5[4,69]. Soil test results reported by the *Irish Farmers Journal* from the Footprint Farms Programme show that 27% of samples fall within the pH range of 6.0 - 6.4, 38% between 6.5 - 6.9, and only 3% have a pH below 6.0[85]. These results highlight the impact of liming efforts to date and emphasise the ongoing importance of maintaining optimal pH levels to support effective P use and availability in Ireland. These results reinforce the effectiveness of liming highlight the importance of continued liming to optimise P use efficiency.

1.1.4. Soil P sorption indicators

Soil P sorption characteristics, such as the 'sorption maximum' and the 'bonding strength between P and soil', are traditionally derived from isothermal batch experiments. Standard isotherm equations used to fit P sorption data include the Temkin, Freundlich, Langmuir and Elovich equations[86]. Of these, the Freundlich models (Equation 1-1)[74,87] and Langmuir models (Equation 1-2)[18,72,78,88–90] are the most commonly used. The Langmuir model assumes monolayer adsorption onto a homogeneous surface with a finite number of identical binding sites and uniform sorption energies. In contrast, the Freundlich model describes adsorption on heterogeneous surfaces with variable sorption energies and no defined saturation limit, providing an empirical fit for complex soil systems[91].

$$S = k_f C^n,$$

or written in logarithmic form as (Equation 1-1)

$$\log S = \log k_f + n \log C,$$

where S (mg·kg^{-1}) is the P sorbed on solid phase; k_f (L·kg^{-1}) is the adsorption constant; C (mg·L^{-1}) is the P concentration in soil water; and n is the affinity constant (n<1), which is equivalent to a binding energy (L·mg^{-1}).

$$S = \frac{k \cdot Smax \cdot C}{1 + k \cdot C}, \quad \text{(Equation 1-2)}$$

where S (mg·kg^{-1}) is the total amount of P sorbed; C (mg·L^{-1}) is the concentration of P in solution after equilibration; *Smax* is the P sorption maximum (mg·kg^{-1}); and k is a binding affinity constant (L·mg^{-1}). Phosphate sorption maximum (*Smax*) has been widely used to estimate soil capacity to adsorb P.

For a range of Irish grassland soils derived from different parent materials, Daly et al. (2015)[18] and Dunne et al. (2021)[90] have demonstrated that the Langmuir model provided a good fit for describing soil P sorption dynamics. The maximum P sorption capacity of a soil (Smax) is an important parameter influencing soil P sorption behaviour. In P-deficient conditions, soils with high Smax replenish plant-available P pools slowly during the 'build-up' phase[62]. As soil P concentration approaches its Smax, less P can be sorbed by the soil, resulting in higher soluble P concentrations and greater risk of P loss via runoff or leaching[92]. However, the sorption parameters derived from these models are empirical in nature[66]. Soil P sorption maximum varies with pH and ionic strength, yet the Langmuir model does not account for these variations mechanistically. Wang et al. (2016)[93] noted that the value of the maximum sorption capacity determined using the Langmuir isotherm equation (Smax) does not represent the P sorption maximum in reality.

The maximum buffer capacity (MBC, mg·kg^{-1}) (Equation 1-3), derived from the Langmuir sorption parameters, is used to determine the soil P sorption capacity by measuring its ability to resist changes in P concentration in solution. A higher MBC reflects a stronger resistance to changes in solution P

concentration (or buffering capacity) and a greater P sorption capacity[94]. The MBC is a variable, soil-specific property that is influenced by factors such as soil types (e.g., clay soils)[95], land use (e.g., wetlands restoration)[96], management practices (e.g., long-term fertilisation)[97], and soil amendments (e.g., biochar addition)[98].

$$\mathrm{MBC} = \mathrm{Smax} \times \mathrm{k}, \qquad \text{(Equation 1-3)}$$

where *Smax* (mg·kg^{-1}) and *k* (L·mg^{-1}) are the Langmuir sorption parameters.

An advantage of the batch technique is the ease of sample handling (straightforward soil-solution separation). It also allows precise control of soil-to-solution ratios and equilibration, enabling accurate, reproducible measurement of P sorption. However, it is time-consuming and costly for routine use. To save time and lower analytical costs, single-point sorption values have been investigated as indicators of sorption parameters, simplifying isotherm modelling by reducing data collection from multiple concentrations (e.g., six points) to a single measurement.

The P Sorption Index (PSI) (Equation 1-4), which evaluates soil sorption capacity using a single added P concentration (e.g., 150 mg P per 100 g soil), was first introduced by Bache and Williams (1971)[99] as an alternative to estimating Smax. The PSI is often used to estimate soil P sorption capacity in the calculation of the degree of soil P saturation (DPS)[93,100]. The DPS, defined as the ratio of sorbed P to the soil P sorption capacity (PSC) of the soil ($\mathrm{DPS} = \frac{\text{Sorbed P (mg·kg}^{-1})}{\text{PSC (mg·kg}^{-1})} \times 100\%$)[101], indicates the potential ability of a soil to desorb soil P and is used as a risk indicator of P losses from agricultural land to surface/ground water [102]; a higher DPS indicates a greater loss potential.

$$\mathrm{PSI} = \frac{\mathrm{X}}{\log \mathrm{C}}, \qquad \text{(Equation 1-4)}$$

where *C* (mg·L^{-1}) is the final P concentration at equilibrium, *X* is P sorbed (mg·kg^{-1}). PSI is expressed in L·mg^{-1}.

According to sorption mechanisms (Section 1.1.3), PSC also estimated empirically using extractable metal concentrations[102]. For example, in non-calcareous sandy soils, PSC is strongly correlated with oxalate-extractable iron (Fe_{ox}) and aluminium (Al_{ox}), and exhibits a linear relationship with the sum of

Al_{ox} and Fe_{ox} (Equation 1-5)[103,104]. Oxalate extraction is not routinely used in most labs. Mehlich-3-based methods have also been proposed as alternatives for estimating PSC (Equation 1-6)[102,105,106].

$$\mathrm{PSC_{OX}} = \alpha(\mathrm{Fe_{ox}} + \mathrm{Al_{ox}}), \qquad \text{(Equation 1-5)}$$

where α is an empirical parameter calculated as the ratio of PSC to the corresponding to the extractable metal concentrations (mmol·kg^{-1}); PSC is expressed in mmol·kg^{-1}.

$$\mathrm{PSC_{Mehlich-3}} = \alpha(\mathrm{Fe_{Mehlich-3}} + \mathrm{Al_{Mehlich-3}}), \qquad \text{(Equation 1-6)}$$

For example, van der Zee et al. (1988)[107] estimated $\mathrm{PSC} \approx 0.5 \times (\mathrm{Fe_{ox}} + \mathrm{Al_{ox}})$ and developed the DPS ($\mathrm{DPS} = \frac{\text{Oxalate-extractable P}}{\text{Oxalate-extractable Al+Fe}} \times 100\%$) to estimate the potential ability of a soil to desorb soil P. Maguire et al. (2001)[108] evaluated DPS values of acidic soils from Northern Ireland and derived $\mathrm{PSC} = 0.57 \cdot \mathrm{Al_{ox}} + 0.28 \cdot \mathrm{Fe_{ox}} + 1.28 \cdot \mathrm{LOI} + 24.2$ (R^2=0.95). The regression coefficient for Al_{ox} (α_1 = 0.57) was approximately twice that for Fe_{ox} (α_2 = 0.28), suggesting that Al_{ox} played a more dominant role than Fe_{ox} in determining soil P sorption capacity. Additionally, the presence of a positive intercept (β = 24.2) in the regression equation indicated that a significant portion of P sorption may be attributed to soil properties beyond Al_{ox} and Fe_{ox}, such as SOM, which was found to be weakly but positively correlated with P sorption capacity.

In Australia, the P buffering capacity (PBC) (Equation 1-7), which measures the soil's ability to resist changes in solution P concentration when P is added or removed[109,110], and the P buffer index (PBI) (Equation 1-8), which quantifies the soil's capacity to immediately remove P from solution[110,111], are commonly used to guide optimal fertilisation on highly weathered soils.

$$\mathrm{PBC} = \frac{\Delta \mathrm{X}}{\Delta \mathrm{C}}, \qquad \text{(Equation 1-7)}$$

where ΔX is the change in P sorbed (mg·kg^{-1}); ΔC is the change in P concentration (mg·L^{-1}).

$$\mathrm{PBI} = \frac{\text{P sorbed}}{\text{Equilibrium P concentration}}, \qquad \text{(Equation 1-8)}$$

P sorbed (mg·kg^{-1}); Equilibrium P concentration (mg·L^{-1}).

1.1.5 Integrating soil P dynamics with test-based recommendations

Low P use efficiency results in unnecessary economic costs and increases the risk of P loss due to legacy-P accumulation, both of which are closely linked to soil P availability and its behaviour under varying soil properties. Growing concerns over the long-term sustainability of P management in agroecosystems has led to increased interest in monitoring soil P dynamics and integrating P sorption indicators with soil test results to improve fertiliser recommendations. For example, Amrani et al. (1999)[95] used MBC to assess P buffering in calcareous soils and recommended incorporating it into P requirement models to optimise yields under different soil types.

In Australia, Colwell-P is widely used to relate soil P levels to crop yield responses. Across different field soils, a strong positive relationship has been observed between P fertiliser requirements and PBC values[17]. Burkitt et al. (2002)[109] developed the PBI, a simple measurement of PBC, to describe a soil's P fixation capacity, with a higher PBI value indicating greater P sorption and a higher Colwell-P level being required to meet crop P requirements. They proposed coupling PBI with the Colwell-P test to optimise fertiliser use and support sustainable practices. Building on this, Moody (2007)[112] quantitatively linked PBI to crop yield response by adjusting critical Colwell-P values (Colwell-P at 90% maximum yield) for five crops (annual medics, soybean, potato, wheat, and temperate pasture) to enhance the precision and sustainability of P management by accounting for variation in soil P sorbing characteristics.

In Ireland, Dunne et al. (2021)[90] assessed soil P sorption capacity employing Langmuir sorption isotherms. Visualisation of the P sorption isotherms showed that they were predominantly *C*-shaped (rapid P sorption with no clear plateau) and *L-non-strict* (gradually decreasing sorption rates) and corresponded to high Mehlich-3-aluminium (M3-Al) levels (702.5 - 2,460.8 $mg \cdot kg^{-1}$) and low M3-Al levels (2.5 - 698.5 $mg \cdot kg^{-1}$), respectively. Dunne et al. (2021)[62] also developed a conceptual framework for P management by defining sorption classes and integrating them with modified P indices. Hall et al. (2023)[113] assessed P sorption characteristics by available Al concentrations, determined by Mehlich-

3 extraction (M3-Al). Kriged variogram outputs showed a significant negative correlation with bioavailable P indices. Areas with high M3-Al concentrations (≥702.5 $mg \cdot kg^{-1}$) predominantly exhibited low plant-available P values (<5 $mg \cdot L^{-1}$). Conversely, regions with low Al concentrations (≤ 697 $mg \cdot kg^{-1}$) showed moderate to high plant-available P concentrations, suggesting a higher likelihood of P remaining in the solution phase and posing a risk to local watercourses. Their mapping of soil P indices alongside M3-Al concentrations yielded a visual depiction of soil P behaviour, offering valuable insights that facilitate enhanced decision-making processes for P management and water quality preservation. However, unlike PBI[112], the important sorption parameter (Langmuir) Smax does not have a defined threshold to indicate a change-point in sorption behaviour.

Additionally, a significant limitation of routine laboratory isotherm testing is that it conflicts with the requirement for soil testing programmes to be timely and cost-effective. Exploring faster and more cost-effective methods for routinely assessing soil P sorption behaviour is essential for supporting farmers in adopting sustainable practices. Some efficient alternatives to wet chemical extractions for assessing soil P sorption dynamics include anion exchange resin membranes[114], diffusive gradients in thin-films[31], X-ray fluorescence spectroscopy[115]. In particular, infrared (IR) spectroscopy has emerged as a cost-effective, non-destructive method for the rapid soil characterisation, enabling the simultaneous prediction of multiple physical, chemical, and biological properties from a single spectrum[116–118].

1.2 Soil Infrared Spectroscopy

1.2.1 Theory of IR spectroscopy and applications in soil analysis

Spectroscopy, known as 'the light of knowledge', studies the spectra resulting from the interaction of electromagnetic radiation (EMR) with matter, which exhibits both wave and particle properties[119,120]. EMR, as a wave, is characterised by its propagation speed (c), wavelength (λ), and frequency (ν). Its particle nature, associated with photons, is described by the energy (E). Table 1-1 summarises the relationships among various parameters associated with EMR. These quantities (c, λ, ν, and E) provide a scale for the electromagnetic spectrum (EMS), which is divided into regions based on energy transitions in matter (Figure 1-2).

IR spectroscopy measures the absorption of EMR in the wavelength range of 800 to 100,000 nm. This range is typically divided into near-infrared (NIR, 800–2500 nm), mid-infrared (MIR, 4000–400 cm^{-1}), and far-infrared (FIR, 25 µm – 1 mm). Of these, NIR, along with visible to near-infrared (Vis–NIR, 350–2500 nm), and MIR are the most widely used in soil spectroscopy, while FIR has seen limited application due to its low-energy radiation and strong water vapour interference[121,122]. As a type of vibrational spectroscopy, IR spectroscopy is based on molecular vibrational changes. When IR light irradiates a substance, specific wavelengths are selectively absorbed, exciting vibrations of covalent bonds within the molecule if (i) their natural vibrational frequency matches the incident radiation and (ii) the absorption induces a change in the molecular dipole moment[119,123]. The resulting spectrum displaying characteristic peaks (i.e., position, shape and intensity) reflects the presence and concentration (Beer–Lambert Law) of molecular functional groups, thereby supporting compound identification[120,124].

Table 1-1. Parameters associated with electromagnetic radiation.

Quantity	Relationships	Explanations
Energy (E/J)	$E = h\nu$ • h (Planck's constant) = 6.6×10^{-34} J·s^{-1} • ν is the frequency	The energy of EMR is directly proportional to its frequency (or wavelength as $E = h\frac{c}{\lambda}$). The phenomenon is known as Planck's law.
Frequency (ν/Hz)	$\nu = \frac{c}{\lambda}$ • c is the velocity of radiation • λ is the wavelength	Frequency of a beam of radiation is determined by the source and invariant. It is defined as the number of waves that can pass through a point in 1 s. It is expressed in cycles per second or in hertz (1 Hz = 1 cycle·s^{-1}).
Velocity (c/m·s^{-1})	$c = \nu\lambda = \frac{\nu}{\bar{\nu}}$ • c (the velocity of radiation) = 3×10^8 m·s^{-1} through a vacuum/air • $\bar{\nu}$ is the wavenumber	Velocity of radiation depends upon the composition of the medium through which it passes. All types of EMR travel with the same velocity and require no medium for their propagation.
Wavelength (λ/nm)	$\lambda = \frac{c}{\nu}$	The distance between two adjacent crests or troughs in a particular wave. It can be expressed in angstrom unit (Å) or nanometre (1 Å = 0.1 nm).
Wavenumber ($\bar{\nu}$/cm^{-1})	$\bar{\nu} = \frac{1}{\lambda}$	It is the reciprocal of wavelength λ and is expressed in per centimetre. It is defined as the total number of waves that can pass through a space in 1 cm.

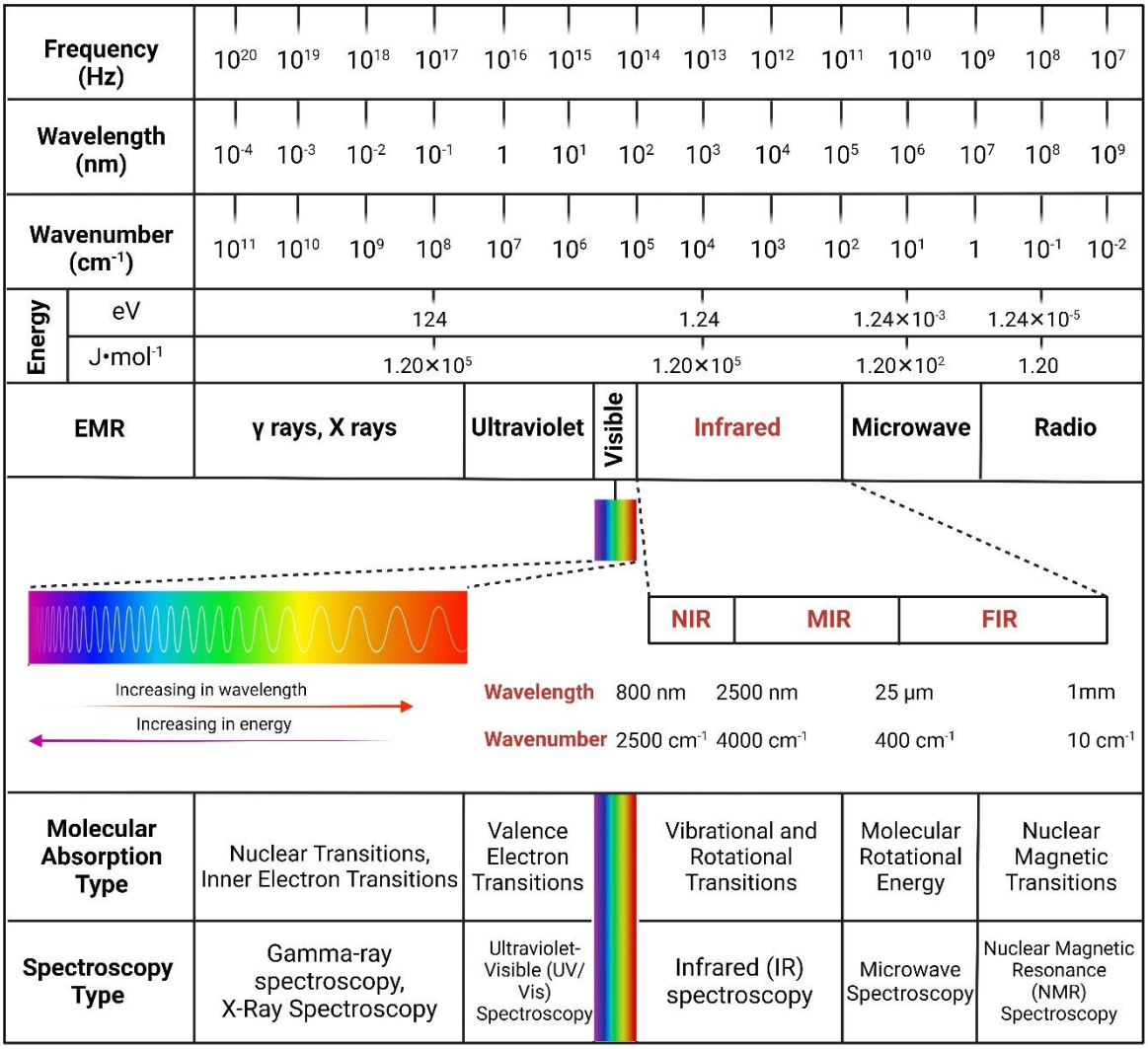


Figure 1-2. Electromagnetic spectrum[125–127].

Early soil spectroscopy was developed using NIR spectroscopy with monochromator-based scanners[128]. With the development of Fourier-transform infrared (FTIR) spectrometers (Michelson interferometers), IR spectroscopy using transmission and reflectance measurement modalities has become increasingly popular for soil characterisation (detailed in Section 1.2.5). Transmission IR spectroscopy using pressed potassium bromide matrix tablets was the earliest method for collecting FTIR spectra of soil minerals and organic components[129,130]. However, limitations include time-consuming sample preparation, adjustment of path length and concentration for suitable intensity, and limited applicability in field soil studies have make it less widely used for quantitative soil analysis today[129–134]. Nowadays, soil spectroscopy focuses on measuring surface-reflected electromagnetic energy, where internal and external reflectance methods are used to capture spectral information that serves as a proxy for various soil properties. Attenuated total reflectance (ATR) spectroscopy utilising the phenomenon of total internal reflection resulting in an evanescent wave[134]. Extensive studies using ATR for investigating SOM and ions (e.g. nitrate, ammonium *etc*.) have been summarised in Tinti et al. (2015)[133].

Diffuse reflectance infrared Fourier-transform spectroscopy (DRIFTS), particularly mid-infrared DRIFTS (mid-DRIFTS), is commonly used in the soil spectroscopy studies[135–142]. It is generally considered advantageous in analysing many soil properties due to its ability to allow electromagnetic energy to penetrate and interact effectively with the soil matrix, capturing detailed information about soil constituents[133,143]. However, the signal intensity and spectral resolution in DRIFTS are influenced by sample surface geometry (e.g., particle size and porosity) and topography (i.e., surface shape), as described by the Kubelka-Munk theory (Equation 1-9)[130,133]. For example, Leue et al. (2010)[144] found that both mid-DRIFT signal intensity and spectral resolution decrease with increasing quartz particle size, whereas DRIFT spectra of particles <70 µm can characterise quartz molecular structure. Soriano-Disla et al. (2014)[116] reviewed studies and recommended that ball-milling dried soil samples to <100 µm is necessary to reduce specular reflectance and maximise homogeneity for MIR spectroscopy analysis.

$$f\left(R_\infty\right) = \frac{(1-R_\infty)^2}{2R_\infty} = \frac{k}{s}, \qquad \text{(Equation 1-9)}$$

where R_∞ is the absolute reflectance of the layer; k is the molar absorption coefficient; and s is a scattering coefficient, depending on particle size and distribution.

Although ball-milling/grinding is important for optimising the reproducibility and accuracy of mid-DRIFTS predictions, recent studies have investigated omitting this step to reduce time and cost, with most research focusing on carbon stock estimation (e.g., total carbon [TC] and soil organic carbon [SOC]) (Table 1-2). For example, Bachion de Santana and Daly (2022)[145] compared MIR-based predictions of various soil properties (e.g., TC, SOC, clay, bulk density [BD], *etc*.) using ball-milled (<0.100 mm) and sieved-only (<2 mm) samples and found no significant differences in accuracy. The comparison of sample preparation conditions by Breure et al. (2022)[146] similarly supported that MIR spectroscopy can reliably predict SOM, clay, and pH without the need for extensive preparation, demonstrating its potential to reduce both time and cost.

Dunne et al. (2020)[147] applied MIR spectroscopy to predict soil P sorption indicators and used ball-milled samples. However, little is known about the effects of grinding on MIR spectra for soil P sorption prediction. Homogenising soil samples helps prevent specular reflection from large particles. However, this sample preparation step limits the advancement of portable MIR devices intended for on-site soil analysis.

Table 1-2. Comparison of soil properties predicted using laboratory-based mid-DRIFTS spectroscopy with and without the grinding step in sample preparation.

Soil parameter	Sample preparation and soil particle size	Prediction accuracy	Ref.
TC (%)	Ball-milled (< 0.100 mm)	R^2_{Val} = 0.92	145
TC (g·kg^{-1})	Ball-milled (< 0.200 mm)	R^2_{CV} = 0.93	148
TC (%)	Ball-milled (< 0.100 mm)	R^2_{CV} = 0.97	149
TC (%)	Sieved (<2 mm)	R_{Val}^2 = 0.93	145
TC (g·kg^{-1})	Sieved (<2 mm)	R_{CV}^2 = 0.98	148
SOC (g·kg^{-1})	Ball-milled (< 0.200 mm)	R^2_{Val} = 0.77	150
SOC (%)	Ball-milled (< 0.100 mm)	R^2_{CV} = 0.85	151
SOC (%)	Ball-milled (< 0.100 mm)	R^2_{Val} = 0.92	145
SOC (%)	Sieved (<2 mm)	R^2_{Val} = 0.93	152
SOC (%)	Sieved (<2 mm)	R^2_{Val} = 0.93	145
SOC (%)	Sieved (<2 mm)	R^2_{Val} = 0.94	153
Total Nitrogen (TN, %)	Ball-milled (< 180 µm)	R^2_{Val} = 0.85	118
N (%)	Ball-milled (< 0.100 mm)	R^2_{Val} = 0.94	145
N (%)	Dried, sieved (<2 mm)	R^2_{Val} = 0.93	145
Clay (%)	Ball-milled (< 0.100 mm)	R^2_{Val} = 0.87	145
Clay (%)	Dried, sieved (<2 mm)	R^2_{Val} = 0.88	145
Clay (g·kg^{-1})	Dried, sieved (<2 mm)	R^2_{Val} = 0.91	152
Sand (%)	Ball-milled (< 0.100 mm)	R^2_{Val} = 0.82	145
Sand (g·kg^{-1})	Ball-milled (< 0.200 mm)	R^2_{Val} = 0.90	150
Sand (%)	Dried, sieved (<2 mm)	R^2_{Val} = 0.80	145
Sand (g·kg^{-1})	Dried, sieved (<2 mm)	R^2_{Val} = 0.87	152
pH (H_2O)	Ball-milled (< 0.200 mm)	R^2_{Val} = 0.54	150
pH (H_2O)	Ball-milled (< 0.100 mm)	R^2_{Val} = 0.87	145
pH ($CaCl_2$)	Ball-milled (< 0.100 mm)	R^2_{Val} = 0.88	145
pH (H_2O)	Ball-milled (< 0.100 mm)	R^2_{Val} = 0.97	146
pH (H_2O)	Dried, sieved (<2 mm)	R^2_{Val} = 0.81	152
pH (H_2O)	Dried, sieved (<2 mm)	R^2_{Val} = 0.86	145
pH ($CaCl_2$)	Dried, sieved (<2 mm)	R^2 = 0.89	145
BD (g·cm^3)	Ball-milled (< 0.100 mm)	R^2_{Val} = 0.78	145
BD (g·cm^3)	Ball-milled (< 0.100 mm)	R^2_{Val} = 0.81	154
BD (g·cm^3)	Dried, sieved (<2 mm)	R^2_{Val} = 0.79	145

1.2.2 IR spectroscopy for soil characterisation and P dynamics prediction

1.2.2.1 NIR

NIR spectroscopy (800 nm – 2500 nm) is based on the overtones of vibrational modes involving functional groups such as O–H, C–H, N–H, sulphate ion (SO_4^{2-}), and carbonate ion (CO_3^{2-}), as well as on combination bands associated with molecules like H_2O and CO_2. These vibrational transitions correspond to specific wavelength regions, including the first overtone (1400–1900 nm), second overtone (1000–1400 nm), third overtone (800–1000 nm), and combination bands (1900–2500 nm). Absorptions in the visible region (350–800 nm) are primarily associated with iron-containing minerals such as hematite and goethite[124]. When combined with the near-infrared region, visible and near-infrared (Vis–NIR) spectroscopy (350–2500 nm) has been widely applied in fields for soil analysis[116,155].

In general, absorption features in the (Vis-)NIR region are characterised by broad, overlapping, and relatively weak signals, making soil (Vis-)NIR spectra relatively more difficult to interpret compared to MIR spectra[156]. Nevertheless, (Vis-)NIR contains useful information for soil analysis, particularly due to its instrument portability, ease of use in remote sensing, and minimal sample preparation requirements[157]. Table 1-3 summarises some soil properties that are well predicted by (Vis-)NIR spectroscopy, such as soil moisture content (R^2 = 0.86 - 0.97), total nitrogen [TN] (R^2 = 0.80 - 0.97) and SOC (R^2 = 0.60 - 0.98), where R^2 represents the coefficient of determination, indicating the proportion of variation in the reference data explained by the spectral model. In addition, (Vis-)NIR spectroscopy has shown good predictability for secondary soil properties, such as soil pH[158], mineralogical properties (e.g., clay mineral composition, clay content, and mineral weathering/alteration degree)[159,160], soil texture and structure[157,160], electrical conductivity (EC)[161], cation-exchange capacity (CEC)[161,162], lime requirement (LR)[117], BD[163], *etc*.

However, soil P has no direct vibrational response in (Vis-)NIR. Using NIR spectroscopy, a very weak correlation between predicted and actual soil P was reported by Krischenko et al. (1991)[164] (R^2 = 0.42), Chang et al. (2001)[162] (R^2 = 0.40), Thomasson et al. (2001)[165] (R^2 = 0.49), and He et al. (2005)[166] (R^2 =

0.46). Nevertheless, the literature also reports some successful cases, especially in the Vis-NIR region. For example, Daniel et al. (2003)[167] used an artificial neural network to predict available-P from reddish-brown and black-clayey soils in Thailand, achieving R^2_{Val} = 0.86 from the laboratory-based Vis-NIR spectra. In studies with limited soil texture variation (e.g., sandy soils), Bogrekci et al. (2004)[168] and Bogrekci and Lee (2005)[169] reported strong correlations between Vis-NIR spectral data and soil P concentrations (Mehlich-1 P), with R^2_{Val} values of 0.87 and 0.92, respectively. The significant influence of soil texture on the accuracy of P prediction using Vis-NIR was also evidenced by Mouazen et al. (2010)[170]. Similarly, Reda et al. (2020)[171] used NIR (1100–2500 nm) to predict available P (P-Olson), achieving improved accuracy (R^2_{Val} = 0.77, and RPD_{Val} = 1.90) by applying advanced support vector regression combined with a variable selection algorithm of genetic algorithm–partial least squares. When samples were grouped based on texture (e.g., clayey and loamy soils), prediction accuracy further improved, reaching R^2_{Val} = 0.83 and RPD_{Val} = 3.11 for P-Olsen prediction.

Additionally, various forms of P in agricultural soils can be indirectly predicted if they are collinear with NIR-active properties such as soil C and N. For example, Serrano et al. (2021)[172] predicted P_2O_5 levels correlated with P fraction bound to SOM, achieving R^2_{Val} =0.76. Total P (TP) was predicted with R^2_{CV} = 0.76 by Abdi et al. (2012)[173] and R^2_{CV} = 0.88 by Reda et al. (2020)[171]. Studies also demonstrated moderate success (a median R^2 value of 0.65) in predicting P sorption using (Vis-)NIR spectroscopy in laboratory settings[116,174]. Recena et al. (2019)[175] used Vis-NIR (400-2500 nm) and achieved good estimates of PBC (Section 1.1.4, describing the capability of the soil to replenish P in solution) (R^2_{CV} = 0.78). However, their predictions of isotherm-derived parameters showed less optimising accuracy, with R^2_{CV} = 0.66 for Freundlich P sorption capacity and R^2_{CV} = 0.42 for Langmuir maximum P sorption capacity (Smax). Using the same chemometrics technique (partial least squares regression) with Vis–NIR spectroscopy (350–2500 nm), Saidi et al. (2025)[176] reported comparable prediction accuracies for PBC (R^2_{Val} = 0.68) and Langmuir Smax (R^2_{Val} = 0.48). However, they also attained higher accuracies using artificial neural networks, with R^2_{Val} = 0.85 for PBC and R^2_{Val} = 0.65 for Smax.

Table 1-3. Laboratory-based prediction accuracies for soil properties using (Vis-)NIR spectroscopy.

Parameter	**Wavenumber ranges**	**Accuracy**	**Ref.**
Soil water (gravimetrically, %)	[Vis-NIR] 350 – 2500 nm	RPD_{Val} = 1.95	163
Soil water (gravimetrically)	[Vis-NIR] 400 – 2498 nm	RPD_{CV} = 2.98	177
Soil water (gravimetrically, %)	[NIR] 12,000 - 3800 cm^{-1}	RPD_{CV} = 4.88	178
Soil water (gravimetrically, g/g)	[NIR] 700 – 2500 nm	R^2_{Val} = 0.85	161
TN ($g \cdot kg^{-1}$)	[NIR] 400 – 2498 nm	RPD_{CV} = 3.99	177
TN ($g \cdot kg^{-1}$)	[Vis-NIR] 350 – 2500 nm	R^2_{Val} = 0.80	179
TN (%)	[Vis-NIR] 350 – 2500 nm	R^2_{Val} = 0.84	180
TN ($g \cdot kg^{-1}$)	[Vis-NIR] 350–1070 nm	R^2_{Val} = 0.97	181
TC ($g \cdot kg^{-1}$)	[Vis-NIR] 400 – 2498 nm	RPD_{CV} = 3.10	177
Log_{10}(SOC)	[Vis-NIR] 350 – 2500 nm	RPD_{Val} = 2.17	163
SOC ($g \cdot kg^{-1}$)	[Vis-NIR] 400 – 2498 nm	RPD_{CV} = 3.26	177
SOC ($g \cdot kg^{-1}$)	[NIR] 12,000 - 3800 cm^{-1}	RPD_{CV} = 5.75	178
SOC ($dag \cdot kg^{-1}$)	[NIR] 810 – 2400 nm	R^2_{Val} = 0.60	127
SOM ($g \cdot kg^{-1}$)	[Vis-NIR] 450 – 2400 nm	R^2_{Val} = 0.75	182
SOM ($g \cdot kg^{-1}$)	[Vis-NIR] 350 – 2500 nm	R^2_{Val} = 0.80	179
SOM ($g \cdot kg^{-1}$)	[Vis-NIR] 350 – 2500 nm	R^2_{Val} = 0.82	180
SOM ($g \cdot kg^{-1}$)	[Vis-NIR] 350–1070 nm	R^2_{Val} = 0.98	181
Clay (%)	[NIR] 400 – 2498 nm	RPD_{CV} = 2.02	177
Clay (%)	[Vis-NIR] 350 – 2500 nm	RPD_{Val} = 2.35	163
Clay (%)	[NIR] 700 – 2500 nm	R^2_{Val} = 0.72	161
Clay (%)	[Vis-NIR] 350 – 2500 nm	R^2_{Val} = 0.77	183
Clay ($g \cdot kg^{-1}$)	[NIR] 1100 – 2500 nm	R^2_{CV} = 0.82	175
Clay (%)	[Vis-NIR] 350 – 2500 nm	R^2_{Val} = 0.84	180
Soil pH (H_2O)	[NIR] 12,000 - 3800 cm^{-1}	RPD_{CV} = 1.90	178
pH (H_2O)	[Vis-NIR] 350 – 2500 nm	RPD_{Val} = 2.28	163
Soil pH ($CaCl_2$)	[Vis-NIR] 400 – 2498 nm	RPD_{CV} = 3.26	177
Soil pH (H_2O)	[NIR] 810 – 2400 nm	R^2_{Val} = 0.54	127
Soil pH ($CaCl_2$)	[NIR] 810 – 2400 nm	R^2_{Val} = 0.57	127
Soil pH (H_2O)	[NIR] 700 – 2500 nm	R^2_{Val} = 0.71	161
Soil pH (H_2O)	[NIR] 1100 – 2300 nm	R^2_{Val} = 0.74	184
Soil pH (KCl)	[Vis-NIR] 350 – 2500 nm	R^2_{Val} = 0.82	179
pH (H_2O)	[Vis-NIR] 350 – 2500 nm	R^2_{Val} = 0.86	179

Continued Table 1-3

Parameter	Wavenumber ranges	Accuracy	Ref.
CEC ($cmol_{(+)} \cdot kg^{-1}$)	[Vis-NIR] 350 – 2500 nm	RPD_{Val} = 2.13	163
CEC ($cmol_{(+)} \cdot kg^{-1}$)	[Vis-NIR] 400 – 2498 nm	RPD_{CV} = 3.10	177
CEC ($cmol_{(+)} \cdot kg^{-1}$)	[NIR] 12,000 - 3800 cm^{-1}	RPD_{CV} = 3.46	178
CEC (%)	[Vis-NIR] 350 – 2500	R^2_{Val} = 0.60	180
CEC ($cmol_{(+)} \cdot kg^{-1}$)	[Vis-NIR] 350 – 2500	R^2_{Val} = 0.74	183
CEC ($cmol_{(+)} \cdot kg^{-1}$)	[NIR] 1100 – 2500 nm	R^2_{CV} = 0.75	175
CEC ($cmol_{(+)} \cdot kg^{-1}$)	[Vis-NIR] 421 – 1745	R^2_{CV} = 0.92	185
BD ($g \cdot cm^{-1}$)	[Vis-NIR] 350 – 2500 nm	RPD_{Val} = 1.87	163
BD ($g \cdot cm^{-1}$)	[Vis-NIR] 350 – 2500 nm	R^2_{Val} = 0.35	183
BD ($g \cdot cm^{-1}$)	[Vis-NIR] 350 – 2500 nm	R^2_{Val} = 0.63	179
LR ($mg \cdot ha^{-1}$)	[NIR] 810 – 2400 nm	R^2_{Val} = 0.50	127
Log_{10}(TP)	[Vis-NIR] 350 – 2500 nm	RPD_{Val} = 1.75	163
TP ($mg \cdot kg^{-1}$)	[NIR] 1100 – 2500 nm	R^2_{CV} = 0.26	175
TP ($g \cdot kg^{-1}$)	[Vis-NIR] 350 – 2500 nm	R^2_{Val} = 0.61	179
TP ($g \cdot kg^{-1}$)	[Vis-NIR] 250 – 2500 nm	R^2_{CV} = 0.66	186
TP (%)	[Vis-NIR] 350 – 2500 nm	R^2_{Val} = 0.66	180
Log_{10}(Available-P)	[Vis-NIR] 350 – 2500 nm	RPD_{Val} = 1.39	163
Olsen Extractable-P ($mg \cdot kg^{-1}$)	[NIR] 1100 – 2500 nm	R^2_{CV} = 0.04	175
Available-P ($g \cdot kg^{-1}$)	[Vis-NIR] 350 – 2500 nm	R^2_{Val} = 0.43	180
Available-P ($g \cdot kg^{-1}$)	[Vis-NIR] 350 – 2500 nm	R^2_{Val} = 0.56	179
Available-P ($g \cdot kg^{-1}$)	[Vis-NIR] 450 – 2500 nm	R^2_{Val} = 0.65	182
Available-P ($g \cdot kg^{-1}$)	[Vis-NIR] 450 – 1100 nm	R^2 = 0.86	167
Langmuir Smax ($mg \cdot kg^{-1}$)	[Vis-NIR] 400–2500 nm	R^2_{CV} = 0.42	175
Freundlich binding energy ($L \cdot mg^{-1}$)	[NIR] 1100 – 2500 nm	R^2_{CV} = 0.58	175
Langmuir Smax ($mg \cdot kg^{-1}$)	[Vis-NIR] 350 - 2500 nm	R^2_{Val} = 0.65	176
Freundlich sorption maximum ($mg \cdot kg^{-1}$)	[Vis-NIR] 400–2500 nm	R^2_{CV} = 0.66	175
MBC	[Vis-NIR] 350 - 2500 nm	R^2_{Val} = 0.71	176
PBC	[Vis-NIR] 400–2500 nm	R^2_{CV} = 0.78	175
PBC	[Vis-NIR] 350 - 2500 nm	R^2_{Val} = 0.85	176

1.2.2.2 MIR

Absorption in MIR spectroscopy arises from the fundamental bands of molecular vibrations and can be approximately divided into four regions: the X–H stretching region (4000–2500 cm^{-1}), the triple-bond region (2500–2000 cm^{-1}), the double-bond region (2000–1500 cm^{-1}), and the fingerprint region (1500–600 cm^{-1})[134]. For example, absorptions around 3700–3600 cm^{-1} are related to O–H bonds in clay minerals (kaolinite, smectite/illite), and those around 2950–2850 cm^{-1} are associated with C–H stretching vibrations in SOM. Absorptions within 1830–1650 cm^{-1} correspond to C=O stretching vibrations in SOM, including functional groups such as ketones, esters, and carboxylic acids (e.g., –COOH, near 1750 cm^{-1}). Silicon (Si)-bearing minerals (e.g., quartz), which lack overtone features in the NIR region, exhibit strong absorption peaks in the MIR region (e.g., fundamental O-Si-O stretching peaks near 1100 – 1000 cm^{-1})[133,145,187].

Table 1-4 highlights studies that apply the advantages of laboratory-based MIR spectroscopy for quantitative analysis of various soil properties. Compared to (Vis-)NIR spectroscopy, which is based on weaker vibrational modes and produces broad, overlapping bands that hinder the resolution of specific chemical constituents, MIR spectroscopy has higher molecular sensitivity and stronger, more distinct absorption features that allow fingerprinting of specific chemical bonds[188,189]. Laboratory-based MIR often yield more robust predictions of various soil properties than (Vis-)NIR, such as soil C contents (R^2 = 0.77 – 0.99, with an median R^2 = 0.93 compare to 0.80 for [Vis-]NIR), clay content (R^2 = 0.79 – 0.96, median R^2 = 0.90 *vs*. 0.80 for [Vis-]NIR), soil pH (R^2 = 0.70 – 0.94, median R^2 = 0.87 *vs*. 0.73 for [Vis-]NIR) and CEC (R^2 = 0.34 - 0.95, average R^2 = 0.84 *vs*. 0.75 for [Vis-]NIR) (Table 1-3 and 1-4).

As in the (Vis-)NIR region, soil P has no direct absorption features in the MIR region due to the low dipole moment of P–O bonds. For example, Ma et al. (2019)[190] used diffuse reflectance (DRS), ATR, and photoacoustic (PAS) MIR techniques to predict available P in paddy soils; however, none of the predictions met the weakest performance criterion (RPIQ < 2), with R^2_{Val} = 0.50 (DRS), R^2_{Val} = 0.49 (ATR), and R^2_{Val} = 0.45 (PAS), respectively. The poor predictability of soil P from MIR has also been reported

in other studies, with R^2 values ranging from 0.07 to 0.58 (Table 1-4). However, soil P sorption capacity is expected to be predictable as a secondary property. A review by Soriano-Disla et al. (2014)[116] concluded that studies predicting P sorption using MIR spectroscopy achieved a median R^2 = 0.83. For residual P after equilibration, similar prediction accuracies were reported by Minasny et al. (2009)[191] and Ramaroson et al. (2023)[192], using equilibrium concentrations of 40 mg·P·L^{-1} and 60 mg·P·L^{-1}, with R^2_{Val} values of 0.79 and 0.77, respectively. A slightly higher accuracy for P sorption, assessed by residual P content, was achieved by Janik et al. (2009)[193], with R^2_{Val} values ranging from 0.84 to 0.87. For highly weathered Australian soils, Burkitt et al. (2006)[194] successfully developed MIR models using ground samples to predict the PBI (Section 1.1.4, measuring the ability of soils to adsorb added P), achieving R^2 values ranging from 0.82 to 0.87[195,196]. Forrester et al. (2015) [197] further evaluated benchtop (7800–450 cm^{-1}) and portable (4000–800 cm^{-1}) MIR spectrometers using unground soil samples (<2 mm) for PBI prediction, achieving similar accuracy with R^2_{CV} values of 0.82 and 0.87, respectively.

However, for Irish soils, P current studies using MIR spectroscopy have yielded less accurate predictions of P sorption capacity. Dunne et al. (2020)[147] used MIR spectroscopy (4000–450 cm^{-1}) to predict various P sorption indicators from ball-milled samples (<0.25 mm), achieving the highest accuracy for Langmuir Smax (isotherm 0 to 50 mg·L^{-1} range, R^2_{Val} = 0.67), while predictions for Freundlich parameters (Freundlich binding energy with R^2_{Val} = 0.34 – 0.39, Freundlich sorption maximum with R^2_{Val} = 0.25 – 0.49) or MBC (Section 1.1.4 , R^2_{Val} = 0.53 – 0.60) were less accurate.

Table 1-4. Laboratory-based prediction accuracies for soil properties using MIR spectroscopy.

Parameter	Wavelength ranges	Sample Preparation	Accuracy	Ref.
TN	4,000–600 cm^{-1}	<500µm	R^2_{CV} = 0.75	198
TN	4,000–600 cm^{-1}	<250µm	R^2_{CV} = 0.76	198
TN	4,000–600 cm^{-1}	<120µm	R^2_{CV} = 0.80	198
N (%)	4,000–600 cm^{-1}	<2mm	R^2_{Val} = 0.93	145
N (%)	4,000–600 cm^{-1}	Ball-milled	R^2_{Val} = 0.94	145
TC (%)	4,000–600 cm^{-1}	Ball-milled	R^2_{Val} = 0.92	145
TC ($g \cdot kg^{-1}$)		Ball-milled	R^2_{CV} = 0.93	148
TC (%)	4,000–600 cm^{-1}	<2mm	R^2_{Val} = 0.93	145
TC	4,000–600 cm^{-1}	<500µm	R^2_{CV} = 0.93	198
TC	4,000–600 cm^{-1}	<250µm	R^2_{CV} = 0.94	198
TC	4,000–600 cm^{-1}	<120µm	R^2_{CV} = 0.95	198
TC ($g \cdot kg^{-1}$)		Ball-milled	R^2_{Val} = 0.95	199
TC (%)		Ball-milled	R^2_{CV} = 0.97	149
TC ($g \cdot kg^{-1}$)		<2mm	R^2_{CV} = 0.98	148
Inorganic C ($g \cdot kg^{-1}$)	4,000–400 cm^{-1}	Ball-milled	R^2_{Val} = 0.98	200
SOC ($g \cdot kg^{-1}$)		Ball-milled	R^2_{Val} = 0.77	150
SOC ($g \cdot kg^{-1}$)	4,000–400 cm^{-1}	<500µm	R^2_{Val} = 0.78	201
SOC	4,000–600 cm^{-1}	<250µm	R^2_{CV} = 0.85	198
SOC	4,000–600 cm^{-1}	<500µm	R^2_{CV} = 0.91	198
SOC	4,000–600 cm^{-1}	<120µm	R^2_{CV} = 0.91	198
SOC (%)	4,000–600 cm^{-1}	Ball-milled	R^2_{Val} = 0.92	145
SOC (%)		<2mm	R^2_{Val} = 0.93	152
SOC (%)	4,000–600 cm^{-1}	<2mm	R^2_{Val} = 0.93	145
SOM (%)	4,000–400 cm^{-1}	Ball-milled	R^2_{Val} = 0.94	200
SOM (%)	4,000–600 cm^{-1}	Ball-milled	R^2_{Val} = 0.99	202
pH	4,000–400 cm^{-1}	<500µm	R^2_{Val} = 0.70	201
pH (H_2O)	4,000–700 cm^{-1}	<80µm	R^2_{CV} = 0.75	127
pH ($CaCl_2$)	4,000–700 cm^{-1}	<80µm	R^2_{CV} = 0.86	127
pH (H_2O)	4,000–600 cm^{-1}	<2mm	R^2_{Val} = 0.86	145
pH ($CaCl_2$)	4,000–600 cm^{-1}	Ball-milled	R^2_{Val} = 0.88	145
pH ($CaCl_2$)	4,000–600 cm^{-1}	<2mm	R^2_{Val} = 0.89	145
pH		Ball-milled	R^2_{Val} = 0.94	202

Continued Table 1-4

Parameter	Wavelength ranges	Sample Preparation	Accuracy	Ref.
CEC ($mmol_{(+)}·kg^{-1}$)	4,000–700 cm^{-1}	<80µm	R^2_{CV} = 0.34	203
CEC ($cmol·kg^{-1}$)	4,000–600 cm^{-1}	<2mm	R^2_{Val} = 0.82	145
CEC	4,000–600 cm^{-1}	<250µm	R^2_{CV} = 0.83	198
CEC ($cmol·kg^{-1}$)	4,000–600 cm^{-1}	Ball-milled	R^2_{Val} = 0.85	145
CEC	4,000–600 cm^{-1}	<500µm	R^2_{CV} = 0.91	198
CEC	4,000–600 cm^{-1}	<120µm	R^2_{CV} = 0.95	198
Clay (%)	4,000–400 cm^{-1}	<500µm	R^2_{Val} = 0.79	201
Clay (%)	4,000–600 cm^{-1}	Ball-milled	R^2_{Val} = 0.87	145
Clay (%)	4,000–600 cm^{-1}	<2mm	R^2_{Val} = 0.88	145
Clay	4,000–600 cm^{-1}	<250µm	R^2_{CV} = 0.92	198
Clay	4,000–600 cm^{-1}	<500µm	R^2_{CV} = 0.94	198
Clay	4,000–600 cm^{-1}	<120µm	R^2_{CV} = 0.96	198
Sand ($dag·kg^{-1}$)	4000–700 cm^{-1}	<80µm	R^2_{CV} = 0.74	127
Sand (%)	4,000–400 cm^{-1}	<500µm	R^2_{Val} = 0.79	201
Sand (%)	4,000–600 cm^{-1}	<2mm	R^2_{Val} = 0.80	145
Sand (%)	4,000–600 cm^{-1}	Ball-milled	R^2_{Val} = 0.82	145
Sand	4,000–600 cm^{-1}	<250µm	R^2_{CV} = 0.93	198
Sand (%)	4,000–400 cm^{-1}		R^2 = 0.95	196
Sand	4,000–600 cm^{-1}	<500µm	R^2_{CV} = 0.95	198
Sand	4,000–600 cm^{-1}	<120µm	R^2_{CV} = 0.95	198
LR ($Mg·ha^{-1}$)	4000–700 cm^{-1}	<80µm	R^2_{CV} = 0.75	127
LR ($t·ha^{-1}$)	4000–450 cm^{-1}	Ball-milled	R^2_{Val} = 0.76	204
LR ($t·ha^{-1}$)	4,000–400 cm^{-1}		R^2 = 0.86	196
BD ($g·cm^{-3}$)	4,000–600 cm^{-1}	Ball-milled	R^2_{Val} = 0.78	145
BD ($g·cm^{-3}$)	4,000–600 cm^{-1}	<2mm	R^2_{Val} = 0.79	145
BD ($g·cm^{-3}$)	4,000–600 cm^{-1}	Ball-milled	R^2_{Val} = 0.81	154
Exchangeable-Al ($mmol_{(+)}·kg^{-1}$)	4000–700 cm^{-1}	<80µm	R^2_{CV} = 0.43	127
Al ($g·kg^{-1}$)	4,000–600 cm^{-1}	Ball-milled	R^2_{Val} = 0.72	145
Oxalate Extractable-Al (%)	4000–1000 cm^{-1}	<180 µm	R^2_{Val} = 0.82	118
Al ($g·kg^{-1}$)	4,000–600 cm^{-1}	<2mm	R^2_{Val} = 0.83	145
M3-Extractable-Al ($mg·kg^{-1}$)	4000–1000 cm^{-1}	<180 µm	R^2_{Val} = 0.95	118
Bicarbonate Available-P	4000–250 cm^{-1}	Ball-milled	R^2_{CV} = 0.07	196

Continued Table 1-4

Parameter	Wavelength ranges	Sample Preparation	Accuracy	Ref.
Available-P (Colwell-P, mg·kg^{-1})	4000–700 cm^{-1}	<80µm	R^2_{CV} = 0.20	127
Available-P (Bray-P, mg·kg^{-1})	4000–500 cm^{-1}	<100 µm	R^2_{Val} = 0.28	193
Available-P (M3-P, mg·kg^{-1})	4000–1000 cm^{-1}	<180 µm	R^2_{Val} = 0.29	118
Bicarbonate Available-P	4,000–400 cm^{-1}	<200 µm	R^2_{Val} = 0.35	191
Available-P (Olsen-P mg·kg^{-1})	4000–1000 cm^{-1}	<180 µm	R^2_{Val} = 0.37	118
Available-P (kg·ha^{-1})	4,000–400 cm^{-1}	<500µm	R^2_{Val} = 0.38	201
Available-P (Bray-1-P, mg·kg^{-1})	4000–1000 cm^{-1}	<180 µm	R^2_{Val} = 0.41	118
Available-P (Oxalate-P, mg·kg^{-1})	4000–1000 cm^{-1}	<180 µm	R^2_{Val} = 0.43	118
P_2O_5	4000–600 cm^{-1}	<500µm	R^2_{CV} = 0.47	198
P_2O_5	4000–600 cm^{-1}	<250µm	R^2_{CV} = 0.52	198
P_2O_5	4000–600 cm^{-1}	<120µm	R^2_{CV} = 0.56	198
Available-P (Bray-1-P, mg·kg^{-1})	4000–400 cm^{-1}	<200 µm	R^2_{Val} = 0.58	191
Freundlich binding energy (L·mg^{-1})	4000–450 cm^{-1}	<250 µm	R^2_{Val} = 0.39	147
Freundlich sorption maximum (mg·kg^{-1})	4000–450 cm^{-1}	<250 µm	R^2_{Val} = 0.49	147
MBC (L·kg^{-1})	4000–450 cm^{-1}	<250 µm	R^2_{Val} = 0.60	147
PSI (L·kg^{-1})	4000–450 cm^{-1}	<250 µm	R^2_{Val} = 0.65	147
Langmuir Smax (mg·kg^{-1})	4000–450 cm^{-1}	<250 µm	R^2_{Val} = 0.67	147
P sorption capacity remaining (mg·kg^{-1})	4000–450 cm^{-1}	<250 µm	R^2_{Val} = 0.67	147
P-sorption (mg·kg^{-1})	4000–400 cm^{-1}	<200 µm	R^2_{Val} = 0.77	192
P-sorption (mg·kg^{-1})	4000–400 cm^{-1}	<200 µm	R^2_{Val} = 0.79	191
PBI	4000–800 cm^{-1}	<2mm	R^2_{CV} = 0.82 - 0.87	197
P-sorption (mg·kg^{-1})	4000–500 cm^{-1}	<100 µm	R^2_{Val} = 0.84 – 0.87	193
P-sorption (mg·kg^{-1})	4000–250 cm^{-1}	Ball-milled	R^2_{CV} = 0.87	196
P retention (%)	7498–600 cm^{-1}	<180 µm	R^2_{Val} = 0.89	118

1.2.4 Comparison of NIR and MIR for predicting soil P sorption indicators

Soil P sorption capacity can be predicted using IR-based models, as P sorption is closely related to the affinity of phosphate anions for soil components that exhibit characteristic responses in IR spectroscopy (e.g., Al/Fe oxides/hydroxides, SOM)[117]. In general, MIR spectroscopy has demonstrated better performance in predicting P sorption capacity, with an average reported R^2 of 0.83[116] and the highest achieved validation accuracy of R^2 = 0.89[118] (Table 1-4). In contrast, (Vis-)NIR spectroscopy has been less successful for predicting P sorption indicators, often yielding R^2 values below 0.70 (Table 1-3), with an average R^2 of 0.65 reported by Soriano-Disla et al. (2014)[116]. For example, Ramaroson et al. (2023)[192] compared MIR (4000-400 cm^{-1}) and NIR (1100-2500 nm) spectroscopy for predicting PSI (Section 1.1.4, a surrogate for P sorption capacity derived from full isotherms) and achieved accuracy of R^2_{Val} = 0.77 from MIR and R^2_{Val} = 0.65 from NIR. For isotherm-derived parameters, the accuracy of predicting Langmuir P adsorption capacity (Smax) using partial least squares regression on Vis-NIR spectra is also lower compared to MIR, with R^2_{Val} = 0.42–0.48 for Vis–NIR[175,176] and R^2_{Val} = 0.67 for MIR[147].

However, with regard to the prediction of Langmuir P sorption capacity, Smax, which is intended to be the predictive indicator for Irish soils in this study, MIR spectroscopy was found to be less successful compared to the median validation accuracy reported in Soriano-Disla et al. (2014)[116] (R^2 = 0.83). The accuracy of R^2_{Val} = 0.67 achieved by Dunne et al. (2020)[147] using benchtop MIR on dried, ball-milled soils was considered suitable only for rough screening purposes. Saidi et al. (2025)[176] compared the performance of a linear model (partial least squares regression) and a non-linear model (artificial neural networks) for predicting Smax, with the R^2_{Val} improving from 0.48 to 0.65. Consequently, this thesis aimed to improve MIR-based Smax prediction accuracy by exploring various machine learning techniques (Section 1.3.5).

1.2.5 Infrared spectrometers

1.2.5.1 Dispersive and Fourier-transform Infrared spectrometers

NIR spectroscopy has gained popularity in soil research due to improvements in grating manufacturing techniques[134]. The most significant advances in IR spectroscopy, though, have come about as a result of the introduction of Fourier-transform spectrometers[134]. At the core of an FTIR spectrometer is the Michelson interferometer, consisting of two perpendicular mirrors, one fixed and one movable along a direction perpendicular to the plane. The reflected beams are recombined to generate a signal as a function of the path length difference, producing an interferogram, which is then measured and transformed into an IR spectrum by a computer using Fourier transformation[134,205].

Compared to a dispersive instrument using monochromator, an FTIR spectrometer offers the advantages of Fellgett (higher signal-to-noise ratio in less time), Jacquinot (higher energy flux for faster scanning), and Connes (internal laser reference without the need for external calibration), typically providing faster, more stable scans and higher signal-to-noise ratio (SNR)[134,205,206]. The Fellgett (multiplex) advantage arises because all wavelengths are measured simultaneously, resulting in a higher signal-to-noise ratio and shorter acquisition time. The Jacquinot (throughput) advantage results from the absence of entrance and exit slits, allowing greater radiant energy to reach the detector and enabling faster scanning. The Connes advantage stems from the use of an internal laser reference, which provides precise wavelength calibration and long-term spectral stability without the need for external calibration. However, the NIR wavelength range of 800–2500 nm allows dispersive elements to effectively resolve the spectrum while meeting spectral resolution requirements[207]. The absence of interferometers makes NIR spectrometers relatively easy to miniaturise for use in a portable version. Even today, both Fourier-transform NIR (FT-NIR) and dispersive NIR (DG-NIR) spectrometers are used, with DG-NIR being the superior choice in many cases[208–210].

1.2.5.2 Instrumentation used in this thesis

In this study, both benchtop and handheld spectrometers were used for MIR scanning. The used benchtop spectrometer was the FTIR Spectrometer INVENIO S® (Bruker, Massachusetts - USA), equipped with an HTS-XT microplate reader extension for diffuse reflectance measurements (hereafter referred to as 'Bruker') (Figure 1-3). The Bruker spectrometer has been widely used in previous studies to predict various soil parameters with high accuracy, including soil C and N contents, clay, pH, CEC, LM, BD, and P sorption indices, *etc*.[145,147,154,188,211,212].

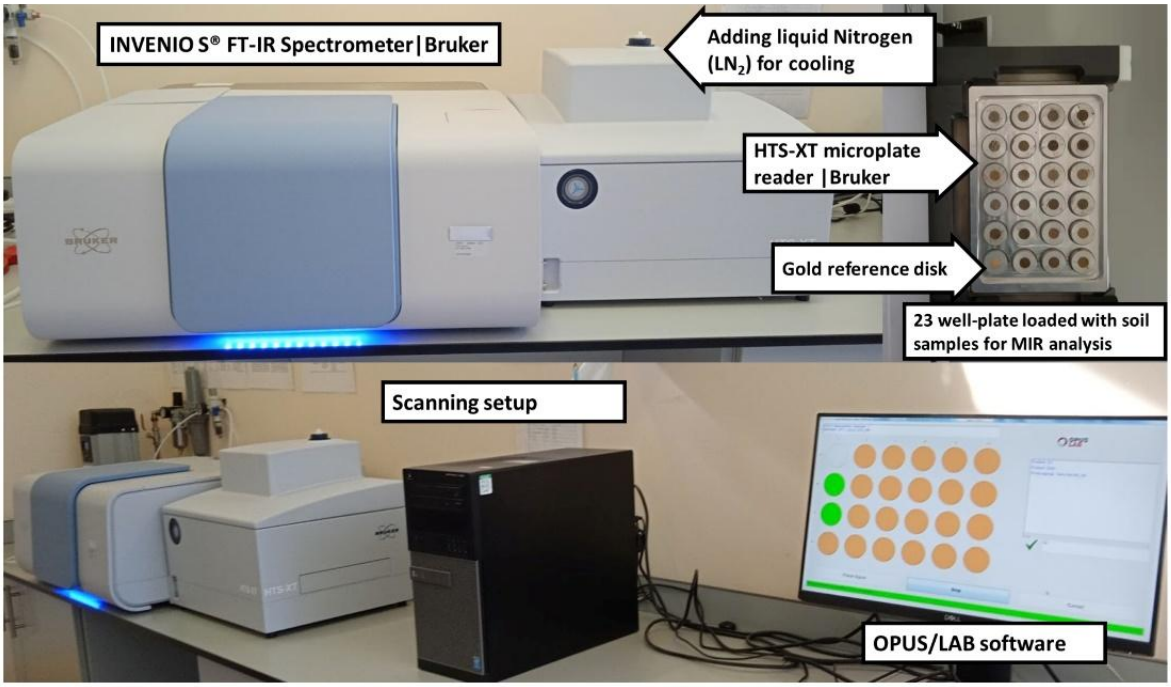


Figure 1-3. Scanning setup for the Bruker benchtop MIR spectrometer in the laboratory.

The used handheld MIR spectrometer was the Agilent 4300 FTIR (Agilent Technologies, Inc., Santa Clara, CA, USA), configured in diffuse reflectance mode (DRIFTS) with Happ-Genzel apodization (hereafter referred to as 'Agilent') (Figure 1-4). Shi et al. (2023)[213] and Mammadov et al. (2024)[214] used the same Agilent spectrometer as employed in this thesis. By scanning air-dried, <2mm sieved soil samples, Shi et al. (2023)[213] obtained good estimations for SOM (RPD = 1.10 - 5.80) and TN (RPD = 1.70 - 4.39); Mammadov et al. (2024)[214] reported superior predictions for $CaCO_3$, pH, SOC, and sand content.

Hutengs et al. (2019)[215] also used the same Agilent device to prediction SOM from MIR spectra collected both directly in the field and from pre-treated (dried and ground) samples in laboratory scanning, achieving an accuracy of R^2_{rdCV} = 0.63 and R^2_{rdCV} = 0.86, respectively. The potential of the handheld same Agilent spectrometer for predicting LR in soil samples with surface heterogeneity and moisture variability was also investigated by Metzger et al. (2021)[216].

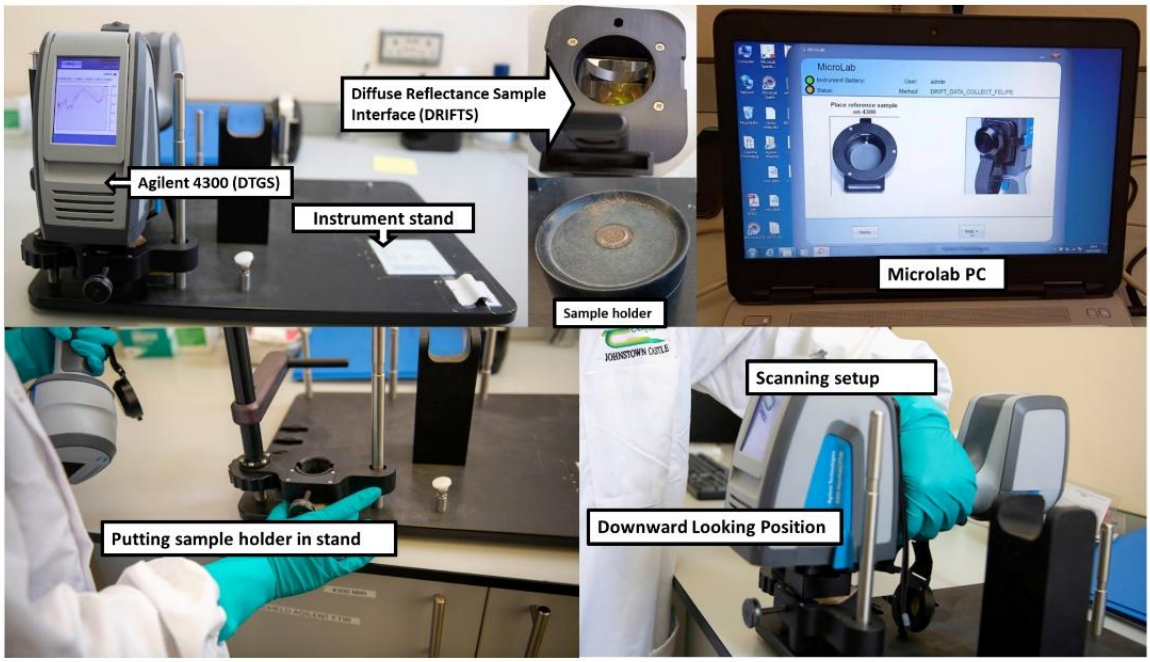


Figure 1-4. Scanning setup for the Agilent handheld MIR spectrometer in the laboratory.

The portable NIR spectrometers used in this study included the IR-S-G1 (abbreviated as 'Inno', 900–1700 nm) and the NeoSpectra scanner (Si-Ware Systems) (abbreviated as 'Neo', 1350–2550 nm) (Figure 1-5). The Neo spectrometer is a low-cost, handheld FT-NIR device based on micro-electro-mechanical systems (MEMS) technology, fabricated using the silicon-based micromachining technique[206,217]. In contrast, the Inno spectrometer employs an optical bandpass filter based on a linear variable filter (LVF), which achieves wavelength selection through a continuous change in optical coating thickness along its wedged geometry[206]. Several studies have used the Neo spectrometer to evaluate its accuracy in predicting soil properties, supporting its potential application in soil science[218–

[222]. For example, Metzger et al. (2024)[223] used the Neo spectrometer for in-field scanning to predict CEC (RPIQ = 5.55, R^2 = 0.90), clay (RPIQ = 4.73, R^2 = 0.95) and extractable magnesium (RPIQ = 2.95, R^2 = 0.78).

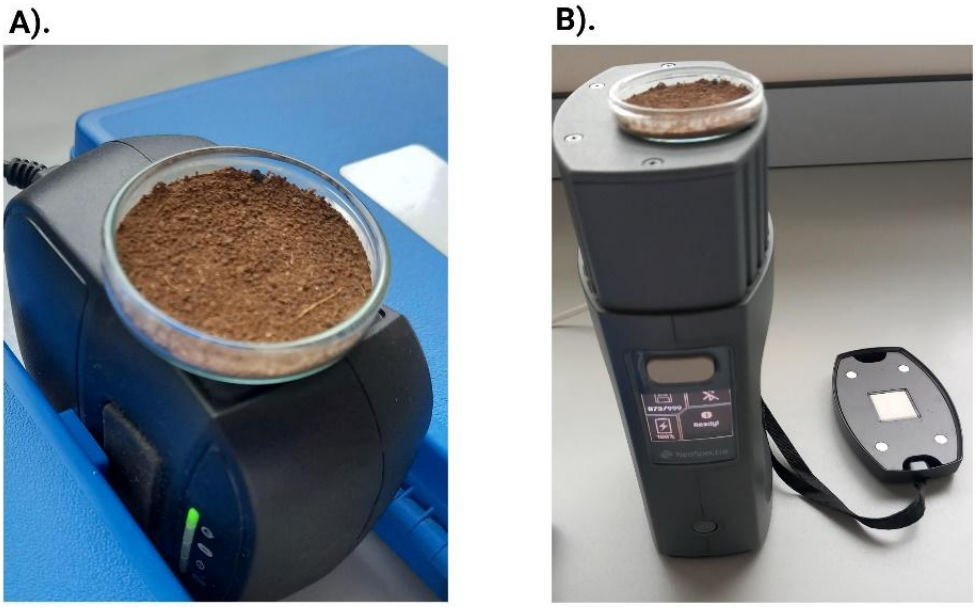


Figure 1-5. Scanning setup for the (A) Inno and (B) Neo handheld NIR spectrometers in the laboratory.

1.2.6 Applications of handheld IR spectroscopy for in situ soil analysis

Applying handheld instruments further reducing sample transfer and preparation steps while enhancing spatial and temporal resolution. Handheld NIR spectrometers, which are less sensitive to moisture interference and allow direct scanning of moist, unprepared soils, are more commercially affordable and widely available for field applications. Studies used portable (Vis-)NIR devices for on-site soil analysis, including the prediction of soil carbon contents (e.g., R^2_{Val} = 0.27 – 0.81)[215,224–227], clay contents (R^2_{Val} = 0.78)[228], pH (R^2_{Val} = 0.80)[226] and soil N (R^2_{Val} = 0.76 – 0.97)[225,226,229]. However, in situ Vis–NIR estimates of soil P dynamics have resulted in unreliable accuracy. Wenjun et al. (2014)[226] applied a portable Vis–NIR spectrometer (350-2500 nm) to estimate available-P, yet the prediction accuracy was limited (R^2_{Val} = 0.43) compared to the laboratory-based models (R^2_{Val} = 0.43 – 0.86).

Portable MIR spectrometers offer a more affordable alternative to spectroscopy laboratories equipped with benchtop FTIR spectrometers coupled with high-throughput sampling accessories[143,230]. Although they typically limit the SNR, spectral range and stability compared to benchtop systems, many studies have found that portable MIR instruments perform comparably to benchtop spectroscopy in laboratory settings[151,197,215,231–234]. Reeves et al. (2010)[231] found that a handheld MIR spectrometer (SOC-400, 4000–400 cm^{-1}) performed comparably to a benchtop spectrometer (Digilab FTS-7000, 4000–400 cm^{-1}) in predicting soil TC (R^2_{Val} = 0.91 [handheld] *vs.* 0.90 [benchtop]) and TN (R^2_{Val} = 0.88 *vs.* 0.86) from scans of air-dried and ground samples. Forrester et al. (2015)[197] compared benchtop (4000–400 cm^{-1}) and handheld (4000–600 cm^{-1}) MIR spectrometers for predicting PBI using dried, unground samples. Although the prediction accuracy declined from R^2_{CV} = 0.87 (benchtop) to R^2_{CV} = 0.82 (handheld), the results were still comparable. Soriano-Disla et al. (2017)[232] compared a benchtop MIR spectrometer (PerkinElmer Frontier, 4000–450 cm^{-1}) with two handheld instruments, the ExoScan-4100 and FlexScan-4200 (both 4000–750 cm^{-1}) for predicting various soil properties from dried and <2 mm sieved samples, and comparable prediction accuracies were also observed, including TC (R^2_{Val} = 0.94 [ExoScan], 0.95 [FlexScan], *vs.* 0.95 [benchtop]), SOC (0.67 and 0.68 *vs.* 0.64), clay (0.76 and 0.81 *vs.* 0.78), pH (0.84 and 0.83 *vs.* 0.82), and CEC (0.82 and 0.81 *vs.* 0.79). For BD, although prediction accuracy was relatively low, it remained similar across instruments (0.39 and 0.40 *vs.* 0.42). For LR, Metzger et al. (2021)[235] also concluded that handheld MIR spectroscopy can achieved comparable accuracy to benchtop measurements, with R^2_{Val} = 0.91 (handheld) *vs.* 0.87 (benchtop) on dried and sieved samples.

Although handheld MIR spectrometers support the transition of MIR spectroscopy from laboratory to field, they remain less widely adopted than portable Vis–NIR instruments for in situ soil analysis. For example, Metzger et al. (2021)[216] applied the Agilent-4300 for on-site soil scanning; however, the detected signal was too weak to yield meaningful spectra. Alongside environmental factors (e.g., rainfall, vegetation cover), surface roughness, and soil heterogeneity, soil moisture is a significant factor affecting soil MIR spectroscopy.

Soil water results in a broad absorption and specular reflection peaks that distort the spectrum and overlap spectral features[236]. Within the MIR region, the fundamental H-O-H stretching vibration for water produces a strong, broad, stretching vibration in the MIR near 3450–3250 cm^{-1} and a weaker bending vibration near 1630 cm^{-1} [236]. These vibrations overlap with features such as clay O–H stretching (~3700 – 3620 cm^{-1})[237]. In the NIR region, the overtone and combination bands of water absorption (e.g., at 1400 and 1900 nm, from the first overtones of O–H stretching vibrations and combinations of H–O–H bending and O–H stretching vibrations) are significantly less intense than their fundamental vibrations in the MIR region[237,238]. Consequently, although MIR spectroscopy generally yields more accurate predictions than NIR under laboratory-pretreated conditions, the presence of soil moisture poses a greater challenge to MIR, limiting its suitability for use with fresh/wet soil samples[215,237,239]. For example, Reeves (2010)[239] compared portable MIR and NIR spectrometers for predicting soil C on naturally moist samples and concluded that MIR calibrations were less robust than those of NIR. Soriano-Disla et al. (2014)[116] reviewed the use of handheld (Vis-)NIR devices under field conditions, with median prediction accuracies of TN R^2 = 0.84. However, Yin et al. (2023)[240] used a handheld MIR spectrometer (4000–650 cm^{-1}) to assess TN in situ in soils from typical black soil areas of China, achieving only R^2_{Val} = 0.71.

The wet MIR spectroscopy for soil parameter predictions have been calibrated under laboratory conditions using re-wetted samples. For example, Dhawale et al. (2015)[241] employed a prototype portable MIR spectrometer (1811 - 898 cm^{-1}) to predict SOM and soil texture (sand and clay content) using re-wetted (air-dried, <2 mm sieved) soil samples. The instrument achieved consistent accuracies for SOM (R^2_{Val} = 0.62 for re-wetted *vs.* R^2_{Val} = 0.58 for air-dried) and soil texture (sand: R^2_{Val} = 0.82 *vs.* 0.90; clay: R^2_{Val} = 0.91 *vs.* 0.88), suggesting that moisture-specific calibration enables portable MIR spectrometers to effectively predict SOM and texture under the moisture condition. Similarly, Seidel et al. (2022)[237] conducted a re-wetting experiment to calibrate handheld MIR spectroscopy for SOC prediction and reported comparable R^2 = 0.81 for re-wetted compared to R^2 = 0.88 for air-dried samples. However, re-wetting pre-treated samples (e.g., sieved and ground), which have altered soil

structure (e.g., aggregates, pore space), can lead to unrepresentative moisture distributions compared to field conditions.

Using a portable MIR device in situ, Ji et al. (2016)[242] employed the same MIR device (1811–898 cm^{-1}) previously tested by Dhawale et al. (2015)[241] on laboratory re-wetted samples. Their predictions for SOM (R^2_{CV} = 0.86), BD (R^2_{CV} = 0.85), and CEC (R^2_{CV} = 0.83) from in situ scanning were comparable to those obtained from laboratory spectroscopy on air-dried, <2 mm sieved samples (R^2_{CV} = 0.90, 0.92, and 0.81, respectively). Hutengs et al. (2019)[215] used the handheld MIR device of Agilent-4300 (4000-650 cm^{-1}) to predicted SOC in situ and achieved an R^2_{Val} = 0.63. Greenberg et al. (2022)[243] used the Agilent-4300 for in situ MIR scanning. Following extensive calibration procedures, 'satisfactory' prediction results (RPIQ ≥ 1.89) were obtained for SOC ($RPIQ_{Val}$ = 3.3), TN ($RPIQ_{Val}$ = 3.5), and clay content ($RPIQ_{Val}$ = 2.1). Yin et al. (2023)[240] also used the Agilent-4300 to assess TN and TP in situ, achieving an R^2_{Val} of 0.71 for TN. However, the prediction accuracy for TP only achieved an R^2_{Val} = 0.22. The infeasibility of using handheld MIR spectroscopy in situ to estimate available-P is consistent with the conclusion of Ji et al. (2016)[242], who predicted M3-P with R^2_{CV} = 0.43.

Most in situ IR spectroscopy studies have focused on primary soil properties (e.g., soil C and N contents). Few have attempted to study soil P dynamics. Although indicators of P sorption capacity are predictable, their estimation using portable IR spectroscopy under in situ conditions remains poorly understood.

1.3 Interpreting soil IR spectroscopy: Chemometrics and Machine Learning

Although interest in soil spectroscopy dates back to the 1930s, advances in analytics and computing have popularised its use over the past two decades[244]. With the development of chemometrics and machine learning (ML) techniques, challenges such as the high dimensionality and collinearity of spectral data have been effectively addressed.

Chemometrics is a subfield of chemistry. Linear latent projection methods such as Principal Component Analysis (PCA) and Partial Least Squares Regression (PLSR), have been foundational in chemometrics for decades[245]. Machine learning techniques, such as support vector machines (SVM), random forests (RF), and Cubist, tend to excel at handling large datasets, flexibly capturing non-linear relationships, and remaining robust to overfitting and noise[246].

1.3.1 Spectral Pre-processing

Interpreting IR spectra is generally challenged by unavoidable contamination from factors such as inhomogeneous particle size, variable optical path length, and differences in measurement conditions (e.g., scanning time and resolution)[189,247,248]. Pre-processing is therefore often recommended as the first step in spectral data analysis to reduce interferences from instrumental noise (e.g., detector noise and laser fluctuations), physical scattering (e.g., fluorescence and Mie scattering), and systematic errors (e.g., calibration drift), and to enhance the SNR[248–250].

Common methods include multiplicative scatter correction (MSC) for light scattering; standard normal variate (SNV) for optical path variation; and Savitzky–Golay filtering (SG) for baseline correction and feature enhancement[251,252] (Figure 1-6). Some advanced pre-processing techniques (e.g., deep learning-based) can provide modest improvements in predictive accuracy, however, they risk simultaneously compromising model transparency and interpretability[253]. In practice, these techniques are often applied sequentially to address different interference types and optimise pre-processing results[254–258]. However, excessive pre-processing can also lead to information loss or introduce biases.

On the other hand, applying pre-processing techniques does not necessarily guarantee improved performance of multivariate models[147,259–262]. Barra et al. (2022)[259] examined the impact of pre-processing techniques of MSC and SNV on soil MIR and NIR spectral data and found that their effectiveness varied depending on soil composition and spectral characteristics. While MSC improved the prediction accuracy for SOM content (R^2 increasing from 0.72 to 0.81), it had a negligible effect on soil texture predictions. In contrast, SNV preprocessing resulted in a decrease in model accuracy for high-clay-content soils, reducing R^2 from 0.78 to 0.65 and increasing the root mean square error by 15%. Salazar et al. (2023)[262] scanned air-dried, <2mm sieved soil samples with a portable NIR spectrometer (1350 - 2500 nm) to predict broad indicators of soil fertility. Their results showed that using raw spectral data, rather than pre-processed spectra, yielded better performance for several soil properties, including TN, SOC, pH, clay content and exchangeable Ca^{2+}. Dunne et al. (2020)[147] also observed that raw MIR spectra chieved a higher accuracy for predicting Langmuir Smax (in the 0 – 25 $mg \cdot L^{-1}$ P region, R^2_{Val}=0.60) than SG (2nd derivative) pre-processed spectra (R^2_{Val}=0.57).

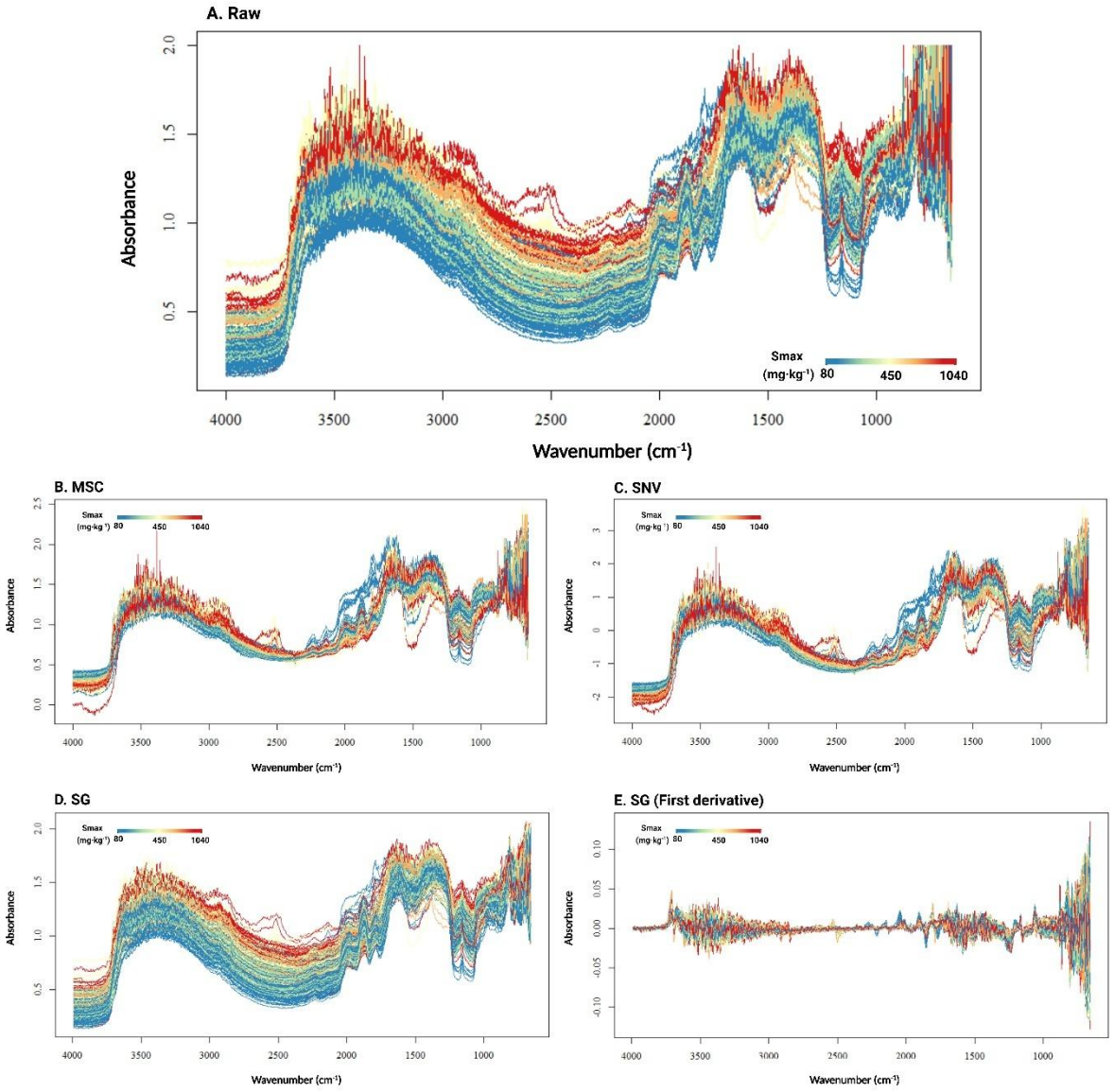


Figure 1-6. Effect of different preprocessing tools on MIR spectra (portable Agilent 4300, dried <2 mm samples, laboratory-scanned; scaled by the Langmuir Smax value). A: raw spectra, B: MSC (to minimise light scattering in spectral data), C: SNV (to remove light scattering in spectral data), D: SG (to smooth spectral signals without distorting their underlying), E: first derivative (to enhance enhances resolution and corrects baseline shifts).

1.3.2 Spectral transfer and moisture correction

The common pre-processing techniques (e.g., Section 1.3.1) are also applied to reduce moisture-related spectral noise. For example, Gao et al. (2021)[263] evaluated the effectiveness of various pre-processing techniques (e.g., first- and second-order derivatives, MSC, SNV, SG, and their combinations) in reducing moisture-induced spectral variation to improve soil water content prediction using wet NIR spectra (900–1700 nm) across a moisture range of 6.13% to 37.83% (median 26.27%). The SG smoothing filter was found to be the most effective at reducing high-frequency spectral noise, improving prediction accuracy from R^2_{Val} = 0.976, RPD = 3.707 (raw spectra) to R^2_{Val} = 0.995, RPD = 10.339 (SG pre-processed). Similarly, Seidel et al. (2022)[237] applied SG filtering (with a second-order polynomial) to pre-process wet MIR spectra and achieved an R^2 = 0.87 for predicting SOC in soils maintained at a constant gravimetric water content (GMC) of 15%. Metzger et al., (2021)[216] applied a pre-processing step using a 15-point moving average smoothing filter to correct wet MIR spectra from field-scanned samples (GMC < 15%) for LR prediction, reducing the deviation from laboratory reference values to from < 10 $t \cdot ha^{-1}$ to <5 $t \cdot ha^{-1}$. However, they also found that at higher field moisture levels (e.g., above the permanent wilting point, where plants cannot extract water), simple pre-processing became insufficient to correct moisture effects in soil property predictions.

To mitigate the distorting effects of water, mathematical techniques have been developed. For example, Ji et al. (2015)[264] applied the direct standardization (DS) algorithm to transform field-acquired Vis–NIR spectra (with soil volumetric water content of 40–50%, average 46.2%) into spectra that resemble laboratory-acquired spectra (air-dried and ground). Using PLSR on the DS-corrected field spectra, the prediction accuracy for SOM improved from R^2_{Val} = 0.25 to R^2_{Val} = 0.69. Other spectral transfer approaches such as continuous wavelet transform (CWT)[265] and piece-wise direct standardization (PDS)[266–268]. However, studies have evaluated spectral transfer approaches and found them to be less effective than orthogonality-based corrections such as orthogonal signal correction (OSC) and external parameter orthogonalisation (EPO). For example, de Santana et al. (2019)[266] reported that the PDS corrected moist NIR spectra (GMC = 5–35% w/w) still resulted in limited

prediction performance for SOM, with an R^2_{Val} = 0.46. The less satisfactory of PDS for moisture correction was also observed by Munnaf and Mouazen (2022)[268], who compared spectral transfer methods (DS and PDS) with orthogonality-based techniques (OSC and EPO), for Vis–NIR moisture correction in SOC prediction. Compared PLSR on non-corrected spectra yielded an R^2 = 0.58, DS-PLSR and PDS-PLSR yielded R^2 = 0.26 and 0.58, respectively, while orthogonality-based corrections (OSC-PLSR and EPO-PLSR) improved SOC prediction to R^2 = 0.61 (OSC-PLSR) and 0.60 (EPO-PLSR).

Especially for EPO, which works by decomposing and removing spectral variation caused by external parameters while preserving information relevant to the target property[269], it is one of the most commonly applied approaches (Equation 1-10). Metzger et al. (2024)[270] reported its use in over 50% of the reviewed studies due to its ease of understanding and straightforward application.

$$\mathbf{X} = \mathbf{XP} + \mathbf{XQ} + \mathbf{R}, \qquad \text{(Equation 1-10)}$$

where moist soil spectra (**X**) can be decomposed into a combination of a useful part (**XP,** related to chemical responses), an external part (**XQ**, containing irrelevant information affected by water content and other external parameters), and an independent residual (**R**).

In this review, most successful applications of EPO were employed for predicting soil C contents from wet spectra, particularly in the (Vis-)NIR region[238,266,270–275]. For example, Minasny et al. (2011)[271] applied EPO pre-processing coupled with PLSR to NIR spectra of re-wetted soil samples (GMC = 5–18%) for SOC prediction, improving R^2_{Val} from 0.59–0.77 to 0.82–0.89, comparable to air-dried <2 mm sieved samples (R^2_{Val} = 0.84). Using re-wetted Vis–NIR spectra, Wijewardane et al. (2016)[272] applied EPO for moisture correction and coupled it with SVM, significantly improving prediction accuracies for TC (R^2_{Val} from 0.06 to 0.70), SOC (R^2_{Val} from 0.12 to 0.75), and inorganic C (R^2_{Val} from 0.08 to 0.70). de Santana et al. (2019)[266] also applied EPO-PLSR to NIR spectra of re-wetted soil samples (5–35% w/w) and achieved SOM prediction accuracy (R^2_{Val} = 0.86, RPD_{Val} = 2.02) equivalent to that of air-dried samples (R^2_{Val} = 0.86, RPD_{Val} = 2.59). Yu et al. (2022)[273] applied EPO-PLSR to moist Vis–NIR spectra from

both re-wetted archives and in situ soils, achieving similar SOC prediction accuracy (R^2_{Val} = 0.56 [re-wetting] and 0.57 [in situ]), suggesting that laboratory re-wetted samples provide a cost-effective alternative for developing EPO matrices.

Other commonly studied soil properties using EPO correction of wet (Vis-)NIR spectra also include N[223,276] and clay content[270,272,274,276–278]. It has been less frequently studied for other parameters such as soil pH[272], CEC[272], and $CaCO_3$[274]. However, not all of these properties have been successfully predicted from EPO-pre-processed wet (Vis-)NIR spectra. For example, Wijewardane et al. (2016)[272] reported that EPO correction failed to improve predictions of soil pH and CEC, although this was likely due to the poor performance of the initial dry-ground models used to construct the EPO transfer.

Fewer studies have explored the use of EPO to correct moisture effects in IR spectra for predicting soil P dynamics. Although Wijewardane et al. (2016)[272] applied EPO to Vis–NIR spectra for soil extractable-P (M3-P) prediction, no improvement in accuracy was observed. Additionally, to the best of our knowledge, few studies have evaluated the effectiveness of EPO in eliminating water in the MIR region.

1.3.3 Chemometrics: PCA and PCR in soil spectral modelling

Principal Component Analysis (PCA) is a widely applied multivariate technique used to reduce data dimensionality by transforming the original correlated variables into a smaller set of uncorrelated principal components (PCs). Each PC is derived from an eigenvector of the covariance matrix, with the corresponding eigenvalue representing the proportion of variance explained. Notably, a PCA performed on a *p*-variate dataset yields p components, while dimensionality reduction is achieved only when the first *q* (< *p*) components are considered sufficient to capture most of the multivariate variation, typically based on the cumulative percentage of variance explained by these *q* components[279–281]. As shown in (Equation 1-11), PCA decomposes the original spectral matrix **X** into multiplication of loading (**P**) and score (**T**) matrices, along with a residual matrix (**E**).

$$\mathbf{X} = \mathbf{T} \cdot \mathbf{P}^{\mathrm{T}} + \mathbf{E}, \qquad \text{(Equation 1-11)}$$

where **X** is an $n \times m$ matrix containing the spectral data, with n representing the number of spectra (i.e., samples) and m the number of spectral bands (wavelengths). The score matrix **T** = **X·P**, which projects **X** into a lower-dimensional space, has dimensions $n \times a$, where a is the number of PCs. PCA aims to determine the loading matrix **P**, with dimensions $a \times m$, such that the score vectors $\mathbf{t_1} \ldots \mathbf{t_a}$ are arranged in descending order of variance and remain orthogonal.

In environmental and agricultural studies, PCA is widely used to identify significant indicators and classify soil into categories, thereby simplifying soil quality assessment and supporting targeted management[282,283]. In soil spectroscopy, PCA bridges spectral and wet chemical data, facilitating interpretation and model development[147,154,284,285]. For example, Dunne et al. (2020)[147] applied PCA to MIR spectra and found that loading plots revealed spectral features associated with soil properties influencing soil P dynamics, such as positive peaks related to SOM at 2892.9 cm^{-1}, 2857.1 cm^{-1}, and within 1821.4–1607.1 cm^{-1}; and peaks at 892.9 cm^{-1} and 750.0 cm^{-1} associated with Fe oxyhydroxides. Shi et al. (2023)[154] also applied PCA to MIR spectra to investigate the relationship between spectral features and soil bulk density. The loading plot of PC1 (29% explanation of the variance) identified absorption bands around 3700 cm^{-1} and 1000–900 cm^{-1}, attributed to clay minerals and indicative of soil texture, which were strongly correlated with BD.

Principal Component Regression (PCR) combines PCA and multiple linear regression by first extracting principal component scores (**T**) from spectral data (**X**) and then regressing these scores against soil properties (**Y**) (Equation 1-12)[279]. PCR is often used for (Vis-)NIR spectral calibration and prediction (Table 1-5).

$$\mathbf{Y} = \beta\mathbf{T} + \mathbf{E} \qquad \text{(Equation 1-12)}$$

where β denotes the vector of regression coefficients, found by solving $\beta = \frac{\mathrm{X^T Y}}{(\mathrm{X^T X})}$. **E** is a matrix of residuals, the same size than **Y**.

Table 1-5. Summary of studies predicting various soil attributes using PCR in IR soil spectroscopy**.**

Soil parameters	**Spectral region**	**n_{Cal} | n_{Val}**	**Accuracy**	**Ref.**
TC ($g \cdot kg^{-1}$)	VIS–NIR (400–2498 nm)	30 | 119	R^2_{Val} = 0.87	162
SOC (%)	NIR (700–2500 nm)	121 | 40	R^2_{Val} = 0.68	161
SOC (%)	UV–VIS–NIR (250–2500 nm)	121 | 40	R^2_{Val} = 0.76	161
TN ($g \cdot kg^{-1}$)	VIS–NIR (400–2498 nm)	30 | 119	R^2_{Val} = 0.85	162
Clay (%)	VIS–NIR (400–2498 nm)	30 | 119	R^2_{Val} = 0.67	162
Clay (%)	NIR (700–2500 nm)	121 | 40	R^2_{Val} = 0.75	161
pH (H2O)	VIS–NIR (400–2498 nm)	30 | 119	R^2_{Val} = 0.57	162
pH (H2O)	NIR (700–2500 nm)	121 | 40	R^2_{Val} = 0.70	161
pH (H2O)	UV–VIS–NIR (250–2500 nm)	121 | 40	R^2_{Val} = 0.71	161
CEC ($mmol_{(+)} \cdot kg^{-1}$)	UV–VIS–NIR (250–2500 nm)	121 | 40	R^2_{Val} = 0.64	161
CEC ($mmol_{(+)} \cdot kg^{-1}$)	NIR (700–2500 nm)	121 | 40	R^2_{Val} = 0.67	161
CEC ($cmol_{(+)} \cdot kg^{-1}$)	VIS–NIR (400–2498 nm)	30 | 119	R^2_{Val} = 0.81	162
Ca ($mmol_{(+)} \cdot kg^{-1}$)	NIR (700–2500 nm)	121 | 40	R^2_{Val} = 0.72	161
Ca (exch.) ($cmol_{(+)} \cdot kg^{-1}$)	VIS–NIR (400–2498 nm)	30 | 119	R^2_{Val} = 0.75	162
Fe (free) (%)	NIR (700–2500 nm)	121 | 40	R^2_{Val} = 0.49	161
Fe (free) (%)	UV–VIS–NIR (250–2500 nm)	121 | 40	R^2_{Val} = 0.52	161
Mehlich3-Fe ($mg \cdot kg^{-1}$)	VIS–NIR (400–2498 nm)	30 | 119	R^2_{Val} = 0.64	162

1.3.4 Partial Least Squares Regression

Partial Least Squares Regression (PLSR), developed by Wold (1983)[286], combines PCA and multiple regression to predict dependent variables (e.g., soil properties) by extracting orthogonal latent variables (LV, e.g., from spectral data), and is thereby also known as Projection to Latent Structures Regression[279]. A good explanation for a LV is available in Dunn (2025)[287] and Sansana et al. (2021)[281]. Unlike PCR, which performs spectral reduction and regression separately, PLSR integrates both steps by extracting LVs that maximise covariance between predictors (**X**, as in PCR) and responses (**Y**)[284]. The decomposition of **X** (Equation 1-11) and **Y** (Equation 1-13) is referred to as the 'outer relation' of PLSR. During calibration, the algorithm maximises the covariance between the **X** scores (**t**) and the **Y** scores (**U**) (Equation 1-14). This optimisation forms a linear regression between **t** and **U**, known as the 'inner relation', allowing the prediction of new, unknown samples.

($\mathbf{X} = \mathbf{T} \cdot \mathbf{P}^{\mathrm{T}} + \mathbf{E}$, Equation 1-11)

$$\mathbf{Y} = \mathbf{U} \cdot \mathbf{Q}^{\mathrm{T}} + \mathbf{F} \qquad \text{Equation 1-13}$$

where **X**, **T**, **P** and **E** have defined in Equation 1-11 (Section 1.3.3). **U** and **Q** are the score and loading matrices for **Y**, respectively. **F** contains its residuals**.**

A regression between **X** and **Y** is obtained by:

$$\mathbf{U} = \mathbf{T} \cdot \beta \qquad \text{Equation 1-14}$$

where β is the vector of regression coefficients of the linear model. Substituting this relationship from the original model, predictions are obtained by $\mathbf{Y} = \mathbf{U} \cdot \mathbf{Q}^{\mathrm{T}} = \mathbf{T} \cdot \beta \cdot \mathbf{Q}^{\mathrm{T}}$

In the review by Soriano-Disla et al. (2014)[116], PLSR was stated to be the most widely used multivariate method for calibrating soil spectral data. Table 1-6 highlights some of the most commonly studied soil properties that have been effectively predicted using PLSR. Additionally, similar to PCA, PLSR loadings are informative for qualitative interpretation, particularly in MIR spectra where well-resolved peaks often correspond to spectral features of components related to the target property[127,193,288].

However, studies have found that PLSR may offer limited predictive performance[202,289–291]. PLSR assumes a linear relationship between soil properties and reflectance behaviour. However, Deiss et al., (2020)[289] found that non-linear models (SVM with optimised hyperparameters) outperformed PLSR in predicting soil properties from Midwestern USA soils, including clay (RMSEP values, SVM *vs.* PLSR = 3.22 *vs.* 4.58%), sand (5.24 *vs.* 7.70%), pH (0.36 *vs.* 0.44), total organic carbon (0.19 *vs.* 0.29%), and permanganate-oxidizable carbon (74 *vs.* 123 mg·kg^{-1}). de Santana et al. (2022)[202] also reported that, compared to other non-linear ML techniques (e.g., Cubist: RMSEP = 1.08–2.14%), PLSR yielded the poorest prediction accuracy (RMSEP = 1.87–2.85%) for SOM in highly variable Irish topsoils (5–20 cm) using MIR spectroscopy.

Table 1-6. Summary of studies predicting various soil attributes using PLSR in IR soil spectroscopy.

Soil parameters	**Spectral region**	**n_{Cal} | n_{Val}**	**Accuracy**	**Ref.**
Biomass (N) (mg·kg^{-1})	NIR (1100–2300 nm)	180	R^2_{CV} = 0.71	184
N organic (total) (%)	MIR (4000-400 cm^{-1})	183	R^2_{CV} = 0.86	117
TN (topsoil) (%)	VIS–NIR (350–2500 nm)	114	R^2_{Val} = 0.85	292
TN (g·kg^{-1})	VIS–NIR (400–2498 nm)	76 | 32	R^2_{Val} = 0.86	293
TN (subsoil) (%)	VIS–NIR (350–2500 nm)	114	R^2_{Val} = 0.88	292
TN (mg·kg^{-1})	NIR (1100–2300 nm)	180	R^2_{CV} = 0.94	184
TN (%)	VIS–NIR (350–2500 nm)	59	R^2_{CV} = 0.95	294
TN (mg·kg^{-1})	NIR (1100–2498 nm)	120 | 59	R^2_{Val} = 0.95	295
TC (g·kg^{-1})	NIR (1100–2498 nm)	177 | 60	R^2_{Val} = 0.86	200
TC (g·kg^{-1})	VIS–NIR (400–2498 nm)	76 | 32	R^2_{Val} = 0.91	293

Continued Table 1-6

Soil parameters	Spectral region	n_{Cal} | n_{Val}	Accuracy	Ref.
TC ($g \cdot kg^{-1}$)	MIR (4000-400 cm^{-1})	177 | 60	R^2_{Val} = 0.95	200
TC ($g \cdot kg^{-1}$)	NIR (1100–2498 nm)	120 | 59	R^2_{Val} = 0.96	295
TC (%)	VIS–NIR (350–2500 nm)	59	R^2_{CV} = 0.97	294
C (inorg.) ($g \cdot kg^{-1}$)	NIR (1100–2498 nm)	177 | 60	R^2_{Val} = 0.87	200
C (inorg.) ($g \cdot kg^{-1}$)	VIS–NIR (400–2498 nm)	76 | 32	R^2_{Val} = 0.96	293
C (inorg.) ($g \cdot kg^{-1}$)	MIR (4000-400 cm^{-1})	177 | 60	R^2_{Val} = 0.98	200
SOC (subsoil) (%)	VIS–NIR (350–2500 nm)	114	R^2_{Val} = 0.73	292
SOC (acidified soil) ($g \cdot kg^{-1}$)	NIR (1100–2498 nm)	177 | 60	R^2_{Val} = 0.80	200
SOC (%)	VIS–NIR (400–2490 nm)	140 | 60	R^2_{Val} = 0.82	296
SOC (topsoil) (%)	VIS–NIR (350–2500 nm)	114	R^2_{Val} = 0.82	292
SOC ($g \cdot kg^{-1}$)	NIR (1100–2498 nm)	177 | 60	R^2_{Val} = 0.82	200
SOC ($g \cdot kg^{-1}$)	VIS–NIR (400–2498 nm)	76 | 32	R^2_{Val} = 0.89	293
SOC (%)	MIR (4000-400 cm^{-1})		R^2 = 0.92	297
SOC (%)	MIR (4000-400 cm^{-1})	188	R^2_{CV} = 0.93	117
SOC ($mg \cdot kg^{-1}$)	NIR (1100–2300 nm)	180	R^2_{CV} = 0.94	184
SOC ($g \cdot kg^{-1}$)	MIR (4000-400 cm^{-1})	177 | 60	R^2_{Val} = 0.97	200
SOM ($g \cdot dm^{-3}$)	VIS–NIR (462.0–2448.3 nm)		RMSEP = 5.40	298
SOM (%)	MIR (4000-400 cm^{-1})	31	R^2_{CV} = 0.98	299
Clay (%)	UV–VIS–NIR (250–2450 nm)	59	$RMSEP_{CV}$ = 2.9	300
Clay ($g \cdot kg^{-1}$)	VIS–NIR (462.0–2448.3 nm)		RMSEP = 77.38	298
Clay (%)	VIS–NIR (400–2490 nm)	140 | 60	R^2_{Val} = 0.19	296
Clay (%)	MIR (4000-400 cm^{-1})	88	R^2_{CV} = 0.79	117
Clay (%)	MIR (4000-500 cm^{-1})		R^2 = 0.87	297
Sand ($g \cdot kg^{-1}$)	VIS–NIR (462.0–2448.3 nm)		RMSEP = 90.63	298
Sand (%)	VIS–NIR (400–2500 nm)	319	R^2_{Val} = 0.70	301
Sand (%)	VIS–NIR (400–2490 nm)	140 | 60	R^2_{Val} = 0.24	296

Continued Table 1-6

Soil parameters	Spectral region	n_{Cal} | n_{Val}	Accuracy	Ref.
Sand (%)	MIR (4000-400 cm^{-1})	88	R^2_{CV} = 0.94	117
pH	VIS–NIR (400–2490 nm)	140 | 60	R^2_{Val} = 0.40	296
pH	MIR (4000-400 cm^{-1})		R^2 = 0.72	297
pH	NIR (1100–2498 nm)	120 | 59	R^2_{Val} = 0.73	295
pH	NIR (1100–2300 nm)	180	R^2_{CV} = 0.74	184
CEC ($cmol_{(+)}\cdot kg^{-1}$)	MIR (4000-400 cm^{-1})	183	R^2_{CV} = 0.88	117
Carbonate (%)	MIR (4000-400 cm^{-1})		R^2 = 0.95	297
LR ($t\cdot ha^{-1}$)	NIR (700–2500 nm)	183	R^2_{CV} = 0.73	117
LR ($t\cdot ha^{-1}$)	MIR (4000-400 cm^{-1})	183	R^2_{CV} = 0.86	117
LR ($t\cdot ha^{-1}$)	MIR (4000-450 cm^{-1})	495 | 160	R^2_{Val} = 0.89	204
BD ($g\cdot cm^{-3}$)	MIR (4000 – 600 cm^{-1})	537 | 134	R^2_{Val} = 0.79	154
Al (exch.) ($cmol\cdot kg^{-1}$)	MIR (4000-400 cm^{-1})	183	R_{CV}^2 = 0.64	117
Total P ($mg\cdot kg^{-1}$)	NIR (1100–2500 nm)	104 | 42	R^2_{Val} = 0.26	192
Total P ($mg\cdot kg^{-1}$)	MIR (4000 – 400 cm^{-1})	104 | 41	R^2_{Val} = 0.60	192
P (ppm)	VIS–NIR (400–2490 nm)	140 | 60	R^2_{Val} = 0.71	296
[1] P_{resin} ($mg\cdot kg^{-1}$)	NIR (1100–2500 nm)	104 | 42	R^2_{Val} = 0.01	192
P_{resin} ($mg\cdot kg^{-1}$)	MIR (4000 – 400 cm^{-1})	104 | 41	R^2_{Val} = 0.01	192
P (avail.) ($mg\cdot kg^{-1}$)	MIR (4000-400 cm^{-1})	186	R_{CV}^2 = 0.07	117
Mehlich3-P ($kg\cdot ha^{-1}$)	Vis-NIR (350-2500 nm)	401	R_{CV}^2 = 0.26	302
[2] P_{rem} ($mg\cdot kg^{-1}$)	NIR	104 | 42	R^2_{Val} = 0.65	192
[3] Langmuir Smax ($mg\cdot kg^{-1}$)	MIR (4000-450 cm^{-1})	147 | 49	R^2_{Val} = 0.67	147
P_{rem} ($mg\cdot kg^{-1}$)	MIR (4000 – 400 cm^{-1})	104 | 41	R^2_{Val} = 0.77	192

[1] P_{resin}: P available.

[2] P_{rem}: P remaining in solution (60 $mg\cdot P\cdot L^{-1}$ (as KH_2PO_4) in 0.01M $CaCl_2$ added at a 1:10 soil:solution ratio).

[3] Langmuir Smax: Langmuir sorption maximum in the 0 - 50 $mg\cdot P\cdot L^{-1}$ region.

1.3.5 Machine learning

1.3.5.1 Cubist

The Cubist algorithm is based on Quinlan's M5 algorithm, which combines decision trees with multivariate linear regressions at terminal nodes ('leaves')[303]. The rule-based structure '*IF* (condition) – *THEN* (linear regression model) – *ELSE* (apply next rule)' directs samples into low variability subsets, enhancing the accuracy of local linear regressions[163,304]. To optimise Cubist, a nearest-neighbour approach is often used to select the most similar training samples[154,202,279,305]. In R programming, the optimal combination of two key hyperparameters, 'committees' (number of sequential Cubist models) and 'neighbours' (extent of adjustment using similar samples) can be identified using a grid search approach[154,202,279,305]. However, the risk of overfitting should be carefully considered during model application[306].

Table 1-7 highlights some studies that applied the Cubist algorithm to predict soil properties. Cubist is considered as an alternative to PLSR for calibrating soil spectroscopic data[306], offering easily interpretable rule-based models and often achieving better predictive performance for soil properties. For example, as summarised in Table 1-6, using IR spectroscopy, PLSR generally yielded R^2_{Val} values of 0.73–0.97 for SOC, 0.19–0.87 for clay, and 0.40–0.74 for soil pH, whereas Cubist achieved higher accuracies with ranges of (SOC:) 0.76–1.00, (clay:) 0.76–0.96, and (soil pH:) 0.53–0.88, respectively. Minasny et al. (2008)[307] were among the first to apply Cubist for soil spectroscopy modelling using MIR, achieving 'good' (RPD > 2) predictions for TC (RPD_{Val} = 6.25), CEC (RPD_{Val} = 2.95), and exchangeable Ca and Mg (RPD_{Val} = 2.46 and 2.78, respectively), but reported BD to be unpredictable by Cubist (R^2_{Val} = 0.26). Several recent studies have demonstrated that Cubist can successfully predict soil BD using IR spectroscopy[145,154,212]. For example, Haghi et al. (2021)[308] used Cubist to predict soil BD, achieving R^2_{Val} of 0.85 and 0.88 using NIR and MIR spectroscopy, respectively. However, few studies have explored the use of Cubist models for predicting soil P dynamics.

Table 1-7. Summary of studies predicting soil attributes using Cubist in IR soil spectroscopy.

Soil parameters	**Spectral region**	**n_{Cal} | n_{Val}**	**Accuracy**	**Ref.**
Log_{10}(TN) (%)	Vis-NIR (350 – 2500 nm)	7449	RPD_{Val} = 2.11	163
SOC (Organic soils) ($g \cdot kg^{-1}$)	Vis-NIR (400 – 2500 nm)	n_{Val} = 368	R^2_{Val} = 0.76	309
SOC (Grassland) ($g \cdot kg^{-1}$)	Vis-NIR (400 – 2500 nm)	n_{Val} = 1383	R^2_{Val} = 0.87	309
SOC (Woodland) ($g \cdot kg^{-1}$)	Vis-NIR (400 – 2500 nm)	n_{Val} = 1564	R^2_{Val} = 0.89	309
SOC (%)	Vis-NIR (350 – 2500 nm)	n_{Val} = 3585	R^2_{Val} = 0.89	285
SOC ($g \cdot kg^{-1}$)	Vis-NIR (350 – 2500 nm)	3738 | 7475	R^2_{Val} = 0.96	310
SOC (%)	MIR	33991 | 8498	R^2_{Val} = 1.00	212
SOM ($g \cdot kg^{-1}$)	Vis-NIR (350 – 2500 nm)	160	R^2_{CV} = 0.57	258
SOM ($g \cdot kg^{-1}$)	MIR (4000 – 650 cm^{-1})	160	R^2_{CV} = 0.83	258
Clay (%)	Vis-NIR (350 – 2500 nm)	15205	RPD_{Val} = 2.35	163
Clay (%)	Vis-NIR (350 – 2500 nm)	n_{Val} = 5177	R^2_{Val} = 0.71	285
Clay (%)	MIR	26274 | 6569	R^2_{Val} = 0.96	212
Sand (%)	Vis-NIR (350 – 2500 nm)	n_{Val} = 3525	R^2_{Val} = 0.57	285
pH	Vis-NIR (350 – 2500 nm)	160	R^2_{CV} = 0.53	258
pH (H_2O)	Vis-NIR (350 – 2500 nm)	n_{Val} = 4389	R^2_{Val} = 0.62	285
pH	MIR (4000 – 650 cm^{-1})	160	R^2_{CV} = 0.75	258
pH (H_2O)	MIR	27959 | 6990	R^2_{Val} = 0.88	212
CEC ($cmol_{(+)} \cdot kg^{-1}$)	MIR (4000 – 650 cm^{-1})	160	R^2_{CV} = 0.49	258
CEC ($cmol_{(+)} \cdot kg^{-1}$)	Vis-NIR (350 – 2500 nm)	n_{Val} = 2826	R^2_{Val} = 0.73	285
CEC ($cmol_{(+)} \cdot kg^{-1}$)	MIR	29270 | 7317	R^2_{Val} = 0.99	212
$CaCO_3$ (%)	Vis-NIR (350 – 2500 nm)	n_{Val} = 739	R^2_{Val} = 0.77	285
$CaCO_3$ (%)	MIR	9668 | 6569	R^2_{Val} = 0.98	212
BD ($g \cdot cm^{-3}$)	Vis-NIR (350 – 2500 nm)	1745	RPD_{Val} = 2.20	163
BD ($mg \cdot m^{-3}$)	MIR	93 | 39	R^2_{Val} = 0.26	307
BD ($g \cdot cm^{-3}$)	MIR	13852 | 3463	R^2_{Val} = 0.67	163
BD ($g \cdot cm^{-3}$)	MIR (4000 – 600 cm^{-1})	537 | 134	R^2_{Val}= 0.78	154

Continued Table 1-7

Soil parameters	Spectral region	n_{Cal} | n_{Val}	Accuracy	Ref.
BD ($g \cdot cm^{-3}$)	NIR (1100 - 2500 nm)	520 | 130	R^2_{Val} = 0.85	308
BD ($g \cdot cm^{-3}$)	MIR (4000 – 400 cm^{-1})	520 | 130	R^2_{Val} = 0.88	308
Log_{10}(Total-P) (%)	Vis-NIR (350 – 2500 nm)	5647	RPD_{Val} = 1.75	163
TP ($g \cdot kg^{-1}$)	Vis-NIR (350 – 2500 nm)	259 | 120	R^2_{Val} = 0.66	311
Log_{10}(Available-P) ($mg \cdot kg^{-1}$)	Vis-NIR (350 – 2500 nm)	5069	RPD_{Val} = 1.39	163

1.3.5.2 Random Forest

Random Forest (RF) is an ensemble learning method that builds decision trees using bootstrap aggregation (bagging) and random feature selection at each node split[279,284,312]. Generally, two-thirds of the calibration samples are randomly selected for training each tree (in-bag), and the remaining third (out-of-bag, OOB) estimates model performance[202]. Random feature selection mitigates overfitting by reducing data dimensionality, selecting a subset of variables ('mtry'), typically one third of the total predictors, at each node for splitting[279,312]. For example, de Santana et al. (2022)[202] explained that in a spectral dataset with 1500 variables, only 500 variables are randomly selected for consideration at each node.

RF is increasingly being used in soil mapping and spectroscopy because of its ability to handle high-dimensional data with relatively limited samples and its robustness to noise and irrelevant variables[313]. de Santana et al. (2018)[314] compared RF with PLSR for predicting various soil quality parameters using Vis–NIR and found that RF significantly statistically outperformed PLSR with higher R^2_{Val} for SOM (0.80 [RF] *vs.* 0.75 [PLSR]), clay (0.80 *vs.*0.71), sand (0.82 *vs.*0.74) and CEC (0.88 *vs.*0.81). Additional advantages of RF in their studies included fewer samples being identified as outliers and the absence of negative predicted values.

However, Rossel and Behrens (2010)[284] compared several data mining algorithms for modelling soil Vis–NIR spectra (e.g., PLSR, SVM, multivariate adaptive regression splines and artificial neural

networks) and found that RF was the least accurate for predicting SOC (R^2 = 0.71), clay (R^2 = 0.77), and pH (R^2 = 0.63); for example, PLSR yielded higher R^2 values of 0.82 (SOC), 0.83 (clay), and 0.73 (pH). Moreover, as a black-box model, RF can overlook pedological knowledge and is often difficult to interpret, potentially resulting in misleading outcomes[280,313]. In this review, compared to other chemometric/ML techniques, RF is the least commonly applied in soil IR spectroscopy (Table 1-8).

Table 1-8. Summary of studies predicting various soil attributes using RF in IR soil spectroscopy.

Soil parameters	**Spectral region**	**n_{Cal} | n_{Val}**	**Accuracy**	**Ref.**
TN ($g \cdot kg^{-1}$)	NIR (1000 – 2500 nm)	119	R^2_{CV} = 0.44	315
TN ($g \cdot kg^{-1}$)	MIR (4000-400 cm^{-1})	119	R^2_{CV} = 0.79	315
TOC ($g \cdot kg^{-1}$)	NIR (1000 – 2500 nm)	119	R^2_{CV} = 0.64	315
TOC ($g \cdot kg^{-1}$)	MIR (4000-400 cm^{-1})	119	R^2_{CV} = 0.89	315
SOC (%)	MIR	33991 | 8498	R^2_{Val} = 0.99	212
SOM ($g \cdot dm^{-3}$)	Vis-NIR (400 – 2500 nm)	2514 | 1241	R^2_{Val} = 0.80	316
SOM (%)	Vis-NIR (400 – 2450 nm)	100	R^2_{CV} = 0.71	317
Clay (%)	Vis-NIR (400 – 2450 nm)	100	R^2_{CV} = 0.66	317
Clay ($g \cdot kg^{-1}$)	Vis-NIR (400 – 2500 nm)	417 | 208	R^2_{Val} = 0.80	316
Clay ($g \cdot kg^{-1}$)	MIR (4000-400 cm^{-1})	119	R^2_{CV} = 0.90	315
Clay (%)	MIR	26274 | 6569	R^2_{Val} = 0.93	212
Sand ($g \cdot kg^{-1}$)	Vis-NIR (400 – 2500 nm)	398 | 208	R^2_{Val} = 0.82	316
pH	MIR (4000-400 cm^{-1})	119	R^2_{CV} = 0.72	315
pH (H_2O)	MIR	27959 | 6990	R^2_{Val} = 0.82	212
CEC ($mmol \cdot kg^{-1}$)	MIR (4000-400 cm^{-1})	119	R^2_{CV} = 0.77	315
CEC ($mmol_{(+)} \cdot dm^{-3}$)	Vis-NIR (400 – 2500 nm)	2485 | 1214	R^2_{Val} = 0.87	316
CEC ($mmol \cdot kg^{-1}$)	MIR	29270 | 7317	R^2_{Val} = 0.97	212
$CaCO_3$ (%)	MIR	9668 | 2417	R^2_{Val} = 0.97	212
BD ($g \cdot m^{-3}$)	MIR (4000-400 cm^{-1})	119	R^2_{CV} = 0.52	315
BD ($g \cdot cm^{-3}$)	Vis-NIR (400 – 2500 nm)	1809 | 653	R^2_{Val} = 0.62	318

Continued Table 1-8

Soil parameters	Spectral region	n_{Cal} | n_{Val}	Accuracy	Ref.
BD ($g \cdot cm^{-3}$)	MIR (4000 – 600 cm^{-1})	537 | 134	R^2_{Val} = 0.80	154
[1] Extractable-Al (%)	MIR	18314 | 4578	R^2_{Val} = 0.83	212
[2] Extractable-Fe (%)	MIR	17054 | 4264	R^2_{Val} = 0.69	212

[1] Al (%): Dithionite citrate extractable aluminium.

[2] Fe (%): Acid oxalate extractable iron.

1.3.5.3 Support Vector Machine

Support Vector Machine (SVM) is a supervised learning algorithm used for classification and regression[319]. It constructs an optimal hyperplane (hard- or soft-margin) that minimises prediction error. Kernel functions (e.g., linear, polynomial, sigmoid, radial basis function [RBF]) enable non-linear modelling by projecting data into higher-dimensional spaces, where the hyperplane maximises the margin from support vectors[320,321]. Among various kernel functions, the RBF kernel is the most widely used in soil spectral modelling[284,289,298,322].

In R programming, applying SVM involves tuning key hyperparameters: (i) 'Gamma', which controls the influence of individual training instances and balances fitting accuracy with model smoothness; (ii) 'Cost', which penalises errors and controls model complexity, with higher values increasing the risk of overfitting; and (iii) 'Epsilon', which defines an insensitive zone where errors within a threshold are not penalised[322,323].

In soil spectroscopy, SVM has been widely used for soil texture and soil quality discrimination[324,325] and various soil property predictions[145,154] (Table 1-9). Many studies have found that SVM outperforms other multivariate models in predicting various soil properties such as soil BD[154], SOC[326], CEC[327], soil pH[289], and soil texture[289]. However, Behrens and Scholten (2006)[328] found that SVM underperformed compared to other ML approaches in digital soil mapping. The study by Meliho et al. (2023)[329] likewise reported that SVM yielded the lowest accuracy for predicting soil organic carbon stocks, compared to

Cubist and RF models. Therefore, the performance of ML varies with soil conditions. Rather than seeking a universally superior algorithm for soil property prediction, it is more important to identify suitable modelling techniques and apply proper tuning.

Table 1-9. Summary of studies predicting various soil attributes using SVM in IR soil spectroscopy.

Soil parameters	**Spectral region**	**n_{Cal} | n_{Val}**	**Accuracy**	**Ref.**
TN ($g \cdot kg^{-1}$)	Vis-NIR (700 – 2500 nm)	501	R^2_{CV} = 0.50	330
TN ($g \cdot kg^{-1}$)	NIR (1000 – 2500 nm)	119	R^2_{CV} = 0.66	315
TN ($g \cdot kg^{-1}$)	Vis-NIR (350 – 2500 nm)	64 | 32	R^2_{Val} = 0.75	331
TN ($g \cdot kg^{-1}$)	MIR (4000 - 650 cm^{-1})	501	R^2_{CV} = 0.79	330
TN ($g \cdot kg^{-1}$)	MIR (4000 - 400 cm^{-1})	119	R^2_{CV} = 0.82	315
TN ($g \cdot kg^{-1}$)	Vis-NIR (350 – 2500 nm)	410 | 175	R^2_{Val} = 0.86	332
TOC ($g \cdot kg^{-1}$)	NIR (1000 – 2500 nm)	119	R^2_{CV} = 0.82	315
TOC ($g \cdot kg^{-1}$)	MIR (4000 - 400 cm^{-1})	119	R^2_{CV} = 0.88	315
SOC ($g \cdot kg^{-1}$) (lab/fresh)	Vis-NIR (350 – 2500 nm)	66 | 32	R^2_{Val} = 0.72	333
SOC ($g \cdot kg^{-1}$) (field)	Vis-NIR (350 – 2500 nm)	309 | 104	R^2_{Val} = 0.81	334
SOC (%)	Vis-NIR (350 – 2500 nm)		R^2_{CV} = 0.82	284
SOC	Vis-NIR (370 – 2150 nm)	98 | 42	R^2_{Val} = 0.84	335
SOC ($g \cdot kg^{-1}$)	Vis-NIR (350 – 2500 nm)	66 | 32	R^2_{Val} = 0.86	333
SOC (%) (total)	Vis-NIR (400 – 2500 nm)	109 | 40	R^2_{Val} = 0.89	336
SOM (%)	Vis-NIR (350 – 2500 nm)	5161 | 1998	$RMSEP_{Val}$ = 4.80	337
SOM (%)	Vis-NIR (350 – 1700 nm)	70 | 32	R^2_{Val} = 0.75	338
SOM ($g \cdot dm^{-3}$)	Vis-NIR		R^2 = 0.78	298
SOM ($g \cdot kg^{-1}$)	Vis-NIR (350 – 2500 nm)	410 | 175	R^2_{Val} = 0.88	332
SOM (%)	Vis-NIR (400 – 1700 nm)	189 | 37	R^2_{Val} = 0.91	339
Clay ($g \cdot kg^{-1}$)	NIR (1000 – 2500 nm)	119	R^2_{CV} = 0.62	315
Clay (%)	Vis-NIR (350 – 1700 nm)	70 | 32	R^2_{Val} = 0.64	338
Clay (%)	Vis-NIR (350 – 2500 nm)		R^2_{CV} = 0.83	284

Continued Table 1-9

Soil parameters	Spectral region	n_{Cal} | n_{Val}	Accuracy	Ref.
Clay ($g \cdot kg^{-1}$)	MIR (4000 - 400 cm^{-1})	119	R^2_{CV} = 0.89	315
Clay ($g \cdot kg^{-1}$)	Vis-NIR		R^2 = 0.91	298
Sand ($g \cdot kg^{-1}$)	Vis-NIR		R^2 = 0.93	298
pH	NIR (1000 – 2500 nm)	119	R^2_{CV} = 0.34	315
pH (H_2O)	Vis-NIR (350 – 2500 nm)		R^2_{CV} = 0.72	284
pH	MIR (4000 - 400 cm^{-1})	119	R^2_{CV} = 0.80	315
CEC ($mmol \cdot kg^{-1}$)	NIR (1000 – 2500 nm)	119	R^2_{CV} = 0.78	315
CEC ($mmol \cdot kg^{-1}$)	MIR (4000 - 400 cm^{-1})	119	R^2_{CV} = 0.79	315
BD ($g \cdot cm^{-3}$)	MIR (4000 – 600 cm^{-1})	537 | 134	R^2_{Val} = 0.81	154
TP ($g \cdot kg^{-1}$)	Vis-NIR (700 – 2500 nm)	501	R^2_{CV} = 0.08	330
TP ($g \cdot kg^{-1}$)	Vis-NIR-MIR	501	R^2_{CV} = 0.15	330
TP ($g \cdot kg^{-1}$)	MIR (4000 - 650 cm^{-1})	501	R^2_{CV} = 0.21	330
TP ($g \cdot kg^{-1}$)	Vis-NIR (350 – 2500 nm)	410 | 175	R^2_{Val} = 0.76	332
P-available ($mg \cdot kg^{-1}$)	Vis-NIR (700 – 2500 nm)	501	R^2_{CV} = 0	330
P-available ($mg \cdot kg^{-1}$)	MIR (4000 - 650 cm^{-1})	501	R^2_{CV} = 0.02	330
P-available ($mg \cdot kg^{-1}$)	Vis-NIR-MIR	501	R^2_{CV} = 0.09	330
P-available (field)	Vis-NIR (350 – 2500 nm)	184	R^2_{CV} = 0.36	340
Olsen P ($mg \cdot 100g^{-1}$)	Vis-NIR (400 – 1700 nm)	189 | 37	R^2_{Val} = 0.81	339

1.3.6 Conclusion on chemometrics and machine learning methods: Soil P dynamics predictions

Reviewing Table 1-5 to Table 1-9, the performance of ML varies with soil conditions. IR spectroscopy combined with various chemometrics/ML techniques has effectively predicted soil carbon contents (e.g., TC, SOC, SOM), typically using PLSR (R^2 ranging from 0.73 with Vis–NIR for subsoil SOC to 0.98 with MIR for SOM, median R^2 = 0.92), generally outperforming Cubist (median R^2 = 0.88), RF (0.80), and SVM (0.83). Similarly to soil nitrogen, also a primary IR-active property, PLSR effectively predicted TN (median R^2 = 0.87), resulting in better performance than other ML techniques such as RF (0.62) and SVM (0.77) in this review. For secondary soil properties such as clay content and soil pH, IR spectroscopy has also shown good predictive performance. However, PLSR generally performed less well than other ML techniques, with median R^2 = 0.79 for clay and 0.73 for pH, compared to Cubist (median R^2 = 0.84), RF (0.85), and SVM (0.83) for clay, and RF (median R^2 = 0.77) and SVM (0.72) for soil pH. Moreover, PLSR, the most widely used multivariate method for calibrating soil spectral data, is often used as a benchmark for evaluating the efficiency of more advanced nonlinear ML techniques.

For soil P dynamics assessed by IR spectroscopy, prediction accuracy for TP is more variable and generally less reliable. In this review, the median R^2 values for PLSR, Cubist, and SVM were 0.60, 0.66, and 0.18, respectively. Notably, SVM showed highly variable performance, with R^2 ranging from 0.08 to 0.76. Fewer studies have investigated soil P sorption capacity and most used PLSR. For example, Dunne et al. (2020)[147] employed PLSR to predict Langmuir Smax from MIR spectroscopy, but concluded that the method was suitable only for rough screening (R^2_{Val} = 0.67). However, Saidi et al. (2025)[176] found that using a non-linear model (artificial neural networks) improved Smax prediction from Vis–NIR spectra, increasing R^2_{Val} from 0.48 (PLSR) to 0.65. Consequently, it can be hypothesised that advanced ML techniques could also enhance MIR-based Smax prediction.

1.4 Conclusions

Sustainable soil P management is significant for balancing food security and environmental protection. Regular analysis (e.g., at least every five years) of soils with long-term P fertilisation is important for managing potential P accumulation. Soil P sorption behaviour varies according to soil's inherent physical and chemical properties. Although previous studies have proposed integrating soil P sorption capacity with STP levels to improve P management[62,109], soil P sorption behaviour groups (e.g., classified based on Smax) have not yet been established for Irish agricultural soils.

Using IR spectroscopy in tandem with chemometrics/machine learning techniques offers a faster and more cost-effective approach to support farmers in nutrient management and soil monitoring. Recent studies in a range of soil types (e.g., agricultural[197], tropical weathered[192], wetland[341], and forest soils[342]) have shown that IR spectroscopy predicts P sorption indicators (e.g. PBI, PSI) with moderate to high success. For Smax, an important traditional P sorption metric, relatively few studies have validated its prediction. In Dunne (2021)[343], Smax was predicted using benchtop MIR spectroscopy under laboratory conditions (samples dried at 40 °C and ball-milled to <0.25 mm), but with the conclusion that the method was suitable only for rough screening (using PLSR achieved R^2_{Val} = 0.67).

Moreover, current spectroscopy studies for soil P sorption are mostly conducted under laboratory assays, e.g., using benchtop devices (stable signal output, consistent power supply, and high spectral resolution) to scan dried and finely ground samples. The high cost of establishing a spectroscopy laboratory and the reliance on laboratory-based calibration limit the full utilisation of handheld spectroscopy with high temporal and spatial resolution. To the best of our knowledge, although Forrester et al. (2015)[197] and Saidi et al. (2025)[176] used handheld MIR and Vis-NIR spectrometers to predict PBI (R^2_{CV} = 0.82) and Smax (R^2_{V} = 0.65) from <2 mm dried samples, suggesting the feasibility of in situ analysis, the measurements were conducted under laboratory conditions. Considering the interference caused by soil sample heterogeneity and soil water content, especially for MIR, few

studies have tested the use of handheld NIR/MIR spectroscopy for on-site assessment of soil P sorption capacity.

It is clear that while handheld technologies have the potential to monitor P sorption capacity in the field, this has not yet been effectively demonstrated in field conditions. Therefore, considering the gaps identified above, the aim of this study was to transfer IR spectroscopy from the laboratory to the field for rapid and relatively cost-effective on-site prediction of P sorption capacity, with reference to a refined Soil Index System for P management that accounts for soil P sorption behaviour. In detail, this study aimed to classify the P sorption behaviour of Irish agricultural soils to refine the current Irish agronomic Soil Index System for long-term sustainable P management. Different chemometrics and machine learning techniques were applied, and the best-performing models for predicting Smax were selected to develop benchtop MIR spectroscopy as an accurate and reliable alternative to the wet chemical method for determining soil P sorption capacity. Although laboratory-based spectroscopy was used, this study further employed soil samples exhibiting heterogeneity and water interference with the aim of developing calibration models to support reliable handheld MIR/NIR on-site prediction of soil P sorption behaviour.

1.5 Aims and Objectives

This research addressed the limitations of the current agronomic Soil Index System for sustainable P management, focusing on the development of IR spectroscopy to enable on-site, real-time assessment of soil P sorption capacities in the future. The main objectives were (i) to improve the prediction accuracy of Smax using benchtop spectroscopy, with the potential to serve as an alternative to the conventional wet chemical method; (ii) to progress from laboratory-based spectroscopy to handheld on-site testing of P sorption capacity, with reference to a refined Soil Index System that integrates soil P sorption characteristics to support agronomic decision-making for long-term, site-specific sustainable P management.

By integrating P sorption behaviour into the agronomic Soil Index System; quantifying it using IR spectroscopic methods for rapid and accurate assessment; and progressing towards field analysis using mathematical techniques to account for soil heterogeneity and moisture interference, it is hoped that this study can be incorporated into agronomic guidelines and applied in practice to optimise productivity while improving environmental sustainability.

The first stage of this work identified a change-point in sorption behaviour (e.g., Langmuir Smax) across Irish agricultural soils that allowed Irish soils to be classified according to their soil P sorption behaviour (Smax) and integrating it with soil P testing, this work aimed to contribute to the development of a refined agronomic Soil Index System to support site-specific and long-term sustainable P management in agricultural ecosystems.

The second stage of this work aimed to use IR spectroscopy as an alternative method for quantitative/qualitative prediction of soil P sorption behaviour, enabling efficient assessment of Langmuir P sorption capacity for practical applications. Building on the work of Dunne (2021)[343], which used benchtop MIR spectroscopy in tandem with PLSR to achieve a rough screening for Smax prediction, this thesis analysed various chemometrics/(non-linear) machine learning techniques (e.g., RF, Cubist, SVM) to improve the accuracy of soil P sorption capacity predictions for more precise

quantitative and qualitative assessments. Recognising the influence of sample heterogeneity and moisture on spectral quality and model performance during the transition from laboratory to field conditions[116,274], this study assessed the impact of sample heterogeneity using unground samples, and evaluated moisture interference using moist soil spectra from a controlled re-wetting experiment. These efforts aimed to calibrate (handheld) IR spectroscopy as a rapid, cost-effective alternative to traditional isothermal extractions to support site-specific and sustainable soil P management. The main sub-objectives included (i) improving prediction accuracy using benchtop/handheld MIR spectroscopy by exploring the best-performing chemometrics/machine learning algorithms (e.g., PLSR, RF, Cubist, SVM), (ii) evaluating simplified sample preparation steps (e.g., excluding ball milling) for MIR spectroscopy, and (iii) calibrating EPO models to mitigate the effects of soil moisture. Step by step, this work has progressed toward the calibration of highly accurate and robust soil spectroscopy models, accounting for soil heterogeneity and moisture variability, and ultimately supporting the future use of handheld IR devices for in situ soil P sorption analysis.

The technical roadmap of this research is illustrated in Figure 1-7, summarising the identified gaps, main research questions, hypotheses, objectives, and key deliverables aimed at progressing from laboratory-based spectroscopy to on-site soil testing for sustainable P management.

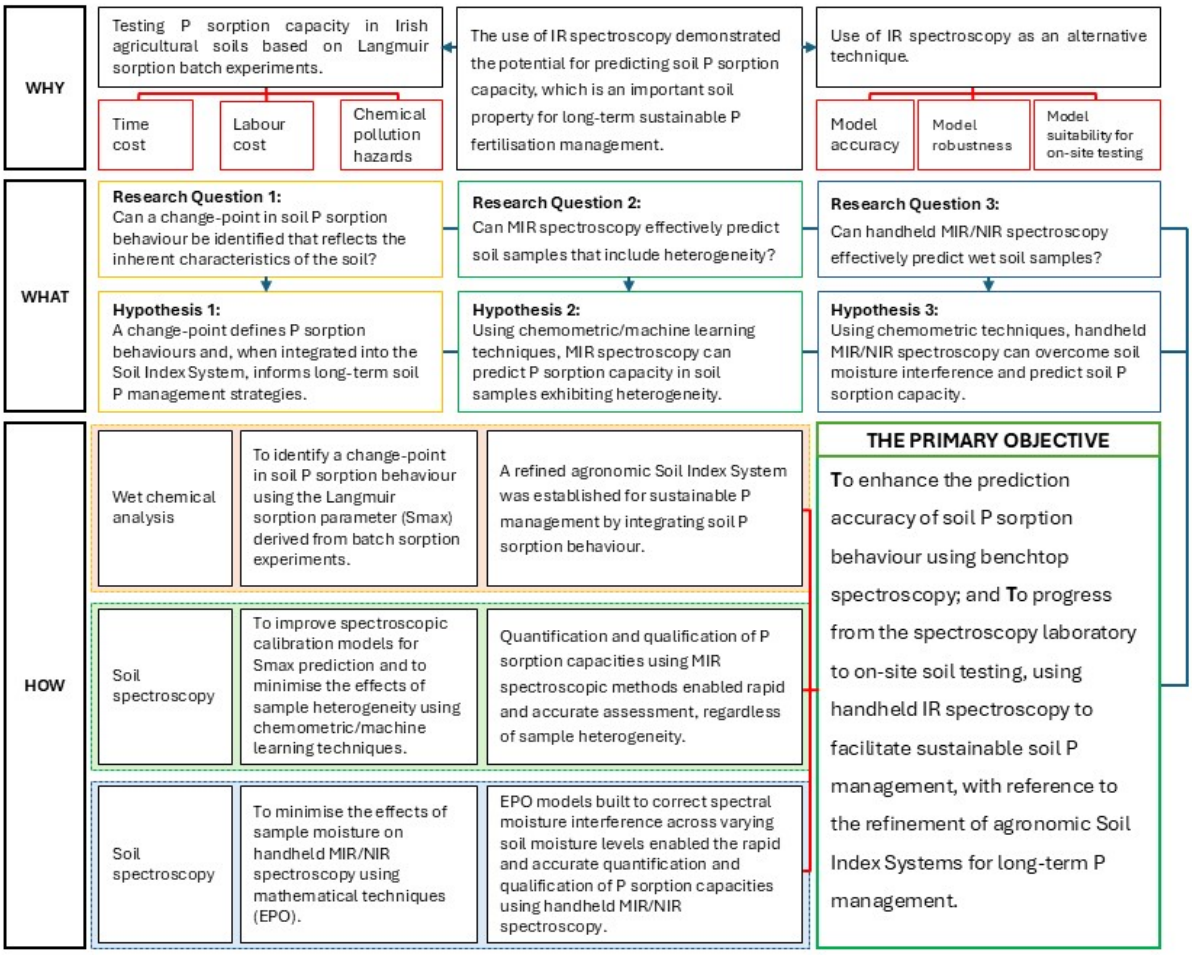


Figure 1-7. Technical roadmap for the research project.

1.6 Thesis Structure

Figure 1-8 illustrates the thesis roadmap. Chapter 1 introduces the background to this work, including the current knowledge on soil P sorption dynamics, soil spectroscopy, and chemometrics/machine learning modelling, and the research objectives and aims of the thesis. Chapters 2, 3 and 4 are the experimental chapters. Chapter 5 delivers conclusions from the overall work and highlights future research questions. The aims and main conclusions from each of the experimental chapters are:

Chapter 2 examined soil P dynamics using the standard batch sorption technique. Soil P sorption behaviour was specifically characterised using the Langmuir-derived sorption maximum capacity (Smax, $mg \cdot kg^{-1}$). Quantitative methods, including principal component analysis and segmented regression, were used to identify the change-point in Smax as an indicator of inherent soil properties. Soils with Smax values smaller or larger than the identified change-point were categorised as low- and high-sorbing soils, with expectations with low sorption capacity being prone to P losses, whereas with high sorption capacity tend to accumulate legacy-P/labile-P. By integrating these P dynamics, the Soil Index System was refined through the inclusion of P sorption classes, thereby enabling more targeted and sustainable site-specific P management.

Chapter 3 established MIR spectral libraries for ball-milled and <2 mm soil samples scanned with benchtop and handheld spectrometers. A comprehensive workflow for interpreting soil spectroscopy was summarised, incorporating spectral pre-processing, outlier detection, and testing of the best-performing chemometrics/machine learning models. As expected, based on prior knowledge of spectrometer resolution and instrument specifications, the most accurate predictions for homogenised soil samples were achieved using the benchtop MIR spectrometer. However, comparison of results from ball-milled and sieved-only samples scanned by the benchtop spectrometer revealed similarly high accuracy in Smax prediction, meaningfully suggesting the potential to omit the grinding step in sample preparation. In addition, a significant finding was that

the handheld MIR instrument suggested the potential to discriminate soil P sorption classes from sieve-only dried soils.

Chapter 4 further explored the application of the handheld MIR device for scanning moist soils. Building on the understanding that moisture has a less significant impact on NIR spectra, the handheld NIR spectrometer was also employed to develop models for predicting soil P sorption. External parameter orthogonalisation, a commonly used technique to minimise moisture-related spectral distortions, was applied to the wet spectra. However, this study further explored the efficiency of EPO in correcting for moisture effects, both with and without incorporating explicit moisture information, to identify a more streamlined and effective correction approach. An encouraging finding was that, under low soil moisture conditions (e.g., gravimetric water content [GMC] ≤ 20%) with known moisture information, handheld MIR spectroscopy showed potential for predicting Smax following EPO correction. Furthermore, under moderate soil moisture levels (e.g., GMC ≤ 40%), both handheld MIR and NIR spectroscopy demonstrated the ability to discriminate soil P sorption behaviour when moisture information was available.

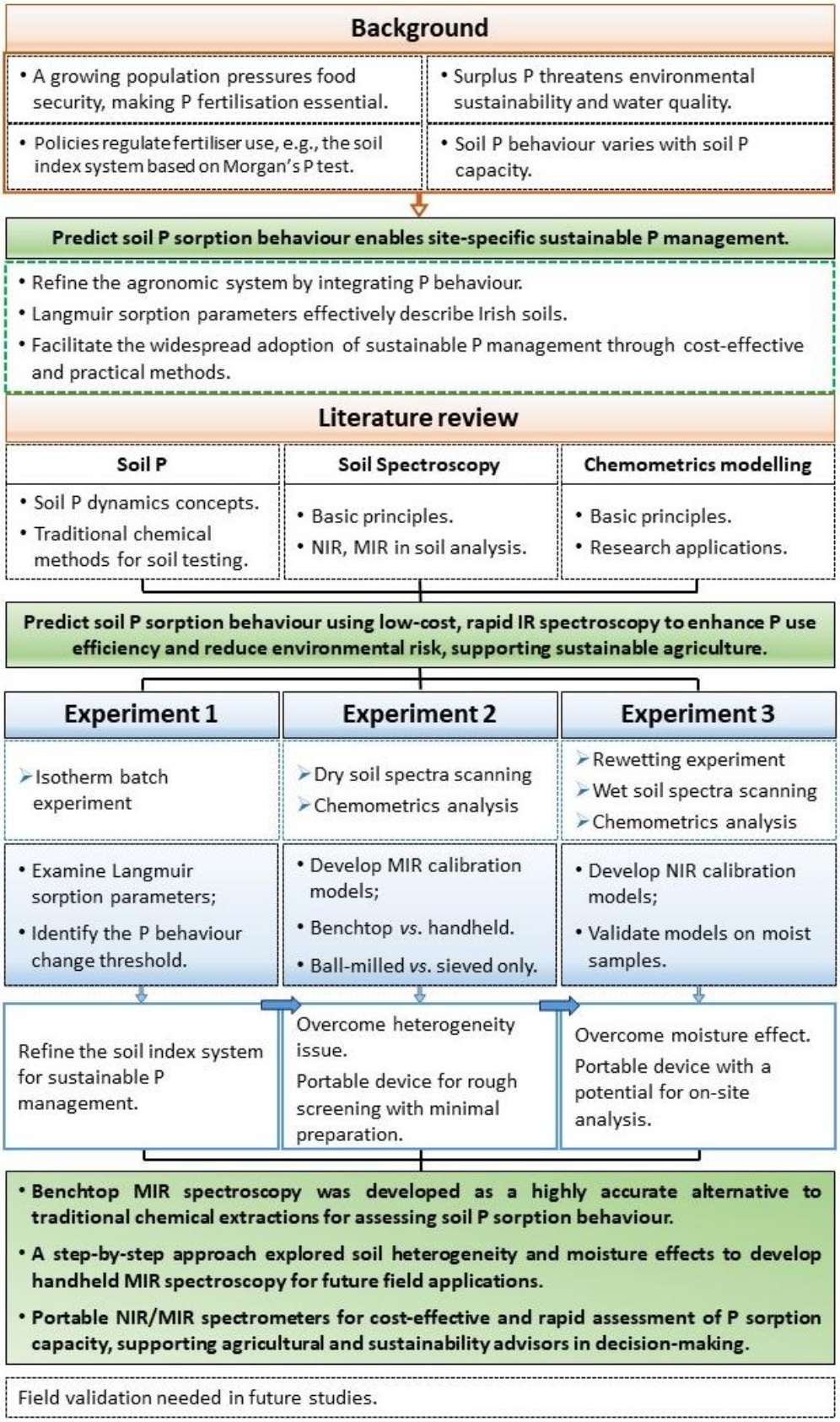


Figure 1-8. Thesis structure summary.

Chapter 2

Integrating soil phosphorus sorption capacity with agronomic indices to improve sustainable phosphorus use in agriculture[1]

[1] Part of this chapter has been accepted for publication. Manuscript details:
Yang, S., White, B., Regan, F., Kent, N., Hall, R.L., Daly, K., 2024. Integrating soil phosphorus sorption capacity with agronomic indices to improve sustainable P use in agriculture. *J. Soils Sediments*.
https://doi.org/10.1007/s11368-024-03900-z

Abstract

Purpose: Phosphorus (P) sorption processes in soils can influence P plant-availability and influence 'build-up' and 'drawdown' P cycles. Current fertiliser recommendations do not take these processes into account. This study aimed to integrate soil P sorption behaviour and P agronomic-indices to strengthen P management recommendations.

Methods: Mineral soil covering 35,716 km^2 of Ireland was characterised by P status (Morgan's P), and Langmuir sorption parameters of P sorption maximum capacity (Smax, $mg \cdot kg^{-1}$) and binding affinity (k, $L \cdot mg^{-1}$).

Results: A segmented regression between Smax and Mehlich-3 extractable aluminium (M3-Al) (R^2 = 0.49) identified a significant change-point at Smax = 450.03 $mg \cdot kg^{-1}$, allowing soils to be classified into 'low' ($SL_{M3\text{-}Al}$) and 'high' ($SH_{M3\text{-}Al}$) P sorbing classes. Sorption parameters in $SL_{M3\text{-}Al}$ did not change with soil P status; however, in $SH_{M3\text{-}Al}$ soils, sorption parameters significantly correlated with P status. High sorbing soils that are P-deficient (Index 1 and 2) will 'fix' P and take longer to build-up plant available P to a value for agronomic production (Index 3). Low P sorbing soils at high P status (Index 4) will decline to Index 3 at faster rates than high P sorbing soils. These soils ($SL_{M3\text{-}Al}$) are at higher risk of soluble P losses to water because of lower binding energies.

Conclusion: Efficient P fertiliser use can be more effective if soils are delineated into 'low' and 'high' sorbing soils coupled with soil P status. By integrating P sorption capacities with agronomic soil P indices, fertiliser advice and water quality measures can be targeted and more effective.

Graphical abstract

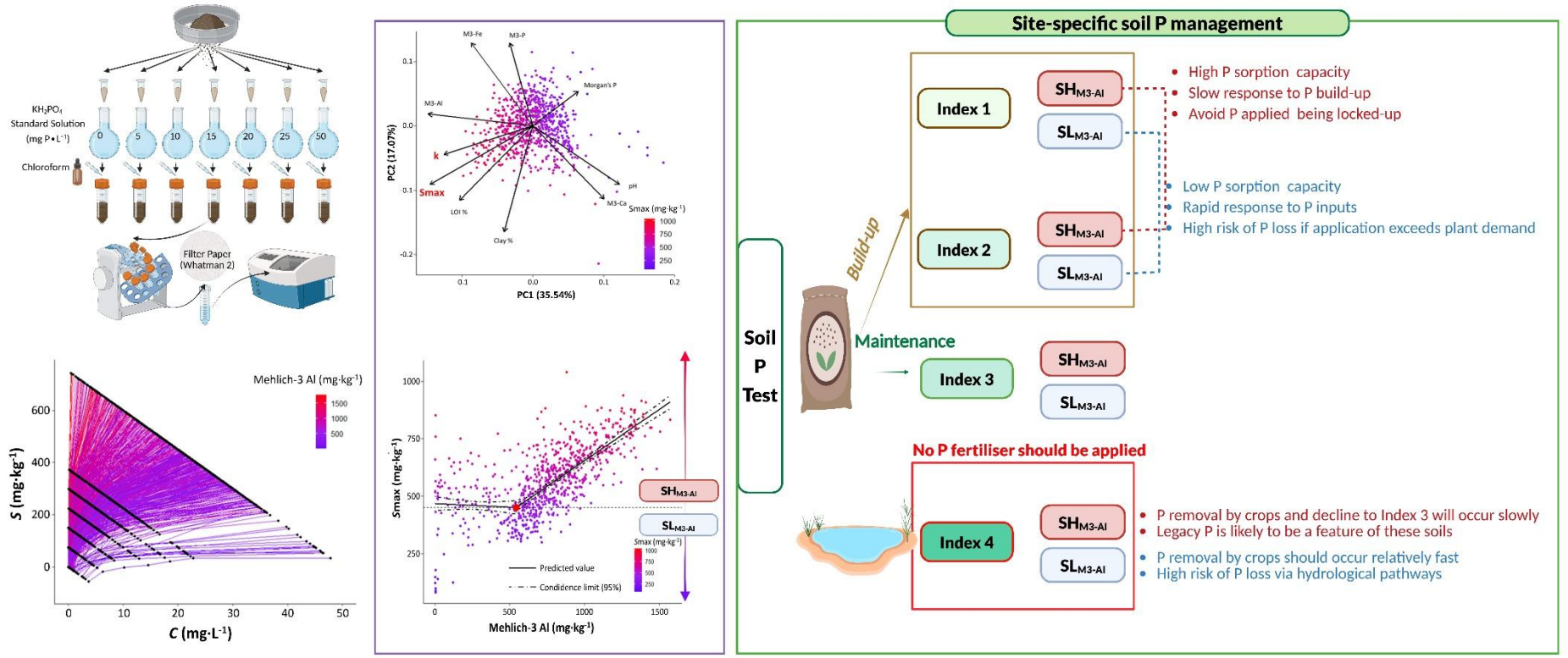


Figure 2-1. Graphical abstract for Chapter 2.

2.1 Introduction

Phosphorus (P) is an essential primary nutrient for various physiological and biochemical functions in plants, significantly impacting crop health and yields in agriculture production[344]. The concentration of total P ranges in soils from 220 to 6800 kg·P·ha^{-1}, in which less than 1 % of orthophosphates (e.g., $H_2PO_4^-$ and HPO_4^{2-}) are directly available to plants[5]. To optimise primary production and sustain today's agricultural consumption, the application of soluble inorganic P fertilisers (e.g., monoammonium phosphate, diammonium phosphate, *etc*.) rapidly replaces P removed in products such as meat, milk and crops[10]. Continuous excessive fertilisation leads to high soil P accumulation, causing diffuse environmental losses[345]. Legacy-P, defined as the accumulated P associated with long-term P fertiliser applications[30], is sorbed to soil clay minerals or stored in secondary P minerals[345]. Furthermore, legacy-P includes labile-P, which is held by the soil and is likely to undergo a change of state to become available[31]. Agricultural soils with high legacy-P/labile-P have reduced buffering capacity, and newly applied P may be more susceptible to surface runoff.

Soil testing for P is designed to provide information on plant-available P in the soil. It has been widely adopted in agronomic decision-making to establish response thresholds for effective fertiliser management[346]. However, there is a need to better understand soil P behaviour, specifically sorption processes in which P is adsorbed from the solution phase and bound to soil particles[69,113]. The rate and quantity of P sorption can be influenced by soil status (e.g., the concentration of plant-available P in soil solution) and soil characteristics or types (e.g., clay content, free iron [Fe] and aluminium [Al] oxides and hydroxides, and pH, *etc*.)[18,62,69]. In the soil-solution-plant system, soil organic matter (SOM) can also influence P sorption via various (abiotic and biotic) mechanisms, i.e., competitive binding to positively charged sites on clays and metal oxides, phosphatase activity, *etc*.[347,348].

Various P sorption indicators have been investigated to describe soil P dynamics, including P buffering capacity (PBC)[17,110], P buffering index (PBI)[109], P sorption capacity (PSC)[104,349], degree of P saturation (DPS) index[350,351], *etc*. Daly et al. (2015)[18] and Dunne et al. (2021)[90] found that P sorption maximum

capacity (Smax, mg·kg^{-1}) and binding affinity (k, L·mg^{-1}) determined using the Langmuir isotherm equation can describe the mechanisms of P sorption dynamics for a range of Irish grassland soils derived from different parent materials.

In Ireland, current recommendations for fertiliser P use are based on soil P indices based on Morgan's P (Soil P Index System): 1 (deficient), 2 (low), 3 (agronomic optimum) and 4 (excessive), guiding 'build-up', 'maintenance' and 'draw-down' practices[5] (Table 2-1). However, the Soil Index system for P management does not take P sorption processes or other controlling factors into account[352]. For example, the labile-P fixed through the adsorption phenomenon with oxides of Fe and Al (of clay fraction) is not captured.

Table 2-1. The P Index system for agronomic advice depends on the amount of available P in soil, extracted by Morgan's solution (10% sodium acetate at pH 4.8)[353].

Soil P Index	Soil P Ranges (mg·L^{-1})		Index Description	Response to Fertilisers
	Grassland crops	Other crops		
1	0.0 - 3.0	0.0 - 3.0	Very Low	Definite
2	3.1 - 5.0	3.1 - 6.0	Low	Likely
3	5.1 - 8.0	6.1 - 10.0	Medium	Unlikely/Tenuous
4	Above 8.0	Above 10.0	Sufficient/Excess	None

The aim of this study was to improve decision support tools for fertiliser P management. To achieve this, soil P sorption behaviour was integrated with traditional soil P testing and agronomic indices to strengthen P management recommendations. Soil P behaviour was described by the Langmuir sorption parameters, P sorption maximum capacity (Smax, mg·kg^{-1}) and binding affinity (k, L·mg^{-1}) in a large population of Irish soils. Sorption values were examined alongside other attributes to derive classes of sorption behaviour that could be coupled with traditional agronomic P indices commonly used to guide fertiliser applications.

2.2 Materials and methods

2.2.1 Soil archive

Mineral topsoil samples (5-20cm depth, loss-on-ignition ≤ 20%) from agricultural lands used in this thesis are from the Terra Soil project (https://www.gsi.ie/en-ie/programmes-and-projects/tellus/projects/terra-soil/Pages/default.aspx) (Figure 2-2). It is a joint research effort between Geological Survey Ireland and Teagasc, with samples collected through the Tellus Programme (https://www.gsi.ie/en-ie/programmes-and-projects/tellus/Pages/default.aspx), funded by the Department of the Environment, Climate and Communications. Tellus is a national programme (2011 – 2024) that collects geochemical and geophysical data across the island of Ireland (https://arcg.is/i954j). A total of over 42,000 samples from 20,660 different locations were collected at a typical density of one site per 2 × 2 km (4 km²) in rural areas and at a density of one site per square kilometre (1 km²) in the peri-urban areas of Dublin and Galway. Five composite samples were collected per site within a 20 m² square grid[354]. Samples were dried at 30°C, disaggregated and sieved to 2 mm. Chemical (e.g., Mehlich-3 tested aluminium, calcium, cobalt, copper, iron, magnesium, potassium, and phosphorus; Morgan's tested potassium and phosphorus; cation-exchange-capacity; loss-on-ignition, etc.) and physical properties (e.g., particle size, bulk density) of samples were examined[355]. These datasets are available at GSI (2025a, 2025b)[356,357].

Non-peat samples were selected using a geospatially random approach in GIS, and a total of 984 samples remained in the archive for this study[358]. As the archived soil samples were also used for other projects, a total of 876 samples (<2mm) with more than 2g remaining were available for isothermal equilibrium tests. After excluding samples with from non-agricultural land uses, including rural settlements (e.g., golf course, parks, playing fields, and roads/verges), dunes, mixed woodlands, and forests, 769 samples remained. In this work, a total of 747 mineral soil samples (with organic matter content determined by loss-on-ignition ≤ 20%) showing a strong and reliable fit to the Langmuir

equation ($R^2 \geq 0.90$) were considered to have accurate and consistent sorption data, and their parameters were used for subsequent analyses.

All sample preparation and analysis were carried out in accordance with the Tellus programme prior to this study[355]. The existing soil library, summarised in Table 2-2, includes measurements of organic matter content (by loss-on-ignition)[113], soil pH (calcium chloride slurry method)[113], particle size distribution (clay <0.002 mm, silt 0.002–0.05 mm, sand 0.05–2 mm)[358], and soil texture class (UK-ADAS system)[358]. Elemental data comprise Morgan's P[113] and Mehlich-3 extractable phosphorus (M3-P)[113], aluminium (M3-Al), iron (M3-Fe), and calcium (M3-Ca). Total concentrations of phosphorus and aluminium were determined by inductively coupled plasma mass spectrometry (ICP-MS) following aqua regia digestion (ICPar-P and ICPar-Al)[113].

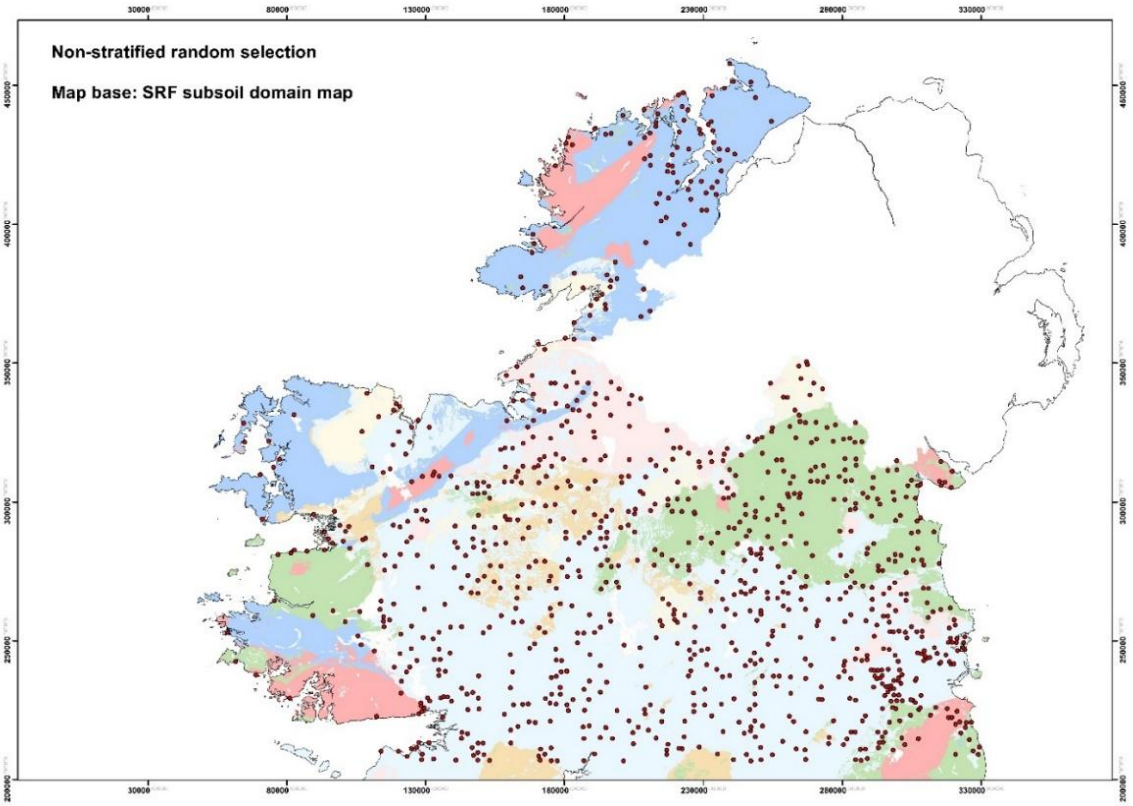


Figure 2-2. Map of the 1,000 samples randomly selected from Tellus Soil Project (10,000 samples) (non-peat sites).

Table 2-2. Summary statistics of general soil properties of the agricultural mineral soil samples (n = 747) archived from the Tellus programme[113,359].

Soil properties	Min	Max	Median	Mean
pH ($CaCl_2$)	3.50	8.15	5.60	5.69
LOI (%)	2.46	20.00	10.65	11.06
M3-Al ($mg \cdot kg^{-1}$)	0.17	1572.33	755.14	750.12
M3-Ca ($mg \cdot kg^{-1}$)	2.50	41609.22	1603.32	2162.72
M3-Fe ($mg \cdot kg^{-1}$)	2.50	718.95	322.27	324.61
M3-P ($mg \cdot kg^{-1}$)	0.97	340.81	46.85	56.10
Morgan's P ($mg \cdot L^{-1}$)	0.71	107.00	4.18	6.05
ICPar-P ($mg \cdot kg^{-1}$)	120.00	3700.00	840.00	892.36
Sand (w/w %)	10.00	99.00	37.00	38.52
Silt (w/w %)	0.00	65.00	38.00	37.65
Clay (w/w %)	1.00	53.00	24.00	23.83

2.2.2 P sorption isotherms

P sorption isotherm analysis was conducted using the standard batch technique by Nair et al. (1984)[360] and Pautler and Sims (2000)[361]. Electrolyte solution with calcium chloride ($CaCl_2$) concentration of 0.01 mol·L^{-1} was made by dissolving $CaCl_2$ in Milli-Q water. The standard monopotassium orthophosphate stock solution (500 mg·L^{-1} KH_2PO_4) was prepared in a volumetric flask by dissolving KH_2PO_4 in the electrolyte solution. The phosphorus working solutions (0, 5, 10, 15, 20, 25, 50 mg P·L^{-1}) was prepared by direct dilution in volume flasks (1.0 L) of the P stock solution with electrolyte solution and two drops of Chloroform (to inhibit microbial activity). 30 mL of phosphorus working solution and 2.0 g (± 0.1 g) soil were placed into a 50-ml centrifuge tube and shaken in an end-over-end shaker (Heidolph Reax overhead shakers - REAX 20/12) at 4 rpm for 24 hours at room temperature. The soil suspension was filtered through Whatman 2 filter papers. The filtrate was further colourimetrically analysed by (Gallery™ Plus) (Figure 2-3).

The total amount of P sorbed (*S*, mg kg^{-1}) was calculated by the difference of P concentration between the initial solution (C_0, mg·L^{-1}) and in the filtrate (C_e, mg·L^{-1}), as described in the following (Equation 2-1. Langmuir isotherm model was used to describe P dynamics, P sorption maximum capacity (Smax, mg·kg^{-1}) and binding affinity (*k*, L·kg^{-1}) with the following (Equation 2-2).

$$\mathrm{S} = (\mathrm{C}_0 - \mathrm{Ce}) \times \frac{\mathrm{V}}{\mathrm{m}}, \qquad \text{(Equation 2-1)}$$

where *V* (L) is the volume of solution, and *m* (kg) is the mass of soil. C_0 (mg·L^{-1}) is the initial concentration of the P stock solution. C_e (mg·L^{-1}) is the concentration of P in the solution after 24 hours of equilibration on the end-over-shaker.

$$\frac{\mathrm{Ce}}{\mathrm{S}} = \frac{1}{\mathrm{kSmax}} + \frac{\mathrm{Ce}}{\mathrm{Smax}}, \qquad \text{(Equation 2-2)}$$

where *Smax* (mg·kg^{-1}) is the maximum sorption of P, and *k* is a binding affinity constant (L·kg^{-1}). The linear form of the Langmuir equation describes a linear relationship between $\frac{Ce}{S}$ and C_e. From this relationship, the slope ($\frac{1}{\mathrm{Smax}}$) is used to calculate the maximum P sorption capacity (Smax = $\frac{1}{\mathrm{slope}}$),

while the intercept ($\frac{1}{k \cdot Smax}$) is used to estimate the P binding affinity constant ($k = \frac{1}{Intercept \cdot Smax} = \frac{Slope}{Intercept}$).

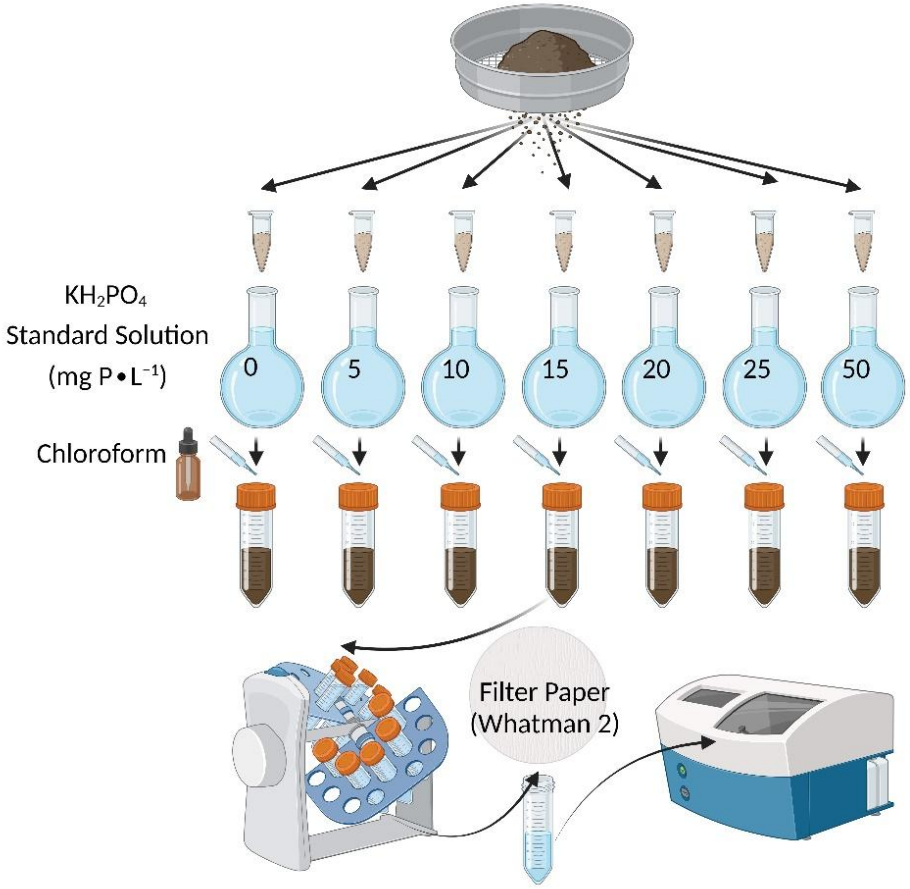


Figure 2-3. Schematic diagram of isothermal batch extraction steps.

Analytical replicates were not performed for the sorption isotherm experiments primarily due to the large number of soil samples, the shared use of laboratory facilities, and the broad concentration range (0 to 50 mg·P·L^{-1}) included in the experimental design. The concentration range and analytical procedures employed have been previously validated in replicated studies, confirming the reproducibility and robustness of the method. While analytical replication is valuable for quantifying random variability and detecting potential systematic errors, the present study aimed to extend the application of this validated approach across a large and diverse soil dataset. In future work, incorporating analytical replicates where feasible would be valuable for further assessing measurement uncertainty and improving methodological comparability.

2.2.3 Statistical analysis

All statistical analyses and figures drawn (ggplot2 package) were conducted using R version 4.2.1 (R Core Team, 2022). Spearman's rank correlation coefficients (r_s-values) (package Hmisc in R) and Principal Component Analysis (PCA) (package factoextra in R) were used for the dimensionality reduction of the original dataset[362]. A two-piece segmented regression examined the relationships between Smax and the M3-Al (package segmented in R)[363,364]. The coefficient of determination (R^2) was used to evaluate the goodness of fit of the segmented regression across the full dataset. The Wilcoxon rank-sum/Mann-Whitney test was used for the P dynamics classification comparisons[365].

2.3 Results

2.3.1 Descriptive statistics for soil data

All samples (n = 747) were classed as mineral agricultural soils (LOI% ranged from 2.46 to 20.00 %). A 'total' P concentration for soil ICPar-P ranged from 120 to 3700 $mg \cdot kg^{-1}$. Plant-available P tested by Morgan's solution resulted in a smaller range of 0.71 to 107 $mg \cdot L^{-1}$. According to the soil P Index based on Morgan's extractable P, 28%, 33%, 20% and 19% of the soil samples were classified as P Index 1, Index 2, Index 3 and Index 4, respectively. Soil tests for plant-available P were measured using the Mehlich-3 extractant, a stronger extraction reagent than the Morgan's P solution. The extractable M3-P test ranged from 0.97 to 340.81 $mg \cdot kg^{-1}$.

2.3.2 Isotherm results

The linear version of the Langmuir model provided a fit for the soils in this study ($R^2 \geq 0.90$). The calculated Langmuir P sorption parameters were described following soil P indices (Table 2-3, Figure 2-4). The Langmuir parameter of P sorption maximum capacity, Smax, ranged from 81.31 to 1041.30 $mg \cdot kg^{-1}$ with a median value at 564.54 $mg \cdot kg^{-1}$; and binding affinity, k, ranged from 0.02 to 2.49 $L \cdot mg^{-1}$ with a median value at 0.45 $L \cdot mg^{-1}$.

The relationship between the Langmuir parameters Smax and k is shown in Figure 2-5, where a higher maximum P sorption capacity generally corresponds to a greater P binding affinity. In this study, Smax was selected as the representative indicator of P sorption behaviour, as it reflects the overall sorption capacity and is less sensitive to experimental variability than the affinity constant k.

Table 2-3. Statistical summary of the Langmuir parameters (Smax, mg·kg^{-1} and k, L·mg^{-1}) according to the soil P indices based on Morgan's extractable P.

Soil P index	n	Langmuir parameters			
		Smax (mg·kg^{-1})		k (L·mg^{-1})	
		Range (Min - Max)	Median	Range (Min - Max)	Median
Index 1	210	346.88 - 1041.30	613.29	0.08 - 2.49	0.61
Index 2	247	288.37 - 940.03	555.19	0.11 - 2.46	0.45
Index 3	150	285.68 - 869.60	547.72	0.05 - 1.93	0.42
Index 4	140	81.32 - 884.49	501.20	0.02 - 1.42	0.34

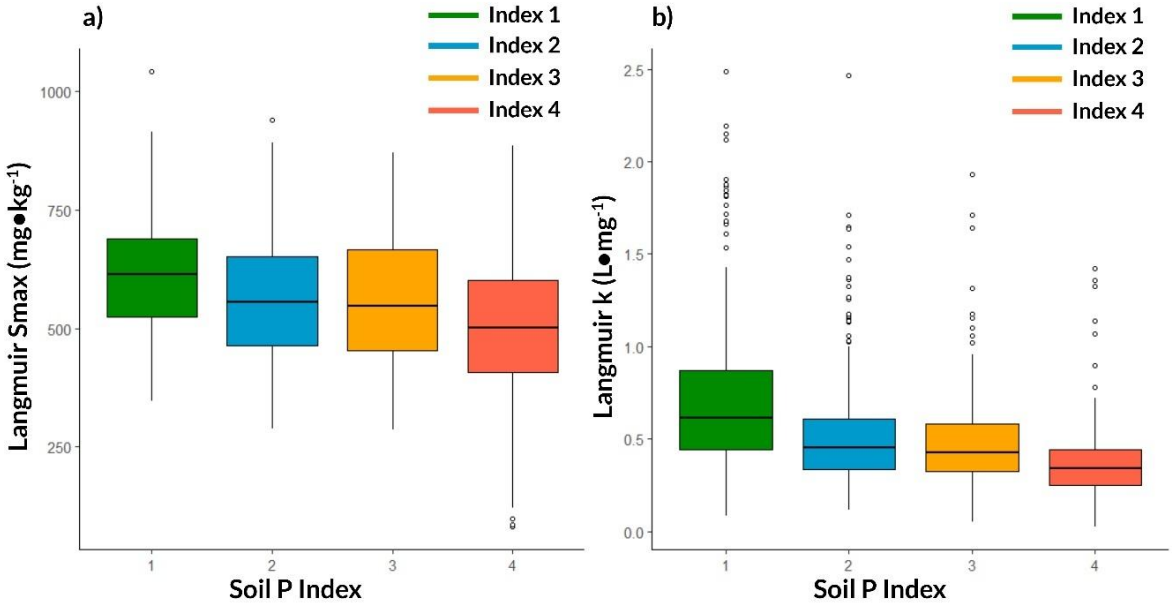


Figure 2-4. Comparison of Langmuir parameters among soils classified by soil P indices based on Morgan's extractable P.

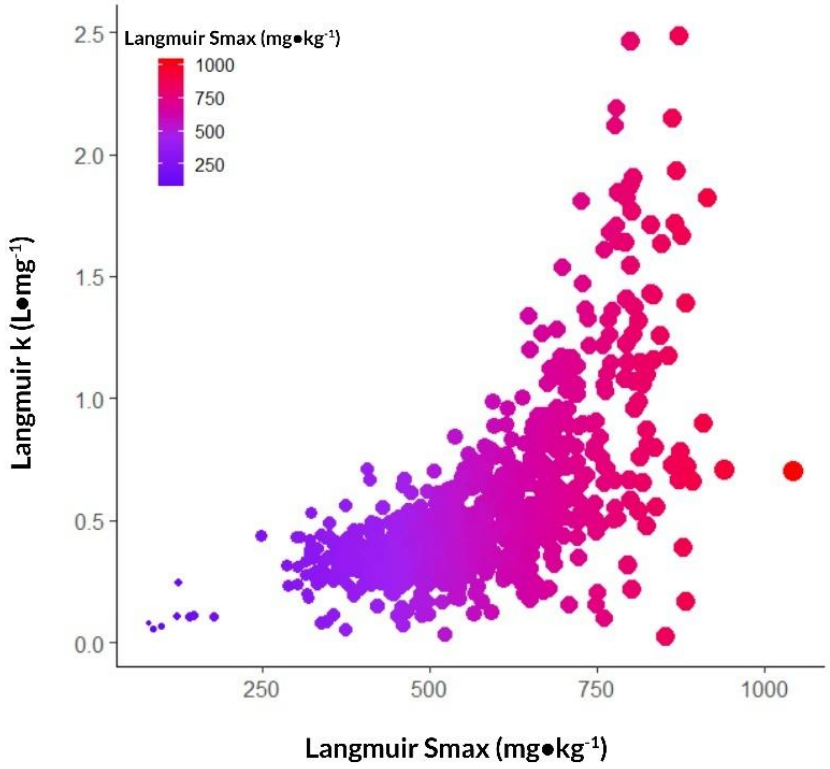


Figure 2-5. The relationship between Smax and k.

2.3.3 Reduction of the dimensionality of soil property information

Spearman's rank correlation coefficients (due to non-parametric) were used to select the soil attributes that significantly influence P sorption behaviour (Langmuir parameters of Smax and k-values) (Table 2-4). Morgan's extractable P values were inversely correlated with both of the Langmuir sorption parameters, Smax (r_s = −0.65, P < 0.05), k (r_s = −0.74, P < 0.01). Whereas M3-Al provided strong positive correlations with Smax (r_s = 0.85, P < 0.05) and k (r_s = 0.95, P < 0.001). The correlation scatter plots in Figure 2-6 further illustrate the relationships among the Langmuir parameters and the screened soil properties, showing that both parameters were strongly associated with soil organic matter and M3-Al, while inversely related to Morgan's extractable P.

Principal component analysis (PCA) was also applied to ascertain the dimensionality reduction of the original dataset. After Spearman's rank correlations and PCA, a biplot for the selected featured soil properties is presented in Figure 2-7.

Figure 2-8 depicts the raw P sorption isotherms according to the (a) soil P Index system and (b) M3-Al gradients. Visual assessment of Figure 2-8 (a), Index 1 soils (depicted as green isotherms) were predominantly distributed at the top, and Index 4 soils (depicted as red isotherms) were predominantly distributed at the bottom. In Figure 2-8 (b), higher M3-Al concentrations corresponded with higher Smax values.

Table 2-4**.** Spearman's rank correlation matrix of r_s-values significant at P = 0.05 derived from soil pH, % loss-on-ignition (LOI%), Mehlich-3 extractable P, Al, Fe, and Ca (mg·kg^{-1}), Morgan's P (mg·L^{-1}), ICPar-P (mg·kg^{-1}), total sand, silt, clay percentage (%), and soil P sorption parameters included Langmuir sorption maximum capacity (Smax, mg·kg^{-1}) and binding affinity (k, L·mg^{-1}). NS denoted non-significant correlation at P = 0.05.

	Smax	**k**	**pH**	**LOI**	**M3-Al**	**M3-Ca**	**M3-Fe**	**M3-P**	**Morgan's P**	**ICPar-P**	**Sand**	**Silt**	**Clay**
Smax	1												
k	0.91	1											
pH	-0.72	-0.89	1										
LOI	0.92	0.77	NS	1									
M3-Al	0.85	0.95	-0.91	0.65	1								
M3-Ca	-0.59	-0.74	0.81	NS	-0.88	1							
M3-Fe	0.79	0.80	-0.87	0.6	0.88	-0.82	1						
M3-P	NS	NS	NS	NS	NS	NS	NS	1					
Morgan's P	-0.65	-0.74	0.71	NS	-0.69	0.66	-0.57	NS	1				
ICPar-P	NS	NS	NS	0.56	NS	NS	NS	NS	NS	1			
Sand	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	1		
Silt	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	-0.86	1	
Clay	0.6	NS	NS	0.7	NS	NS	NS	NS	NS	NS	-0.90	0.64	1

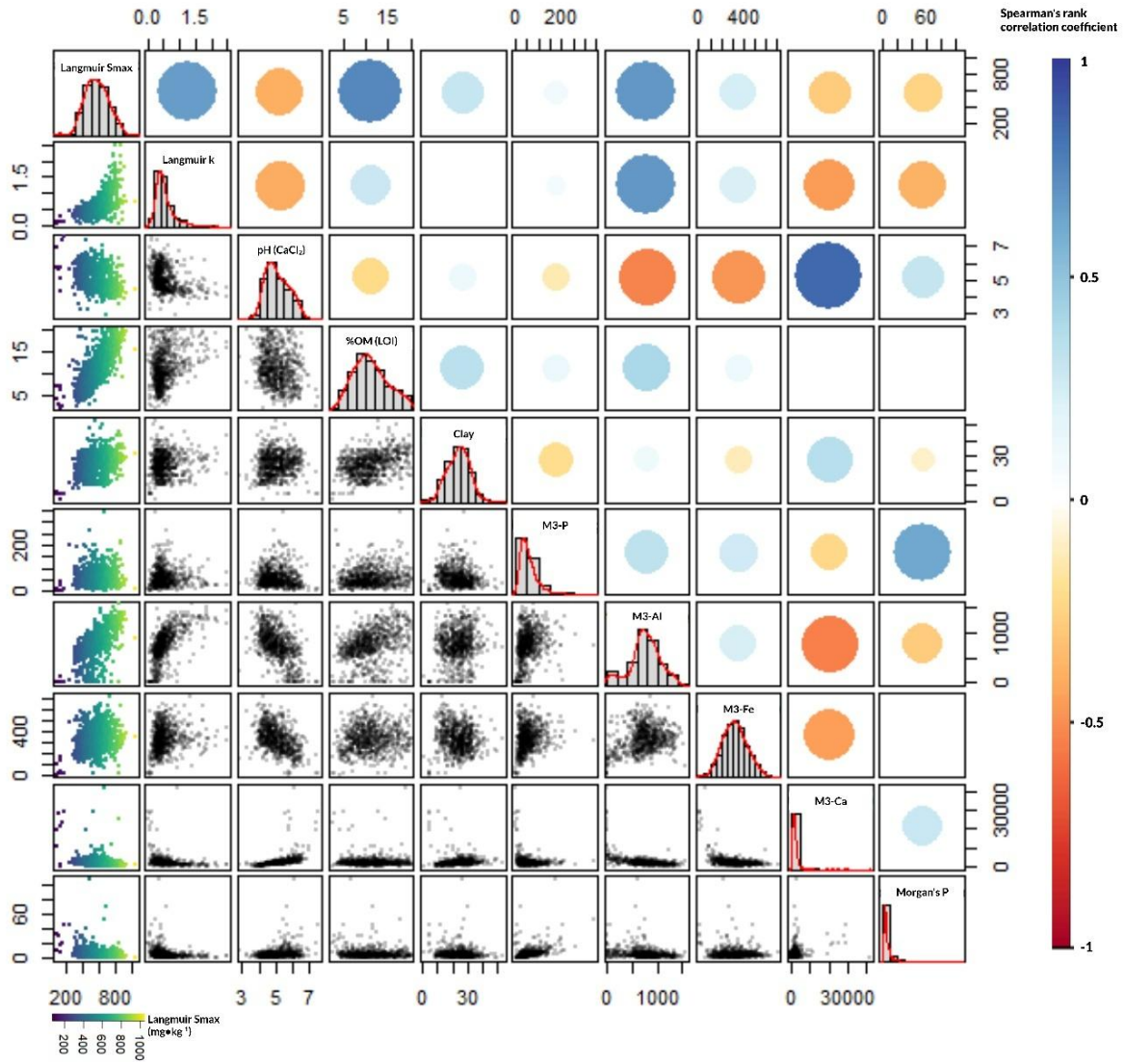


Figure 2-6. Correlation scatter plots summarising the relationships between the Langmuir parameters and the screened soil properties.

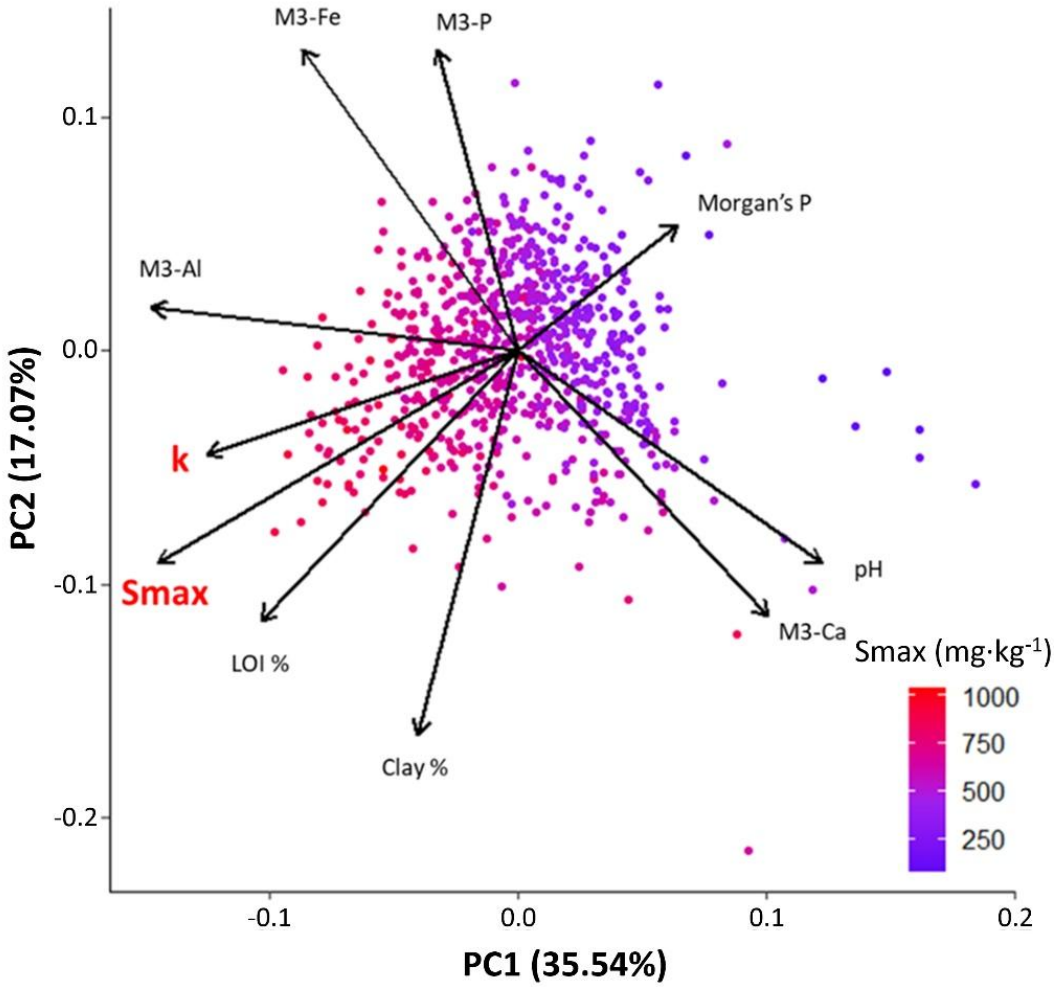


Figure 2-7. PCA biplot. Samples coloured by Langmuir Smax values (ranged from 81.32 to 104.30 $mg \cdot kg^{-1}$). The Smax values on the first principal component (PC1) tended to increase from left (PC1 < 0) to right (PC1 > 0). The angles between the vectors indicated how the soil properties correlated with each other, with a 90° angle indicating no correlation between two soil properties.

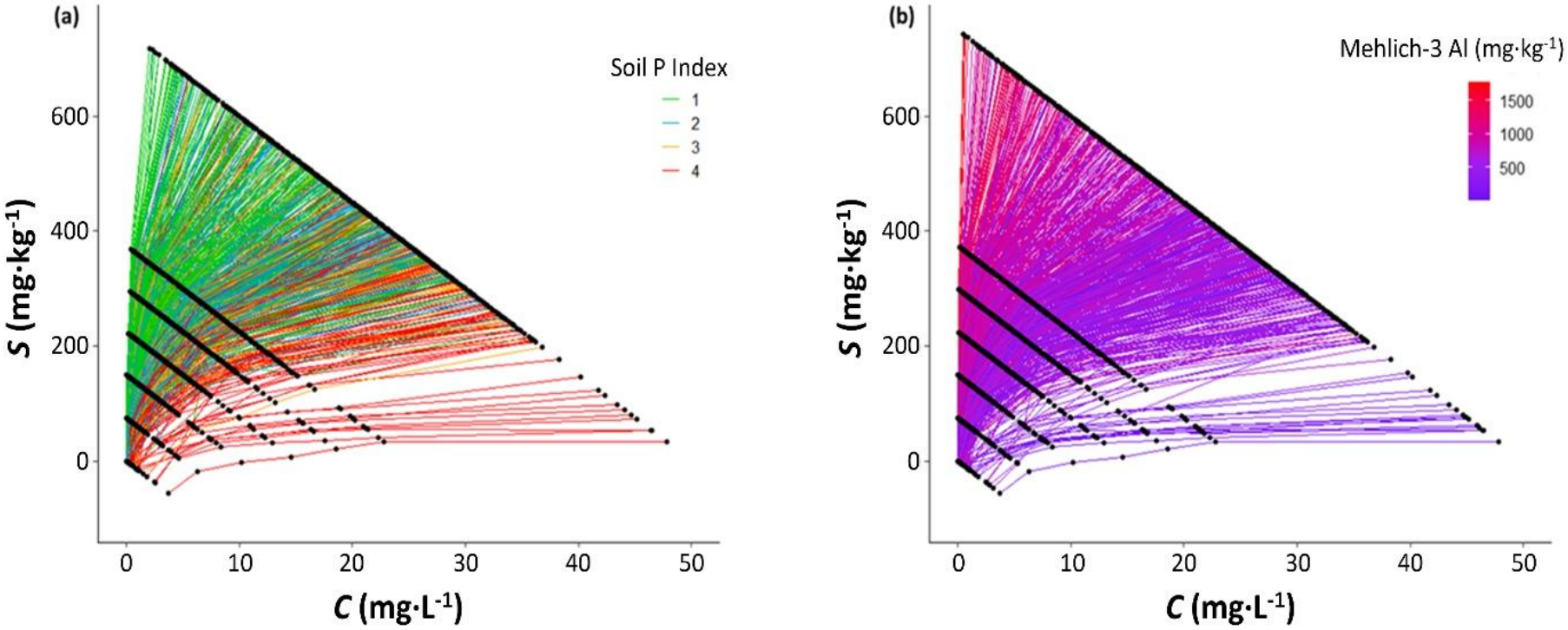


Figure 2-8. P sorption isotherms for the mineral samples from agricultural used lands depicted following (a) Morgan's P index system: Index 1 (green), Index 2 (blue), Index 3 (orange), and Index 4 (red) isotherms; (b) soil extractable Mehlich-3 extractable Al gradients. The parameter *C* (mg·L^{-1}) is the final concentration of P after equilibration, and *S* (mg·kg^{-1}) is the amount of P sorbed after equilibration.

2.3.4 Phosphorus sorption dynamics classification into 'low' and 'high' capacity

2.3.4.1 Principal Component Analysis

Principal component analysis, excluding Langmuir sorption parameters (Smax and k), was applied to the dataset to avoid overfitting. PC1 explained 32.97% of the variation in the data. When samples were coloured by a gradient of Smax values, although no clear groups for soil P sorption behaviour were clustered, PC1=0 roughly separated Smax into 'low' and 'high' ranges. To the left of PC1 (PC1 < 0), samples were more likely to have a high P sorption maximum, whereas to the right of PC1 (PC1 > 0), samples were more likely to perform at a low Smax values (Figure 2-9 [a]).

Figure 2-9 (b) presents a scatter plot of PC1 versus Smax. A linear relationship was observed. Results for the model, Smax = a + (b x PC1), are shown in Table 2-5. PC1 explained 35% of the variation in Smax and was negatively correlated. When PC1 = 0, the threshold value for P sorption behaviour, denoted as '$Smax_{T1}$', was identified at Smax = 566.94 (± 8.45) $mg \cdot kg^{-1}$. Soils with Smax values above $Smax_{T1}$ were classified as 'high' P sorption capacity (SH_{PC1}). Otherwise, soils with Smax < 566.94 $mg \cdot kg^{-1}$ were classified as 'low' P sorption capacity (SL_{PC1}). Median Smax and k values in SL_{PC1} soils were 464.34 $mg \cdot kg^{-1}$ and 0.36 $L \cdot mg^{-1}$, respectively. Median Smax and k values in SH_{PC1} soils were 669.84 $mg \cdot kg^{-1}$ and 0.63 $L \cdot mg^{-1}$, respectively (Table 2-6). Wilcoxon rank-sum tests (non-normality) were carried out to test for significant differences in medians of the P dynamics between the two P dynamics subgroups. P-values for Smax ($W = 0$, $P < 0.001$) and k ($W = 23532$, $P < 0.001$) between the SL_{PC1} and SH_{PC1} were significant.

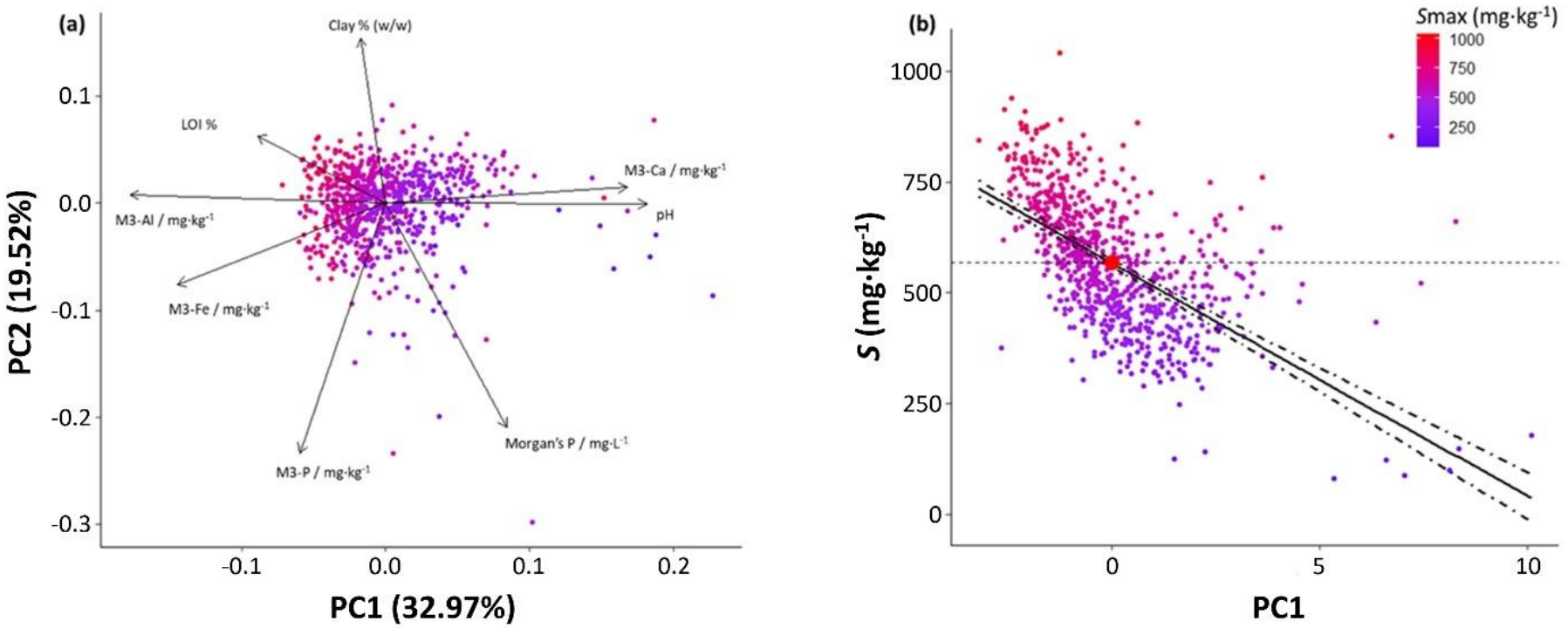


Figure 2-9. (a) Principal component 1 (PC1) explained 32.97% of the variation in the data. It was constructed from the following weighted soil attributes: PC1 = 0.5pH – 0.24LOI – 0.49M3-Al + 0.46M3-Ca – 0.4M3-Fe – 0.16M3-P + 0.23Morgan's P – 0.05Clay. (b) Relationship between PC1 and Langmuir P sorption maximum capacity (Smax) for mineral soil samples (R^2 = 0.35). When PC1 = 0, Smax = 566.94 (± 8.45) mg·kg^{-1}. Samples are coloured by Smax gradients.

Table 2-5**.** Parameter estimates derived from the (i) Smax against PC1 linear regression and (ii) Smax against Mehlich-3 soil extractable Al (M3-Al) segmented regression fitted to the mineral soil data from agricultural used lands (n =747). Thresholds ($Smax_{T1}$ and $Smax_{T2}$) for maximum P sorption (Smax, $mg \cdot kg^{-1}$) are highlighted in bold, below and above which P sorption behaviour differed significantly.

		Estimate	Standard error	t-value	*P*	95% confidence limit	
						Lower	Upper
	Intercept	566.96	4.30	131.75	<0.0001	558.52	575.41
Figure 2-5 [b]	Slope	-52.56	2.65	-19.82	<0.0001	-57.76	-47.35
	Value of Smax at PC1 = 0 ($Smax_{T1}$)	**566.96**	4.30	131.75	<0.0001	558.52	575.41
	Intercept	467.18	15.60	29.95	<0.0001	451.58	482.78
	Slope 1	-0.03	0.05	-0.69	0.49	-0.12	0.06
Figure 2-6	x = c	542.72	34.88	15.56	<0.0001	474.24	611.20
	Value of Smax at change-point ($Smax_{T2}$)	**450.03**	14.53	31.00	<0.0001	421.77	478.82
	Slope 2	0.48	0.02	22.54	<0.0001	0.41	0.49

Table 2-6**.** Summaries of the Langmuir sorption parameters for soils classified into 'low' and 'high' sorption dynamics according to $Smax_{T1}$ and $Smax_{T2}$, derived by PCA and segmented analysis, respectively.

P dynamics classification threshold	**P sorption subgroups**	**n**	**Langmuir parameters**					
			Smax (mg·kg^{-1})			**k (L·kg^{-1})**		
			Range (Min-Max)	Median	*P*-value	Range (Min-Max)	Median	*P*-value
$Smax_{T1}$ = 566.94 mg·kg^{-1}	SL_{PC1}	376	81.32 - 566.92	464.34	*W* = 0	0.04 - 0.84	0.36	*W* = 23532
	SH_{PC1}	371	567.01 - 1041.30	669.84	*P* < 0.001	0.02 - 2.49	0.63	*P* < 0.001
$Smax_{T2}$ = 450.03 mg·kg^{-1}	$SL_{M3\text{-}Al}$	154	81.32 - 447.96	387.02	*W* = 0	0.05 - 0.71	0.33	*W* = 19156
	$SH_{M3\text{-}Al}$	593	450.27 - 1041.30	603.79	*P* < 0.001	0.02 - 2.49	0.51	*P* < 0.001

2.3.4.2 Segmented regression analysis

M3-Al was the most dominant soil property influencing soil P sorption behaviour (Figure 2-7). A nonlinear relationship was observed in the scatter plotted M3-Al versus Smax. A broken-line regression model (package 'segmented' in R) was used to account for the piecewise linear relationship[363,364]. In Figure 2-10, a broken-line regression (R^2 = 0.49) was fitted to Smax and M3-Al, and a significant breakpoint (P < 0.001) was found at M3-Al concentration of 542.72 (± 34.88) mg·kg^{-1}, above which Smax in soil increased linearly with increasing M3-Al (slope 2 = 0.48, P < 0.001). Parameter estimates for this relationship are presented in Table 2-5.

M3-Al explained more variation in Smax compared to PC1. In accordance with the breakpoint for M3-Al acting on Smax, the threshold for P sorption behaviour classification was defined at Smax= 450.03 (± 14.53) mg·kg^{-1}, above which soils were classified as 'high' P sorption capacity ($SH_{M3\text{-}Al}$) soils, otherwise (Smax < 450.03 mg·kg^{-1}) soils were classified into 'low' P sorption capacity ($SL_{M3\text{-}Al}$) soils. Median Smax and k values of $SL_{M3\text{-}Al}$ were 387.02 mg·kg^{-1} and 0.33 L·mg^{-1}, respectively. Median Smax and k values of $SH_{M3\text{-}Al}$ were 603.79 mg·kg^{-1} and 0.51 L·mg^{-1}, respectively. Significant differences (P < 0.001) in P sorption behaviour were tested between $SL_{M3\text{-}Al}$ (n = 154) and $SH_{M3\text{-}Al}$ (n = 593) (Table 2-6).

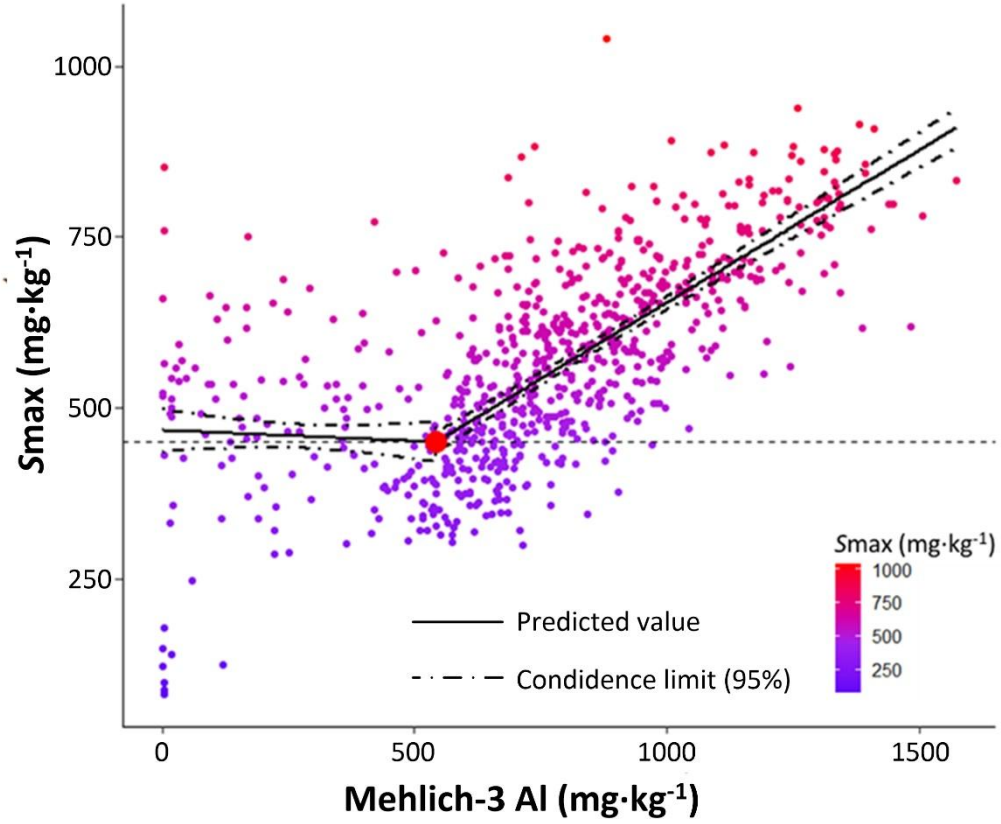


Figure 2-10. A broken-line regression was fitted between M3-Al (mg·kg^{-1}) and Langmuir maximum sorption capacity (Smax, mg·kg^{-1}) for mineral soils. Samples are colour-coded by Smax gradients. The red dot represents Smax behaving breakout, located at M3-Al = 542.72 mg·kg^{-1} and Smax = 450.03 mg·kg^{-1}. Soils with Smax < 450.03 mg·kg^{-1} were classified as $SL_{M3\text{-}Al}$ (n = 154), and soils with Smax ≥ 450.03 mg·kg^{-1} were classified as $SH_{M3\text{-}Al}$ (n = 593).

2.3.5 Coupling P sorption dynamics classes with the soil P indices

The P dynamics classes based on Smax = 450.03 mg·kg^{-1} (change-point derived from segmented regression with M3-Al) described above were significant in exploring soils with 'low' and 'high' P sorption capacity. Combining the P dynamics classification with the P Index System, Figure 2-11 presents the median values of (a) Smax and (b) k for $SL_{M3\text{-}Al}$ and $SH_{M3\text{-}Al}$ soils at each soil index. Large vertical gaps between $SH_{M3\text{-}Al}$ soils (red solid lines) and $SL_{M3\text{-}Al}$ soils (blue dashed lines) were observed, suggesting that the Langmuir sorption parameters (Smax and k) between $SL_{M3\text{-}Al}$ and $SH_{M3\text{-}Al}$ also differed significantly across soil P indices. Summary statistics of soil sorption capacity comparison between $SL_{M3\text{-}Al}$ and $SH_{M3\text{-}Al}$ soils for each soil P index is listed in Table 2-7.

The P saturation status affected P behaviour differently between the $SL_{M3\text{-}Al}$ and $SH_{M3\text{-}Al}$ soils (Figure 2-11). For the $SL_{M3\text{-}Al}$ soils, P sorption dynamics were less affected by soil P status; no statistically significant differences were found among different soil P indices. However, for the $SH_{M3\text{-}Al}$ soils, the Smax and k values were lower at higher P indices. Index 1 $SH_{M3\text{-}Al}$ soils had the highest P sorption capacity (median values for Smax = 621.76 mg·kg^{-1} and k = 0.63 L·mg^{-1}). Index 4 $SH_{M3\text{-}Al}$ soils had the lowest sorption capacity (median values for Smax = 572.71 mg·kg^{-1} and k = 0.37 L·mg^{-1}). For P-deficient $SH_{M3\text{-}Al}$ soils, Index 1 soils showed significantly higher Langmuir sorption parameters than Index 2 soils, with median Smax values of 621.39 vs. 593.89 mg·kg^{-1} and k values of 0.64 vs. 0.49 L·kg^{-1}.

A modified P Index system (Table 2-8) integrates P sorption capacity classifications with the original soil P Index to enhance the characterisation of soil P behaviour. A spatial distribution of the classified high- and low-sorbing soil samples based on their sampling locations is shown in Figure 2-12. While this map visualises the sampling coverage of sorption classes, it does not represent continuous spatial variation in soil P sorption capacity. To better understand field- and landscape-scale patterns, the application of GIS-based spatial interpolation or predictive modelling to generate continuous maps of P sorption capacity is recommended for future studies[345].

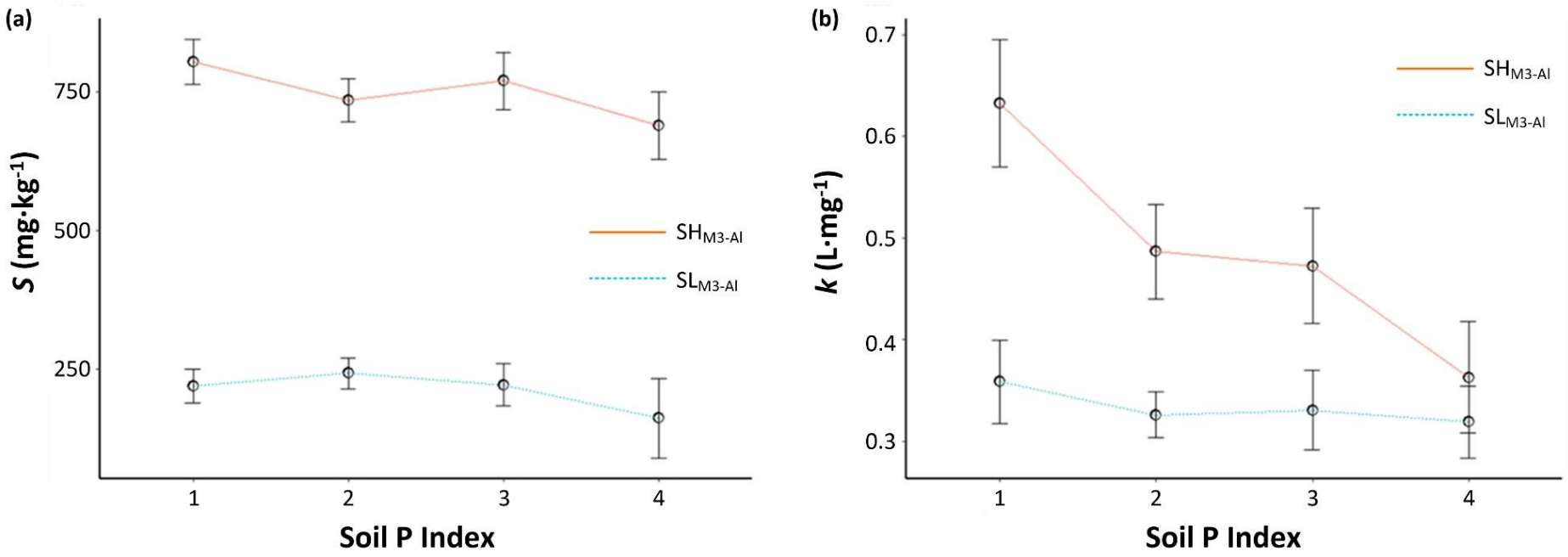


Figure 2-11. Median values of Langmuir parameters of (a) Smax and (b) k under the soil P Index system for $SL_{M3\text{-}Al}$ (blue dashed line, n = 154) and $SH_{M3\text{-}Al}$ (red solid line, n = 593) subgroups. Error bars represent 95% confidence intervals. Soil P sorption behaved significantly differently between the $SL_{M3\text{-}Al}$ and $SH_{M3\text{-}Al}$ subgroups at each Morgan's P index. In the horizontal comparison, no significant difference was observed in the Langmuir parameter Smax or k of the $SL_{M3\text{-}Al}$ soils under different Morgan P indices (blue dashed lines); while in the $SH_{M3\text{-}Al}$ soils, although indices 2 and 3 soils had the same Smax and k values, the Smax and k values of Index 1 soils were significantly higher than those of Index 4 soils.

Table 2-7. $SL_{M3\text{-}Al}$ and $SH_{M3\text{-}Al}$ in accordance with soil P indices based on Morgan's extractable P.

Soil P index	n	Langmuir parameters		Soil pH	LOI (%)	Clay (w/w %)	Morgan's P	Mehlich-3 information			
		Smax ($mg \cdot kg^{-1}$)	k ($L \cdot mg^{-1}$)				($mg \cdot L^{-1}$)	M3-Al ($mg \cdot kg^{-1}$)	M3-Fe ($mg \cdot kg^{-1}$)	M3-Ca ($mg \cdot kg^{-1}$)	M3-P ($mg \cdot kg^{-1}$)
		Median	Median	Median	Median	Median	Median	Median	Median	Median	Median
P sorption sub-group: $SL_{M3\text{-}Al}$ (Smax < 450.03 $mg \cdot kg^{-1}$)											
1	19	388.09	0.36	5.70	6.57	20.00	2.37	651.32	240.35	1418.11	25.68
2	47	397.21	0.33	6.00	6.64	22.00	3.83	609.00	243.56	1897.24	35.68
3	37	388.97	0.33	6.20	6.44	24.00	6.19	584.50	224.36	1969.85	49.10
4	51	364.80	0.32	6.20	7.37	18.00	12.2	527.23	268.02	2208.95	74.50
P sorption sub-group: $SH_{M3\text{-}Al}$ (Smax ≥ 450.03 $mg \cdot kg^{-1}$)											
1	191	621.39	0.63	5.35	11.10	26.00	2.35	881.59	309.48	1314.53	27.61
2	200	593.89	0.49	5.40	11.07	24.00	3.88	818.88	338.96	1449.91	49.08
3	113	607.78	0.47	5.53	12.50	25.00	6.09	818.37	353.92	1639.21	67.68
4	89	575.57	0.36	6.00	12.30	25.00	10.70	736.64	353.03	2254.59	100.74

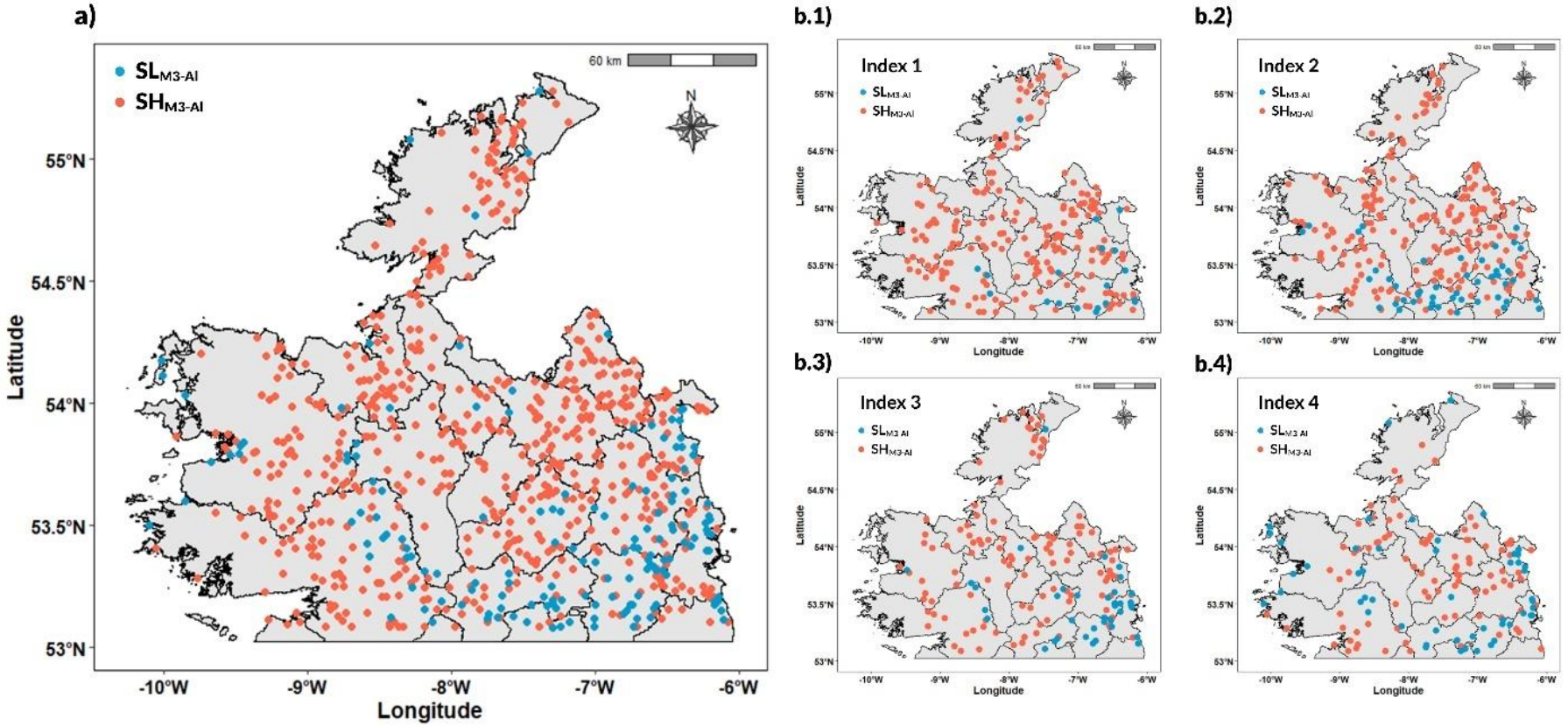


Figure 2-12. Spatial distribution of classified $SL_{M3\text{-}Al}$ and $SH_{M3\text{-}Al}$ soil samples based on sampling locations. Panels (b.1–b.4) show the distribution of sorption classes within each soil P Index (1–4).

Table 2-8**.** Modified soil P Index integrating P dynamics classification.

Soil P Index	Soil P Ranges (mg·L^{-1})		Index Description	P dynamics classes	Response to Fertilisers and Soil Management Advice based on sorption category
	Grassland crops	Other crops			
1	0.0 - 3.0	0.0 - 3.0	Very Low	$SL_{M3\text{-}Al}$	Definite responses to P inputs.
				$SH_{M3\text{-}Al}$	Definite responses to P inputs, build up to target Index 3 will be take longer due to high P fixing and sorption capacities. Slower response to changes in P inputs compared to $SL_{M3\text{-}Al}$ soils. Split applications and liming can help to avoid P being locked-up.
2	3.1 - 5.0	3.1 - 6.0	Low	$SL_{M3\text{-}Al}$	Likely responses to P inputs.
				$SH_{M3\text{-}Al}$	Likely responses to P inputs. However, build up to target Index 3 will be take longer compared to soils in SL category. Similar soil management for soils in this scenario as Index 1-$SH_{M3\text{-}Al}$ soils.
3	5.1 - 8.0	6.1 - 10.0	Medium	$SL_{M3\text{-}Al}$ & $SH_{M3\text{-}Al}$	Unlikely/Tenuous responses to P added. Little difference in P management between the $SL_{M3\text{-}Al}$ and $SH_{M3\text{-}Al}$ subcategories. Applications of P fertiliser to replace P removed.
4	Above 8.0	Above 10.0	Sufficient/ Excess	$SL_{M3\text{-}Al}$	No P fertiliser should be applied. Water quality measures on these soils to decline to Index 3, P removal by crops should occur relatively fast. The low P binding affinity and high P status increases the risk of P losses where these soils interact with hydrological pathways. Landscape or critical source area management that slows water movement in these soils and help to decrease the risk for P losses.
				$SH_{M3\text{-}Al}$	No P fertiliser should be applied. P removal by crops and decline to Index 3 will occur slowly. Legacy P is likely to be a feature of these soils.

2.4 Discussion

2.4.1 The classification of P dynamics

In order to classify P sorption behaviour, both PCA (PC1 versus Smax) and segmented regression analyses (M3-Al versus Smax) were conducted separately. The threshold value derived from the PCA was Smax = 566.94 mg·kg^{-1}, whereas a slightly lower value of Smax = 450.03 mg·kg^{-1} was derived from the segmented regression. The statistical analyses illustrated that the segmented regression (R^2 = 0.49) had a higher coefficient of determination compared to the value obtained using PCA (R^2 = 0.35) (Figure 2-9 and Figure 2-10). The PCA results indicated that M3-Al (loading on PC1 = -0.45) was the predominant factor influencing the P sorption behaviour (loadings on PC1 = -0.44 for Smax, and -0.38 for k) (Figure 2-7), in agreement with previous studies[113,147]. Based on these results, soil P sorption dynamics for Irish soils were classified into $SL_{M3\text{-}Al}$ and $SH_{M3\text{-}Al}$ using the Smax threshold ($Smax_{T2}$ = 450.03 mg·kg^{-1}) obtained from the segmented regression analysis using M3-Al.

For the population of Irish soils in this study, a positive correlation between Smax and k was observed (r_s = 0.88, P < 0.001), indicating that soils with higher P sorption maximum capacities were associated with more free sorption sites, enhancing P fixation and robust storage[68,88,366,367]. The rate of sorption can be understood from the shape of the sorption isotherms[90]. The $SL_{M3\text{-}Al}$ soils also contained less M3-Al, and their isotherms were predominantly distributed at the bottom in Figure 2-8 (b), with a gradual decrease in sorption with increasing added P ('L strict-shaped' concave isotherms). This implied that the $SL_{M3\text{-}Al}$ soils exhibited a slower rate of P sorption, compared to the $SH_{M3\text{-}Al}$ soils with higher M3-Al, whose isotherms were mainly located near the top of Figure 2-8 (b) ('C-/L non-strict-shaped' isotherms), indicating of 'fast rate' sorption. Therefore, the $SL_{M3\text{-}Al}$ soils were characterised as soils with low maximum P sorption capacity and weak binding affinity (k values ranged from 0.05 - 0.71 L·mg^{-1}, median k = 0.33 L·mg^{-1}). The $SH_{M3\text{-}Al}$ soils were characterised as soils with high maximum P sorption capacity and strong binding affinity (k values ranged from 0.02 - 2.49 L·mg^{-1}, median k = 0.51 L·mg^{-1}) (Figure 2-11).

2.4.2 Low P sorption capacity soils ($SL_{M3\text{-}Al}$)

The $SL_{M3\text{-}Al}$ soil class described soils with inherently low sorption capacities. There were no significant differences in sorption values across soil P indices from deficient to excessive soil P levels, indicating that P sorption behaviour was not affected by soil P status (Figure 2-11, Table 2-7). In the $SL_{M3\text{-}Al}$ soils, the exchanges of P between the pools are likely to take place at a 'fast rate' because of fewer binding sites and lower capacity[62]. P-deficient $SL_{M3\text{-}Al}$ soils (Index 1 and 2) were likely to replenish the P-liable pool at the same 'fast rate' in response to P inputs in the 'build-up' phase. Similarly, on Index 4 soils where it was necessary to 'draw-down' excess P, $SL_{M3\text{-}Al}$ soils would also likely take a shorter time for STP to decline to an agronomic optimum value, compared to the $SH_{M3\text{-}Al}$ soils. However, due to low P sorption capacity and weak binding energies, there was less leeway for $SL_{M3\text{-}Al}$ soils. Good practices in P fertilisation are critical in protecting water quality from diffuse inputs (Table 2-8).

2.4.3 High P sorption capacity soils ($SH_{M3\text{-}Al}$)

The $SH_{M3\text{-}Al}$ soil class described soils with inherently high sorption capacities and were also characterised by high k values. The exchange of P between the pools would likely take place at a 'slow rate'. Soils in this class were characterised by more binding sites for labile-P, with greater capacity for storage. However, Soil P status affected sorption behaviour in these soils. Within this class, there was a significant effect of P status on sorption values and $SH_{M3\text{-}Al}$ soils at high P status were observed to have relatively lower Langmuir Smax and k values, as shown in Figure 2-11 and Table 2-7. This is supported by previous studies that demonstrated weaker binding energies in P-saturated soils[18,69]. The P-deficient $SH_{M3\text{-}Al}$ soils had the highest Smax and k values, and compared to the $SL_{M3\text{-}Al}$ soils at similar P status, they would likely take longer to 'build-up' from deficient to optimum for plant availability and crop uptake of P. Furthermore, the Langmuir P sorption parameters of Index 1-$SH_{M3\text{-}Al}$ soils (median values of Smax = 621.39 $mg \cdot kg^{-1}$, k = 0.63 $L \cdot mg^{-1}$) were significantly higher than those of Index 2-$SH_{M3\text{-}Al}$ soils (median values of Smax = 593.89 $mg \cdot kg^{-1}$, k = 0.49 $L \cdot mg^{-1}$), indicating that Index 1 soils were expected to be more aggressive in 'locking-up' the added P than Index 2 soils. Split

applications help to reduce the added P being locked-up[62]. Hall et al. (2023)[113] mapped soils in the northern half of Ireland, and their results showed that large areas of low P index (Index 1 and 2) were overlain by soils with high P sorption characteristics (with high soil M3-Al). Therefore, adjusting soil pH, such as liming, becomes essential in improving P fertilisation efficiency[69].

The 'draw-down' of the soil P Index from Index-4 to 3 is a water quality measure for agriculture under current legislation in Ireland (S.I. No. 113 of 2022 and S.I. No. 393 of 2022). The Index 4-$SH_{M3\text{-}Al}$ soils are expected to take longer to decline excess P than the Index 4-$SL_{M3\text{-}Al}$ soils owing to their greater capacity for storage and high binding energies. These soils are likely to hold Legacy-P amounts, which is a major concern for this group of soil in hydrologically active areas of the landscape. Soils with high legacy P sorbed to soil matrix are a significant threat to water quality because often traditional soil testing does not delineate their large sorption capacity from other soils[368]. Therefore, after separating Index 4 soils into $SL_{M3\text{-}Al}$ and $SH_{M3\text{-}Al}$ subgroups, regulations from policy provisions such as the Nitrates Action Programme and Water Framework Directive measures will be more effective. On the other hand, great agronomic and economic benefits can be gained from a good utilisation of legacy-P[30,369,370]. Farmers can grow crops with high P demand, such as potatoes, corn, hemp, *etc*., to help take up the excess STP.

Differentiating soils of similar P status but different intrinsic P sorption capacities can inform site-specific P management recommendations. This information can be incorporated into agronomic indices for P, as presented in Table 2-8. The existing P Index for Ireland is augmented with sorption categories to improve decision support for P management on soils by integrating the current soil P indices with the P sorption classes. Table 2-8 presents potential P management strategies that could be targeted at specific sorption categories under P deficient (Index 1, 2) and excess (Index 4) scenarios.

2.5 Conclusions and recommendations

Extractable Al strongly influences soil P sorption behaviour in mineral agricultural soils in Ireland. The single-parameter segmented regression analysis with M3-Al (R^2 = 0.49) defined a significant change-point for Smax = 450.03 mg·kg^{-1}. Based on which, soil P dynamics were classified into soils with Smax values below ($SL_{M3\text{-}Al}$) and above ($SH_{M3\text{-}Al}$) the threshold at Smax = 450.03 mg·kg^{-1}. The classification for P dynamics successfully distinguished soils with inherently 'low' and 'high' P sorption capacities. Significantly different P sorption behaviour was observed between the $SL_{M3\text{-}Al}$ and $SH_{M3\text{-}Al}$ soils for each soil P index. $SL_{M3\text{-}Al}$ soils, with low Langmuir Smax and k values, are likely to have a 'fast rate' responding to P changes, whereas $SH_{M3\text{-}Al}$ soils are likely to act at a 'slow rate' and take longer before a response is observed. The implication arising from this work is that we can augment the agronomic P indices with sorption categories to improve decision support for P management on soils. A modified soil P index system is presented in Table 2-8. This allows for soil-type specific recommendations for P management on agricultural mineral soils;

- Index 1- and 2-$SH_{M3\text{-}Al}$ soils (Definite/likely responses to P inputs at a 'slow rate'): Split applications and liming can help to avoid P being locked-up.
- Water quality policy needs to consider a separation between Index 4-$SL_{M3\text{-}Al}$ soils ('fast rate' of 'drawing-down') and Index 4-$SH_{M3\text{-}Al}$ soils ('slow rate' of 'drawing-down'). A modified P index system with soil P dynamics can help assess the risk of P loss and thereby mitigate its environmental impact.
- Future research is suggested to investigate long-term field experiments to verify the response of soils classified as 'low' and 'high' P sorbing classes to P fertiliser amendments under growing conditions, and to calibrate these responses against crop yield and fertiliser recovery data to develop quantitative recommendations for farm nutrient management.

Chapter 3

Comparing the potential of benchtop and handheld mid-infrared spectrometers for predicting soil phosphorus sorption capacity and evaluating the influence of sample preparation[2]

[2] Part of this chapter has been accepted for publication. Manuscript details:
Yang, S., White, B., de Santana, F.B., Hall, R.L., Daly, K., 2024. Comparing the potential of benchtop and handheld mid-infrared spectrometers for predicting soil phosphorus (P) sorption capacity and evaluating the influence of sample preparation. Spectrochim. *Acta Part A Mol. Biomol. Spectrosc*. 322, 124856. https://doi.org/10.1016/j.saa.2024.124856

Abstract

Purpose: Advice for phosphorus (P) fertilisation is typically based on extractive soil tests, without accounting for P sorption. However, traditional assessment by the Langmuir isotherm experiment is time-consuming, labour-intensive, and costly. The objective of this work was to overcome these limitations and accurately predict the isothermal parameter of P sorption maximum capacity (Smax, $mg \cdot kg^{-1}$) from MIR spectroscopy.

Methods: The same soil samples from the Terra archive (Chapter 2) were used, with reference Smax values applied to classify them into 'low' and 'high' sorption capacity groups based on the threshold of 450.03 $mg \cdot kg^{-1}$. This study developed MIR spectral libraries using both benchtop and handheld MIR spectrometers by scanning soil samples in two particle sizes, <0.100 mm (ball-milled) and <2 mm. Chemometrics and machine learning techniques, including partial least squares (PLS), Cubist, support vector machine (SVM) regression, and random forest (RF), were applied to the four spectral libraries and evaluated for predicting Smax.

Results: Benchtop spectral libraries at both particle sizes yielded 'excellent models', with SVM achieving high Smax prediction accuracy ($RPIQ_{Val}$ = 4.50 for ball-milled and 4.25 for <2 mm samples). In comparison, handheld spectral libraries exhibited higher noise and lower resolution. Ball-milled samples improved Smax prediction accuracy attributable to greater homogeneity, with the Cubist model on raw spectra achieving an 'approximate quantitative model' ($RPIQ_{Val}$ = 2.74). However, for <2 mm samples, the best-performing Cubist model achieved only a 'fair' performance ($RPIQ_{Val}$ = 2.23), with potential to distinguish between 'low' and 'high' Smax classes.

Conclusion: The results indicate that benchtop MIR spectroscopy can accurately predict Langmuir Smax without requiring ball-milling, demonstrating its potential as a rapid alternative for assessing soil P sorption capacity. In contrast, handheld MIR spectroscopy provides only approximate quantitative predictions for ball-milled samples, and for <2 mm samples, it is suitable only for classifying soils into low and high sorption capacity groups.

Graphical abstract

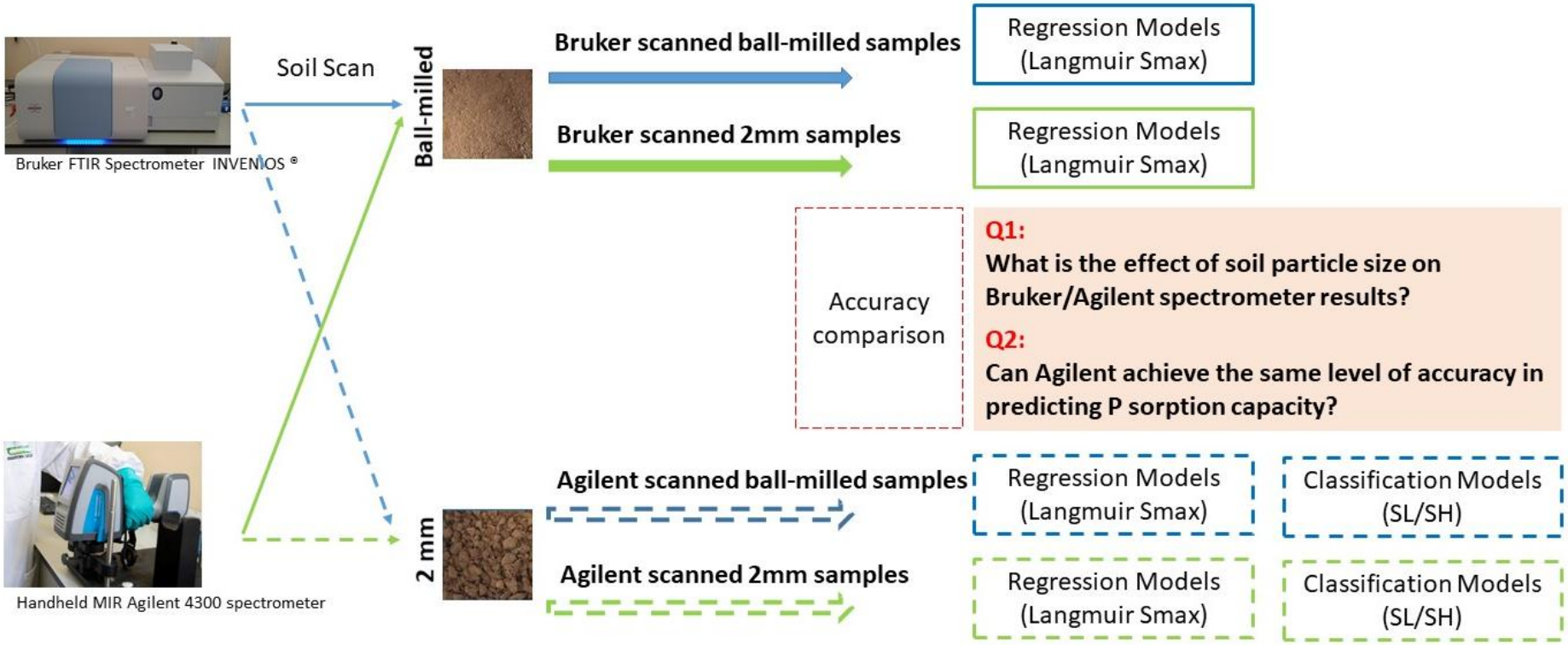


Figure 3-1. Graphical abstract for Chapter 3.

3.1 Introduction

Soil phosphorus (P) requires rational management to support sustainable agricultural production and protect the environment. The Irish National Standard for agronomic P recommendations is based on the Soil Index System, with critical plant-available P indices determined by Morgan's reagent extraction[371]. To gain insight into soil P dynamics with different parent materials, the Langmuir parameters of P sorption maximum capacity (Smax, $mg \cdot kg^{-1}$) and binding affinity constant (k, $L \cdot mg^{-1}$) derived from Langmuir sorption isotherms can be highly effective[18].

In Chapter 2, the Langmuir parameter 'Smax' was identified as a discriminant for the differentiation of soils into two significantly different sorption behaviours, based on their intrinsic sorption capacity, namely (i) soils with low P sorption capacity (SL soils, Smax < 450.03 $mg \cdot kg^{-1}$), characterised by a 'fast rate' of P exchange between pools, and (ii) soils with high P sorbing capacity (SH soils, Smax ≥ 450.03 $mg \cdot kg^{-1}$), where there is rapid P sorption to the soil particles and a 'slow-rate' of soil P build-up or draw-down. However, the routine chemical isotherm experiment is time-consuming, labour-intensive, generates chemical waste, and comes with financial costs.

Mid-infrared spectroscopy (MIR, 4000 - 400cm^{-1} or 2500 – 25000 nm) is sensitive to molecular information and is well-established in the field of soil science. Combined with chemometrics/machine learning algorithms, MIR spectroscopy has been investigated as a rapid and cost-effective technique for estimating multiple soil properties simultaneously from a single spectrum[145]. Successful predicted soil properties include various forms of soil carbon, pH, cation exchange capacity (CEC), texture, nutrient content, *etc*.[116]. For example, soil organic components are related to absorbance between 3100 -2800 cm^{-1}, and studies have successfully used MIR spectroscopy to predict soil total carbon, organic carbon (OC), organic matter (OM) and bulk density (BD)[154,298]. Soil mineral components are related to the absorbance between 2600 - 1650 cm^{-1}, where interpretation can enable good estimation of various soil mineral components, including extractable magnesium (Mg), calcium (Ca), aluminium (Al), *etc*.[188,372]. A summary of soil components with corresponding wavenumbers of spectral

absorption in the MIR regions can be found in Soriano-Disla et al. (2014)[116] and Bachion de Santana and Daly (2022)[145].

Plant-available P, based on the principle of transitions in the vibrational and rotational states of a molecule, is not a spectrally active component in soils for absorbance detection by MIR[191]. Although studies showed poor prediction for plant-available P[162,191,193,372], MIR spectra nonetheless contain useful information for predicting soil P sorption capacity that is highly correlated with SOC, clay mineralogy and CEC[191,373,374]. For example, Minasny et al. (2009)[191] applied the Cubist algorithm to predict P sorption from MIR spectra, achieving a coefficient of determination (R^2) of 0.79. Gruszczyński et al. (2022)[375] predicted P sorption index using the generalised regression neural network model, achieving a R^2 = 0.78. Ng et al. (2022)[118] used memory-based learning and well predicted P retention using the New Zealand method (R^2 = 0.89). Dunne et al. (2020)[147] also found that Langmuir parameter of Smax can be calibrated to a satisfactory standard (R^2 = 0.67) with MIR and used for rough screening.

Advancements in portable MIR handheld Fourier-transform infrared (FTIR) spectrometer technology has made it feasible to shift MIR soil analysis from the laboratory to the field[232]. Additional benefits include enabling spatially dense on-site sampling and accelerating analysis times, as well as yielding cost benefits by eliminating the need for transportation and sample pre-treatment (i.e., drying and grinding)[231]. However, to the best of our knowledge, most studies of MIR spectroscopy in soil P dynamics focused on laboratory-prepared ball-milled samples using benchtop spectrometers[118,147,191,192]. Forrester et al. (2015)[197] evaluated the performance of the handheld Agilent 4100 spectrometer in predicting soil P buffering index by partial least-squares regression (PLSR) cross-validation. In their comparison of MIR spectra obtained from both the benchtop (Perkin Elmer Inc., Waltham, MA, USA) and handheld (Agilent-4100) instruments, they observed a reduction in accuracy for PBI prediction with the handheld instrument (R^2 = 0.82) compared to the benchtop spectrometer (R^2 = 0.87). Nevertheless, the Agilent-4100 maintained a similar level of PBI prediction accuracy. Additionally, Forrester et al. (2015)[197] investigated the impact of soil grinding on spectra from the benchtop spectrometer. Although a significant reduction in overall spectral intensity for

finely ground soil compared to unground soil, similar cross-validation models were identified for both sample pre-treatments, with R^2 values of 0.87 for finely ground soil and 0.89 for unground soil. However, the effect of soil heterogeneity on handheld spectroscopy was not addressed.

The primary objective of this study was to assess and compare the performance of benchtop and handheld MIR spectrometers in predicting Smax values across a diverse range of agricultural mineral soils. Additionally, the study also aimed to evaluate the impact of sample particle sizes on the accuracy of Smax predictions using both benchtop and handheld MIR spectrometers, suggesting potential reductions in machinery costs and time associated with the elimination of grinding. To meet the objectives, soil samples were scanned with both spectrometers in two different particle sizes (i. dried, sieved; ii. dried, sieved, and ball-milled), resulting in four spectral libraries. The comparison of Smax prediction by the optimal chemometrics/machine learning algorithm across these different heterogeneous soil samples obtained from both spectrometers was conducted. Furthermore, the potential of the handheld MIR spectrometer to classify P sorbing classes was also investigated. Finally, this study also determined the most affordable and convenient method for understanding soil P dynamics based on the accuracy of Smax prediction from the four spectral libraries.

3.2 Materials and methods

3.2.1 Soil archive and reference analysis

The same soil archive (www.tellus.ie, Geological Survey Ireland 2020), comprising 737 mineral topsoil samples (0.05–0.20 m depth) used in Chapter 2, was employed for spectral scanning. Owing to limited sample volume, ten <2 mm samples could not be scanned by both instruments, as 2 g of the <2 mm fraction had been used for chemical extraction (Section 2.2.2), resulting in 737 common samples across the four spectral libraries (Bruker-BM, Bruker-2 mm, Agilent-BM, and Agilent-2 mm).

Soil samples were previously oven-dried at 40 °C and mechanically sieved to a size <2 mm before analysis. Soil P sorption capacity was previously assessed using routine Langmuir isotherms, as

summarised in Table 3-1. A proportion of the <2 mm samples were further ball-milled (with a single ball having a frequency of 23 times per second for a duration of 90 seconds) to a particle size nominally 95% < 0.032 mm[145].

Table 3-1. P sorption dynamics were described by P sorption maximum capacity (Smax, $mg \cdot kg^{-1}$) and soil-phosphate binding affinity (k, $mg \cdot L^{-1}$) from the routine Langmuir isotherm test.

Langmuir parameters	**Range (min - max)**	**Median**	**Mean**	**SD**
All mineral soil samples from agricultural land (n = 737)				
Smax ($mg \cdot kg^{-1}$)	81.32 - 1041.30	564.41	566.58	144.04
k ($L \cdot mg^{-1}$)	0.02 - 2.49	0.45	0.55	0.35
SL soils[1] (n = 152)				
Smax ($mg \cdot kg^{-1}$)	81.32 - 447.96	387.02	373.72	68.87
k ($L \cdot mg^{-1}$)	0.05 - 0.71	0.33	0.33	0.11
SH soils[2] (n = 585)				
Smax ($mg \cdot kg^{-1}$)	450.27 - 1041.30	602.27	616.69	112.80
k ($L \cdot mg^{-1}$)	0.02 - 2.49	0.50	0.60	0.37

[1] SL soils (defined as Langmuir parameter of Smax < 450.03 $mg \cdot kg^{-1}$) are low sorbing soils with low P maximum sorption capacity and weak binding affinity.

[2] SH soils (defined as Langmuir parameter of Smax ≥ 450.03 $mg \cdot kg^{-1}$) are high sorbing soils with high P maximum sorption capacity and strong binding affinity.

3.2.2 MIR spectral data acquisition

The four MIR spectral libraries scanned using benchtop and handheld spectrometers for the ball-milled and <2mm samples collected in this study are described in the following:

3.2.2.1. Benchtop spectral libraries

The benchtop spectrometer, Bruker FTIR INVENIO S® (Massachusetts, USA; hereafter 'Bruker'), was operated in diffuse reflectance mode (DRIFTS) to collect MIR spectra from both ball-milled and <2 mm soil samples. The instrument was calibrated daily using a reference compound (acetaminophen, USP specification) before the first scan and again in the evening. Background spectra were acquired using a gold reference disk. Spectra were collected in the MIR range (4,000–600 cm^{-1}) at a resolution of 1.43 cm^{-1}, with 32 scans averaged to maximise the signal-to-noise ratio.

The obtained benchtop MIR spectral libraries are referred to as 'Bruker-BM' for ball-milled samples and 'Bruker-2mm' for <2 mm samples.

3.2.2.2. Handheld spectral libraries

The handheld Agilent 4300 FTIR spectrometer (Agilent Technologies, Inc., Santa Clara, CA, USA) was operated in diffuse reflectance mode (DRIFTS) to collect MIR spectra from both ball-milled and <2 mm soil samples, referred to as 'Agilent-BM' and 'Agilent-2mm', respectively.

The handheld spectrometer was mounted on a stand in a benchtop configuration. Soil samples were placed in 10 mm stainless steel sample cups, and the surface was levelled using a razor blade. The spectrometer was lowered until the diffuse reflectance attachment made full contact, confirmed by a characteristic MIR soil spectrum on the real-time display. Spectra were collected with the MIR range of 4000–650 cm^{-1} at a resolution of 4 cm^{-1}. A background scan was performed every 15 minutes using a gold reference disk. Each spectrum, recorded as pseudo-absorbance, was averaged from 32 scans and exported as a text file.

3.2.3 Pre-processing framework

The reflectance MIR spectra from Bruker scans were converted to pseudo-absorbance, $\log(\frac{1}{\text{Reflectance}})$. Reflectance MIR spectra from Agilent scans were automatically converted to pseudo-absorbance by Agilent MicroLab software. Spectral regions showing atmospheric CO_2 signals (2398–2281 cm^{-1}) were removed from the spectrum[202]. The spectral libraries pre-processing treatment framework is summarised in Figure 3-2.

3.2.3.1 Benchtop MIR spectral pre-processing

To preserve a broad range of soil properties, all samples (n = 737; Smax range: 81.32–1041.30 $mg \cdot kg^{-1}$) were included in the chemometric and machine learning analysis.

Four commonly used pre-processing algorithms for vibrational spectroscopy were applied separately to the raw Bruker-BM and Bruker-2mm spectral libraries, including (i) standard normal variate (SNV), (ii) multiplicative signal correction (MSC), (iii) Savitzky–Golay smoothing with first derivative (SG-1), and (iv) second derivative (SG-2). Each pre-processing method was tested with partial least squares regression, and the optimal pre-processing algorithm was identified by the lowest cross-validation root mean squared error.

3.2.3.2 Handheld MIR spectral pre-processing

The same pre-processing steps in Section 3.2.3.1 were applied to the raw Agilent-BM and Agilent-2mm spectral libraries (n = 737). However, to improve the accuracy of the regression modelling, additional pre-processing steps were applied, specifying: a) combination of pre-processing algorithms; b) detection and removal of outliers; c) trimming of noisy wavelengths (Figure 3-2).

a). Optimal pre-processing algorithm selection

The additional pre-processing algorithms investigated included: a combination of Savitzky-Golay simple smoothing (zero derivative, SG-0) and SG-1 (SG-0+SG-1); a combination of SNV and SG-1

(SNV+SG-1); and a combination of SNV, MSC and SG-1 (SNV+MSC+SG-1)[376,377]. The same determination was used in selecting the best pre-processing algorithm.

b). Outliers detection and removal

The detection of outliers were based on chemically unrepresentative samples and spectrally unusual samples[378,379]. Langmuir Smax distributions were examined by histogram and box-and-whisker plot. The whiskers were the highest and lowest Smax values that are not outside 1.5 times the interquartile range (IQR). Values outside the whiskers were referred to as 'chemically unrepresentative outliers' and not used. To identify spectral outliers, Principal Component Analysis (PCA) was carried out on the optimally pre-processed Agilent-BM and Agilent-2mm spectral datasets. Hotelling T^2 and Q residues values were calculated. Hotelling T^2-statistic index measured the variation of each sample within the PCA model and the Q residuals shows samples with spectra peaks not contained in the calibration[380]. Samples with both Q residues and Hotelling T^2 values above their critical values at a significance level of probability of 0.01 were considered outliers.

c). Noise wavelengths selection and trimming

The alternative noise regions were trimmed after the Variable Importance in Projection (VIP) filtering. Variable Importance in Projection (VIP) quantifies the contribution of a variable to the model by calculating the variance explained by each variable[381]. The most important wavenumbers were evaluated based on VIP scores calculated by the PLSR models. All wavenumbers with VIP scores higher than 1 (recommended threshold) were considered important. The soil components corresponding to the absorptions were summarised in (Bachion de Santana and Daly (2022)[145].

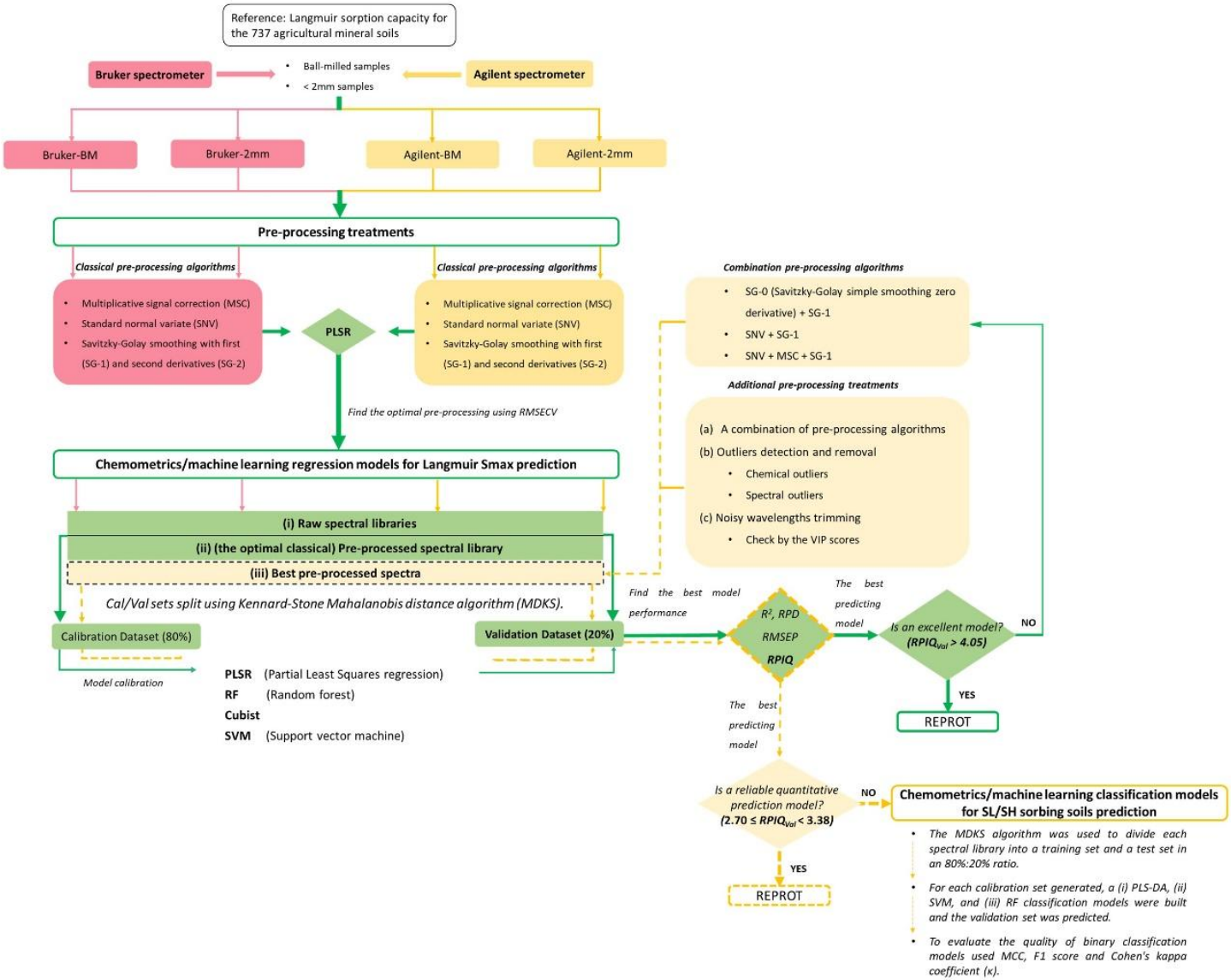


Figure 3-2. A flow chart summarising the spectral pre-processing steps for Bruker and Agilent spectral libraries.

3.2.4 Regression models

Samples in each spectral library were split into a calibration set (80%) and a validation set (20%) according to the Kennard-stone algorithm based on Mahalanobis distance (MDKS)[382,383]. This algorithm selects samples evenly across the spectral feature space, ensuring representative calibration coverage, while the remaining samples provide an independent and unbiased validation set.

For each calibration set generated, a (i) partial least squares regression (PLSR), (ii) Cubist, (iii) support vector machines (SVM), and (iv) random forest (RF) was built and the validation set was predicted (Figure 3-2). PLSR is a linear multivariate regression model. Cubist, SVM (radial basis kernel function) and RF are non-linear multivariate regression models.

The PLSR algorithm, regarded as the standard multivariate regression algorithm, is commonly used in MIR spectroscopy data analysis. The number of latent variables (LVs) used in the PLSR was determined by selecting the number of LVs that yielded the minimum value of the root mean square error of cross-validation (RMSECV). In this study a 10-fold cross-validation was used.

The Cubist algorithm is an extension of Quinlan's M5 model tree. In the 'IF - THEN - ELSE' rule-based structure, samples are grouped into low variability trees based on their similarity. In this study, the optimum combination of the two important hyper-parameters (are 'committees' and 'neighbours') the grid search with 10-fold cross-validation.

The SVM algorithm is a supervised, nonparametric, machine learning technique. The radial basis function (RBF) kernel is the most used in soil spectral libraries[289,298]. Applying the RBF kernel, hyper-parameters of 'Gamma', 'Cost', and 'Epsilon' need to be optimized. In this study, the three hyper-parameters were determined through grid search with 10-fold cross-validation.

The RF algorithm is a decision-tree-based supervised machine learning method to reveal the relationship between Langmuir Smax and the corresponding spectral library[316]. Two-thirds of bootstrap samples within bag samples are chosen to construct each tree, and the remaining one-third

of out-of-bag (OOB) samples are utilised to estimate the model performance of the individual tree, resembling a cross-validation process. The three tuning hyper-parameters (‘*mtry’*, number of trees and minimum node size) were optimised by grid search with 10-fold cross-validation.

3.2.5 Performance comparison between regression models

The best-performing model was identified based on the simultaneous occurrence of a high coefficient of determination (R^2; Equation 3-1), low root mean square error of prediction (RMSEP; Equation 3-2), high ratio of performance to interquartile distance (RPIQ; Equation 3-3), and high ratio of performance to deviation (RPD; Equation 3-4).

$$\mathrm{R}^2 = \frac{\mathrm{V_{res}}(0) - \mathrm{RMSEP}^2}{\mathrm{V_{res}}(0)}, \qquad \text{(Equation 3-1)}$$

$$\mathrm{RMSEP} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}}, \qquad \text{(Equation 3-2)}$$

$$\mathrm{RPIQ} = \frac{\mathrm{IQR}}{\mathrm{RMSEP}}, \qquad \text{(Equation 3-3)}$$

$$\mathrm{RPD} = \frac{\mathrm{SD}}{\mathrm{RMSEP}}, \qquad \text{(Equation 3-4)}$$

where n is the number of samples in the validation set, y_i is the reference value and $\hat{y}_i$ is the predicted value. *IQR* is interquartile range ($Q_3 - Q_1$) of reference values. $V_{res}(0)$ is the mean square error for the reference set ($\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}$). *SD* is standard deviation of the reference set ($\sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}}$), where $\bar{y}$ is the mean value of the reference set.

R^2 and RPD metrics were taken together: ‘*very reliable calibration*’ is with $R^2 > 0.9$ and RPD > 3; ‘*reliable calibration’* is with $0.7 \leq R^2 \leq 0.9$ and $1.75 \leq RPD \leq 3$; ‘*less reliable calibration’* is with $R^2 < 0.7$ and $RPD < 1.75$. Result values of $0.5 \leq R^2 \leq 0.7$ and $1.4 \leq RPD \leq 2$ were considered satisfactory for *rough screening* by Nduwamungu et al. (2009)[384].

Using the RPIQ values, regression models can be classified using the thresholds proposed by Ludwig et al. (2017)[385]: "*Excellent Model (EM)*" with RPIQ > 4.05; "*Good Model (GM)*" with 3.38 ≤ RPIQ ≤ 4.05; "*Approximate Quantitative Model (AQM)*" with 2.70 ≤ RPIQ < 3.38; "*Fair Model (FM)*" (model can distinguishing between high and low values) with 2.02 ≤ RPIQ < 2.70; and "*Non-Reliable Model (NRM)*" for RPIQ < 2.02.

3.2.6 Classification models

Classification models were developed for spectral libraries where regression models failed to predict Langmuir Smax values reliably (Figure 3-2). The aim was to evaluate the extent to which a classification model can be used to discriminate between soils with low (SL) and soils with high (SH) soil P dynamics classes.

Given that the MDKS algorithm systematically encompasses spectral variations across all samples, the simulation of the proportion of SL and SH soils (SL: SH = 152: 585) involved partitioning SL and SH soils within each spectral library into training and test sets at an 80%:20% ratio, respectively. Therefore, the ultimate training dataset (n = 589) consisted of SL (n = 121) and SH (n = 468) soils, and the final gained test dataset (n = 148) consisted of SL (n = 31) and SH (n = 117) soils. For each calibration set generated, (i) PLS-DA, (ii) SVM, and (iii) RF classification models were built and the validation set was predicted. A detailed description of the listed machine learning techniques for classification algorithms can be found in the literature Wadoux et al. (2021)[250].

The procedure used to build the PLS-DA classification model was the same as that used for the PLSR. However, instead of using Langmuir Smax values as y-vectors, discrete variables (P sorbing classes of SL/SH) were used for each sample. SH soils were predicted as class 1 and SL soils were predicted as class −1.

Similar to the SVM regression algorithm, SVM classification algorithm calculated a hyperplane to separate samples from the SL/SH classes using the RBF kernel. The optimisation of Cost function and Sigma hyper-parameters were performed with 10-fold cross-validation.

The procedure used to build the RF classification model was also the same as that used for the RF regression algorithm. Bagging randomisation consists of building each tree (classifier) using a bootstrap replica of the training set with replacement[145]. In this study, a grid search with 10-fold cross-validation was performed to tune hyper-parameters (mtry, number of trees and minimum node size) to find the best model.

3.2.7 Performance comparison between classification models

The imbalance sample sizes for the SL (n = 152) and SH (n = 585) soil samples corresponded to the local soil characteristics[7]. Therefore, the metrics based on the confusion matrix used in this study included F1-score (Equation 3-5), Matthews correlation coefficient (MCC) (Equation 3-6)[145], and Cohen's kappa coefficient (κ) (Equation 3-7)[386]. TP (true positive) are the samples from the SH classified as samples from the SH class; TN (true negative) are the samples from the SL classified as samples from the SL class; FP (False positive) are samples from the SL classified as samples from SH; and FN (false negative) are the samples from SH classified as samples from SL class.

$$F_1\ \text{score} = \frac{2}{\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}}, \qquad \text{(Equation 3-5)}$$

where $Precision = \frac{\text{TP}}{\text{TP+FP}}$, $Recall = \frac{\text{TP}}{\text{TP+FN}}$. Therefore, F1-score was calculated as $F_1\text{score} = \frac{2\times\text{TP}}{2\times\text{TP+FP+FN}}$.

$$\text{MCC} = \frac{\text{TP}\times\text{TN} - \text{FP}\times\text{FN}}{\sqrt{(\text{TP+FP})(\text{TP+FP})(\text{TN+FP})(\text{TN+FN})}}, \qquad \text{(Equation 3-6)}$$

MCC values vary from −1 to 1, with 1 indicating a completely accurate classification, 0 by random prediction, and negative values showing that the agreement is worse than random (− 1 by a completely inaccurate classification).

$$\kappa = \frac{2\times(\text{TP}\times\text{TN} - \text{FN}\times\text{FP})}{(\text{TP} + \text{FP})(\text{FP} + \text{TN}) + (\text{TP} + \text{FN})(\text{FN} + \text{TN})}, \qquad \text{(Equation 3-7)}$$

κ values from −1 to 1, with 1 indicating complete agreement, 0 by chance agreement, and negative values showing that the agreement is worse than random. A higher variability of κ indicates that the performance of some machine learning techniques can be highly influenced by the imbalance in the number of observations belonging to the two classes [386].

3.2.8 Data analysis

All chemometrics were performed in RStudio with R-4.2.1. Regression models for PLS, Random Forest, Cubist and SVM were established by 'pls (version 2.8-1)', 'randomForest (version 4.7–1.1)', 'Cubist (version 0.4.0)', and 'e1071 (version 1.7-11)' packages for R, respectively. classification models for PLS-DA, KNN and DT were established by 'Caret (version 6.0-93)', 'DMwR (version 0.4.1)', and 'rpart (version 4.1.21)' packages for R, respectively. 'Caret (version 6.0-93)' was used for cross-validation and hyper-parameter tuning.

3.3 Results

3.3.1 MIR spectra and qualitative description

Following the framework shown in Figure 3-2 and based on pre-tests identifying the most effective pre-processing techniques (Table 3-2), spectra from the benchtop MIR spectrometer (Bruker-BM and Bruker-2mm) were pre-processed using the first derivative Savitzky-Golay smoothing algorithm (SG-1) with a window size of 11 and a second-order polynomial. Spectra from the handheld MIR spectrometer (Agilent -BM and Agilent -2mm) were pre-processed using a combination of simple smoothing and SG-1 (with a window size of 11 and a second-order polynomial) (SG-0/1).

The four spectral libraries (Bruker-BM, Bruker-2mm, Agilent-BM, and Agilent-2mm) are shown in Figure 3-3 (raw spectra) and Figure 3-4 (pre-processed spectra). In both figures, each spectra was colour coded based on Langmuir Smax values (81.32 - 1041.30 $mg \cdot kg^{-1}$) to visualise the relationship between the spectra and the Langmuir Smax values. The Bruker-BM spectral library exhibited the lowest noise levels both before and after pre-processing, while Agilent-2mm showed a significantly

higher level of spectral noise. Compared to the benchtop spectrometer, the handheld device produced spectra with significantly higher noise. On both instruments, <2 mm samples resulted in poorer spectral quality than ball-milled samples, with reduced spectral features and greater baseline variation, which can be caused by the differences in radiation penetration and light paths due to the differences in particle sizes[197].

Table 3-2. Pre-processing algorithms were applied to each raw spectral library and evaluated using PLSR with cross-validation. The selected method with the lowest RMSECV and highest R^2_{CV}, RPD_{CV}, and $RPIQ_{CV}$ is shown in bold.

Spectral library	Pre-processing	R^2_{CV}	RMSECV	RPD_{CV}	$RPIQ_{CV}$
Bruker-BM	SNV	0.85	55.60	2.59	3.68
	MSC	0.86	52.90	2.72	3.86
	SG-1	**0.87**	**51.86**	**2.78**	**3.94**
	SG-2	0.78	66.96	2.15	3.05
Bruker-2mm	SNV	0.83	58.84	2.45	3.47
	MSC	0.84	58.29	2.47	3.51
	SG-1	**0.86**	**53.79**	**2.68**	**3.80**
	SG-2	0.82	60.65	2.38	3.37
Agilent-BM	SNV	0.86	54.08	2.66	3.78
	MSC	0.84	57.02	2.53	3.58
	SG-1	0.67	83.06	1.73	2.46
	SG-2	0.69	80.46	1.79	2.54
	SG-0 + SG-1	**0.88**	**50.44**	**2.86**	**4.05**
	SNV + SG1	0.67	83.08	1.73	2.46
	SNV + MSC + SG1	0.67	82.93	1.74	2.46
Agilent -2mm	SNV	0.75	71.46	2.02	2.86
	MSC	0.75	72.16	2.00	2.83
	SG-1	0.69	80.32	1.79	2.54
	SG-2	0.68	81.92	1.76	2.50
	SG-0 + SG-1	**0.84**	**57.23**	**2.52**	**3.57**
	SNV + SG1	0.69	79.98	1.80	2.56
	SNV + MSC + SG1	0.69	79.67	1.81	2.57

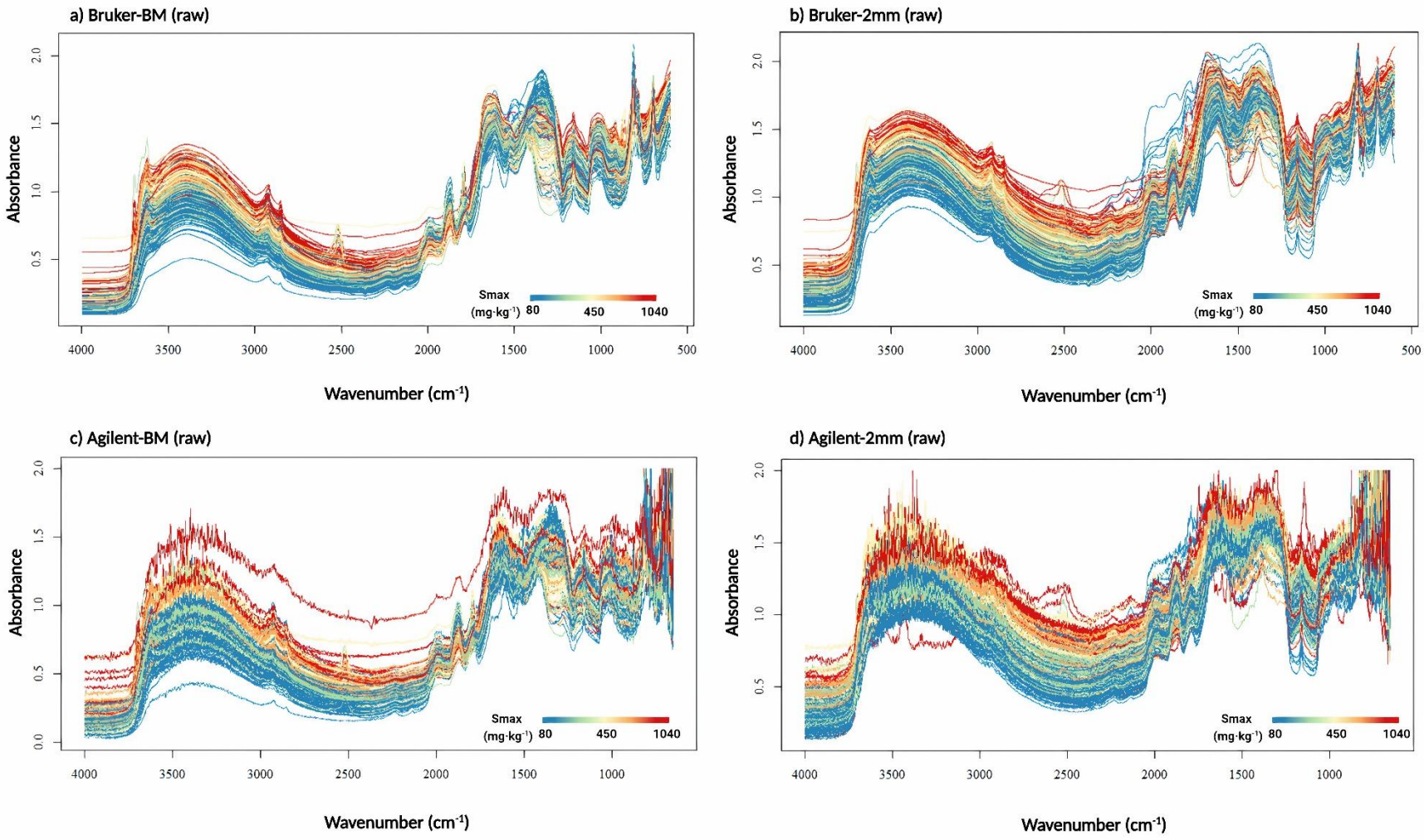


Figure 3-3. Raw spectrograms of the four spectral libraries are plotted in the order of: a). Bruker-BM; b). Bruker-2mm; c). Agilent-BM; d). Agilent-2mm. The spectra are coloured in the reference of the Langmuir Smax ($mg \cdot kg^{-1}$) values.

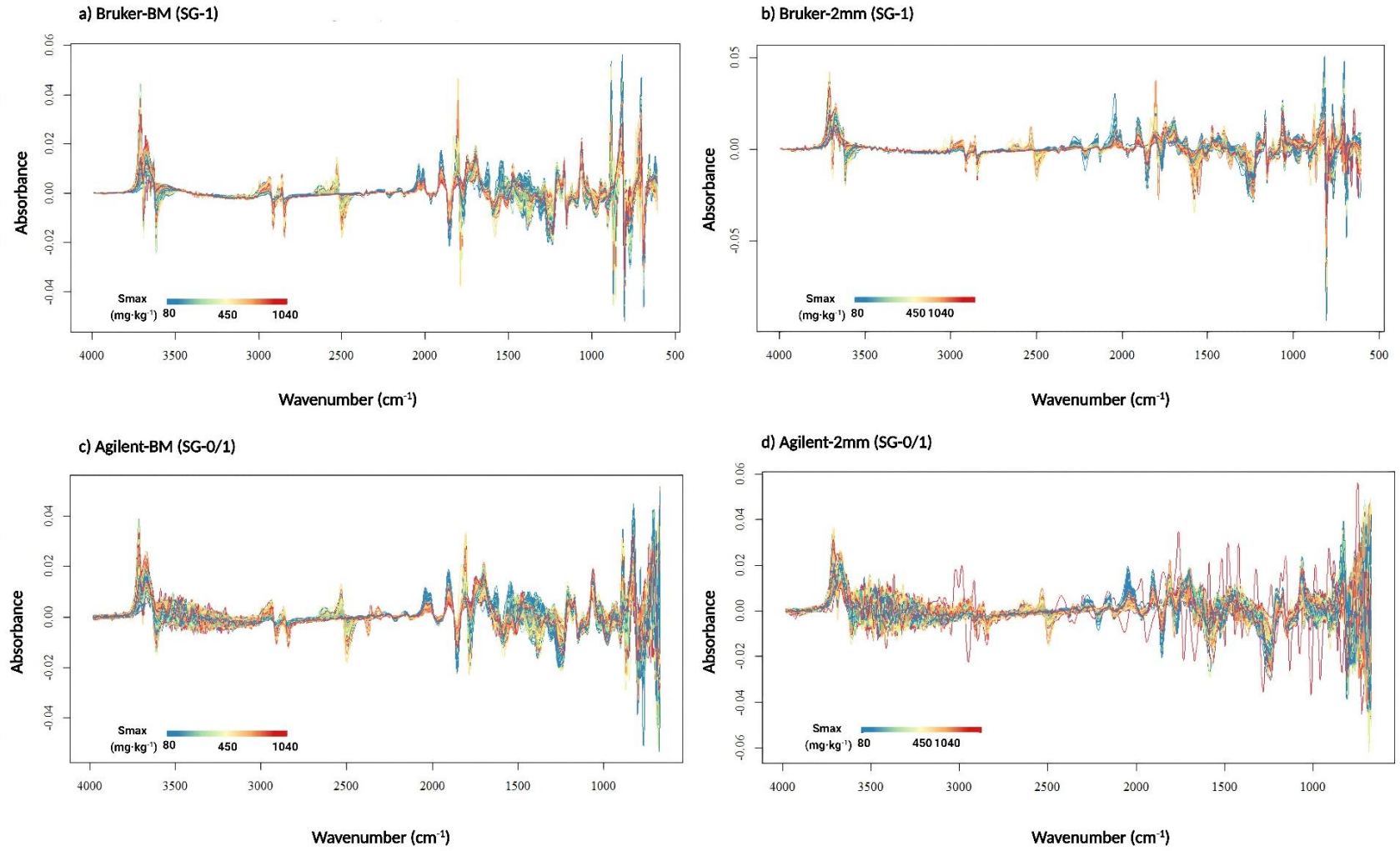


Figure 3-4. Pre-processed spectral libraries coloured based on in the reference of the Langmuir Smax ($mg \cdot kg^{-1}$) values. a). SG-1 pre-processed Bruker-BM; b). SG-1 pre-processed Bruker-2mm; c). The combination SG-0/1 pre-processed Agilent-BM; d). SG-0/1 pre-processed Agilent-2mm.

3.3.2 Chemometrics/machine learning models for benchtop MIR spectral libraries

For Bruker-BM and Bruker-2mm, the four chemometrics/machine learning regression models (PLSR, Cubist, SVM and RF) were developed. The model accuracy results are presented in Table 3-3. Using the raw benchtop MIR spectral libraries, the Cubist algorithm achieved the best RMSEP results for Smax prediction, with RMSEP = 46.46 mg·kg^{-1} for the ball-milled samples (Bruker-BM) and RMSEP = 58.00 mg·kg^{-1} for seived-only samples (Bruker-2mm). These results correspond to a 'Good Model' for the ball-milled samples ($RPIQ_{Val}$ = 3.57) and a 'Fair Model' for the <2 mm samples ($RPIQ_{Val}$ = 2.40).

After SG-1 pre-processing, the accuracy of spectral models using ball-milled and <2 mm soil samples increased significantly. As shown in Table 3-3, the SVM regression algorithm yielded the most accurate models for both the Bruker-BM (RMSEP = 43.07 mg·kg^{-1}) and the Bruker-2mm spectral libraries (RMSEP = 50.90 mg·kg^{-1}). Both Bruker-BM and Bruker-2mm achieved 'Excellent Models' for Smax predction ($RPIQ_{Val}$ = 4.50 and 4.25 for Bruker-BM and Bruker-2mm), presenting minor differences in Figure 3-5.

Therefore, it can be concluded that there were no differences in accuracy between the benchtop MIR spectra models built using ball-milled and <2 mm soils when evaluating Smax. Sieving soil samples to <2 mm was sufficient for the benchtop MIR spectrometer to predict Smax with a high prediction accuracy.

Table 3-3. Summary of the accuracy parameters for each model performing on the Bruker scanned spectral libraries.

Spectral libraries	No. Cal/Val.	Regression modelling algorithms	Calibration				Validation			
			R^2_{Cal}	RPD_{Cal}	RMSEC	$RPIQ_{Cal}$	**R^2_{Val}**	**RPD_{Val}**	**RMSEP**	**$RPIQ_{Val}$**
RAW spectral libraries										
Bruker-BM	589/148	PLSR	0.82	2.14	60.88	2.66	0.87	2.45	51.02	2.93
		Cubist	0.87	2.53	51.29	3.31	**0.89**	**2.67**	**46.46**	**3.57**
		SVM	0.83	2.08	60.11	2.84	0.89	2.40	48.91	3.34
		RF	0.67	1.32	83.33	1.81	0.82	1.68	62.18	2.37
Bruker-2mm	589/148	PLSR	0.77	1.81	70.23	2.33	0.76	1.78	64.51	2.12
		Cubist	0.87	2.47	52.11	3.26	**0.81**	**2.00**	**58.00**	**2.40**
		SVM	0.8	1.94	64.76	2.59	0.76	1.72	64.80	1.97
		RF	0.59	1.03	93.84	1.40	0.67	1.24	77.34	1.50
SG-1 pre-processed spectral libraries										
Bruker-BM	589/148	PLSR	0.69	1.50	80.11	1.99	0.81	1.68	64.43	2.66
		Cubist	0.81	1.95	63.35	2.68	0.87	2.32	51.91	3.47
		SVM	0.9	2.81	46.82	3.81	**0.91**	**2.99**	**43.07**	**4.50**
		RF	0.72	1.48	76.78	2.13	0.86	2.10	55.67	3.06
Bruker-2mm	589/148	PLSR	0.71	1.56	77.55	2.03	0.77	1.66	69.56	2.66
		Cubist	0.79	1.78	66.34	2.27	0.82	1.98	62.22	2.67
		SVM	0.89	2.77	47.28	3.74	**0.88**	**2.82**	**50.90**	**4.25**
		RF	0.66	1.21	85.00	1.69	0.80	1.73	65.87	2.32

* Values in bold highlight the regression models with the best accuracy.

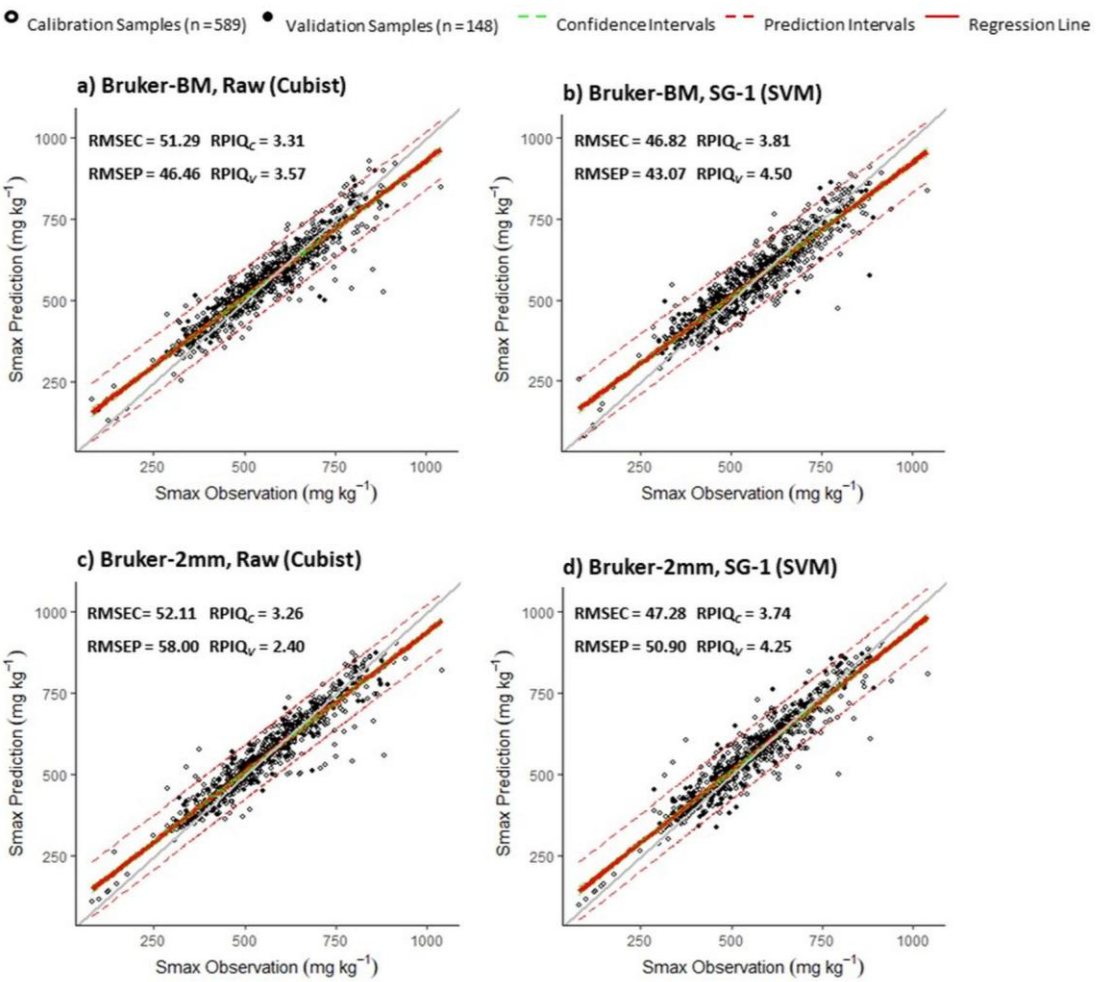


Figure 3-5. Scatter plots showed the best regression algorithms for predicting soil Langmuir Smax on the raw and SG-1 pre-processed Bruker-BM and Bruker-2mm spectral libraries. Auxiliary lines were added to show the prediction performance: ideal line (grey line), regression line (red solid line), confidence interval (green dotted lines) and prediction interval (red dotted lines).

3.3.3 Chemometrics/machine learning models for handheld MIR spectra libraries

3.3.3.1 Additional pre-processing treatments for the handheld MIR spectra

A combination of simple smoothing and SG-1 pre-processing, SG-0/1, was selected for pre-processing the handheld MIR spectra (Agilent-BM and Agilent-2mm). Unrepresentative outliers and spectral noise regions were removed and trimmed after SG-0/1 pre-processing.

Samples with Smax values outside the whiskers (Smax < 157.07 $mg \cdot kg^{-1}$ or Smax > 974.71 $mg \cdot kg^{-1}$) were referred to as unrepresentative 'chemical outliers' (n = 7). The most occurring reference data outliers (6 out of 7 outliers) were SL soils with poor sorption properties from imperfectly/poorly drained sandy soils. The number of spectral outliers was 18 for Agilent-BM and 5 for Agilent-2mm (Figure 3-6). These outliers, identified by Q and T^2 values exceeding the 0.01 significance threshold, were likely due to process failures such as air entrapment. A greater number of spectral outliers were detected in the ball-milled handheld spectra compared with the <2 mm samples, which likely reflects both analytical and material factors. In this study, the ball-milled samples were finely ground (<0.100 mm)[145], producing a more homogeneous powder that improved spectral contact and reduced light scattering. However, the smaller gaps between particles created a flatter soil surface with fewer voids, resulting in more uniform reflectance and spectral values concentrated within a narrower range[387].

The soil chemical information corresponds to the most important wavenumbers selected by VIP scores (> 1), with references to a summary by Bachion de Santana and Daly (2022)[145]. As verified by the VIP scores for the Agilent-BM and Agilent-2mm spectral libraries, the range from 3,500 to 3,000 cm^{-1} (2,857 – 3,333 nm) and 800 to 650 cm^{-1} (12,500 – 15,385 nm) were trimmed from both of the Agilent spectra due to the presence of high instrumental noise (Figure 3-7).

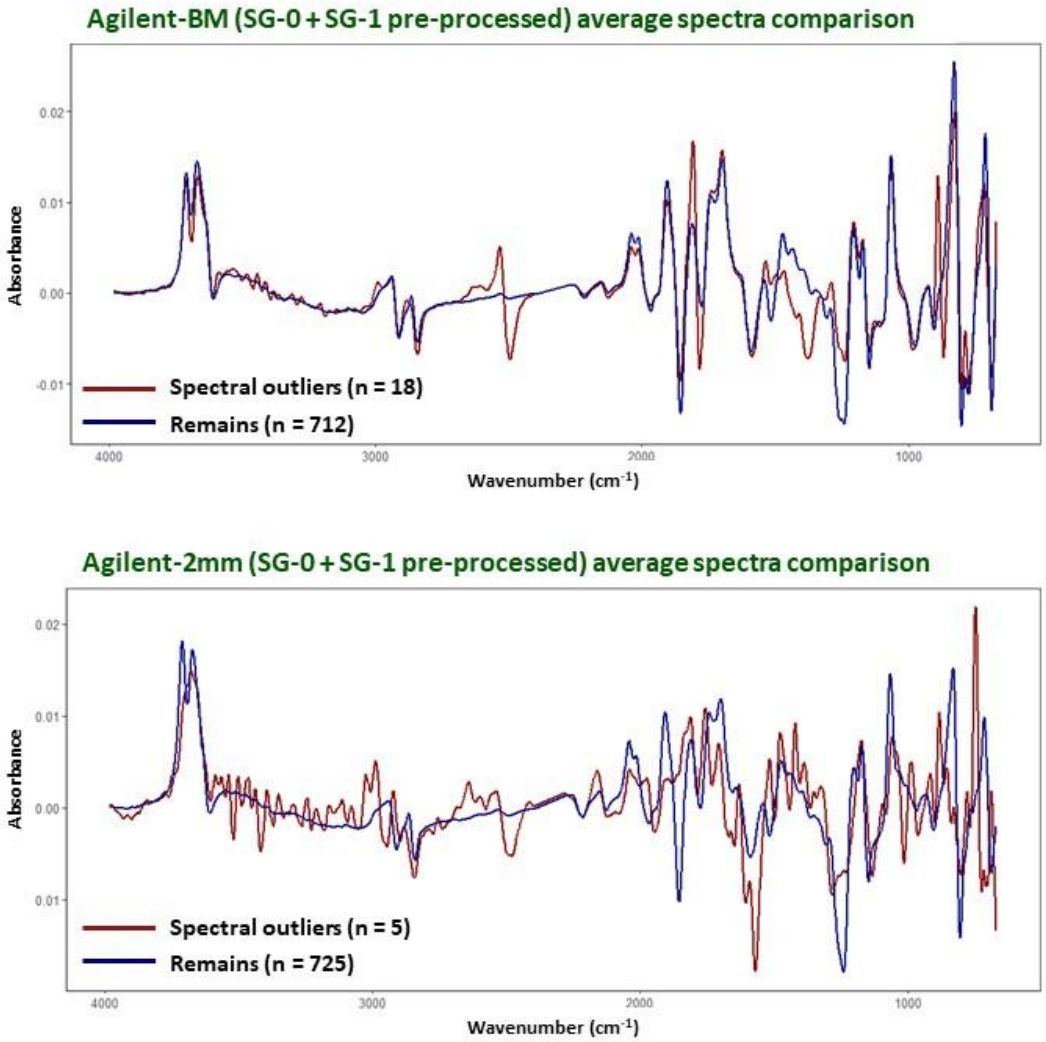


Figure 3-6. Average spectral comparison between detected outlier spectra (red) and remaining spectra (blue) for Agilent-BM and Agilent-2mm spectral libraries. The spectral outliers (n = 18 for the Agilent-BM, n = 5 for the Agilent-2mm) were detected by T^2 values and Q residuals.

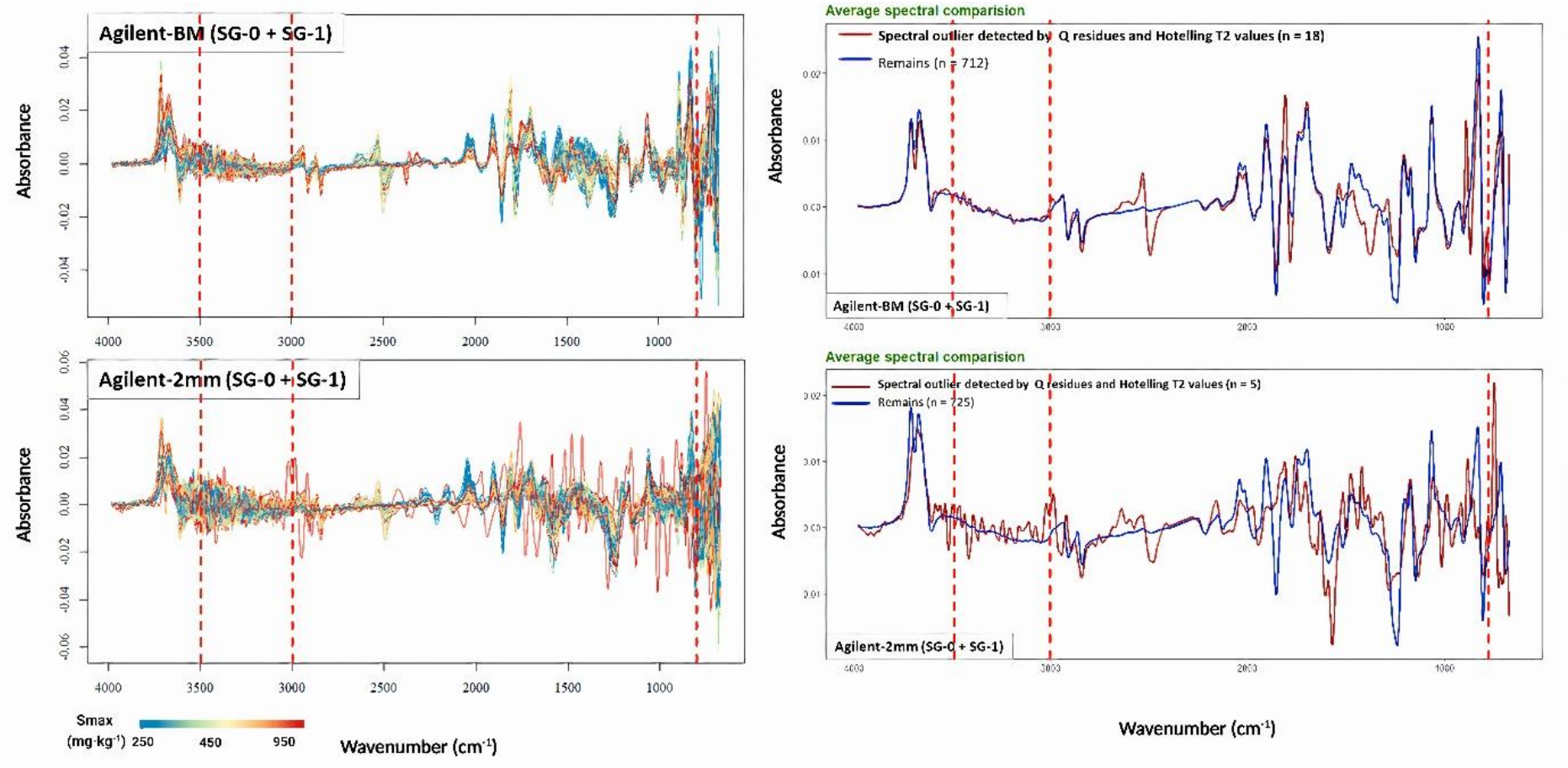


Figure 3-7. The the range from 3,500 to 3,000 cm^{-1} (2,857 – 3,333 nm) and 800 to 650 cm^{-1} (12,500 – 15,385 nm) was trimmed from the handheld MIR spectra (Agilent-BM and Agilent-2mm) due to the presence of high instrumental noise.

3.3.3.2 Regression models using raw handheld MIR spectroscopy

Table 3-4 summarises the regression results from the chemomectrics/ machine learning models (PLSR, Cubist, SVM, RF) applied to the raw handheld MIR spectra (Agilent-BM and Agilent-2mm).

For the raw handheld spectral libraries, the Cubist algorithm out-performed the others for Smax prediction with the lowest RMSEP results for samples with both particle sizes (Agilent-BM, RMSEP = 56.77 and Agilent-2mm, RMSEP = 69.23). Additionally, Cubist on Agilent-BM resulted in a higher R^2_{Val} value of 0.83 and a more confined prediction interval compared to the optimal regression obtained from Cubist on Agilent-2mm (R^2_{Val} = 0.71).

The best regression model to predict Smax from raw Agilent-BM spectra reached an 'Approximate Quantitative Model' ($RPIQ_{Val}$ = 2.74), whereas for 2mm samples on Agilent-2mm, it only achieved a 'Fair Model' ($RPIQ_{Val}$ = 2.23). The effect of soil grinding on the Agilent raw spectroscopy was shown in Figure 3-8 (a) and (d).

3.3.3.3 Regression models on the pre-processed handheld MIR spectroscopy

Table 3-4 summarises the four chemometrics/machine learning results for the Agilent-BM and Agilent-2mm spectral libraries with SG-0/1 spectral pre-processing, and additional pre-processing treatments (further with outliers removal and noisy range trimming).

For Agilent-BM and Agilent-2mm after pre-processing (SG-0/1 solo, and additional pre-processing), SVM yielded the most accurate regression results for predicting Smax. Spectral pre-processing with a combination of SG-0/1 was the most effective management to improve the model for Smax at calibrations (R^2_{Cal} = 0.94, RPD_{Cal} =3.70 for Agilent-BM; R^2_{Cal} = 0.89, RPD_{Cal} = 2.64 for Agilent-2mm). However, according to the validation results, spectra pre-processed with outlier removal and noisy wavenumber trimmed gave better results than spectra pre-processed only with SG-0/1 filtering (Table 3-4). For Agilent-BM (with additional pre-processing treatments), the prediction, which included 712 samples, was satisfactory for rough screening. The best predictive model of Smax by SVM obtained a 'Fair Model' with an $RPIQ_{Val}$ = 2.67. For Agilent-2mm (with additional pre-processing treatments), the

prediction, which included 725 samples, was not reliable to predict the soil Smax value. The best predictive model of Smax by SVM only obtained a 'Non-reliable Model' with an $RPIQ_{Val}$ = 1.83. Overall, for both particle sizes, Smax was best predicted when the spectra were modelled by Cubist on the raw spectral libraries without any pre-processing (Table 3-4 and Figure 3-8).

Table 3-4. Summary of the accuracy parameters for each model performing on the Agilent scanned spectral libraries with both particle sizes, including (1) none (raw spectra); (2) SG-0/1 spectral pre-processing; (4) SG-0/1 spectral pre-processing, outliers removal and trimming.

Spectral libraries	No. Cal/Val.	Regression modelling algorithms	Calibration				Validation			
			R^2_{Cal}	RPD_{Cal}	RMSEC	$RPIQ_{Cal}$	**R^2_{Val}**	**$RPDV_{al}$**	**RMSEP**	**$RPIQ_{Val}$**
RAW spectral libraries										
Agilent-BM	589/148	PLSR	0.78	1.89	67.45	2.57	0.74	1.44	74.80	1.87
		Cubist	0.85	2.24	56.19	3.16	**0.83**	**1.75**	**56.77**	**2.74**
		SVM	0.83	2.12	60.11	2.91	0.82	1.59	58.29	2.63
		RF	0.67	1.28	83.89	1.79	0.73	1.26	78.30	1.76
Agilent-2mm	589/148	PLSR	0.77	1.83	70.31	2.34	0.66	1.32	74.28	1.99
		Cubist	0.82	2.02	63.40	2.64	**0.71**	**1.39**	**69.23**	**2.23**
		SVM	0.85	2.21	57.02	2.95	0.68	1.32	72.52	1.89
		RF	0.64	1.12	90.39	1.41	0.54	1.04	85.67	1.22
SG-0/1 pre-processed spectral libraries										
Agilent-BM	589/148	PLSR	0.76	1.80	71.17	2.40	0.68	1.32	72.99	1.69
		Cubist	0.74	1.57	75.46	2.24	0.68	1.25	73.43	1.52
		SVM	0.94	3.70	3.68	5.26	**0.78**	**1.73**	**60.39**	**2.35**
		RF	0.68	1.30	83.62	1.82	0.67	1.09	76.81	1.52
Agilent-2mm	589/148	PLSR	0.84	2.28	58.74	3.03	0.60	1.28	85.46	1.61
		Cubist	0.81	1.83	65.03	2.45	0.52	1.07	92.79	1.36
		SVM	0.89	2.64	49.38	3.69	**0.61**	**1.28**	**83.83**	**1.68**
		RF	0.67	1.27	84.40	1.79	0.54	1.00	91.82	1.31

Continued Table 3-4

Spectral libraries	No. Cal/Val.	Regression modelling algorithms	Calibration				Validation			
			R^2_{Cal}	RPD_{Cal}	RMSEC	$RPIQ_{Cal}$	**R^2_{Val}**	**$RPDV_{al}$**	**RMSEP**	**$RPIQ_{Val}$**
Additional pre-processed spectral libraries[1]										
Agilent-BM	569/143	PLSR	0.78	1.87	65.26	2.54	0.70	1.42	66.81	1.93
		Cubist	0.80	1.92	62.39	2.60	0.73	1.61	62.75	2.12
		SVM	0.93	3.40	38.07	4.67	**0.76**	**1.82**	**59.18**	**2.67**
		RF	0.70	1.40	76.24	2.09	0.72	1.48	64.65	2.03
Agilent-2mm	580/145	PLSR	0.79	1.94	62.71	2.72	0.75	1.63	65.73	1.72
		Cubist	0.79	1.80	62.82	2.59	0.73	1.48	68.50	1.67
		SVM	0.85	2.25	53.10	3.15	**0.74**	**1.63**	**65.23**	**1.83**
		RF	0.64	1.19	82.45	1.73	0.71	1.30	70.54	1.69

Values in bold highlight the regression models with the best accuracy.

[1] Additional pre-processed spectral libraries are the spectra processed by smoothing and second derivative Savitzky -Golay filtering (SG-0/1), both of the chemically unrepresentative and spectral outliers detection and removal, and trimming (the the range from 3,500 to 3,000 cm^{-1} and 800 to 650 cm^{-1} was trimmed from the Agilent spectra due to the presence of high instrumental noise).

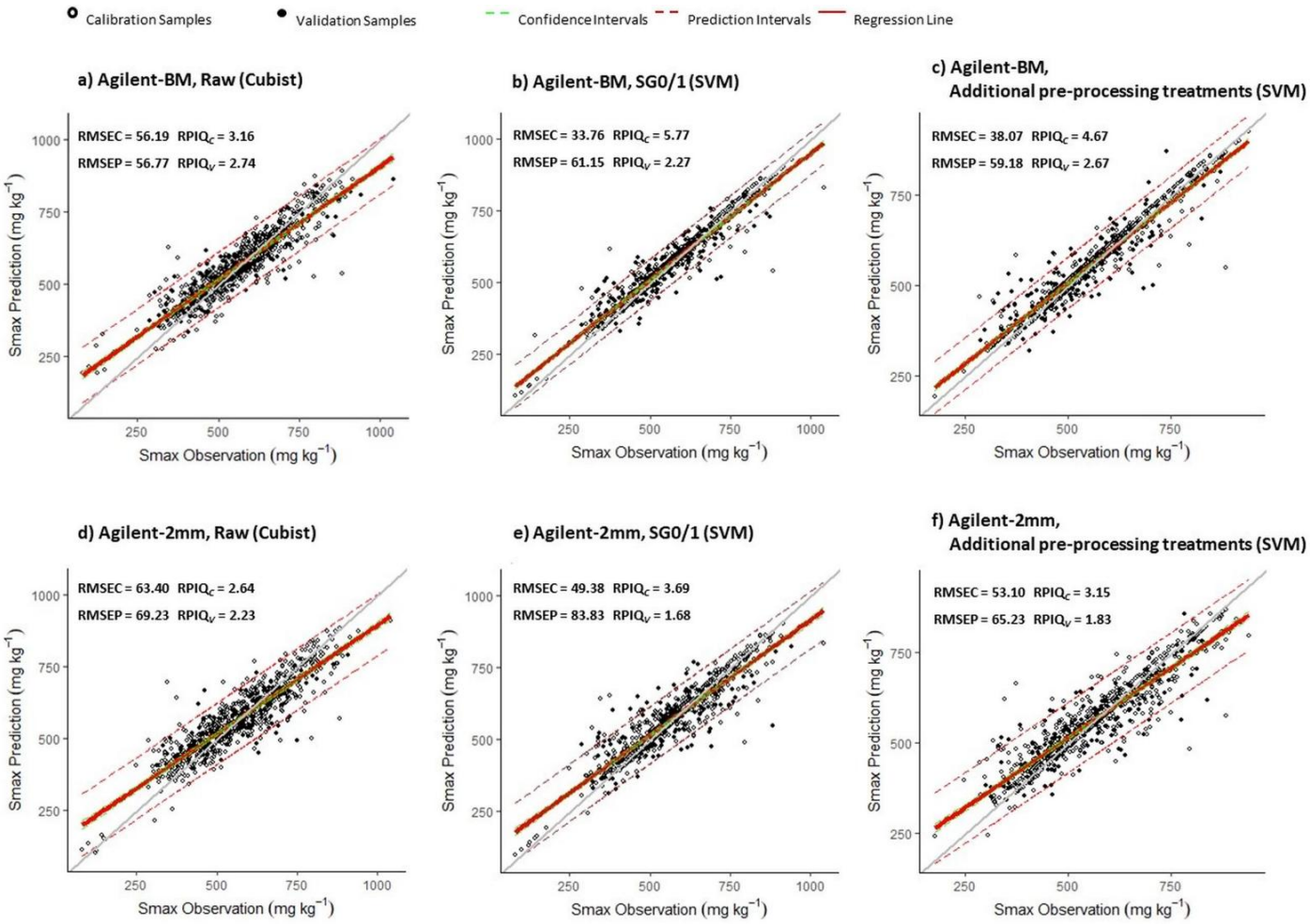


Figure 3-8. Scatter plots showed the best regression algorithms for predicting soil Langmuir Smax on the (i) raw, (ii) SNV pre-processed, (iii) SG-0/1 pre-processed, (iv) SG-0/1 pre-processed, outliers removed and noisy part trimmed Agilent-BM and Agilent-2mm spectral libraries. Auxiliary lines were added to show the prediction performance: ideal line (grey line), regression line (red solid line), confidence interval (green dotted lines) and prediction interval (red dotted lines).

3.3.4 Classification models

The highest accuracy for Smax predicted from the raw Agilent-2mm spectral library only reached a 'Fair Model' ($RPIQ_{Val}$ = 2.23) with the potential for distinguishing between 'low' and 'high' Smax values. Classification analysis was applied to the raw Agilent-2mm spectral library, aiming to distinguish between the 'low' and 'high' soil P sorbing classes (SL/SH soils). Model methods (i) PLS-DA, (ii) SVM and (iii) RF classification were applied to the Agilent-2mm raw spectral library (n = 737). Table 3-5 summarises the F1-scores, MCC values and Cohen's kappa coefficients for all methods in the test set (n = 148, where the number of SL and SH soils was 31 and 117, respectively).

The difference in F1-score between the three classification methods was not substantial, with SVM having the highest F1-score of 0.93 and RF having the lowest F1-score of 0.91. However, the MCC values for evaluating binary classification models resulted in a great variation, with a range of 0.14. Models of PLS-DA and RF gave similar MCC values (0.49 and 0.46 respectively). The SVM algorithm gave the best performance with the highest MCC value of 0.60. The values of Cohen's kappa coefficients had the highest variability with respect to the other measures considered, showing that the performance of RF (κ = 0.35) and PLS-DA (κ = 0.41) were more influenced by the imbalance in the number of observations belonging to the two classes than that of SVM (κ = 0.55, achieving a moderate agreement)[388].

Table 3-5. Summary of the classification methods on the raw Agilent-2mm spectra library with the respective MCC values, F1-scores, and Cohen's kappa from the test set (n = 148).

Spectral library	**Method**	**Test set**						
		TP_{Val}	FP_{Val}	TN_{Val}	FN_{Val}	**F1-score**	**MCC**	**Cohen's kappa**
Agilent-2mm	PLS-DA	116	21	10	1	0.92	0.49	0.41
	SVM	**116**	**17**	**14**	**1**	**0.93**	**0.60**	**0.55**
	RF	117	23	8	0	0.91	0.46	0.35

3.4 Discussion

3.4.1 Handheld vs. benchtop MIR spectrometer

Compared to the benchtop instrument, the Agilent handheld device has many advantages (e.g. smaller size, lower cost, robustness, ease of analysis, portability, *etc*.) that make it a viable option for in situ non-destructive analysis. The quality of the spectra collected with the Agilent handheld spectrometer was inferior than the spectra collected with the Bruker benchtop spectrometer for ball-milled or <2mm samples (Figure 3-3, Figure 3-4). In this study, the results suggest that the handheld spectrometer is suitable for distinguishing soils with low and high P dynamics. However, for predicting P maximum sorption capacity values with high accuracy confidence, the benchtop spectrometer proves to be highly effective, sample ball-milling can be omitted.

The quality of the collected spectra is influenced by the energy throughput, the type of detector and the measurement protocol used, which have been described in Section 1.2.5. In this study, the Bruker spectrometer was equipped with a Mercury Cadmium Telluride (MCT) detector, while the Agilent spectrometer was equipped with a Deuterated Triglycine Sulphate (DTGS) detector. The MCT detector, which is photoconductive, is more sensitive and faster than the thermal DTGS detector, providing a spectrum with a higher signal-to-noise ratio[389]. Compared to the Bruker benchtop spectrometer, which has a spectral range of 4,000-600 cm^{-1} and a resolution of 1.43 cm^{-1}, the Agilent 4300 handheld spectrometer has a narrower spectral range (4000-650 cm^{-1}) and a lower resolution of 4 cm^{-1}. In this study, variations in signal-to-noise ratio can have a stronger effect on model performance than spectral resolution. The Bruker benchtop FTIR typically operates at a signal-to-noise ratio an order of magnitude higher than that of handheld instruments[219], contributing to greater predictive stability and lower baseline noise. The reduced signal-to-noise ratio of the handheld Agilent spectrometer likely increased spectral noise, obscuring weak absorption features and decreasing prediction accuracy. Overall, due to a less stable light source with lower energy throughput and the measurement

of the sample with a larger sample cup surface, the Agilent spectrometer collected spectra of lower quality than the Bruker benchtop spectrometer, showing a significantly noisier appearance.

3.4.2 The optimal regression algorithms

In this study, for both the Bruker-BM and Bruker-2mm spectral libraries, Cubist and SVM exhibited the highest accuracy as regression models to predict Smax on the raw and pre-processed spectral libraries, respectively (Table 3-3, Figure 3-5). This was also true for the Agilent-BM and Agilent-2mm spectral libraries. Specifically, Cubist (for the raw Agilent spectra) and SVM (for the SG-0/1 pre-processed Agilent spectra) were identified as the most accurate models for predicting Langmuir Smax using soil samples with both particle sizes (Table 3-4, Figure 3-8).

The best performance of Cubist and SVM compared to the other models for predicting soil properties in general is in line with the literature[202,289,390]. The reasons for this can be explained by a lower ability of PLSR to generate non-linear relationships between soil spectra and soil P sorption capacity, and by outliers having a greater influence on PLSR than on Cubist or SVM[145]. The RF, Cubist, and SVM models are capable of capturing complex and nonlinear relationships. However, across all spectral libraries, RF exhibited less effective performance in predicting Smax. This reduced performance can be attributed to the RF algorithm's inability to extrapolate beyond the range of the training data. Unlike SVM, or Cubist, which rely on mathematical functions that permit limited extrapolation, the RF model predicts values solely based on the reference data (*y*) contained within the nodes of its decision trees. Consequently, extrapolation beyond the calibration range is not possible for RF.

3.4.3 Benchtop MIR spectral libraries for Smax prediction and particle size effects

Benchtop MIR predictions for Langmuir Smax were very successful for both ball-milled and <2mm soil particle sizes. Pre-processing is essential for the MIR spectra scanned by the Bruker benchtop spectrometer. After SG-1 pre-processing, SVM demonstrated similarly high accuracy in predicting Smax for both ball-milled and <2mm soils.

The Bruker MIR spectra were collected at high spectral resolution (1.43 cm^{-1}) and high signal-to-noise rate. Physical noise will attenuate the reading of the spectral fingerprint region of Smax. Therefore, pre-processing will be essential for Bruker spectra before modelling to remove physical noise and other interfering influences and to enhance the chemical peak from the spectra.

Soil grinding reduces the heterogeneous nature of samples and mitigates other sources of non-systematic variability, thus enhancing the reproducibility and accuracy of predictions derived from the spectra. Our results from the benchtop spectroscopy yielded 'Excellent Model' for both ball-milled and <2mm soil samples. Therefore, when employing MIR benchtop spectroscopy, sieving soil samples to <2mm is adequate in mitigating inter-sample heterogeneity for predicting soil Smax.

3.4.4 VIP analysis

The 'Variable Importance in Projection' or VIP scores are often used for providing valuable insight into the most effective spectral regions by finding informative wavelengths. Figure 3-9 illustrates the informative wavelengths conveying soil chemical information obtained from handheld Agilent spectroscopy for both ball-milled and <2 mm samples.

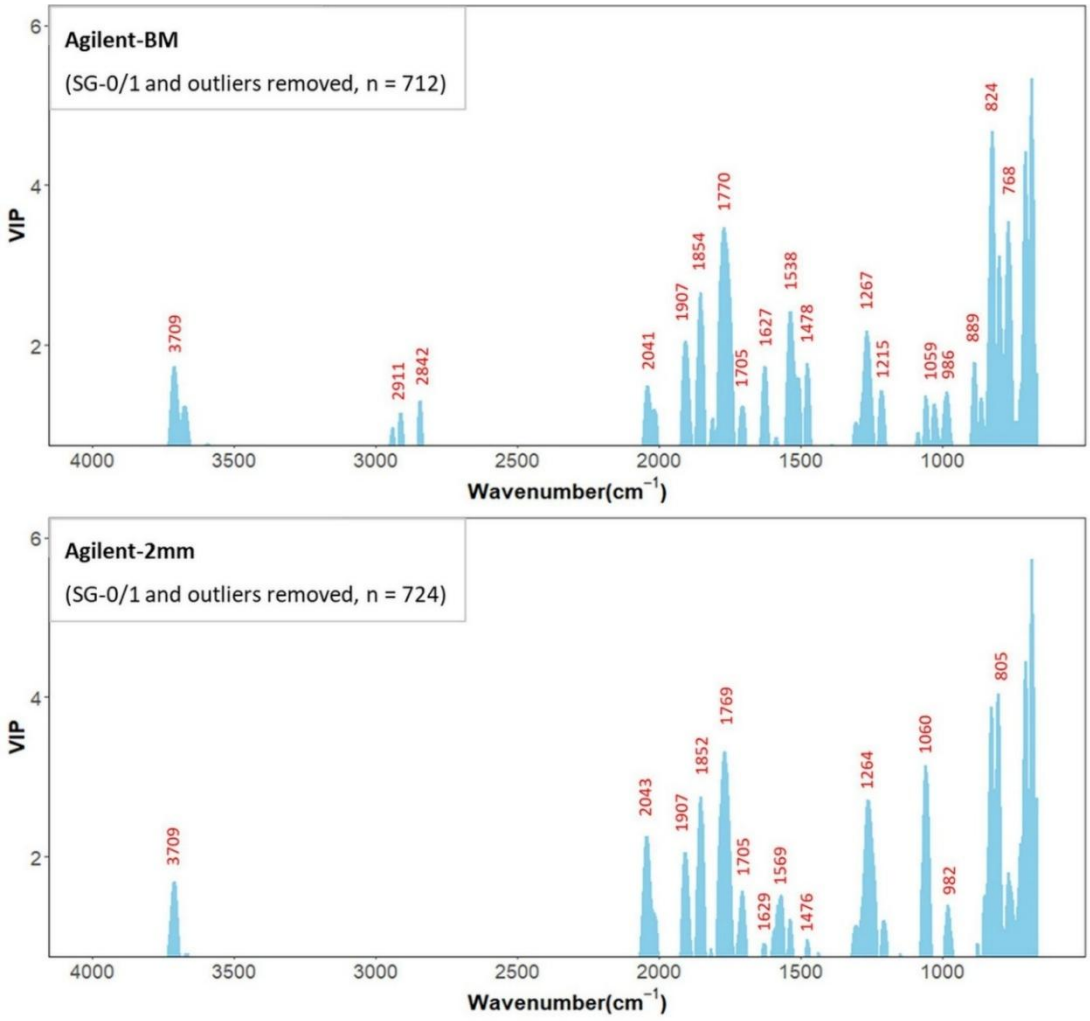


Figure 3-9. Variable importance in projection (VIP) for PLS regression models using ball-milled and <2 mm samples collected by the handheld Agilent spectrometer.

The Agilent-BM and Agilent-2mm spectra showed similar absorption bands. For example, the absorption bands detected in the 3700 - 3600 cm^{-1} range were assigned to clay minerals. The absorption at 2044 cm^{-1} was associated with organic matter (N = C = S). In the range of 2000 – 1820 cm^{-1}, the absorptions were linked to quartz. The intensities of the bands at 1850 cm^{-1} and 1770 cm^{-1} were attributed to samples with low organic carbon content. In the range of 1750 – 1680 cm^{-1}, the absorptions were related to C = O of carboxylic acid and esters. Proteins, phenolics (C–OH), and carbohydrates (C–O) accounted for absorption bands at 1680 and 1525; 1285 and 1100 cm^{-1}, respectively. The band at 1100 cm^{-1} could also be attributed to SiO_2, associated with quartz. The absorption at 815 cm^{-1} could be related to clay minerals (Al-OH) and organic matter (e.g., cellulose).

However, the VIP scores showed that the adsorption bands or spectral intensities were not exactly the same for the ball milled and <2mm sample particle sizes. For example, Agilent-BM spectroscopy revealed the absorption at 2930 – 2850 cm^{-1}, which was attributed to alkyl groups, while this was not observed in the Agilent-2mm. This suggests that the relationships between spectral intensity and Smax may not be the same for ball-milled and <2mm soil samples in the handheld Agilent scanning.

3.3.5 Handheld MIR spectral libraries for Smax prediction and particle size effects

Unlike Bruker spectroscopy, Agilent spectroscopy produced the best predictions for Smax from the raw spectra libraries for both soil particle sizes (Table 3-4, Figure 3-8). The best result for the Agilent achieved on ball-milled samples was an 'Approximate Quantitative' model (RMSEP = 56.77, R_{Val}^2 = 0.83, RPD_{Val} = 1.75, and $RPIQ_{Val}$ = 2.74) to determine Smax using the Cubist regression on the raw Agilent-BM. For 2mm samples, the Cubist model on the raw spectra from the Agilent only reached the potential to distinguish between high and low Smax values (RMSEP = 69.23, R^2_{Val} = 0.71, RPD_{Val} = 1.39, and $RPIQ_{Val}$ = 2.23).

Therefore, for Agilent spectroscopy, ball-milling is an essential step in ensuring a more homogenous mixture of soil micro-aggregates, resulting in a significant improvement in Smax prediction. In comparison to pre-processed spectra that emphasize soil chemical information, raw spectra encompass both chemical and physical soil information. Generally, the quality of chemical information is higher than that of physical information in MIR vibrational spectroscopy. For the Bruker spectrometer, which provides superior spectral quality, it was evident that pre-processed spectra with emphasized chemical information enhanced the accuracy for Smax prediction. However, for the Agilent spectrometer with poorer spectral quality, relying solely on chemical information was insufficient for quantifying the Smax value.

3.3.6 Classification models using handheld MIR spectra

As can be seen in Table 3-5, the Agilent spectrometer was capable of discriminating soils with 'low' and 'high' P sorption characteristics (SL/SH soils). The Agilent-2mm demonstrated high accuracy in discrimination, particularly for high sorption SH soils, achieving an F1-score of 0.93 (precision = 0.87, recall = 0.99). The only misclassified SH soil sample by SVM in the test dataset (FN = 1, Smax = 514.04 $mg \cdot kg^{-1}$), in contrast to the correctly classified SH soils (TP = 116), had a lower organic matter content, with a loss-on-ignition (LOI) of 5.94%, whereas the average LOI for the TP samples was 11.45%.

In contrast, the value of the Matthews correlation coefficient was relatively low (MCC = 0.60). This discrepancy was attributed to the limited number of observations associated with SL soils, consequently reducing the model's ability to recognize soils with low P sorption capacity. It was in agreement with the Cohen's kappa coefficient value (κ = 0.55). According to Gardiner and Radford's (1980)[391] classification, the 17 misclassified SL soils were mostly from the Great Groups classes of Luvisol (7 of the 17) and Surface Water Gley (8 of the 17). 13 of the 17 misclassified SL soils had a clay loam texture. Compared to the correctly classified SL soils (average value of Smax = 373.32 $mg \cdot kg^{-1}$), the misclassified SL soils had a higher P sorption capacity, with an average Smax value of 388.7 $mg \cdot kg^{-1}$. 9 of the FP soils had Smax values above 400 $mg \cdot kg^{-1}$, which were close to the discrimination threshold for distinguishing low and high sorption capacity.

3.3.7 MIR spectroscopy contribution outlook for testing soil health in a package

Healthy soils contribute to ecosystem services, including the achievement of biodiversity, climate neutrality, zero pollution, sustainable farming and achieving the SDGs [392–394]. European Commission (EU) (2021)[392] reported that 60 - 70% of EU soils are in an unhealthy condition and only 56% of watercourses are classified as being of good status due to fertiliser runoff[395]. As a consequence, soil health monitoring is required. As an alternative to traditional chemical analysis, MIR reflectance spectroscopy has been shown to provide good predictions of various physical (e.g. soil texture, bulk density, soil particle size, etc.), chemical (e.g., soil carbon, lime requirement, cation exchange capacity, pH, Al, Fe, P sorption, etc.), and biological soil properties (e.g., organic carbon, total nitrogen, etc.)[116].

Here, MIR spectroscopy is demonstrated for a holistic understanding of soil P dynamics, which is essential for P use optimisation and efficiency. Finally, all indicators can be proposed as an integrated soil health indicator package with P sorption suggestion for soil management from a single scan. A comprehensive understanding of soil health enables effective soil management to ensure fertility and overall soil health, with the ultimate aim of minimising environmental damage and promoting sustainable development.

3.5 Conclusion

In this study, a systematic approach was proposed to improve the reliability of spectral predictions for the understanding of soil P dynamics. Overall, we have demonstrated that the soil P sorption capacity of the Langmuir-derived maximum sorption capacity, Smax, can be predicted by mid-infrared reflectance spectroscopy for the population of agricultural soils in Ireland.

Highly accurate prediction of Smax values by SVM with Savitzky-Golay pre-processing produced excellent models on the Bruker benchtop spectrometer for both ball-mill and <2mm soil particle sizes. Additionally, it is suitable for describing soils with high variability. Where end-users are looking for accurate information on the P capacity of their soils and are expecting accurate Smax values, they may find that there is no difference between the results obtained from the Bruker spectroscopy for the ball-mill and the <2 mm samples. It should be noted though that the benchtop spectrometers with MCT detectors are more expensive compared to handheld ones (~1.5 – 2 time more expensive).

For cost savings compared to benchtop spectroscopy in a laboratory setting, the portable spectrometer is recommended as a time-efficient alternative for routine testing. The scanned spectra are noisier for a lower signal-to-noise ratio. Cubist applied to the raw spectral libraries gave the best Smax prediction as it contains more soil information. Although ball-milled samples can provide approximate Smax values, it is recognised that the results may be less reliable than those obtained from Bruker spectroscopy. To further save time on sample preparation, the Agilent spectrometer can differentiate well between soils with high P sorption capacity. With the comprehensive soil P indices, significant progress has been made towards site-specific P management.

Chapter 4

Mitigating moisture effects in spectral models for determining soil phosphorus sorption capacity using handheld MIR and NIR[3].

[3]Part of this chapter has been submitted for publication. Manuscript details:
Yang, S., White, B., de Santana, F.B., Hall, R.L., Daly, K., 2025. Mitigating moisture effects in spectral models for determining soil phosphorus sorption capacity using handheld MIR and NIR. *Geoderma* 463, 117579. https://doi.org/10.1016/j.geoderma.2025.117579

Abstract

Purpose: Moisture significantly affects soil reflectance spectra. External parameter orthogonalisation (EPO) has been studied as an efficient method for transforming moist spectra.

Methods: This study developed EPO algorithms to eliminate soil moisture effects from handheld MIR (Agilent 4300 FTIR) and NIR spectral models for predicting soil phosphorus sorption capacity (Smax) on wet soils. NIR spectra from InnoSpectra (900-1700 nm) and NeoSpectra (1350-2550 nm) handheld instruments were merged to develop Smax estimation models using dried and sieved soils (n = 686). Of these, 60 representative samples were rewetted to eight gravimetric moisture contents (MC) for EPO development, and 34 samples from the remaining archive were used for validation at low (20%), medium (40%), and high (60%) MC. To address known and unknown moisture conditions, *specific* and *general* EPO models were built using two different matrix calculation approaches, and their effectiveness was compared. Optimal component numbers were determined by comparing RMSEP minimisation and Wilk's Λ maximisation.

Results: *Specific* EPO outperformed *general* EPO for moisture correction, particularly in MIR. At low moisture levels, handheld MIR with *specific* EPO achieved an RPIQ of 3.22 for Smax rough screening. Under medium moisture levels, both handheld MIR and NIR after *specific* EPO correction showed a potential for classification of Smax (MCC of 0.45 and 0.42, respectively).

Conclusion: EPO minimised moisture-induced interference in MIR and NIR spectra. Future in situ handheld MIR and NIR scanning for predicting soil P sorption is recommended under dry soil conditions, with moisture measurement for specific EPO application.

Graphical abstract

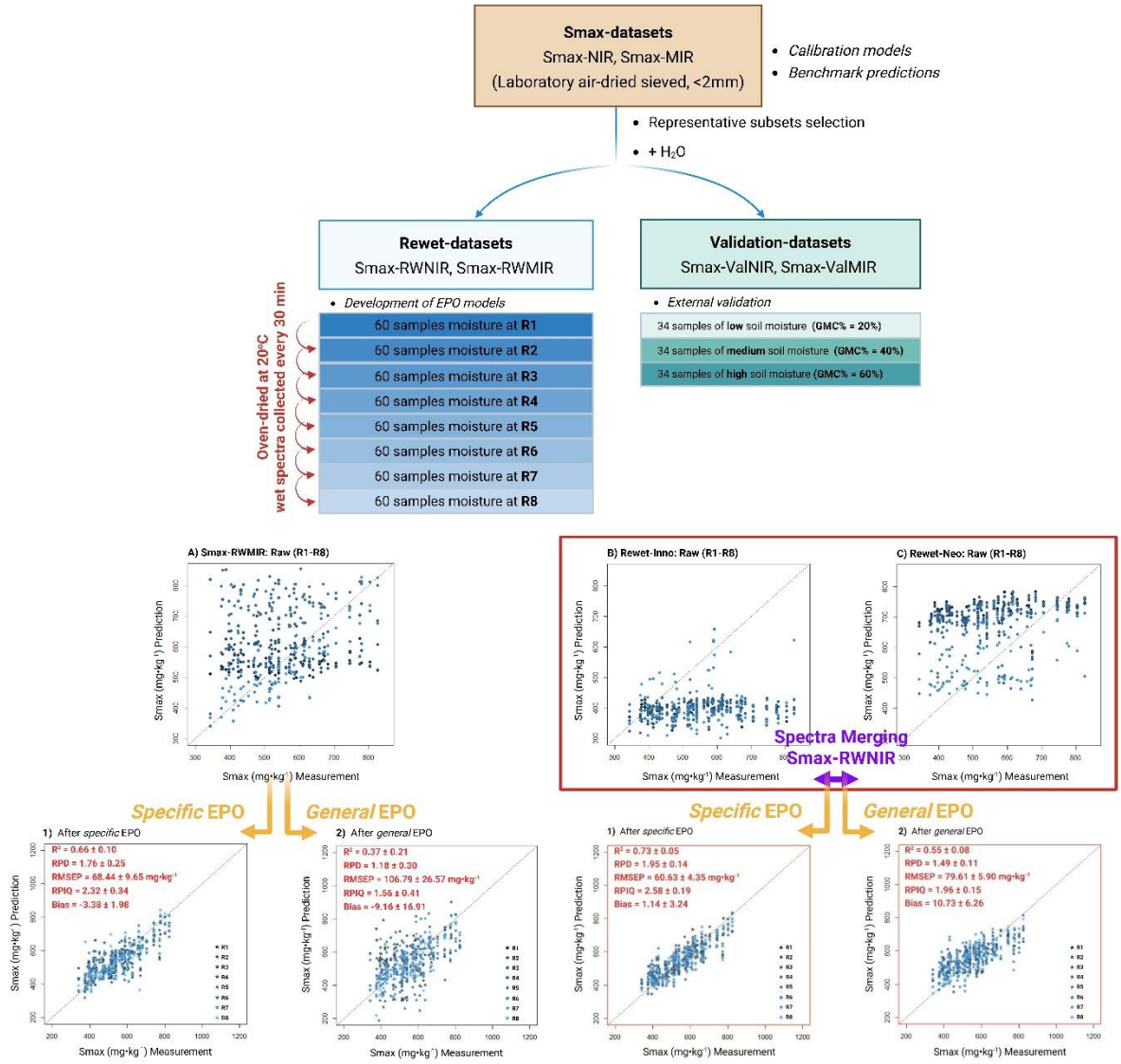


Figure 4-1. Graphical abstract for Chapter 4.

4.1 Introduction

Chapter 2 examined the soil phosphorus (P) sorption capacity, expressed as the Langmuir sorption maximum (Smax, $mg \cdot kg^{-1}$), for samples from the northern half of Ireland, based on laboratory measurements from isotherm batch experiments. Additionally, based on their Smax values, soils were classified as high sorbing (SH, Smax ≥ 450.03 $mg \cdot kg^{-1}$) or low sorbing (SL, Smax < 450.03 $mg \cdot kg^{-1}$), reflecting differences in P dynamics.

To enable time- and cost-efficient, environmentally sustainable analysis with reduced chemical reagent use and laboratory waste , Chapter 3 developed handheld mid-infrared (MIR, 4000–400 cm^{-1}) spectral models using dried, <2 mm sieved-only samples containing heterogeneous soil mastics to predict Smax. The model performance was considered for rough screening purposes, with R^2 = 0.71, RMSEP = 69.23 $mg \cdot kg^{-1}$, and RPIQ = 2.23. Classification of soils into sorption categories (SL/SH) using spectra collected from dried, <2 mm sieved soils with the handheld FTIR device was moderately successful, achieving a Matthews correlation coefficient (MCC) of 0.60. However, the developed MIR models for Smax were based on archived soils that had been dried and sieved without disturbances caused by soil surface roughness, debris, and other environmental factors typically present in field-moist samples. On-site MIR spectra collection of soils without pre-treatment and under field-moist conditions presents challenges for in situ analysis[231].

Near-infrared (NIR, 900 – 2500 nm) reflectance spectroscopy has rapidly become a fast, robust, and cost-effective analytical method for assessing various soil properties, especially with the potential for portable instruments to enable on-the-go spectral soil information collection in the field[239,315]. Studies have used laboratory NIR spectra to estimate soil carbon[309], soil nitrogen[396], and mineralogical properties[160], *etc*. Additionally, several studies have applied (Vis-)NIR for in situ estimation, particularly for soil carbon and nitrogen[397,398]. Similar to MIR, the low dipole moment between P and oxygen limits the success of NIR spectroscopy in detecting inorganic P (e.g., plant available P)[342,399]. However, studies demonstrated moderate success in predicting P sorption using (Vis-)NIR spectroscopy (400 – 2500 nm)

in laboratory settings. Recena et al. (2019)[175] fitted sorption data to the Freundlich and Langmuir models, achieving good estimates of soil P buffer capacity at 1 mg·P·L^{-1} using the Freundlich model (R^2_{CV} = 0.78), whereas predictions for soil maximum P sorption capacity (Smax) based on the Langmuir model were less promising (R^2_{CV} = 0.42). Saidi et al. (2025)[176] also predict the Langmuir Smax of dried, sieved samples using Vis-NIR (350 - 2500 nm), achieving an accuracy of R^2 = 0.65.

For in situ analysis, soil water presents one of the most prominent challenges for both NIR and MIR spectroscopy, causing broad absorption and reflection peaks that distort the spectra and overlap spectral features[236,270,400]. In comparison, MIR is less feasible for non-dried samples[231,384]. The fundamental H-O-H stretching vibration for water produces a strong, broad, stretching vibration in MIR near 3450-3250 cm^{-1} and a weaker bending vibration near 1630 cm^{-1} [236]. Moreover, these vibrations overlap with features such as clay O–H stretching (∼3700 – 3620 cm^{-1})[237]. As a consequence, removing only specific water absorption features will not eliminate moisture effects in the spectra. Additional irregularities and external factors such as weather (e.g., ambient atmospheric conditions) introduce further spectral distortions[215,232,400], complicating the use of in situ spectral collection. Unlike MIR spectroscopy, which excites fundamental vibrations, NIR spectroscopy is based on weaker overtones and combination bands of molecular vibrations, primarily from bonds such as O–H, C–H, and N–H[401]. While moisture interference is generally less severe in NIR than in MIR spectroscopy, due to the higher sensitivity of MIR to water, the overtone and combination bands of water absorption (e.g., at 1400 and 1900 nm) still causes challenges for predictive accuracy[215,236–238].

To mitigate the distorting effects of water, mathematical techniques have been developed, including direct standardisation[264], orthogonal signal correction[267], and wavelet transformation[265], *etc*. Among these, external parameter orthogonalisation (EPO) is one of the most commonly applied approaches, with Metzger et al. (2024)[270] highlighting its use in over 50% of investigated studies due to its ease of understanding and straightforward application[266,271–273]. EPO approach decomposes and excludes the components influenced by external parameters, thereby enabling it to use spectral information associated with the target property[269]. Many studies have reported the successful application of EPO

to wet soil spectra for predicting soil carbon, particularly in the (Vis-)NIR region[266,270,271]. However, to the best of our knowledge, few studies have evaluated the effectiveness of EPO in eliminating water interference for predicting soil P sorption capacity, especially in the MIR region.

The objective of this study was to apply an EPO modelling approach to portable NIR and MIR spectral libraries to predict soil P sorption capacity under moist conditions. The effects of soil moisture on the accuracy of predicting Smax using portable NIR and MIR were examined. The application of EPO models on NIR and MIR to eliminate water influence and improve Smax predictions under different moisture conditions was evaluated. The findings aim to support the practical application of handheld devices for future on-site prediction or classification of Smax values from agricultural soils.

4.2 Materials and methods

4.2.1 Reference laboratory data and spectral libraries

The soil data used in this study were divided into three datasets (Table 4-1). An existing reference laboratory data of Smax from the Langmuir model (n = 686; Smax values ranging from 81.32 – 1041.30 $mg \cdot kg^{-1}$, median of 561.28 $mg \cdot kg^{-1}$), based on topsoil samples (5 – 20 cm) collected in the northern half of Ireland, was available for this study (referred to as the Smax-datasets). This archive (soils were oven-dried at 40 °C and sieved to a size <2 mm) was used to collect spectral libraries from portable MIR and NIR for developing Smax prediction models. The resulting spectral libraries are named Smax-MIR and Smax-NIR for the Smax-datasets.

A total of 60 dried, sieved soils were selected as representative of the soil archive for rewetting, referred to as the Rewet-datasets. The purpose of this dataset was to develop EPO projection matrices for moisture correction through a systematic rewetting experiment (Section 4.2.4). Representative selection was performed using the Kennard-Stone algorithm based on Mahalanobis distance (MDKS)[382]. The Rewet-datasets include Smax values ranging from 342.17 – 825.16 $mg \cdot kg^{-1}$, with a median of 539.47 $mg \cdot kg^{-1}$. Soils were initially rewetted to 100% water saturation and then gradually

oven-dried at 20°C, with soil moisture recorded every 30 minutes across seven intervals. Wet spectra were collected across a wide range of moisture conditions after eight drying durations. Therefore, MIR and NIR spectra for the Rewet-datasets were recorded at each of the eight moisture levels. The Rewet-datasets comprises MIR and NIR spectral libraries, Smax-RWMIR and Smax-RWNIR, across the eight moisture levels.

Soils for validation were randomly selected from the remaining archive (n = 34; Smax values ranging from 320.78 – 799.76 $mg \cdot kg^{-1}$, median of 552.84 $mg \cdot kg^{-1}$), referred to as the Validation-datasets, used to evaluate the performance of the EPO algorithms. Soils were moisturised at low (20% w/w), medium (40% w/w) and high soil moisture (60% w/w) levels, with MIR and NIR spectra recorded at each condition. The resulting spectral libraries are named Smax-ValMIR and Smax-ValNIR for the Validation-datasets.

Table 4-1. Soil P sorption capacity (Smax, mg·kg^{-1}) from the routine Langmuir isotherm test (Chapter 2).

P sorption parameter	Range (min - max)	Median	Mean	SD
(1) Smax-datasets (n = 686, including spectral libraries of Smax-MIR and Smax-NIR)[1]				
Smax (mg·kg^{-1})	81.32 - 1041.30	561.28	565.47	145.28
SL soils[2] (n = 144)				
Smax (mg·kg^{-1})	81.32 - 447.96	386.68	372.61	69.90
SH soils[3] (n = 542)				
Smax (mg·kg^{-1})	450.27 - 1041.30	602.77	616.70	113.59
(2) Rewet-datasets (n = 60, including spectral libraries of Smax-RWMIR and Smax-RWNIR)				
Smax (mg·kg^{-1})	342.17 - 825.16	539.47	548.97	117.71
SL soils (n = 14)				
Smax (mg·kg^{-1})	342.17 - 438.29	405.69	403.27	26.39
SH soils (n = 46)				
Smax (mg·kg^{-1})	453.12 - 825.16	579.61	593.32	96.70
(3) Validation-datasets (n = 34, including spectral libraries of Smax-ValMIR and Smax-ValNIR)				
Smax (mg·kg^{-1})	320.78 - 799.76	552.88	552.84	156.22
SL soils (n = 11)				
Smax (mg·kg^{-1})	320.78 - 445.50	377.55	374.42	36.95
SH soils (n = 23)				
Smax (mg·kg^{-1})	453.40 - 799.76	654.78	638.17	111.61

[1] Some of the archived soils were used after the MIR scanning was completed in 2021, which left less material available for the NIR scanning in 2024.

[2] SL soils (defined as Langmuir parameter of Smax < 450.03 mg·kg^{-1}) are low sorbing soils with low P maximum sorption capacity and weak binding affinity.

[3] SH soils (defined as Langmuir parameter of Smax ≥ 450.03 mg·kg^{-1}) are high sorbing soils with high P maximum sorption capacity and strong binding affinity.

4.2.2 Spectra acquisition

All soils (dried and rewetted) were scanned using three handheld instruments in the benchtop mode. These included (i) the Agilent 4300 Handheld FTIR, Agilent Technologies, Inc., Santa Clara, CA, USA (abbreviated as 'Agilent') in the MIR range of 4,000–650 cm^{-1}; (ii) the InnoSpectra NIR-S-G1, Innospectra Corp., Hsinchu, Taiwan (abbreviated as 'Inno') in the NIR range of 900–1700 nm; (iii) and the NeoSpectra Scanner, Siware Systems Inc., Cairo, Egypt (abbreviated as 'Neo') in the NIR range of 1350–2550 nm. For NIR spectroscopic analysis, spectra from Inno and Neo were combined to cover the full NIR region (900–2500 nm) for chemometrics analysis using a data merging approach.

The MIR spectra were collected using the handheld Agilent spectrometer, configured for 32 co-added scans in diffuse reflectance mode (DRIFTS, 4000-650 cm^{-1}) at a resolution of 4 cm^{-1}. Background readings were collected by scanning a gold coated reference disk every 15 min. Spectra are reported in absorbance units ($\mathrm{A} = \log\left(\frac{1}{\mathrm{Refelctance}}\right)$). The Agilent spectrometer was fixed in the stand and lowered onto the sample cup until the tip of the diffuse reflectance accessory was in close contact with the sample surface**.**

The NIR measurements were undertaken using the handheld Inno and Neo devices in benchtop configurations. The Inno spectrometer operates in the range of 900–1700 nm with a spectral resolution of 3 nm. Each spectrum was the average of 16 scans, featuring a digital resolution of 228 and a PGA gain of 64. The lamp was turned on for 10 minutes before starting the analysis to stabilise the detector's temperature. Background measurements were taken with the lamp remaining lit throughout the analysis. The device was shut down and allowed to cool to room temperature between each validation series[206]. The Neo spectrometer features an InGaAs photodetector, with spectra calculated via Fourier transformation. It covers a spectral range from 1350 to 2550 nm with a resolution of 16 nm, resulting in a total of 257 measured bands.

To cover the full NIR spectral region, following Cevoli et al. (2024)[402], low- and mid-level fusion strategies were applied to integrate the two NIR spectral blocks (from the Inno and Neo devices). As

data fusion is not the main focus of this study, the process is hereafter referred to as spectral merging (Figure 4-2). In the low-level merging approach, spectral data collected from the Inno and Neo spectrometers were directly concatenated to form a combined spectral matrix. To address the overlapping region (1350 – 1650 nm), two strategies, referred to as Treatment 1 and Treatment 2, were implemented. Treatment 1 involved concatenating the full Inno spectra with the trimmed Neo spectra. Treatment 2 involved concatenating the trimmed Inno spectra with the full Neo spectra. For the mid-level merging approach (Treatment 3), data were compressed prior to merging using PCA for feature selection, retaining scores corresponding to the PCs that captured 99% of the variance.

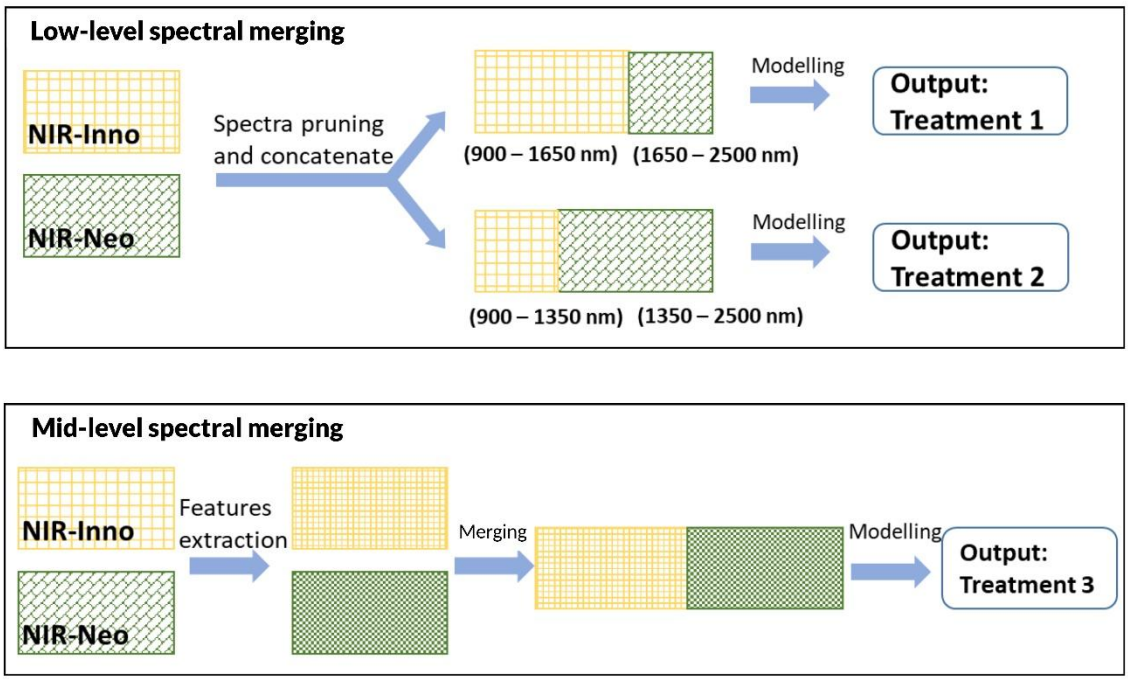


Figure 4-2. General scheme describing low- and mid-level spectral merging approaches.

4.2.3 Chemometrics modelling

Samples in Smax-datasets (Smax-NIR) were split into a calibration set (80 %) and a validation set (20 %) according to the MDKS. Regression models developed for soil Smax values prediction including (i) partial least squares regression (PLSR), (ii) cubist, (iii) support vector machine (SVM, radial basis kernel function), and (iv) random forest (RF). Classification models used to discriminate between soils with

low (SL) and soils with high (SH) soil P dynamics classes including (i) partial least squares discriminant analysis (PLS-DA), (ii) SVM, and (iii) RF classification models. The performance of both regression and classification models was assessed using the validation/test set to ensure reliability and robustness.

An optimal regression performance was identified by evaluating the simultaneous occurrence of a high coefficient of determination (R^2), low root mean square error of prediction (RMSEP), high ratio of performance to interquartile distance (RPIQ), and high ratio of performance deviation (RPD). Additionally, bias was introduced as a metric to assess model performance on the Rewet- and Validation-datasets, with biases closer to zero indicating superior model performance[223] .

The imbalance in sample sizes between SL and SH soil samples reflected local soil characteristics. Classification performance was evaluated using Matthews's Correlation Coefficient (MCC), Cohen's Kappa Coefficient (κ), and the F1-score. MCC and κ values range from −1 to 1, with 1 indicating perfect classification or agreement, 0 representing random prediction or chance agreement, and negative values indicating performance worse than random. The F1-score ranges from 0 to 1, with higher values indicating a better balance between precision and recall, making it particularly useful for evaluating imbalanced datasets.

A detailed description of the listed machine learning techniques and model performance indicators can be found in Chapter 3.2.5 and Chapter 3.2.6. Data analysis was performed in RStudio with R version 4.2.1 (R Core Team, 2025). Spectral visualisation and regression plotting were supported by the R packages 'mdatools 0.14.2', 'rchemo 0.1-3', 'soilspec 0.1.0', 'lattice 0.22-5', and 'ggplot2 3.5.0'. The 'squash 1.0.9' package was used for colour-based visualisation of multivariate data. Regression and classification modelling were conducted using 'pls 2.8-5', 'caret 7.0-1', 'Cubist 0.4.3', 'e1071 1.7-14', 'randomForest 4.7-1.1', and 'kernlab 0.9-32'. Additional packages used for preprocessing and multivariate analysis included 'prospectr 0.2.5', 'pracma 2.4.4', and 'factoextra 1.0.7'.

4.2.4 Rewetting soils

For rewetting of dried soils, a percolation experiment[403] was conducted to determine the initial amount of water required to achieve 100% moisture saturation for each sample (Equation 4-1). Soil samples from the Rewet-datasets (2g ± 0.01) were placed in 50-mm Petri dishes. Deionised water was initially added to achieve 100% saturation (Equation 4-2). The dishes were covered with Parafilm and stored overnight in a cold room to homogenise the moisture. The gravimetric moisture content (GMC) was determined (Equation 4-3), recorded as the moisture content at R1, and the corresponding spectra at the maximum moisture content (R1) were scanned. To obtain the lower moisture contents, the samples were oven-dried at 20°C. Wet spectra were collected every 30 minutes, seven times during gradual oven drying, and the corresponding moisture levels were recorded as GMC from R2 to R8 (Equation 4-2), representing moisture conditions from wettest to driest. Wet spectra from the MIR and NIR instruments were acquired separately due to different scanning protocols (NIR scans were performed bottom-up without sample loss, whereas MIR scans were performed top-down with unavoidable sample loss). The resulting spectral libraries are hereafter referred to as Smax-RWMIR and Smax-RWNIR (at moisture levels R1 to R8, wettest to driest).

For the Validation-datasets, soil samples (2 g ± 0.01) were placed in 50-mm Petri dishes. The rewetting experiment included three soil moisture levels, corresponding to GMC values of 20%, 40%, and 60%, designated as the ‘low’, ‘medium’, and ‘high’ moisture levels. Due to sample mass constraints, the same samples were subsequently used to obtain spectra at progressively higher moisture levels. Deionised water was first added to achieve the low moisture content of 20% w/w. Wet spectra were collected after the samples were sealed with Parafilm and stored overnight in a cold room. Water was then incrementally added, calculated using Equation 4-4, to reach the ‘medium’ and the ‘high’. MIR and NIR spectra were collected at each moisture level following the same protocol.

$$100\%\ \text{Saturation} = \frac{\text{Amount of water after fully absorption (g)}}{\text{Weight of soil sample (g)}} \times 100\% \quad \text{(Equation 4-1)}$$

$$\text{Initially added water} = 100\%\ \text{Saturation} \times \text{weight of soil sample (g)} \quad \text{(Equation 4-2)}$$

$$\text{Soil moisture content (GMC)} = \frac{(\text{Weight}_i - \text{Weight}_{Dry})}{\text{Weight}_{Dry}} \quad \text{(Equation 4-3)}$$

where Weight_i (i = 1 to 8) represents the sample weight at each time interval.

$$\text{Added water} = \text{GMC}_{(low/medium/high)} \times \text{Weight}_{Dry} \quad \text{(Equation 4-4)}$$

where $\text{GMC}_{(low)}$ = 20%, $\text{GMC}_{(medium)}$ = 40%, $\text{GMC}_{(high)}$ = 60%.

However, hysteresis, which refers to differences in soil water distribution at a given moisture content depending on whether the soil has reached that state by drying from a wetter condition or by rewetting from a drier one, may influence both MIR and NIR spectral responses through changes in surface hydration and light scattering. The rewetting procedure applied to the Rewet-datasets was designed to obtain corresponding moist MIR and NIR spectra across a wide range of known moisture levels rather than to follow specific drying or wetting paths. Although an even distribution of water within the soil matrix was considered during sample preparation to minimise spatial variability, this does not eliminate the potential effects of hysteresis, which result from differences in soil water configuration between drying and wetting processes. The potential impact of hysteresis is therefore recognised as a source of uncertainty and should be examined in future investigations.

4.2.5 External parameter orthogonalisation (EPO)

The EPO algorithm aims to eliminate the influence of external environmental parameters on reflectance spectra[269], specifically the removal of moisture effects from soil spectra. The EPO algorithm identifies spectral regions affected by soil moisture and then projects the new reflectance spectra orthogonally to these variations[266,269,273].

Moist soil spectra (**X**) can be decomposed into a combination of a useful part (**XP,** related to chemical responses), an external part (**XQ**, containing irrelevant information affected by water content and other external parameters), and an independent residual (**R**) (Equation 4-5).

$$\mathbf{X} = \mathbf{XP} + \mathbf{XQ} + \mathbf{R} \quad \text{(Equation 4-5)}$$

The procedure to find **XP** is through spectra matrix **D**, which is the difference between the soil spectra matrix with and without external influence. **Q** is estimated through singular value decomposition of **D**, and **XP** is then calculated as **X**(**I** − **Q**), where **I** is the identity matrix. The EPO method proceeds are as follows:

(1) Calculate **XQ** from the difference (**D**) between the moist and dry soil spectra.

(2) Build a PCA using **XQ**, which captures the variation introduced by moisture, and obtain the corresponding scores (**T**) and loadings ($\mathbf{P}^T$).

(3) Project the moist spectra of new samples ($\mathbf{X}_{new}$) onto the previously calculated loadings ($\mathbf{P}^T$) to obtain the corresponding score values ($\mathbf{T}_{new}$).

(4) Multiply $\mathbf{T}_{new} \cdot \mathbf{P}^T$ to obtain $\mathbf{XQ}_{new}$, which is the part of the soil spectra that is caused by moisture.

(5) Based on the assumption that **R** is irrelevant, the useful part of the spectra ($\mathbf{XP}_{new}$) can be calculated by subtracting the original spectra $\mathbf{X}_{new}$ from $\mathbf{XQ}_{new}$. Machine learning algorithms are applied using the $\mathbf{XP}_{new}$ matrix.

4.2.5.1. Development of *specific* vs. *general* EPO algorithms

The difference matrix, **D**, for the EPO algorithm was examined using two approaches, namely EPO models under (1) known soil moisture conditions, referred to as '*specific'* EPO models, and (2) unknown soil moisture conditions, referred to as the '*general'* EPO model.

To construct *specific* EPO models, **D** was calculated for wet soil spectra relative to the dry soil spectra at each of the eight moisture levels (R1 to R8). Consequently, eight *specific* EPO models were developed, each designed to correct wet soil spectra for a specific moisture condition. However,

applying *specific* EPO transformations to appropriate moisture ranges presents practical challenges, such as requiring additional soil moisture data, often obtained through field-based moisture probes. In practical applications, soils typically exist at varying moisture levels, making it essential to address this issue.

To explore the potential of constructing one *general* EPO model capable of correcting soil spectra across various moisture levels without prior knowledge of soil moisture conditions, samples from the Rewet-datasets measured at eight soil moisture levels were combined, totalling 480 samples. A single **D** was calculated for wet spectra relative to the dry soil spectra across the eight moisture levels.

4.2.5.2. Determination the number of components (*g*)

The number of EPO components (***g***) is an important parameter during EPO development. To determine ***g***, two methods were used in this study: (1) minimised RMSEP and (2) Wilk's Λ.

Following the development of calibration models from the Smax-datasets, ***g*** was tested from 1 to 30, generating 30 transformation matrices (**P**) along with corresponding recalibrated Smax-models and transformed moist spectra. The optimum value of ***g*** was determined based on the corresponding minimum RMSEP.

Wilks' Λ is a metric for cluster separation, representing the ratio of inter-group variance to total variance (Equation 4-6). It provides a straightforward method for determining ***g*** by measuring the similarity of the transformed spectra. A value of 0 for Λ reveals a null separation (all the centres of mass are confounded). A value of 1 for Λ expresses a perfect separation (each group exactly matches its centre of mass).

$$\text{Wilk}'\text{s } \Lambda = \frac{\text{Trace (B)}}{\text{Trace (T)}} \qquad \text{(Equation 4-6)}$$

where **B** is the inter-group variance-covariance matrix of the aggregated transformed spectra (i.e. EPO transformed spectra) and **T** is the variance-covariance matrix of the EPO transformed spectra.

A detailed flowchart illustrating the experimental procedures for the Rewet-datasets and Validation-datasets and the implementation of *specific* and *general* EPO transformation procedures are presented in Figure 4-3.

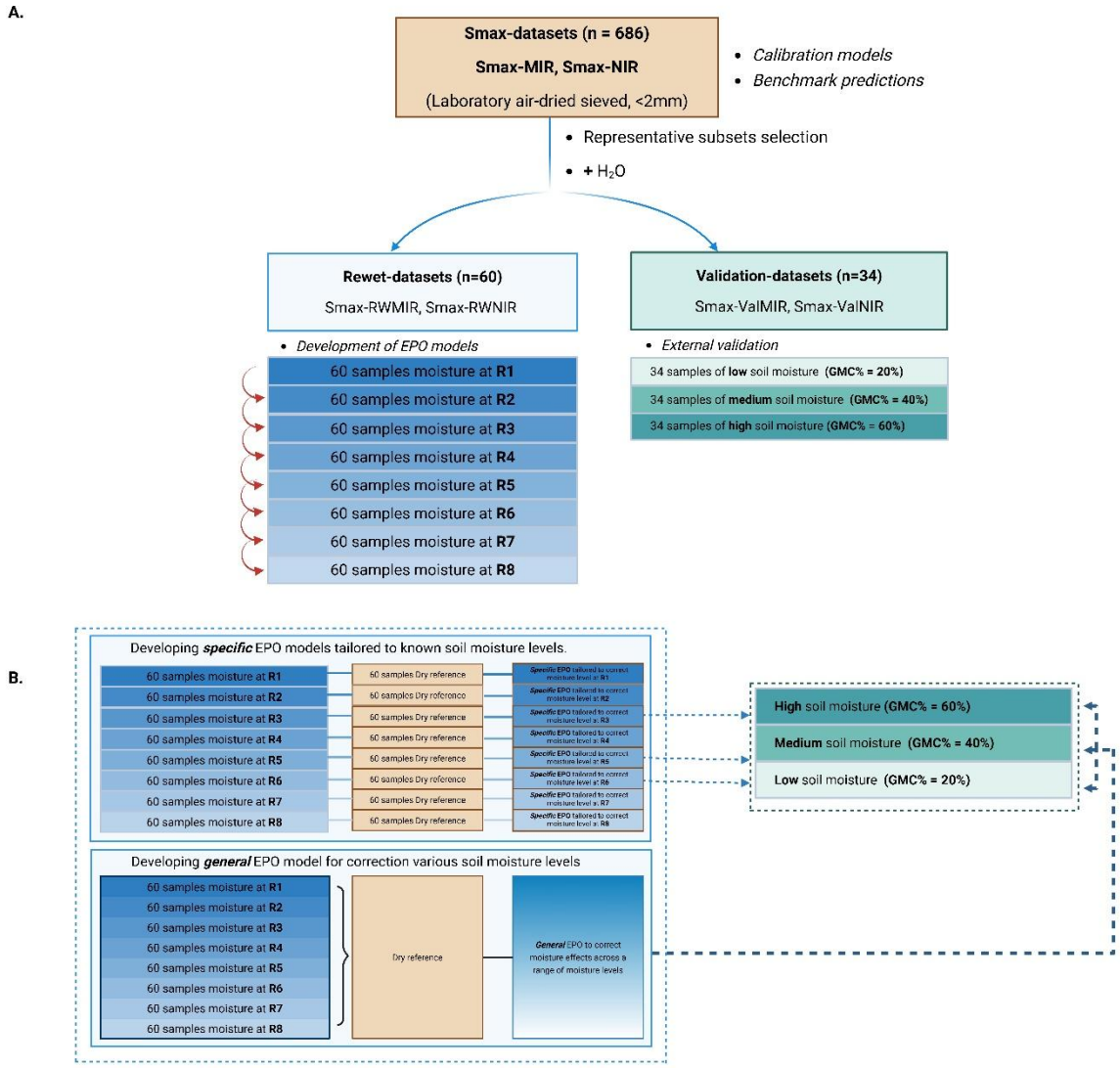


Figure 4-3. Flowchart illustrating: (A) the composition and function of the three datasets (Smax-dataset, Rewet-datasets, and Validation-datasets); and (B) the development of specific and general EPO models using the Rewet-datasets and their application to the Validation-datasets.

4.3 Results

4.3.1 Development of calibration models for Smax prediction and Smax classification

4.3.1.1 Regression models

Figure 4-4 illustrates results of the best-performing algorithms for Smax-MIR and Smax-NIR. MIR spectral regions affected by atmospheric CO_2 signals (2398 – 2281 cm^{-1}) were excluded. The cubist-based model for predicting Smax from handheld MIR of dried and sieved samples established in our previous study (Chapter 3) was applied to the raw Smax-MIR spectra (R_{Val}^2 = 0.81, RMSEP = 63.17 $mg \cdot kg^{-1}$, RPIQ = 3.26). For the Smax-NIR, low-level spectral merging by Treatments 1, concatenating SG1-preprocessed spectra from the Inno (900 – 1650 nm) and Neo (1650 – 2500 nm), was identified as the most effective merging strategy, achieving the highest R_{Val}^2 of 0.69 and lowest RMSEP of 65.66 $mg \cdot kg^{-1}$. Yet, according to the regression model quality standards proposed by Ludwig et al. (2017)[385], even the best-performing NIR model remains classified as a 'Non-Reliable Model' (NRM), with RPIQ < 2.02, indicating limitations in its predictive reliability.

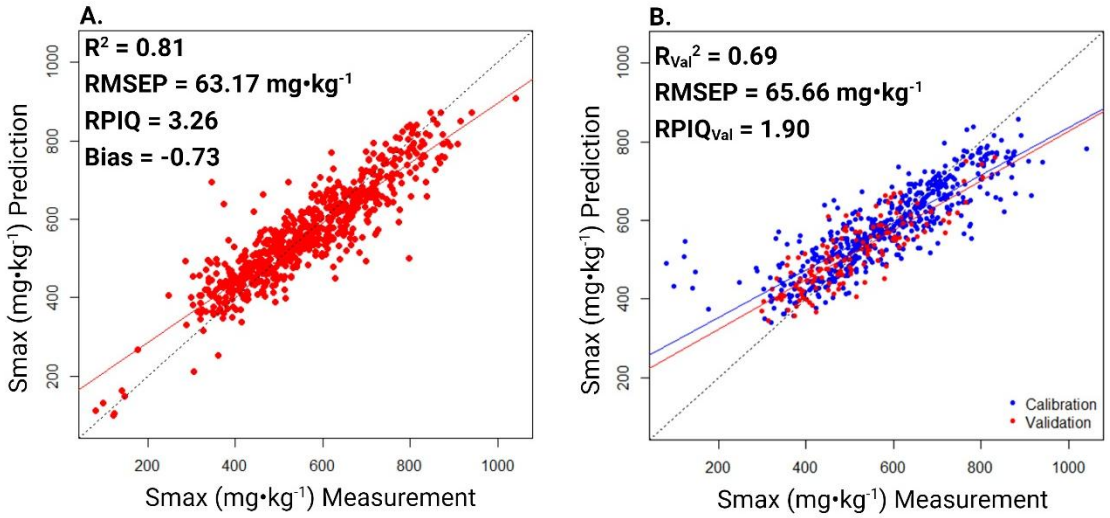


Figure 4-4. Measured and predicted Smax values using Smax-datasets. (A) Results for Smax prediction using Smax-MIR raw spectra with calibration models developed in Chapter 3. (B) Results for Smax prediction using Smax-NIR (low-level spectral merging by Treatment 1) using SVM algorithm. The 1:1 line is plotted in black in each plot. Regression lines are shown in blue for calibration samples and red for validation samples.

4.3.1.2 Classification models

Given that the MDKS algorithm systematically encompasses spectral variations across all samples, Smax-datasets were divided into a training set (n = 548, SL:SH = 115:433) and a test set (n = 138, SL:SH = 29:109). For Smax-MIR, applying the SVM-based classification model established in Chapter 3 yielded an MCC of 0.65 (κ = 0.62, F1-score = 0.94). For Smax-NIR, the best-performing RF-based algorithm achieved MCC_{Test} of 0.63 (κ_{Test} = 0.63, $F1\text{-}score_{Test}$ = 0.92).

Detailed results for all regression and classification algorithms applied to Smax-NIR spectroscopy (including low- and mid-level merging approaches) are provided in the Table 4-2 and Table 4-3.

Table 4-2. Summary of the accuracy parameters for each regression model performing on the Smax-NIR (dried, sieved) spectral library.

Spectral datasets	Regression modelling algorithms	Calibration				Validation			
		R^2_{Cal}	RMSE	RPD_{Cal}	$RPIQ_{Cal}$	R^2_{Val}	RMSEP	RPD_{Val}	$RPIQ_{Val}$
Individual detector									
	PLSR	0.25	125.65	0.59	0.77	0.34	118.2	0.52	0.66
Raw Inno	**Cubist**	0.51	103.49	0.83	1.11	**0.43**	**108.74**	**0.77**	**0.93**
	SVM	0.06	141.41	0.23	0.25	0.08	137.12	0.20	0.24
	PLSR	0.29	122.38	0.64	0.87	0.31	118.87	0.57	0.75
SG1 Inno	Cubist	0.16	135.27	0.28	0.30	0.18	134.70	0.18	0.30
	SVM	0.50	105.05	0.79	1.11	**0.44**	**108.14**	**0.72**	**0.87**
	PLSR	0.57	96.66	1.15	1.58	0.56	85.82	1.02	1.37
Raw Neo	**Cubist**	0.64	92.39	0.98	1.44	**0.68**	**77.31**	**1.08**	**1.51**
	SVM	0.28	125.07	0.58	0.80	0.43	100.75	0.63	0.96
	PLSR	0.47	107.29	0.94	1.3	0.56	88.92	0.85	1.09
SG1 Neo	Cubist	0.56	102.6	0.77	1.42	0.55	94.3	0.64	0.63
	SVM	0.73	78.69	1.34	1.89	**0.66**	**76.52**	**1.29**	**1.66**
Low-level data merging									
	PLSR	0.59	95.15	1.20	1.67	0.59	76.61	1.08	1.54
Treatment 1	Cubist	0.54	107.23	0.66	1.14	0.52	90.80	0.58	0.75
	SVM	0.74	79.59	1.31	1.91	**0.69**	**65.66**	**1.35**	**1.90**
	PLSR	0.55	99.47	1.10	1.60	0.58	79.68	0.95	1.31
Treatment 2	Cubist	0.53	107.07	0.69	1.28	0.49	94.38	0.57	0.65
	SVM	0.76	75.35	1.42	2.16	**0.65**	**69.31**	**1.30**	**1.94**
Mid-level data merging									
	PLSR	0.57	97.39	1.16	1.60	0.59	73.04	1.08	1.40
Treatment 3	Cubist	0.58	98.18	1.00	1.39	0.55	78.99	0.87	1.09
	SVM	0.87	58.57	1.98	3.01	**0.62**	**72.42**	**1.27**	**1.56**

Table 4-3. Summary of the accuracy parameters for each classification model performing on the Smax-NIR (dried, sieved) spectral library.

Spectral datasets	Classification algorithms	Testing datasets (No. SL/SH Samp.= 29/109)				MCC	κ	F1-score
		TP	FP	TN	FN			
Individual detector								
Raw Inno	RF	108	27	2	1	0.17	0.09	0.89
	SVM	109	29	0	0	0.00	0.00	0.88
	PLS-DA	109	29	0	0	0.00	0.00	0.88
SG1 Inno	RF	105	19	10	4	0.42	0.38	0.90
	SVM	107	15	14	2	**0.59**	**0.56**	**0.93**
	PLS-DA	100	18	11	9	0.34	0.33	0.88
Raw Neo	RF	97	18	11	12	0.29	0.29	0.87
	SVM	109	27	2	0	0.24	0.10	0.89
	PLS-DA	109	26	3	0	0.29	0.15	0.89
SG1 Neo	RF	106	13	16	3	0.62	0.60	0.93
	SVM	105	12	17	4	**0.62**	**0.61**	**0.93**
	PLS-DA	109	26	3	0	0.29	0.15	0.89
Low-level data merging								
Treatment 1	**RF**	100	8	21	9	**0.63**	**0.63**	**0.92**
	SVM	92	6	23	17	0.57	0.56	0.89
	PLS-DA	109	22	7	0	0.45	0.33	0.91
Treatment 2	**RF**	99	7	22	10	**0.64**	**0.64**	**0.92**
	SVM	93	6	23	16	0.58	0.57	0.89
	PLS-DA	102	11	18	7	0.59	0.59	0.92
Mid-level data merging								
Treatment 3	**RF**	108	12	17	1	**0.70**	**0.67**	**0.94**
	SVM	95	12	17	14	0.45	0.45	0.88
	PLS-DA	109	19	10	0	0.54	0.45	0.92

4.3.2 Effects of soil moisture on reflectance spectra

Moist MIR and NIR spectra from the Rewet-datasets were collected separately at eight moisture levels (Figure 4-5.A). For Smax-RWNIR, the median soil moisture contents at the eight moisture levels (R1 – R8) were 96%, 84%, 70%, 58%, 45%, 33%, 20%, and 8%. For Smax-RWMIR, the median soil moisture contents at the eight moisture levels (R1 – R8) were 94%, 82%, 68%, 53%, 36%, 18%, 3%, and 1%. In Smax-RWNIR, increased water content predominantly led to increase in total absorbance over the entire spectral ranges, particularly around 1450 nm and 1930 nm (Figure 4-5.B, C, and E). Additionally, the absorption coefficient at 1930 nm was highest in the NIR region, approximately three times greater than at 1450 nm, consistent with Hale and Querry (1973) [404]. Moisture levels R1–R4 exhibited overlapping average spectra, while from R5 (median GMC = 45%) onward, spectra became more distinct with decreasing moisture content. Soil moisture significantly distorted the overall shapes of the MIR spectra (Figure 4-5.D), with wet Smax-RWMIR spectra showing similarities at high moisture levels yet notable deviations from the dried soil reference spectra (red dashed line). For example, the average spectra for R1 – R4 displayed strong absorption peaks around 3700–3200 cm^{-1} and 1650 cm^{-1}. By R5 (median GMC = 36%), the spectra began resembling the reference spectra. From R6 (median GMC = 18%) onward, the absorption features (e.g., 3730 – 3610 cm^{-1}, 1950 – 1750 cm^{-1}, and 1750 – 1210 cm^{-1}) aligned with the reference spectra.

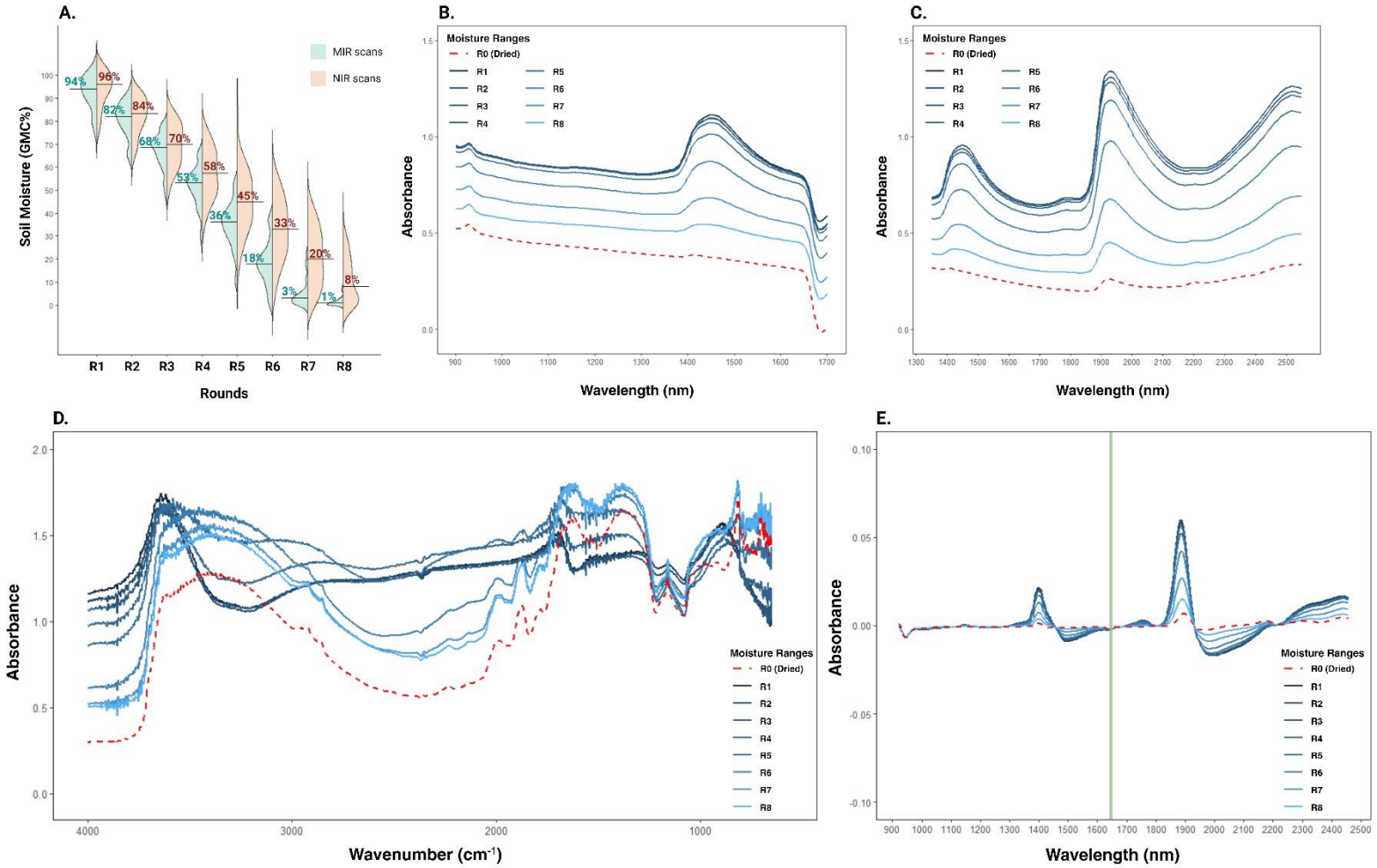
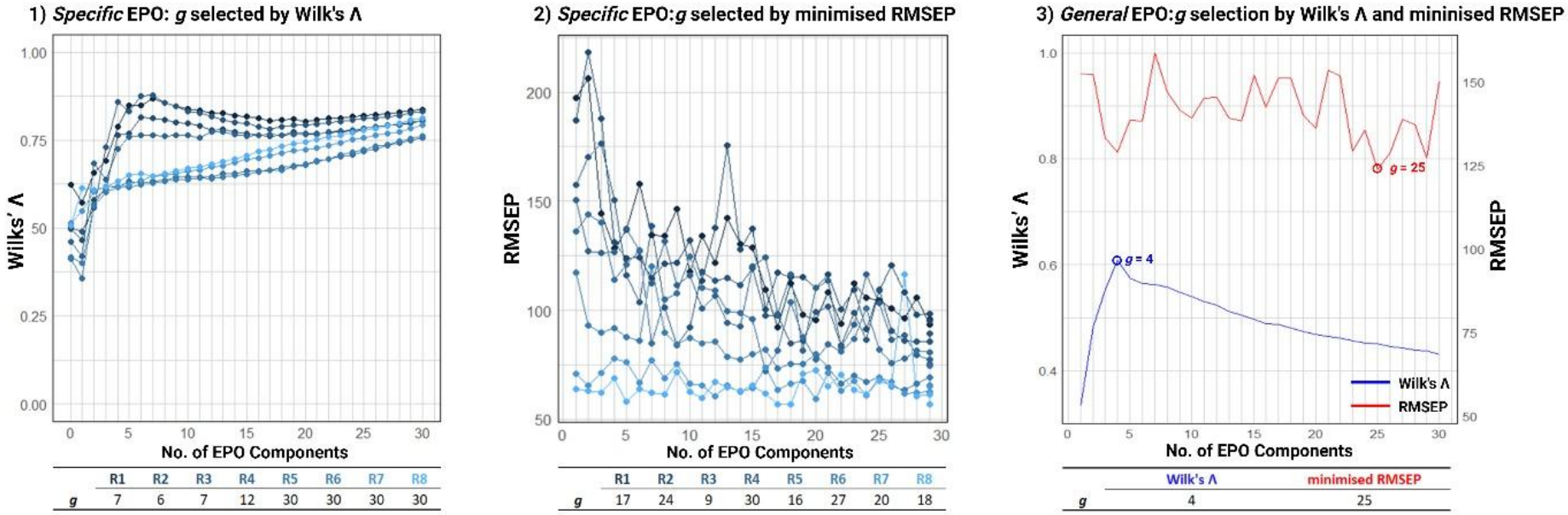


Figure 4-5. (A) Soil moisture values during the rewetting experiment (Rewet-datasets), with the median GMC values for NIR and MIR scans. (B, C) Mean absorbance spectra at different moisture levels for partial NIR regions scanned by Inno and Neo devices, respectively. (E, F) Mean absorbance spectra for Smax-RWMIR and Smax-RW-NIR (obtained using the low-level data merging approach following Treatment 1).

4.3.3 EPO for moisture correction

The optimal ***g*** determination processes and the *specific*/*general* EPO pre-processing algorithms for spectra are demonstrated using their application to Smax-RWMIR at R1 (median GMC of 94%) as an example. Figure 4-6 illustrates the optimal ***g*** selection by (i) Wilk's Λ and (ii) minimised RMSEP for *specific* and *general* EPO for Smax-RWMIR. At R1, the ***g*** values for the *specific* EPO selected based on Wilk's Λ and minimised RMSEP were 7 and 17, respectively.

Figure 4-7 shows the Smax-RWMIR at R1 moisture level (***X*1**). EPO transformed the moist spectra into a space orthogonal to the moisture effect (***X*1***), expressed as: ***X*1*** = ***X*1•P**, where **P** is the transformation matrix. ***X*1*** was processed similarly to the recalibrated Smax-MIR spectra. **P** in Figure 4-7 (a) and (b) highlighted the wavelengths affected by moisture, with large positive (yellow) or negative (purple) values indicating the spectral regions transformed to correct for moisture effects. For example, the three highlighted regions (~3400 cm^{-1}, ~1645 cm^{-1}, and ~2150 cm^{-1}) were correlated with distinctly increased water absorption, corresponding to fundamental vibrations of water. These included a broad feature at ~3400 cm^{-1} (symmetric and asymmetric H–O–H stretching), a peak at ~1645 cm^{-1} (H–O–H bending), and a weaker feature at ~2150 cm^{-1}, attributed to the combination of H–O–H bending and water liberation [237].



	R1	R2	R3	R4	R5	R6	R7	R8
g	7	6	7	12	30	30	30	30

	R1	R2	R3	R4	R5	R6	R7	R8
g	17	24	9	30	16	27	20	18

	Wilk's Λ	minimised RMSEP
g	4	25

Figure 4-6. Optimal ***g***-values for (1, 2) *specific* and (3) *general* EPO determined using Wilk's Λ and minimised RMSEP. Larger Wilk's Λ indicates better sample separation across moisture levels. Smaller RMSEP reflects better ***g***-selection for predicting soil Langmuir Smax using the recalibrated calibration model.

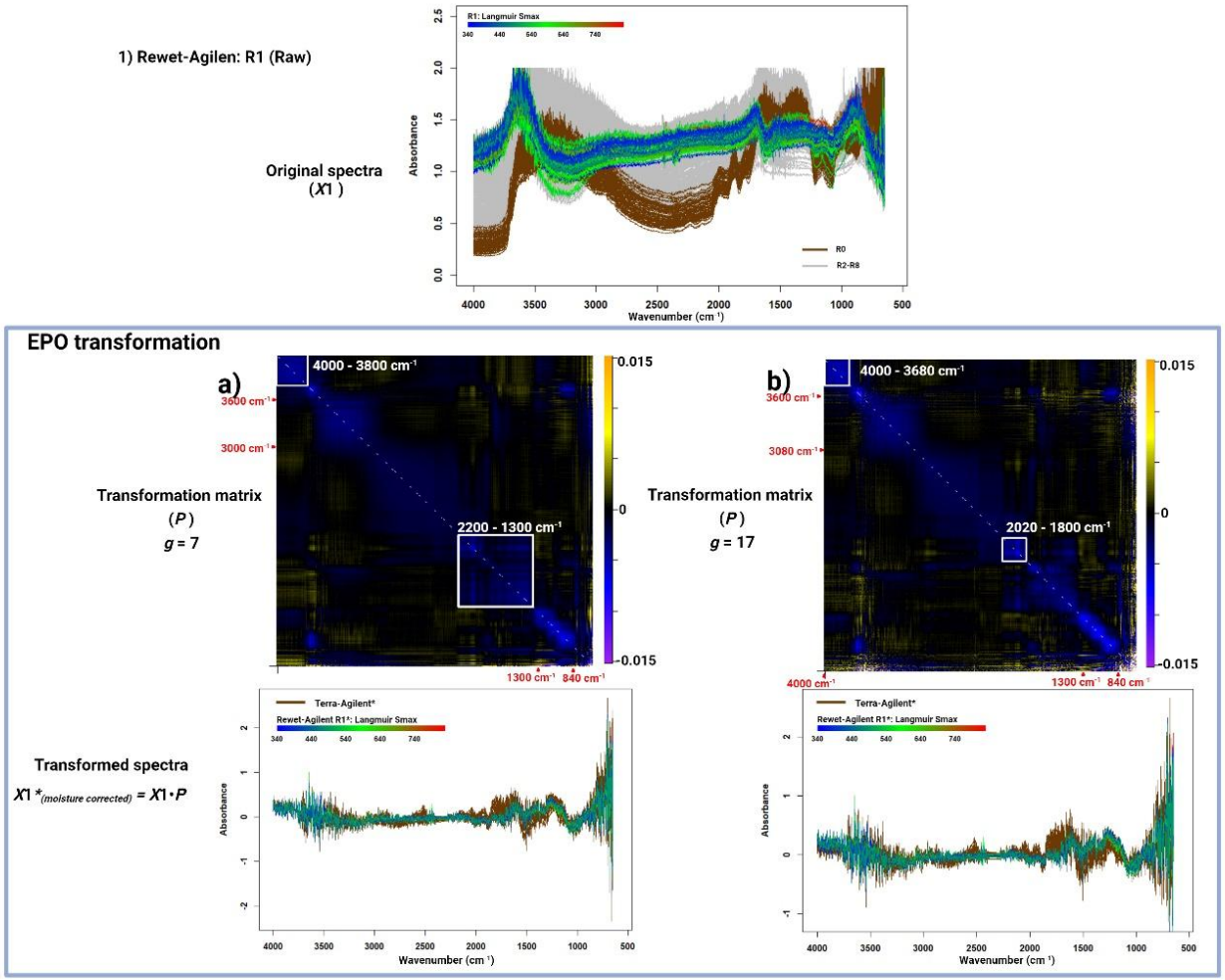
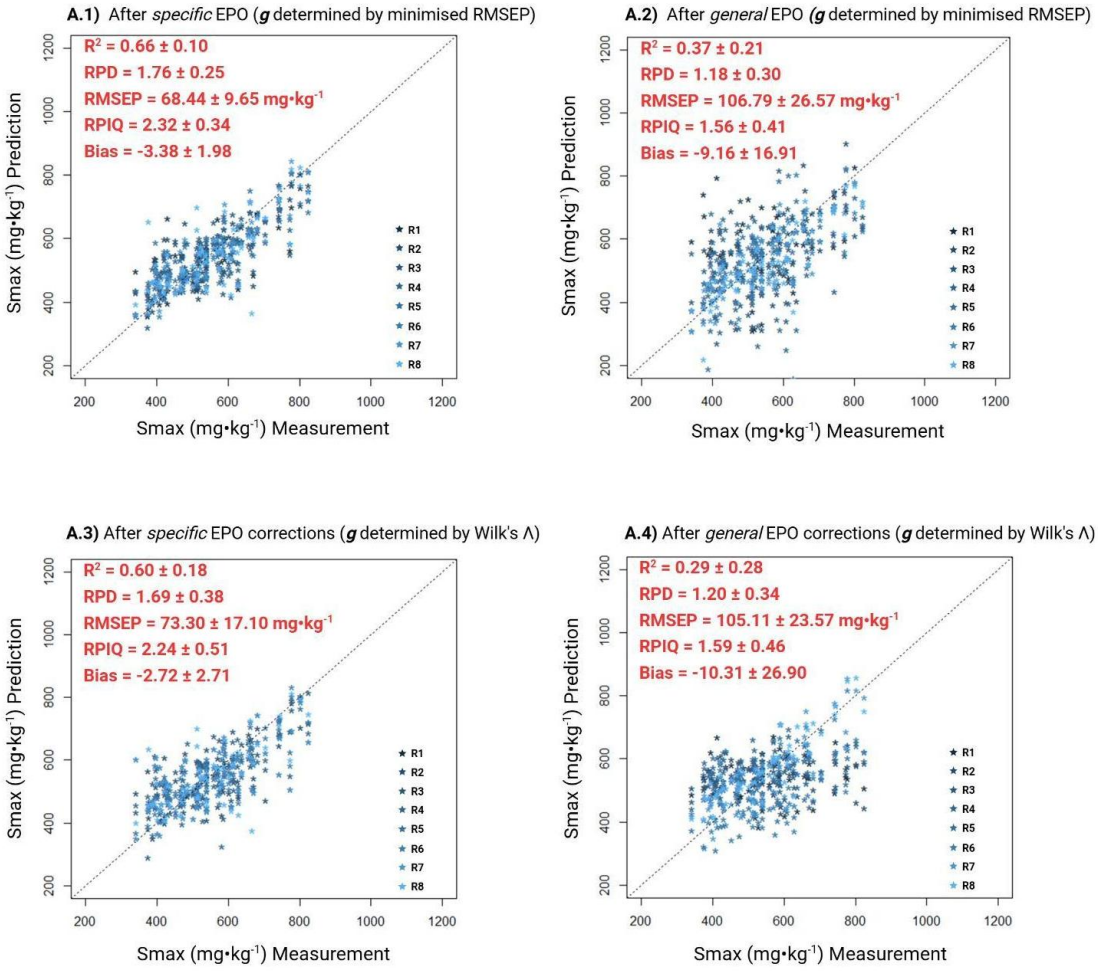


Figure 4-7. EPO algorithm applied to the Smax-RWMIR at R1 spectra. (a) ***g*** determined by Wilk's Λ and (b) ***g*** determined by minimised RMSEP. The axes in projection matrix (***P***) represent the indices of wavenumbers affected by moisture. Regions with large positive or negative values correspond to spectral areas influenced by moisture, which will be transformed.

Results from EPO transforming within MIR region suggest a preference for ***g*** determined by minimised RMSEP method over Wilk's Λ for both *specific* and *general* EPO algorithms (Figure 4-8.A). A key difference between the two ***g*** determination methods was the incorporation of recalibrated Smax-datasets model validation on the transformed spectra in the minimised RMSEP approach. This made ***g*** selection dependent on (i) the interaction between EPO process and model validation, and (ii) the response variable of the Smax values. In contrast, Wilks' Λ determination produced a single ***g*** value, regardless of these factors. The minimised RMSEP approach better balanced noise reduction and preservation of essential spectral information. Therefore, the subsequent development of EPO focused on ***g*** values determined by the minimised RMSEP method. Figure 6.B summarises the prediction results of Smax-RWNIR for soil Smax values following the application of *specific* and *general* EPO transformations.

A. Smax-RWMIR



B. Smax-RWNIR

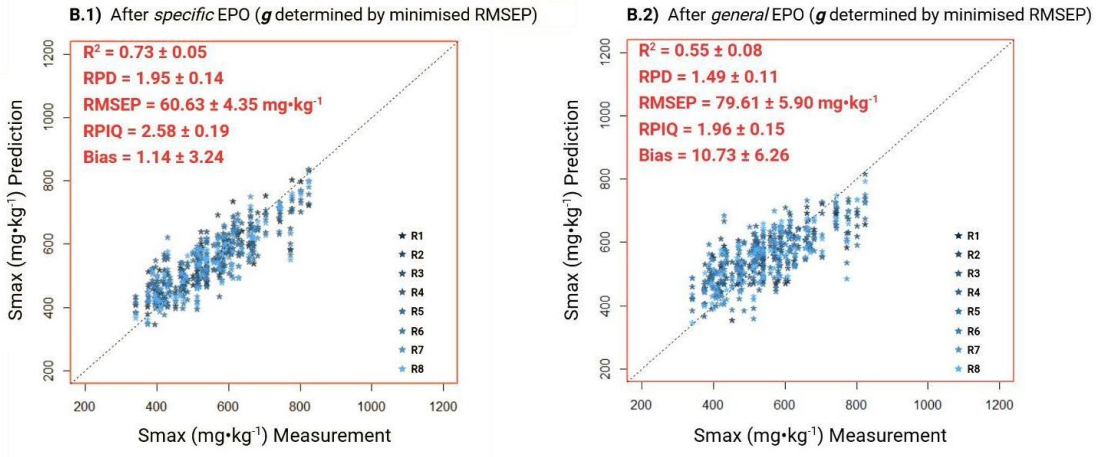


Figure 4-8**.** (A) Regression plots for soil Langmuir Smax prediction for Smax-RWMIR at moisture levels R1–R8 *specific* and *general* EPO corrections, comparing *g* values determined by Wilk's Λ and minimised RMSEP. (B) Regression plots for soil Langmuir Smax prediction for Smax-RWNIR at moisture levels R1-R8 after *specific* and *general* EPO corrections (***g*** values determined by minimised RMSEP). Annotated indicators represent the mean ± standard deviation.

4.3.4 Comparison of specific and general EPO algorithms

Table 4-4 summarises the comparison of the efficiency of *specific* and *general* EPO (***g*** determined by minimised RMSEP) moisture corrections for Smax-RWMIR and Smax-RWNIR.

Specific EPO consistently outperformed *general* EPO across Smax-RWMIR and Smax-RWNIR for moist spectra transformation. Application of *general* EPO resulted in less effective moisture correction, particularly in the Smax-RWMIR, where the average RMSEP was 106.79 mg·kg^{-1} after *general* EPO correction, compared to an average RMSEP of 68.44 mg·kg^{-1} with *specific* EPO correction. These results indicate that applying *specific* EPO, tailored to known soil moisture conditions, can be essential in improving the predictive accuracy of MIR wet spectra, achieving accuracy potentially comparable to that of dry soil samples.

Under *general* EPO transformations, the advantages of NIR over MIR in wet soil spectroscopy were demonstrated. The poor performance of Smax-RWMIR (average of R^2 = 0.37, RMSEP = 106.79 mg·kg^{-1}) emphasises the importance of *specific* EPO transformation in addressing moisture-related variability. In contrast, although prediction accuracy decreased for Smax-RWNIR, the decreases were less pronounced, only a slight increase in RMSEP was observed (from 60.63 mg·kg^{-1} to 79.61 mg·kg^{-1}), highlighting the superior potential of the full NIR range for enhanced robust prediction accuracy in wet soil analysis.

Table 4-4. Summary of prediction results (average [min to max] across R1 to R8) for *specific* and *general* EPO transformations (with ***g*** determined by minimised RMSEP) applied to wet MIR and NIR spectra from the Rewet-datasets. Smax-RWNIR includes concatenated SG1 pre-processed Inno/Neo spectra following low-level merging (Treatment 1) for a full NIR region.

Rewet-datasets	**Accuracy metrics (mean, (min-max))**									
	After *specific* EPO transformations (R1–R8) with optimum *g*					**After *general* EPO transformation with optimum *g***				
	R^2	**RMSEP**	**RPD**	**RPIQ**	**Bias**	**R^2**	**RMSEP**	**RPD**	**RPIQ**	**Bias**
Smax-RWMIR	0.66 (0.52 – 0.66)	68.44 (56.13 – 82.33)	1.76 (1.43 – 2.10)	2.32 (1.88 – 2.78)	-3.38 (-6.15 – 0.51)	0.37 (0.04 – 0.68)	106.79 (71.76 – 145.06)	1.18 (0.81 – 1.65)	1.56 (1.07 – 2.25)	-3.38 (-6.15 – 0.51)
Smax-RWNIR	0.73 (0.69 – 0.82)	60.63 (53.90 - 65.23)	1.95 (1.80 – 2.18)	2.58 (2.38 – 2.87)	1.14 (-3.87 – 6.16)	0.55 (0.42 – 0.67)	79.61 (71.92 – 88.57)	1.49 (1.33 – 1.64)	1.96 (1.75 – 2.15)	10.73 (-1.61 – 16.77)

4.3.5 Classification models in tandem with (specific and general) EPO transformation

Table 4-5 presents the metrics derived from the confusion matrix for wet MIR and NIR based on the recalibrated classification models after *specific* and *general* EPO (***g*** determined by minimised RMSEP) transformations.

Results indicate the significantly greater efficiency of *specific* EPO algorithms over *general* EPO in improving classification accuracies. *Specific* EPO algorithms consistently outperformed *general* EPO algorithms across all datasets, especially for the Smax-RWMIR (MCC = 0.53 *vs*. 0.27 for *specific* and *general* EPO). After *specific* EPO transformations, the recalibrated model applied to the Smax-RWMIR achieved a comparable high accuracy of SL/SH soil classification to that of dry MIR spectroscopy (MCC = 0.60). Under *general* EPO transformations, Smax-RWNIR (average MCC of 0.45) demonstrated greater accuracy and robustness for SL/SH soil classifying compared to MIR (average MCC of 0.27)**.** The observations emphasise the superior effectiveness of the full NIR range in enhancing classification reliability under complex (unknown) moist conditions.

Table 4-5. SL/SH soils classification accuracies (average [min to max] across R1 to R8) using the recalibrated model after *specific* the *general* EPO transformations of the Rewet-datasets with optimal ***g*** from minimised RMSEP. Smax-RWNIR includes concatenated SG1 pre-processed Inno/Neo spectra following low-level merging (Treatment 1) for a full NIR region.

Spectral datasets	**Accuracy metrics (mean, (min-max))**					
	After *specific* EPO transformations (R1–R8) with optimum *g*			**After *general* EPO transformations (R1–R8) with optimum *g***		
	MCC	**κ**	**F1-score**	**MCC**	**κ**	**F1-score**
Smax-RWMIR	0.53 (0.42 – 0.61)	0.48 (0.29 – 0.61)	0.91 (0.89 – 0.92)	0.27 (-0.23 – 0.65)	0.27 (-0.22 – 0.65)	0.85 (0.76 – 0.92)
Smax-RWNIR	0.52 (0.36 – 0.60)	0.48 (0.34 – 0.56)	0.91 (0.88 – 0.92)	0.45 (0.05 – 0.60)	0.42 (0.04 – 0.56)	0.89 (0.83 – 0.92)

4.3.6 Validation

Based on the shifts in trends of the wet MIR and NIR spectra in Figure 4-5, the wet MIR and NIR spectra for the Validation-datasets were rewetted to 'low', 'medium', and 'high' moisture levels, corresponding to GMC values of 20%, 40%, and 60% (Figure 4-9).

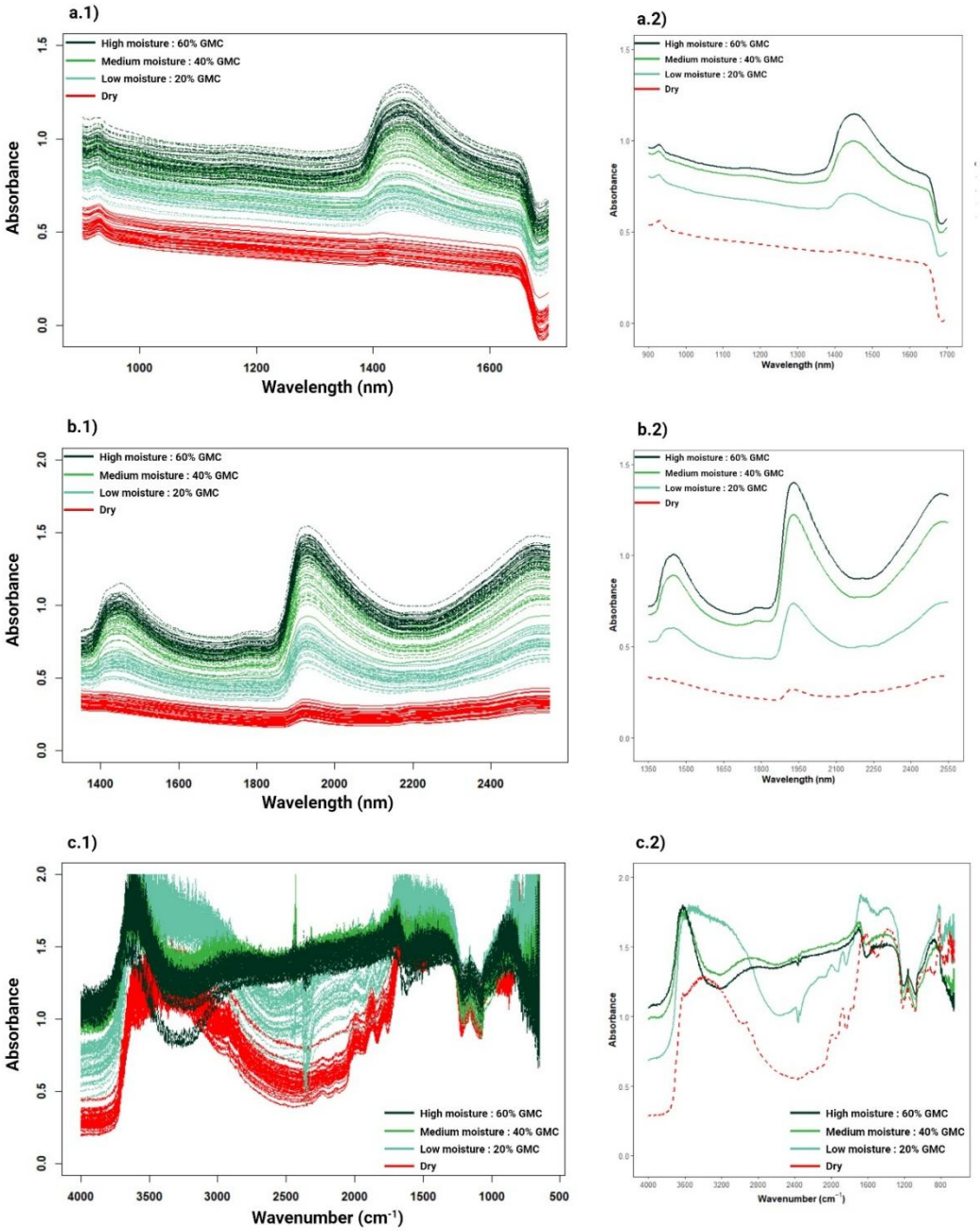


Figure 4-9. Scanned spectra for Validation-datasets at different moisture levels (air-dried, low moisture: 20% GMC, medium moisture: 40% GMC and high humidity: 60% GMC). Panel (a.2), (b.2) and (c.2) show the mean absorbance spectra for Inno, Neo and Agilent scans at the respective moisture levels.

4.3.6.1 Validation of *specific* and *general* EPO transformations for wet IR spectra in Smax prediction

Specific EPO and *general* EPO models, both with optimal ***g*** determined by minimised RMSEP, were validated for moisture correction in Smax prediction using Smax-ValMIR and Smax-ValNIR (Table 4-6).

For the Smax-ValMIR, *specific* EPO demonstrated superior moisture correction efficiency at the low moisture level (RMSEP = 86.65 $mg \cdot kg^{-1}$, RPIQ = 3.22). However, its efficiency decreased at the medium moisture level, where, after *specific* EPO correction, it only showed a potential to discriminate SL/SH soils (RPIQ = 2.29). For wet MIR spectroscopy at the high moisture level, Smax-ValMIR predicted Smax more accurately with a *general* EPO transformation (RMSEP = 133.18 $mg \cdot kg^{-1}$, RPIQ = 2.10) than with a *specific* EPO transformation (RMSEP = 254.80 $mg \cdot kg^{-1}$, RPIQ = 2.09). Nevertheless, for wet soils beyond a low moisture level, the application of EPO corrections to MIR spectroscopy remained inadequate for accurate Smax prediction. In contrast to MIR, the differences between *specific* and *general* EPO algorithms were less pronounced for Smax-ValNIR and demonstrated higher consistency across moisture levels.

Table 4-6. Results for recalibrated models predicting soil Smax values on the Validation-datasets with *specific* and *general* EPO (optimal ***g*** determined by minimised RMSEP) transformations.

Application of the recalibrated Smax-MIR and Smax-NIR models after *specific* and *general* EPO transformation												
Validation-dataset moisture levels	***SPECIFIC* EPO transformations**						***GENERAL* EPO transformations**					
	g [1]	**R^2**	**RMSEP**	**RPD**	**RPIQ**	**Bias**	***g***	**R^2**	**RMSEP**	**RPD**	**RPIQ**	**Bias**
Smax-ValMIR												
Low, GMC = 20%	27	0.75	86.65	1.80	3.22	-3.74		0.53	107.58	1.45	2.60	18.79
Medium, GMC = 40%	16	0.46	124.69	1.26	2.29	-29.51	25	0.24	136.74	1.15	2.09	-19.40
High, GMC = 60%	9	0.38	254.80	0.61	1.10	-223.86		0.27	133.18	1.17	2.10	14.89
Smax-ValNIR												
Low, GMC = 20%	29	0.50	120.12	1.30	2.32	-18.14		0.45	124.07	1.26	2.25	17.20
Medium, GMC = 40%	28	0.29	131.80	1.19	2.12	10.76	26	0.34	132.43	1.18	2.11	12.48
High, GMC = 60%	19	0.56	124.19	1.26	2.25	-0.49		0.48	127.19	1.23	2.20	9.54

[1]: The optimal ***g*** for the *specific* EPO applied to the Validation-datasets at low, medium, and high moisture levels corresponds to the *specific* EPO developed at R3, R5, and R7 for Smax-RWMIR and Smax-RWNIR.

4.3.6.2 Validation of *specific* and *general* EPO transformations for wet IR spectra in SL/SH classify

Table 4-7 presents the classification performance of calibration models for distinguishing SL/SH soils in the Smax-ValMIR and Smax-ValNIR after *specific* and *general* EPO (optimal ***g*** determined by minimised RMSEP) transformations.

Consistent with the findings for Smax predictions, *specific* EPO transformations significantly mitigated the effects of soil water on classification modelling compared to *general* EPO transformations. Furthermore, the accuracy achieved with Smax-ValNIR was less reliable than that obtained with Smax-ValMIR.

MIR is also not recommended for discriminating SL/SH soils in high moisture soils. *Specific* EPO showed better moisture correction efficiency at the low (MCC = 0.44) and medium (MCC = 0.45) moisture levels. However, at the high moisture level, *specific* EPO transformed MIR spectra were less effective for moisture correction. In contrast, *general* EPO corrected for Smax-ValMIR demonstrated lower classification accuracy at low (MCC = 0.36) and medium (MCC = 0.33) moisture contents but showed better consistency across all moisture levels (MCC = 0.31 ± 0.06). However, at the high moisture level, the application of *general* EPO to Smax-ValMIR also resulted in poor SL/SH soil classification (κ=0.12).

Table 4-7. Results for the recalibrated classifying soil P dynamics classes (SL/SH) on the Validation-datasets with *specific* and *general* EPO transformations.

Application of the recalibrated Smax-MIR and Smax-NIR models after *specific* and *general* EPO transformation								
Validation-dataset moisture levels	***SPECIFIC* EPO transformations**				***GENERAL* EPO transformations**			
	g	**MCC**	**κ**	**F1-score**	***g***	**MCC**	**κ**	**F1-score**
Smax-ValMIR								
Low, GMC = 20%	27	0.44	0.44	0.83	25	0.36	0.35	0.82
Medium, GMC = 40%	16	0.45	0.33	0.85		0.33	0.31	0.82
High, GMC = 60%	9	0.00	0.00	0.81		0.25	0.12	0.82
Smax-ValNIR								
Low, GMC = 20%	29	0.33	0.33	0.78	26	- 0.06	- 0.05	0.75
Medium, GMC = 40%	28	0.42	0.37	0.85		0.27	0.26	0.80
High, GMC = 60%	19	0.36	0.23	0.84		0.25	0.22	0.81

4.4 Discussion

4.4.1 Handheld NIR spectroscopy for soil Smax prediction

In this study, Smax was predicted using NIR coupled with the SVM algorithm on dried, sieved soil (R^2_{Val} = 0.69, RMSEP = 65.66 mg·kg^{-1}, $RPIQ_{Val}$ = 1.90), achieving a similar accuracy to that reported by Saidi et al. (2025)[176] (R^2_{Val} = 0.65, $RPIQ_{Val}$ = 1.57) in dried, sieved soils using Vis-NIR (350–1000 nm). Additionally, our results were comparable to other studies that have predicted similar soil P adsorption indices, such as P remaining in solution, as reported by Ramaroson et al. (2023)[192] using NIR (1100–2498 nm) on dried, finely ground soils (R^2_{Val} = 0.65). Similar to MIR coupled with cubist for Smax prediction, SVM demonstrated strong performance as a non-linear technique for capturing the complex relationships between Smax and NIR spectral data. The superior performance of handheld MIR over NIR for Smax prediction (R^2 = 0.81, RMSEP = 63.17 mg·kg^{-1}, RPIQ = 3.26) indicated that MIR generally provided more accurate predictions than NIR for most soil properties[150]. Although the result for Smax-NIR showed a less reliable estimations for Smax ($RPIQ_{Val}$ < 2.02)[385], the RF classification model demonstrated reasonable discrimination between SL and SH soils (MCC_{Test} = 0.63, κ_{Test} = 0.63, and F1-$score_{Test}$ = 0.92).

Smax predictions are achieved through secondary correlations with soil properties controlling P sorption with direct NIR absorptions, such as soil organic matter (SOM) (e.g., ~900 nm and 2320 – 2370 nm)[176], and clay minerals (e.g., Al-OH combination band ~2200 nm)[405]. Therefore, fusing partial NIR spectra from Inno and Neo for soil Smax estimation leveraged a broader range of information and improved accuracy. In this study, low-level data mergings (Treatment 1 and 2) outperformed mid-level feature merging (PCA-PC scores), probably because PCA decomposition led to information loss, reducing the predictive performance and robustness of the fused models. Additionally, Treatment 1 was selected to be better than Treatment 2 (Table 4-2). Overtones and combination bands around 1400 nm (e.g., 1358, 1380, 1404, 1462 nm) were critical for Smax prediction[175]. Treatment 1

encompassed the full Inno spectra (900 – 1650 nm), ensuring continuous spectral coverage between 1350 – 1650 nm without artificial gaps, thereby enhancing model stability and interpretability.

4.4.2 The effect of moisture content in the NIR and MIR regions

The overlapping average MIR and NIR spectra in Figure 3 at high water contents (e.g., median GMC > ~50%) indicated that strong water absorption leads to spectral reflectance saturation[400]. For Smax-RWMIR, as moisture decreased to R5 (median GMC = 36%), absorption features associated with clay mineralogies (3730 – 3610 cm^{-1})[406], quartz and clay mineralogy (1950 – 1750 cm^{-1})[406], and organic matter (1750 – 1210 cm^{-1})[406] converged to the dried soil MIR spectra. Using NIR spectroscopy to detect soil components is advisable to avoid water content exceeding 40% - 60% on a dry basis (db)[407,408]. Additionally, to minimise interference when using the region around 1940 nm for detecting other soil components, Li et al. (2006)[407] recommends maintaining the water content of samples below 20% db. According to the changes in Smax-RWNIR and Smax-RWMIR spectral curve shapes across varying soil moisture levels in Figure 4-5, the observations in this study are consistent with the moisture level recommendations.

4.4.3 EPO development and moisture transformation efficiency

The success of *specific*/*geneal* EPO moisture corrections in MIR and NIR, as shown in Figure 6, is consistent with previous findings[266,409]. However, unlike de Santana et al. (2019)[266], where EPO transformation across different moisture levels in NIR spectra (1100–2500 nm) achieved equal accuracy for SOM prediction in dry and wet soils, Smax predictions from Smax-RWMIR and Smax-RWNIR after EPO corrections remained less accurate than those from dried soil spectra (RMSEP = 63.17 $mg \cdot kg^{-1}$ and 65.66 $mg \cdot kg^{-1}$ for Smax-MIR and Smax-NIR on dried sieved soils, respectively). A likely explanation is that Smax is predicted through secondary correlations rather than direct energy absorption, as in SOM. However, the limitation of EPO in maximising the correction of wet soil spectra is consistent with findings from other studies[274,410]. For example, the EPO developed by Mirzaei et al. (2022)[274] for soil organic carbon (OC) prediction using Vis-NIR (350–2500 nm) also showed lower

accuracy than dry-ground models. This may be because EPO assumes that the wet soil spectral matrix consists of three components: a useful part, a moisture-affected part, and an independent residual part. However, in practical applications, the influence of the independent residual component is ignored.

Regarding the critical parameter affecting the EPO transformation efficiency, ***g***, this study suggests a preference for the (model-coupled) minimised of RMSEP over Wilk's Λ selection approach, which is also a method commonly used in other EPO algorithm studies[266,411]. For example, the average RMSEP for Smax-RWMIR across R1–R8 after *specific* EPO transformations was 68.44 mg·kg^{-1} and 73.30 mg·kg^{-1} with ***g*** determined by minimised RMSEP and Wilk's Λ, respectively. This is because using ***g*** determined by minimised RMSEP ensured that wet soil spectra corrected by EPO achieved good quantitative performance by producing smoother transformed spectra and more parsimonious models[279]. However, the finding differs from the statement reported by Wijewardane et al. (2016)[272], who suggested Wilk's Λ as a more efficient approach for determining ***g***. This may be due to the main difference between the two studies, here the calibration models derived from the soil archive were previously established, whereas Wijewardane et al. (2016)[272] adjusted the tuning parameters of the nonlinear models while selecting the model-coupled EPO ***g***. Additionally, the determination of ***g*** was also influenced by different decision-making perspectives. For example, for Smax-RWMIR at R1, Wilk's Λ suggested ***g*** = 7, whereas minimised of RMSEP suggested ***g*** = 17. This study suggested ***g*** = 17 for higher prediction accuracy (minimised RMSEP) after **P** transformation, whereas Wijewardane et al. (2016)[272] suggested applying a lower ***g*** to simplify the transformation and reduce spectral noise.

4.4.4 Efficiencies for specific verse general EPO algorithms

A well-known advantage for EPO is the potential to transform wet soil spectra without knowing the intensity of the water effect[411]. Results from this study also showed that Smax predictions for Smax-RWMIR and Smax-RWNIR after *general* EPO (***g*** determined by minimised RMSEP) transformations achieved average RMSEP values of 106.79 mg·kg^{-1} and 79.61 mg·kg^{-1}, respectively. However, compared

to the moisture correction efficiency of *specific* EPO, particularly for wet MIR (average RMSEP = 68.44 $mg \cdot kg^{-1}$), *general* EPO exhibited significantly lower transformation efficiency and less accuracy.

The better accuracy of *specific* EPO corrections can be explained by their construction using defined moisture levels (R1 to R8), enabling the calculation of a difference matrix (**D**) to capture moisture-related variability and generate projection matrices that effectively eliminate the impact of soil water. In principle, analogous to principal component analysis, the EPO algorithm is an information removal process to remove the spectral information related to moisture. The precise calculation of the **D** ensures that the variability due to moisture is captured and represented in the projection matrices. By projecting the wet spectra orthogonally to moisture-related variability, EPO minimised overlap with spectral features associated with soil properties. Consequently, after aligning spectra orthogonally to specific moisture variability, the *specific* EPO algorithm effectively eliminated moisture interference while preserving critical soil spectral information for predictive modelling.

Therefore, although (*general*) EPO removes systematic errors caused by soil moisture[266,272,273], a finding from this study is that incorporating soil moisture information into the application of *specific* EPO models is crucial for improving correction efficiency, especially in MIR spectroscopy.

4.4.5 Results on the Validation-datasets

Results from the Validation-datasets (Table 4-6 and Table **4**-**7**) confirmed the feasibility of using wet MIR and NIR under relatively medium to low moisture conditions for Smax prediction, consistent with the indications in Figure 4-5. Specifically, for soils at low (GMC = 20%) moisture level, MIR wet spectra pre-processed with corresponding *specific* EPO transformations resulted reliable Smax predictions (R^2 = 0.75, RPIQ = 3.22). For soils at medium (GMC = 40%) moisture level, MIR wet spectra after *specific* EPO transformations resulted reliable SL/SH soil discrimination (MCC = 0.45). Particularly, the feasibility of MIR spectroscopy for predicting soil properties at low moisture levels after moisture correction is consistent with Metzger et al. (2021)[216] and Seidel et al. (2022)[237]. The limitation of EPO in correcting soil moisture at high levels has also been pointed out in[410,412,413]. Mirzaei et al. (2022) [274]

applied EPO corrections to wet Vis-NIR spectra (350-2500 nm) and found that the prediction accuracies for soil OC, clay and calcium carbonate progressively decreased with increasing moisture content and levelled off at higher moisture contents (e.g., >24%).

The high efficiency of the *specific* EPO emphasised the importance of incorporating soil moisture information to determine appropriate corrections. However, the soil moisture levels suggested in this study, based on GMC, are not optimised for rapid in situ measurements[414,415]. Traditionally, GMC can be determined through laboratory-based destructive analysis by oven-drying fresh soil samples. Soriano-Disla et al. (2014)[116] summarised the feasibility of reflectance spectroscopy for predicting GMC, achieving median R^2 values of 0.83 - 0.89 and the achieved highest values of 0.98 (UV-Vis-NIR), 0.97 (NIR), and 0.91 (MIR), though these results were obtained under laboratory conditions. Recent studies also investigated using other sensors (e.g., time domain reflectometer) for estimating soil moisture (with gravimetric data) in relation to gravimetric measurements[415].

In comparison, under unknown soil moisture conditions (with *general* EPO application) or at a high moisture level (GMC = 60%), NIR demonstrated advantages over MIR for wet samples. Smax prediction after applying *general* EPO for Smax-ValMIR and Smax-ValNIR achieved average R^2 = 0.35 and R^2 = 0.42, respectively. When at the high moisture level, Smax prediction after applying *specific* EPO resulted in MCC = 0 for Smax-ValMIR and MCC = 0.36 for Smax-ValNIR. This can be explained by NIR being less sensitive to soil moisture than MIR, as the overtone and combination bands of water absorption in NIR are much weaker than the fundamental vibrations in MIR. However, the accuracy of Smax prediction or SL/SH soil discrimination remains less reliable for practical applications.

However, overall, the validation results from the Validation-datasets in this study were generally poor. The relatively small sample size may have contributed to these results[413]. Moreover, the effectiveness of EPO in removing moisture is low when applied to soils with high variability in texture or mineral composition[413]. According to the Great Soil Groups, the 34 samples in the Validation-datasets exhibited high variability, including 29% Brown Earth and Luvisol, 24% Surface Water Gley, 12% Peat,

and 3% Alluvial and Rendzina soils. Differences in wetting protocols may have also influenced the results. In the Validation-datasets, soil water was added to GMC levels of 20%, 40%, and 60% w/w, whereas in the Rewet-datasets, samples were first saturated and then oven dried over time.

4.4.6 User application and future work

Results from the Validation-datasets suggested potential for applying handheld MIR and NIR spectroscopy in relatively low to medium moisture soil conditions for classification of samples into low or high Smax. Additionally, under low moisture conditions, handheld MIR even showed the potential for rough screening Smax. As soils at GMC = 20%, an 'Approximate Quantitative Model' for Smax prediction was achieved using MIR (R^2 = 0.75, RPIQ = 3.22), placing it in the same regression model category as the handheld MIR spectroscopy applied to dried, sieved soils[385]. However, the prerequisite for these is the application of *specific* EPO transformations that correspond to the respective moisture level. Therefore, for future in situ applications of handheld MIR and NIR, maintaining dry field conditions and using a water-measuring probe to support *specific* EPO application is recommended.

In future studies, verifying the *specific* and *general* EPO models established in this study for MIR and NIR on a larger sample size and standardisation of soil wetting protocols (e.g., following a wetting and drying approach) could improve the accuracy of validation. The development of moisture prediction models for handheld MIR and NIR spectroscopy is recommended as a next step to reduce costs and simplify field use by minimising the need for additional moisture sensors (with gravimetric moisture data). Before verifying EPO with MIR and NIR in situ spectra, it would be helpful to investigate how many days of dry weather (considering seasonal temperature variations) are required to achieve low soil moisture conditions on-site.

4.5 Conclusions

Following previous research, calibration models were developed to approximate quantitative predictions of soil Smax values and discriminate between SL/SH soils within NIR ranges. This study evaluated the effectiveness of the EPO algorithm in reducing the impact of moisture on MIR and NIR spectra for soil P dynamics estimation. Two methods for determining the optimal number of EPO components (***g***) were compared: model-coupled RMSEP minimisation and Wilk's Λ. Since Wilk's Λ produces smoother spectra using only reflectance data, the discussion focused on results based on EPO development using minimised RMSEP. Two types of EPO methods were developed: (1) *specific* EPO algorithms, designed for transforming wet spectra under known soil moisture conditions, and (2) *general* EPO algorithms, intended for application under unknown soil moisture conditions. The major conclusions drawn from this study are as follows.

1. EPO effectively removes moisture effects from soil MIR and NIR spectra at different moisture levels, particularly at low moisture contents (e.g., GMC ≤ 20%).
2. *Specific* EPO algorithms are more effective than *general* EPO algorithms in removing moisture variation, particularly for wet MIR spectroscopy.
3. Moist MIR using *specific* EPO pre-processing can predict soil Smax values (R^2 = 0.75, RPIQ = 3.22) under known low moisture conditions (e.g. GMC ≤ 20%).
4. Under known low to medium moisture conditions (e.g., GMC ≤ 40%), wet soil MIR and NIR spectra using *specific* EPO pre-processing are expected to classify SL/SH soils (e.g., MCC = 0.45 and 0.42 for MIR and NIR at GMC = 40%, respectively).
5. For in situ handheld MIR/NIR Smax prediction, maintaining dry field conditions and using a water-measuring probe for *specific* EPO application is recommended.

Chapter 5

Project conclusions and future outlook

5.1 Summary and implications

Soil phosphorus (P) is an essential nutrient with critical implications for food security, soil and water quality, and global climate change. However, owing to the limited consideration of soil P dynamics in current agricultural practices in Ireland, which typically rely on time-consuming and labour-intensive wet chemistry-based sorption isotherm experiments, the overarching aim of this thesis was to develop infrared spectroscopy as a practical alternative to monitor soil P. This included improving the prediction accuracy of benchtop MIR spectroscopy and maximising the potential of handheld MIR/NIR devices to enable cost-effective, rapid, and green assessments, enabling the transformation from laboratory-based ex situ analyses to in situ applications for future site-specific and sustainable P management.

Specifically, this study progressed from testing soil P sorption capacities using the traditional isothermal sorption experiment, establishing the reference values and defining behavioural classes for integration into agronomic guidelines (Chapter 2), to improving benchtop MIR predictive accuracy while reducing sample preparation costs (Chapter 3), and finally to demonstrating the potential field applicability of handheld MIR and NIR through calibration for soil heterogeneity and moisture effects (Chapters 3 and 4). Collectively, these findings support the use of MIR spectroscopy for rapid and reliable prediction of P sorption capacity and highlight the potential of handheld spectrometers for practical on-site soil P management in the future.

Building on previous work that used Langmuir sorption parameters to describe P maximum sorption capacity (Smax, $mg \cdot kg^{-1}$) in Irish soils[18,62], Chapter 2 analysed archived samples from the northern half of Ireland using the traditional wet chemical isotherm method. According to Spearman's rank correlation analysis and principal component analysis, mineral agricultural topsoils exhibited a significant change-point in Smax behaviour, with labile aluminium (Mehlich-3 extractable aluminium, M3-Al) identified as a key explanatory variable associated with P sorption capacity. Segmented regression analysis between Smax and M3-Al (R^2 = 0.49) identified a significant change-point in Smax

behaviour at 450.03 mg·kg^{-1}, classifying soils into 'low' sorbing capacity (SL soils, Smax < 450.03 mg·kg^{-1}), which are expected to have a 'fast rate' reaction responding to changes in soil P; and 'high' sorbing capacity (SH soils, Smax ≥ 450.03 mg·kg^{-1}), which are expected to act at a 'slow rate' and take longer before a response is observed. Classifying soil P sorption behaviours addresses the limitation of the current P Index System, which assumes a uniform soil response. For example, under the current agronomic Index System, soils grouped into indices 1 or 2 are described as having low available P, and fertilisation is suggested to support crop growth. When further accounting for inherent soil sorption characteristics, tailored strategies can be applied: for SH soils (definite/likely responses to P inputs at a slow rate), split applications and liming are recommended to prevent P lock-up; whereas for SL soils (definite/likely responses at a fast rate), fertilisation should be timed precisely and applied under optimal soil conditions to match crop demand, thereby maximising efficiency and minimising losses. An output of the work in Chapter 2 is integration of the current agronomic Soil Index System with classified SL/SH soil sorption capacities (Table 2-8), providing soil-type-specific recommendations for future P management in agricultural mineral soils to support more efficient and sustainable fertiliser decisions that maintain productivity while protecting water quality.

However, the wet chemical method utilised in Chapter 2 is time-consuming and labour-intensive, making it unsuitable for analysing large volumes of soil samples in practical applications. Chapter 3 therefore explored the use of MIR spectroscopy as a rapid, highly accurate, reliable, and green alternative for classifying soils sorbing capacity and predicting Smax.

Using laboratory-based benchtop MIR spectroscopy (Bruker device, 4000–400 cm^{-1} on ball-milled samples <0.1 mm), this study improved the prediction accuracy of Smax, achieving performance comparable to that of conventional wet chemistry methods ($RPIQ_{Val}$ = 4.50, with RPIQ > 4.05 defined as an 'Excellent Model'[385]), in comparison to the previous work by Dunne et al. (2020)[147], which only achieved a rough screening for Smax prediction (PLSR, R^2_{Val} = 0.67, $RPIQ_{Val}$ = 1.55) using a benchtop device (Perkin-Elmer Spectrum 400 FTIR instrument with a DRIFT accessory, 4000 to 450 cm^{-1}) on ball-milled samples (<0.25 mm). Specifically, beyond the use of a larger dataset for model training (n = 737

[$n_{Cal}|n_{Val}$ = 589|148] *vs.* n = 196 [$n_{Cal}|n_{Val}$ = 147|49]), Smax prediction was enhanced by applying a systematic approach that combined optimal spectral pre-processing with non-linear machine learning models. For example, support vector machines (SVM) combined with Savitzky–Golay first derivative pre-processing achieved the highest predictive accuracy (R^2_{Val} = 0.91, $RPIQ_{Val}$ = 4.50), outperforming the traditional linear PLSR model based on raw spectra (R^2_{Val} = 0.87, $RPIQ_{Val}$ = 2.93).

Moreover, although ball milling of dried soil samples is a standard practice in most MIR laboratory scanning, omitting this step can further reduce sample processing time and cost. The same spectroscopic approach was also applied to predict Smax from heterogeneous samples (<2 mm, sieved-only). Overall, SVM combined with Savitzky–Golay first derivative pre-processing enabled the benchtop MIR spectroscopy based on either <2mm or ball-milled samples to achieve 'Excellent Models[385] (RPIQ > 4.05)' for accurate Smax prediction ($RPIQ_{Val}$ = 4.50 and $RPIQ_{Val}$ = 4.25, respectively). The comparable accuracy of benchtop MIR models based on ball-milled and <2 mm sieved samples suggests that fine grinding is not essential when optimal chemometric techniques are applied. Therefore, the findings suggest that benchtop MIR spectroscopy enables accurate, reliable, and rapid Smax estimation while reducing analysis time and costs by simplifying the grinding stage of sample preparation.

In addition to improving the accuracy of benchtop MIR-based Smax predictions, another significant finding from Chapter 3 suggests that portable MIR spectroscopy (Agilent, 4000–650 cm^{-1}) can serve as a time-efficient alternative to the traditional wet chemical testing and a cost-effective substitute for benchtop MIR instruments. Despite higher spectral noise, handheld MIR spectroscopy using ball-milled samples achieved an 'Approximate Quantitative Model'[385] for the Smax value prediction (using Cubist on the raw spectra, $RPIQ_{Val}$ = 2.74). Importantly, using heterogeneous samples (<2 mm), handheld MIR spectroscopy demonstrated the ability to distinguish between 'low' and 'high' sorbing classes ($RPIQ_{Val}$ = 2.23). The best SVM classification model for SL/SH discrimination achieved a Matthews correlation coefficient (MCC) of 0.60, where MCC = 1 indicates perfect classification. Therefore, although portable MIR spectroscopy is less accurate than the benchtop instrument for

Smax prediction, its practical advantages (e.g., smaller size, lower cost, robustness, ease of analysis, and portability) and the ability to differentiate SL/SH sorbing soils contribute to the progress of practical, site-specific P management in the future.

Handheld NIR devices (Inno: 900–1700 nm; Neo: 1350–2550 nm) were used to scan sieved-only, dried archived samples from the northern half of Ireland. Since Smax is a secondary property, spectral merging strategies were applied to combine the two NIR spectral blocks, covering the full NIR range (900–2500 nm). Chemometrics/machine learning modelling for Smax prediction followed the same procedures as in earlier analyses. Although the predictive accuracy was considered unreliable (R^2_{Val} = 0.69, and $RPIQ_{Val}$ = 1.90), the best-performing classification model, based on a random forest algorithm, achieved a test set accuracy of 0.63 (κ = 0.63; F1-score = 0.92), suggesting a medium classification potential despite limited quantitative accuracy.

Aiming to transfer the application of handheld spectrometers to field testing, beyond soil heterogeneity, soil moisture further challenges in situ spectroscopy. Chapter 4 conducted a re-wetting (<2mm sieved-only samples) experiment to mimic fresh soils and collecting the handheld wet MIR and NIR spectra. As noted in the review of the related literature (Chapter 1), few studies have investigated the use of wet MIR or NIR spectroscopy for analysing soil P dynamics. Although external parameter orthogonalisation (EPO) is a widely used and effective method for spectral moisture correction, its application to wet MIR spectra has received limited attention. This study developed two types of EPO models based on difference matrix (**D**) calculations: (i) '*specific*' EPO models, using **D** calculated for each of eight known moisture levels (one model per level), and (ii) a *general* EPO model, using a single **D** calculated across all moisture levels to enable correction without prior knowledge of moisture conditions. Although a well-known advantage for EPO is the potential to transform wet soil spectra without knowing the intensity of the water effect[411], this work found that *specific* EPO corrections for moisture interference, especially for wet MIR spectra, performed more effectively than the *general* EPO correction. This can be attributed to the difference matrices built for *specific* EPO models better capturing moisture-related variability and generating projection matrices that more effectively

eliminate the impact of soil water. Therefore, although (*general*) EPO removes systematic errors caused by soil moisture, a conclusion from this work strongly suggests that incorporating soil moisture information into the application of *specific* EPO models is crucial for improving correction efficiency, especially in MIR spectroscopy. Another significant finding is that under low moisture levels (i.e., gravimetric moisture content [GMC] ≤ 20%), handheld MIR with *specific* EPO achieved an RPIQ of 3.22 for rough Smax screening. Under medium soil moisture levels (i.e., GMC ≤ 40%), handheld MIR and NIR, after *specific* EPO correction, showed potential for Smax classification, achieving MCC values of 0.45 and 0.42, respectively. Although this work was a controlled laboratory experiment designed to simulate field conditions, the findings suggest that under relatively dry field conditions, and with information on soil moisture status, handheld MIR and NIR devices have potential for in situ determination of soil P sorption capacity.

Overall, this research aimed to advance the application of spectroscopy to predict soil P characteristics that can be coupled with agronomic advice for sustainable P management of soil. After identifying a change-point in P sorption behaviour classifying soils based on their inherent phosphorus sorption capacities, work in Chapter 3 provided a noteworthy improvement over previous work by using non-linear machine learning techniques to enhance Smax prediction using MIR spectroscopy. Additionally, the successful transfer was demonstrated of MIR spectroscopy from benchtop to handheld devices and from laboratory to potential in situ applications. Chapters 3 and 4 respectively addressed the technical challenges of soil matrix heterogeneity and moisture interference in using handheld MIR spectroscopy for in situ Smax prediction. By progressively addressing soil heterogeneity and moisture effects through laboratory calibrations, this work lays the foundation for future in situ prediction of soil P sorption capacity using handheld MIR devices.

In practice, the findings of this thesis provide important support for the adoption of sustainable P management practices by farmers, which enhance fertiliser efficiency and maintain crop yields while meeting environmental policy requirements in practice. For example, although traditional laboratory determinations provide the highest accuracy and traceability, the time and cost involved make routine

laboratory analyses impractical for regular on-farm monitoring. Using benchtop MIR spectroscopy with drying and sieving preparation can deliver comparable accuracy for predicting soil P sorption capacity. The development of a benchtop MIR spectroscopy laboratory requires relatively high equipment and preparation costs. Nevertheless, it enables the simultaneous prediction and analysis of multiple soil properties and is therefore recommended for advisory or institutional laboratories that conduct routine soil analyses.

Handheld MIR and NIR devices are more cost-effective alternatives to benchtop MIR spectrometers. While benchtop instruments provide the highest analytical performance, MIR/NIR scans of dried and sieved samples can still deliver acceptable quantitative accuracy at a substantially lower cost. For farmers and advisory services seeking to understand how soils respond to phosphorus fertilisation, distinguishing between SL and SH soil groups is generally sufficient for guiding management decisions. In addition, although NIR devices are more commonly used in the field, this study showed that handheld MIR spectroscopy, when applied under low to moderate soil moisture conditions and combined with EPO moisture correction, achieved high prediction accuracy for distinguishing SL and SH soils and demonstrated strong potential for rapid and reliable field screening. When balancing cost, turnaround time, and acceptable uncertainty, handheld MIR can represent a practical and efficient tool for routine advisory soil P management, being less expensive than benchtop MIR while offering greater accuracy and interpretability than handheld NIR and supporting the implementation of sustainable P management strategies.

5.2 Limitations and future perspectives

This thesis used mineral agricultural soil samples from the Terra Project archive (n = 747), which, as shown in Figure 2-2, were collected from the northern half of Ireland. However, as an inherent soil characteristic, the applicability of the suggested change-point in soil P sorption behaviour at Smax = 450.03 mg·kg^{-1} at the national scale remains uncertain. The absence of Smax training data from the southern half of Ireland also raises concerns about the robustness and generalisability of the developed MIR calibration models for predictive performance at the national scale.

An earlier study by Dunne (2021)[343] applied the same wet chemical method to determine Smax in archived soils from the national-scale Irish Soil Information System (ISIS) dataset, albeit with a much smaller dataset and spectral library (n = 224). For a rough comparison, mineral samples were selected (loss-on-ignition ≤ 20%) and fitted using the linear Langmuir model ($R^2 \geq 0.80$), resulting in a sample size of 187. Phosphorus sorption isotherms classified using (a) Giles' isotherm shapes[90] and (b) the P sorption behaviour threshold defined in this study are compared in Figure 5-1. Visually, ISIS soils classified as SL (Smax < 450.03 mg·kg^{-1}; [b] isotherms in blue) mostly correspond to *L strict*-shaped isotherms ([a] isotherms in yellow) and *S*-shaped isotherms ([a] isotherms in green) described by Dunne et al. (2021)[90] as concave, with sorption decreasing as equilibrium P concentration increases; these were described as exhibiting slower P sorption rates, consistent with the 'slow rate' behaviour defined in the SL classification. Similarly, ISIS soils classified as SH (Smax ≥ 450.03 mg·kg^{-1}; [b] isotherms in red) mostly align with the *C*- and *non-strict L*-shaped isotherms ([a] blue and orange) that exhibit no visible sorption plateau and low final P concentrations after equilibration[90]. These were characterised by Dunne et al. (2021)[90] as having rapid P sorption behaviour, consistent with the 'fast rate' defined in the SH classification.

Among the 223 ISIS samples classified by Dunne et al. (2021)[90] (from the full ISIS dataset of 224 samples, with one excluded as unclassifiable), 21 samples (about 9.4%) were described as exhibiting slower P sorption rates (16 strict *L*-shaped and five *S*-shaped isotherms). In Figure 5-1 (b) (based on a

filtered ISIS dataset of 187 samples), 17 samples (about 9.1%) were grouped under the SL classification (Smax range: 247.39–435.44 mg·kg^{-1}; median Smax = 383.10 mg·kg^{-1}). For the described fast-rate sorption behaving soils, 170 (about 90.9%) were grouped as SH (Smax range: 451.77–1036.86 mg·kg^{-1}; median Smax = 693.78 mg·kg^{-1}), compared to the 202 soils (about 90.6%) reported by Dunne et al. (2021)[90] (60 *C*-shaped and 142 *non-strict L*-shaped isotherms). These findings indicate that the P sorption classification method developed in this study, based on archived samples from the northern half of Ireland, may be applicable at the national scale.

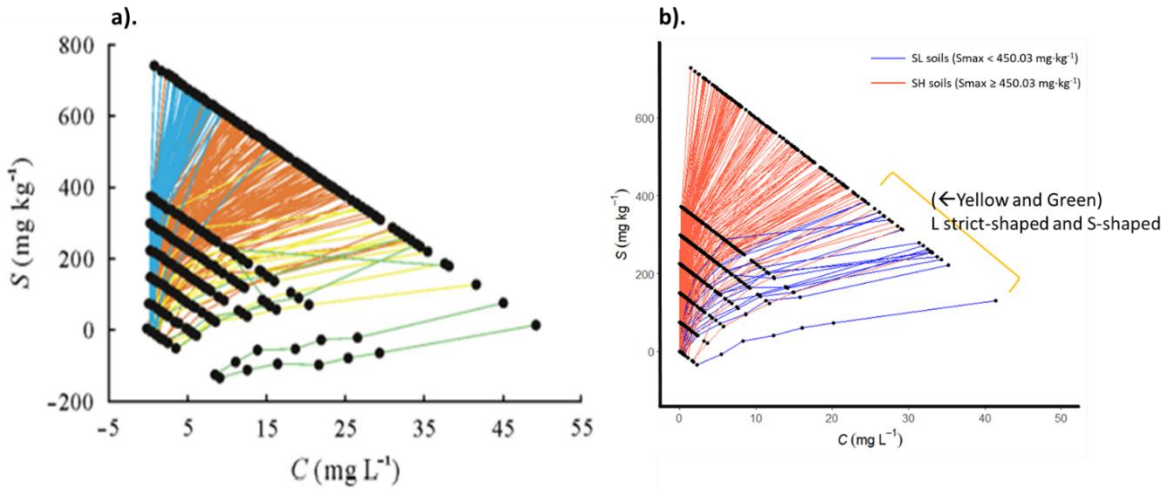


Figure 5-1. Comparison of classification methods for P sorption capacity in the national-scale ISIS-dataset. (A) P sorption isotherms from Dunne et al. (2021)[90], depicted using Giles' types: *C*-shaped (blue), *non-strict L*-shaped (orange), *L*-shaped strict (yellow), and *S*-shaped (green); (B) P sorption isotherms coloured by SL and SH classes based on the Smax threshold identified in Chapter 2 of this work.

Although the consistency in isotherm positions suggests the potential applicability of the P sorption behaviour classes developed from the larger regional Terra dataset (n = 737) to the national scale, their application to the nationally-covered ISIS dataset (n = 187) provides only preliminary evidence of national-scale applicability, as the smaller sample size may limit statistical confidence. For example, the ratio of low- to high-sorbing soils was approximately 0.26 (SL:SH = 154:593) in the northern-region dataset, but only 0.10 (SL:SH = 17:170) in the nationally scaled prediction. Figure 5-2 compares soil properties (A) and Langmuir sorption parameters (B.1 and B.2) between the Terra and ISIS datasets.

The observed median shifts in soil properties, such as for Smax (669.68 $mg \cdot kg^{-1}$ for ISIS vs. 564.54 $mg \cdot kg^{-1}$ for Terra), may be attributed to sampling bias from the smaller size but broader spatial coverage of the ISIS dataset, which contains only about one-quarter the number of samples compared to the northern-based Terra collection. Therefore, for the ISIS dataset, despite Figure 5-1 suggesting an alignment between the sorption behaviour described by Dunne et al. (2021)[90] and the SL/SH classification scheme, future validation using a larger, nationally representative dataset is recommended to confirm the robustness of the SL/SH classification at the national scale.

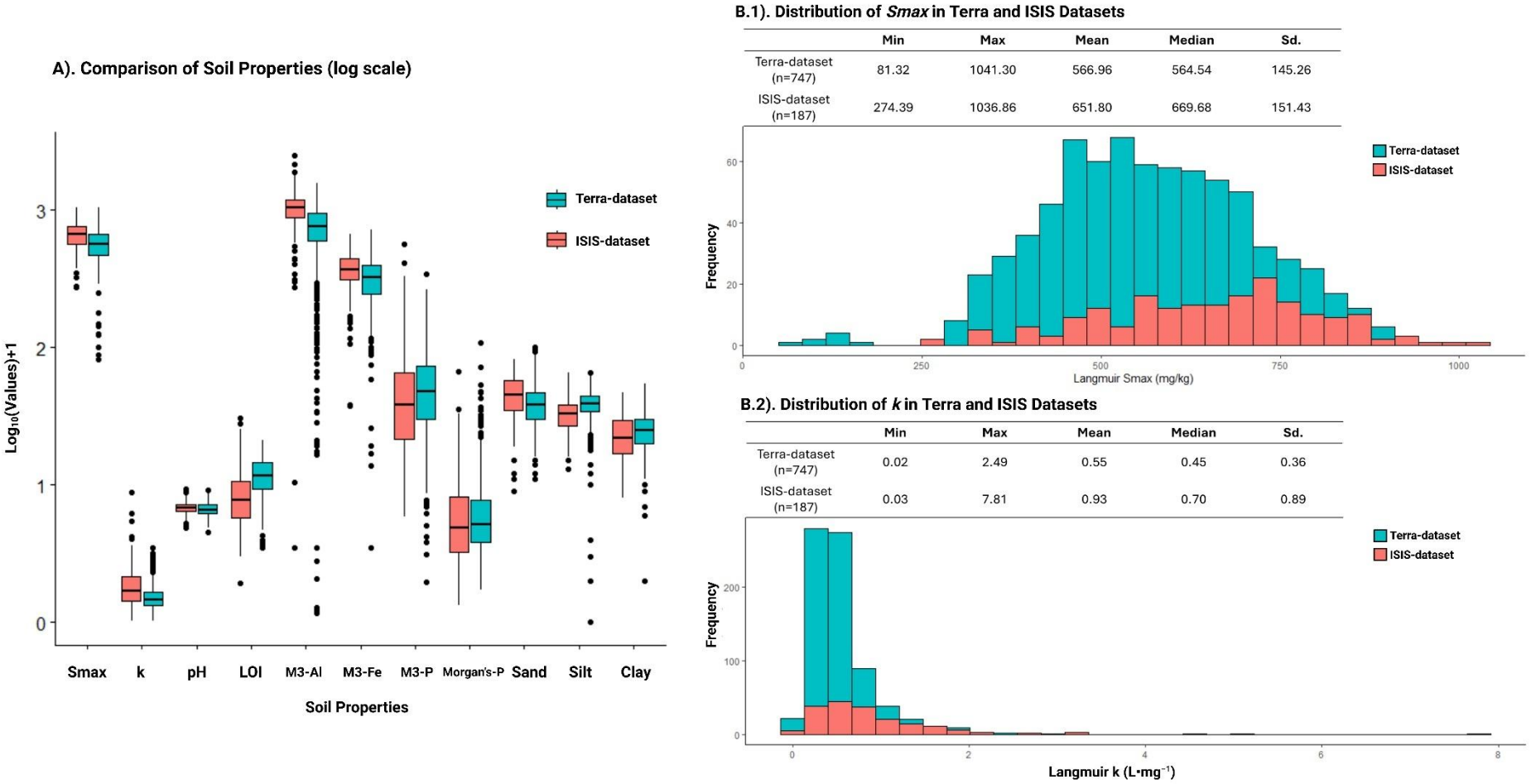


Figure 5-2. Soil information, particularly the Langmuir P sorption parameters Smax (B.1) and k (B.2), collected at the national scale (ISIS-dataset, red) and from the northern region (Terra-dataset, green).

MIR spectral libraries from the benchtop spectroscopy laboratory for the ISIS samples, including both dried, sieved-only and ball-milled (<0.25 mm) samples, have been obtained using the Bruker spectrometer after the work of Dunne (2021)[343]. Therefore, the benchtop MIR calibration models (Chapter 3; Bruker-BM (<0.10 mm) and Bruker-2mm) developed in this study were tested for their generalisability using nationally scaled soil samples (Figure 5-3). Moderate correlations were observed, with R^2 of 0.53 for the ball-milled samples and 0.58 for the sieved-only samples, respectively. The calibration models showed systematic underestimation of Smax values across the national dataset for both sample preparation methods. Significant negative bias was observed: −102.08 $mg \cdot kg^{-1}$ for the ball-milled spectra and −89.46 $mg \cdot kg^{-1}$ for the sieved-only spectra.

Bland–Altman plots (Figure 5-4 A.1 and B.1) compare isothermally assessed Smax values of the ISIS archived soil samples with the predicted Smax values generated by calibration models developed from the northern-region soil spectral libraries. After applying bias corrections (adding a constant offset of 102.08 $mg \cdot kg^{-1}$ for the ball-milled sample predictions; and a proportional correction for the sieved-only sample predictions), the goodness-of-fit plots for the ISIS national-scale dataset showed slight improvement. However, according to the RPIQ values, reflecting model performance relative to the variability of the observed data, values of 1.84 for the ISIS ball-milled and 1.96 for the ISIS sieved-only spectroscopy predictions fall below the commonly accepted reliability threshold (RPIQ < 2.02)[385]. Consequently, despite the high accuracy of the benchtop MIR calibration models developed in this study for northern-region soils, their predictive strength appears limited when applied to soils sampled at the national scale or outside the original calibration domain.

The limited generalisation performance of the region-based MIR calibration models on nationally scaled samples suggests the need for additional calibration using soils from more diverse regions (e.g., the southern half of Ireland) to improve the robustness of benchtop MIR models for accurately predicting soil P sorption capacity across a wider range of soil types in future work.

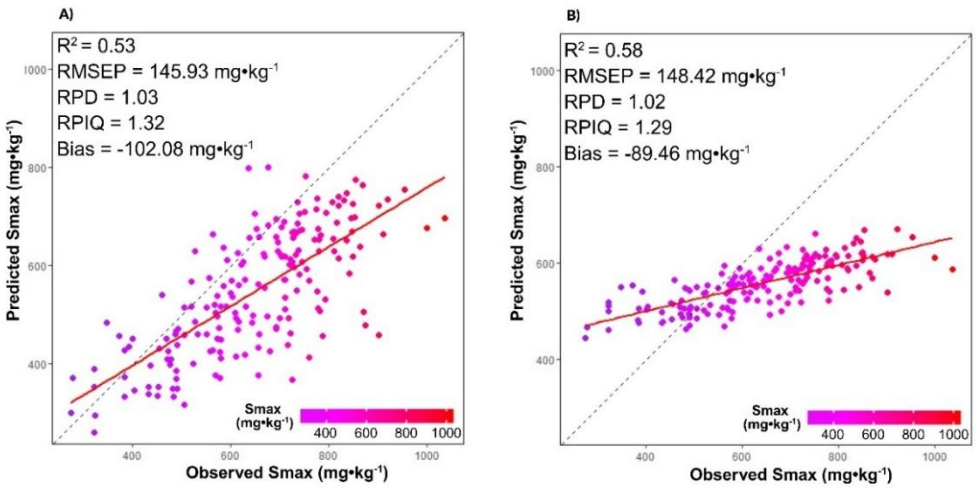


Figure 5-3. Application of the corresponding benchtop MIR calibration model, developed using Terra archive samples from the northern region (as described in Chapter 2) to the ISIS national-scale MIR spectral libraries of dried samples: (A) sieved and ball-milled, and (B) sieved-only.

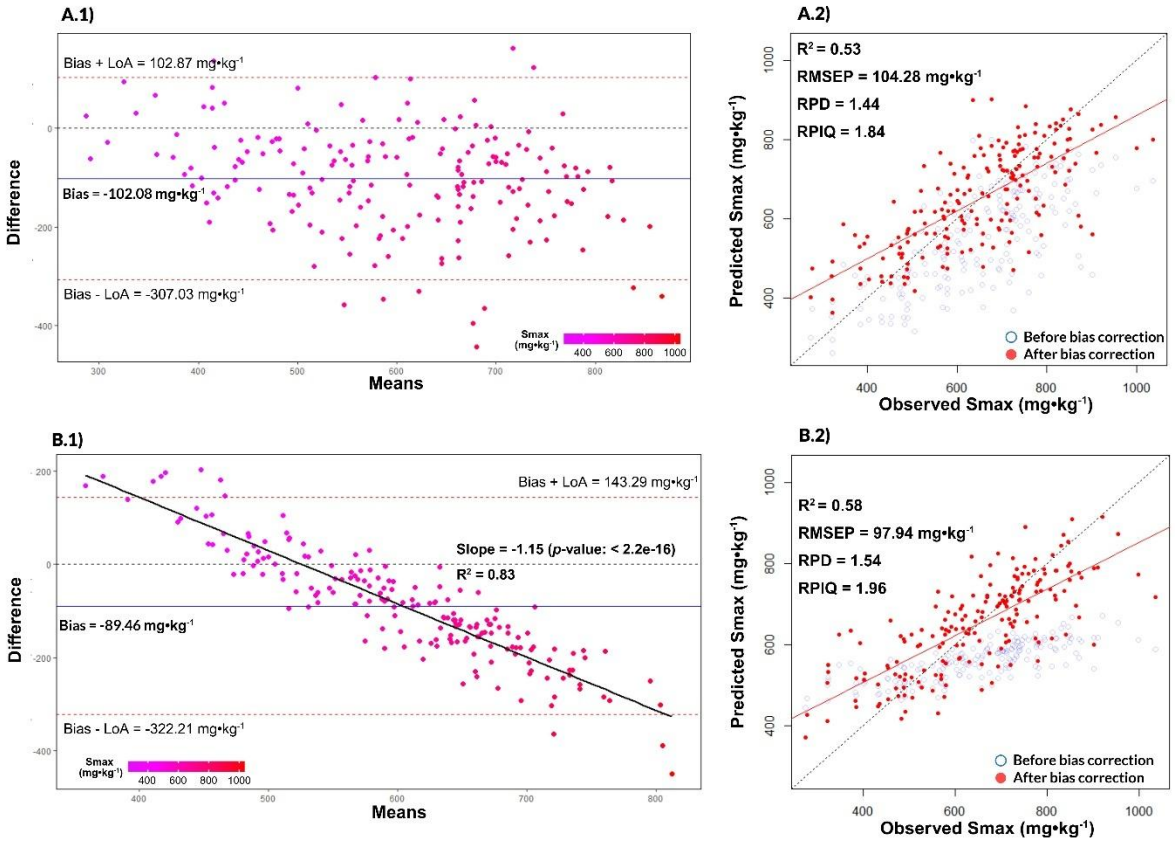


Figure 5-4. (A.1) and (B.1) Bland–Altman plots comparing isothermally assessed and predicted Smax values for the ISIS samples using Terra-based MIR calibration models. A clear proportional bias was observed for the sieved-only samples (B.1), with errors increasing with Smax (slope = –1.15, R^2 = 0.83). The annotation 'LoA' refers to the limits of agreement, indicating the range within which 95% of the differences between predicted and observed values are expected to fall. (A.2) and (B.2) Goodness-of-fit plots for ISIS samples using Terra-based MIR calibration models after applying bias correction.

For handheld MIR spectroscopic analysis, although Chapter 3 aimed to enhance performance by applying chemometric techniques to reduce spectral noise and emphasise key absorption features, the best performing models built from dried samples, whether homogeneous ball-milled or more heterogeneous <2 mm sieved, demonstrated only approximate predictive accuracy or moderate performance in SL/SH soil classification. Similarly to the benchtop MIR spectroscopy for Smax prediction, non-linear machine learning techniques have shown superior performance over PLSR for Smax prediction in this study. As a secondary soil property, P sorption capacity is influenced by multiple soil factors (e.g., Al/Fe oxides, pH, etc.), which affect spectral features in overlapping and non-linear ways. Non-linear models, often equipped with built-in feature selection, are generally better suited to capturing these complex interactions, leading to improved prediction accuracy. Given these findings, artificial neural networks can be considered in future work to further improve handheld MIR spectroscopy accuracy, using a similar methodology to that demonstrated by Saidi et al. (2025)[176].

Additionally, according to the spectral transformation techniques explored in Chapter 4, applying standardisation algorithms, such as direct standardisation (DS) or piecewise direct standardisation (PDS), may have a potential to further improve the predictive performance of handheld MIR for Smax. DS and PDS establish a transformation matrix that aligns spectra from a primary to a secondary instrument. In comparison, DS considers all wavelengths simultaneously, whereas PDS restricts the relationship to local spectral regions, enabling more flexible and potentially nonlinear transformations[247]. Practical applications of instrument standardisation and calibration transfer have been studied in both NIR[416–418] and MIR spectroscopy[419]. For example, Parrott et al. (2022)[419] applied DS and PDS to transfer MIR-ATR spectra from the FTLA2000 (ABB, Québec, Canada) to the MB155 (ABB BOMEN, Clairet Scientific, Northampton, UK). For ethanol and ethyl acetate prediction, RMSEP was reduced from 2.8% and 2.0% to 2.4% and 0.9% with DS, and to 2.0–2.6% and 0.6–0.9% with PDS, respectively. Consequently, a future investigation into applying spectral space transformation and standardisation algorithms to improve handheld MIR spectroscopy for Smax prediction is warranted. High-resolution benchtop MIR spectra from ball-milled samples could be used as a standardisation

reference ('master'), which would enable calibration models to be transferred to handheld devices ('slave'). Larger spectral differences between instruments are expected to result in greater improvements in predictive accuracy following DS/PDS transformation. By applying a transformation matrix[279], handheld spectra (from both ball-milled and sieved-only samples) are expected to be standardised to the benchtop ball-milled reference, with the hypothesis that Smax prediction accuracy will improve. For example, a simple comparison of a DS transformation applied to handheld Agilent-2mm spectra, using benchtop Bruker-BM spectra as the reference, is shown in Figure 5-5. Before DS transformation (A), the Bruker-BM and Agilent-2mm spectra were pre-processed with Savitzky-Golay filter and standard normal variate transformation for baseline correction. However, the Agilent-2 mm spectra (red) still contain a lot of noise, e.g., around 3500-3000 cm^{-1}. After DS transformation (B), visual inspection shows that the Agilent-2mm spectra improved noticeably (e.g., around 3500-3000 cm^{-1}), showing a reduction in noise. Both DS and PDS rely on linear models to link spectral responses between the master and handheld instruments[419]. PDS is expected to outperform DS in correction performance, as it estimates the transformation matrix locally across the spectrum using a moving window approach, allowing it to better account for non-linear variations. Nevertheless, PDS presents challenges, including the need to carefully select the window size, as improper selection may introduce spectral artefacts.

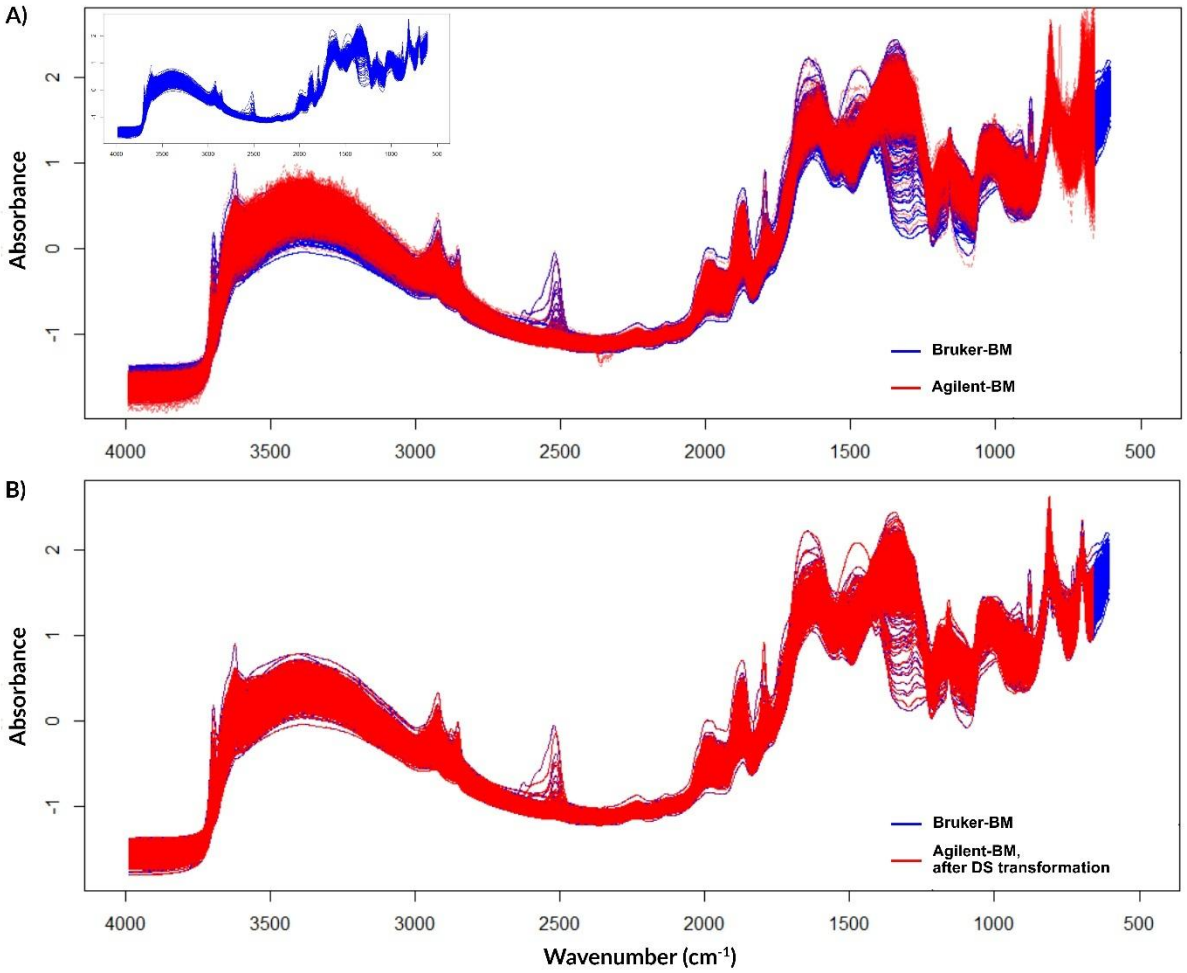


Figure 5-5. Spectral libraries, Bruker-BM (benchtop, ball-milled) and Agilent-2mm (handheld, sieved-only) spectra for 737 mineral soil samples, were derived from Chapter 3. (A) Spectra from Bruker-BM (reference) and Agilent-2mm (subordinate) before DS transformation. (B) Spectra from the reference (Bruker-BM, blue) and DS-corrected spectra (Agilent-2mm, red) from the subordinate instrument.

An important opportunity for future study is conducting field validation, including validation of the determined soil P sorption maximum capacity threshold and the handheld MIR spectroscopic models applied to fresh soil measurements corrected using the developed (*specific*/*general*) EPO models. Field validation of the proposed Smax threshold of 450.03 $mg \cdot kg^{-1}$ and its agronomic relevance may require long-term temporal study involving field trials involving P fertiliser application, monitoring of soil plant-available P levels, and crop response over multiple growing seasons. Although time-intensive, such validation is essential and worthwhile, as it would account for the complexity of soil systems, including plant–soil interactions (e.g., phosphatase activity), farming practices, grazing activities, *etc*. confirming the threshold's effectiveness for sustainable P management under real-world conditions. Additionally, as demonstrated by the potential of handheld MIR spectroscopy to predict soil P dynamics (SL/SH sorbing capacity) in wet soils, once validated for field use, its ease of implementation, cost- and time-effectiveness, and robustness could support the practical application of the refined Soil Index System for P management, contributing to both food security and environmental sustainability.

For example, although soils in Index 1/2 require P fertilisation for build-up, and Index 3 soils require maintenance applications, on-site handheld MIR spectroscopy can readily identify differences in their inherent sorption capacities. This enables tailored management based on P fixation potential and environmental loss risk. Identified SH soils, with high P sorbing capacity, may rapidly adsorb applied P, reducing its immediate availability unless application timing is aligned with plant uptake or soil pH is adjusted (e.g. through liming). In contrast, identified SL soils, with low sorption capacity, retain less P and pose a greater risk of loss to water bodies, particularly in areas prone to runoff.

The realisation of handheld MIR spectroscopy for in situ discrimination of soils into low- and high-P sorption capacity classes marks a major step toward rational and sustainable P management. It offers faster and more cost-effective assessments than traditional laboratory methods, supports site-specific fertiliser recommendations, and improves P use efficiency while reducing environmental risks, contributing to the EU's Zero Pollution Action Plan and Ireland's Nitrates Action Programme.

Bibliography

1. United Nations. Global Issues: Population. https://www.un.org/en/global-issues/population_2024_wpp_2024_advance_unedited_0.pdf (2024).

2. FAO. *Climate Change and Food Security: Risks and Responses*. (Food and Agriculture Organization of the United Nations, 2016).

3. FAO, IFAD, UNICEF, WFP & WHO. *The State of Food Security and Nutrition in the World 2023*. (FAO ; IFAD ; UNICEF ; WFP ; WHO, Rome, Italy, 2023). doi:https://doi.org/10.4060/cc3017en.

4. Kedir, A. J., Nyiraneza, J., Hawboldt, K. A., McKenzie, D. B. & Unc, A. Phosphorus Sorption Capacity and Its Relationships With Soil Properties Under Podzolic Soils of Atlantic Canada. *Front. Soil Sci* **2**, 931266 (2022).

5. Plunkett, M., Wall, D., Sheil, T. & Bolger, J. The Efficient Use of Phosphorus in Agricultural Soils. *The Fertilizer Association of Ireland in association with Teagasc Technical Bulletin* vol. 4 48 (2019).

6. Brogan, J., Crowe, M. & Carty, G. DEVELOPING A NATIONAL PHOSPHORUS BALANCE FOR AGRICULTURE IN IRELAND. *Environmental Protection Agency, Ireland* 44 (2001).

7. Hall, R. *et al.* Geospatial Modelling of Soil Phosphorus Fractions and Sorption Indicators from Heterogeneous Landscapes. *Available SSRN 4425542* (2023).

8. Roy, E. D. *et al.* Soil phosphorus sorption capacity after three decades of intensive fertilization in Mato Grosso, Brazil. *Agric. Ecosyst. Environ.* **249**, 206–214 (2017).

9. McDowell, R. W., Pletnyakov, P. & Haygarth, P. M. Phosphorus applications adjusted to optimal crop yields can help sustain global phosphorus reserves. *Nat. Food* **5**, 332–339 (2024).

10. Bechtaoui, N. *et al.* Phosphate-Dependent Regulation of Growth and Stresses Management in Plants. *Front. Plant Sci* **12**, (2021).

11. Kang, J. *et al.* Understanding how various forms of phosphorus stress affect microbiome

functions and boost plant disease resistance: Insights from metagenomic analysis. *Sci. Total Environ.* **904**, 166899 (2023).

12. CSO. Environmental Indicators Ireland 2024. *Central Statistics Office statistical publication* https://www.cso.ie/en/releasesandpublications/ep/p-eii/environmentalindicatorsireland2024/landuse/?utm_source (2025).

13. EEA. Estimated consumption of manufactured fertilisers. *European Environment Agency.* https://www.eea.europa.eu/en/analysis/maps-and-charts/estimated-consumption-of-manufactured-fertilisers#tab-chart_1 (2015).

14. O'Donnell, C. *et al.* An overview on deficit and requirements of the Irish national soil phosphorus balance. *Sci. Total Environ.* **785**, 147251 (2021).

15. Ott, C. & Rechberger, H. The European phosphorus balance. *Resour. Conserv. Recycl.* **60**, 159–172 (2012).

16. Cordell, D. & White, S. Life's Bottleneck: Sustaining the World's Phosphorus for a Food Secure Future. *Annu. Rev. Environ. Resour.* **39**, 161–188 (2014).

17. Burkitt, L. L., Gourley, C. J. P., Sale, P. W. G., Uren, N. C. & Hannah, M. C. Factors affecting the change in extractable phosphorus following the application of phosphatic fertiliser on pasture soils in southern Victoria. *Aust. J. Soil Res.* **39**, 759–771 (2001).

18. Daly, K., Styles, D., Lalor, S. & Wall, D. P. Phosphorus sorption, supply potential and availability in soils with contrasting parent material and soil chemical properties. *Eur. J. Soil Sci.* **66**, 792–801 (2015).

19. Gript News. High fertilizer prices crippling farmers and driving up costs of food, says Collins. *Gript News**Irish News* https://gript.ie/high-fertilizer-prices-crippling-farmers-and-driving-up-costs-of-food-says-collins/?utm_source (2022).

20. Ahmad, N. *et al.* Environmental implications of phosphate-based fertilizer industrial waste and

its management practices. *Environ. Monit. Assess.* **195**, 1326 (2023).

21. Mishra, C., Nayak, S., Guru, B. C. & Rath, M. Environmental impact and management of wastes from phosphate fertilizer plants. *J. Ind. Pollut. Control* **26**, 57–60 (2010).

22. Lun, F. *et al.* Global and regional phosphorus budgets in agricultural systems and their implications for phosphorus-use efficiency. *Earth Syst. Sci. Data* **10**, 1–18 (2018).

23. Wang, J., Qi, Z. & Wang, C. Phosphorus loss management and crop yields: A global meta-analysis. *Agric. Ecosyst. Environ.* **357**, 108683 (2023).

24. MacDonald, G. K., Bennett, E. M., Potter, P. A. & Ramankutty, N. Agronomic phosphorus imbalances across the world's croplands. *Proc. Natl. Acad. Sci.* **108**, 3086–3091 (2011).

25. Mueller, N. D. *et al.* Closing yield gaps through nutrient and water management. *Nature* **490**, 254–257 (2012).

26. Luo, X. *et al.* The global fate of inorganic phosphorus fertilizers added to terrestrial ecosystems. *One Earth* **7**, 1402–1413 (2024).

27. Asomaning, S. Processes and Factors Affecting Phosphorus Sorption in Soils. in (eds. Kyzas, G. Z. & Lazaridis, N.) (IntechOpen, Rijeka, 2020). doi:10.5772/intechopen.90719.

28. Hall, R. L. Phosphorus speciation, agronomic performance and soil nutrient testing for waste recovered fertiliser use in Northern Irish soil. (Queen's University Belfast, 2020).

29. Shen, J. *et al.* Phosphorus Dynamics: From Soil to Plant. *Plant Physiol.* **156**, 997–1005 (2011).

30. Doydora, S. *et al.* Accessing Legacy Phosphorus in Soils. *Soil Syst.* **4**, 74 (2020).

31. Hall, R. L. *et al.* Phosphorus speciation and fertiliser performance characteristics: A comparison of waste recovered struvites from global sources. *Geoderma* **362**, 114096 (2020).

32. Cerven, V. *et al.* The Occurrence of Legacy P Soils and Potential Mitigation Practices Using Activated Biochar. *Agron J.* **11**, 1–11 (2021).

33. Doody, D. G. *et al. Phosphorus Stock and Flows in the Northern Ireland Food System*. https://www.afbini.gov.uk/sites/afbini.gov.uk/files/publications/RePhoKUs report October 2020x.pdf (2020).

34. Schoumans, O. F. *et al.* Mitigation options to reduce phosphorus losses from the agricultural sector and improve surface water quality: A review. *Sci. Total Environ.* **468**–**469**, 1255–1266 (2014).

35. Liu, L., Zheng, X., Wei, X., Kai, Z. & Xu, Y. Excessive application of chemical fertilizer and organophosphorus pesticides induced total phosphorus loss from planting causing surface water eutrophication. *Sci. Rep.* **11**, 23015 (2021).

36. Ramos, M. C., Lizaga, I., Gaspar, L. & Navas, A. The impacts of exceptional rainfall on phosphorus mobilisation in a mountain agroforestry catchment (NE, Spain). *CATENA* **216**, 106407 (2022).

37. Kleinman, P. J. A. *et al.* Phosphorus and the Chesapeake Bay: Lingering Issues and Emerging Concerns for Agriculture. *J. Environ. Qual.* **48**, 1191–1203 (2019).

38. Pieterse, N. M., Bleuten, W. & Jørgensen, S. E. Contribution of point sources and diffuse sources to nitrogen and phosphorus loads in lowland river tributaries. *J. Hydrol.* **271**, 213–225 (2003).

39. Igwaran, A., Kayode, A. J., Moloantoa, K. M., Khetsha, Z. P. & Unuofin, J. O. Cyanobacteria Harmful Algae Blooms: Causes, Impacts, and Risk Management. *Water, Air, Soil Pollut.* **235**, 71 (2024).

40. Qin, B. *et al.* A Drinking Water Crisis in Lake Taihu, China: Linkage to Climatic Variability and Lake Management. *Environ. Manage.* **45**, 105–112 (2010).

41. Akinnawo, S. O. Eutrophication: Causes, consequences, physical, chemical and biological techniques for mitigation strategies. *Environ. Challenges* **12**, 100733 (2023).

42. Guan, X. *et al.* Reducing the environmental risks related to phosphorus surplus resulting from

greenhouse cucumber production in China. *J. Clean. Prod.* **332**, 130076 (2022).

43. EESP. EU Zero Pollution Action Plan published. *European Sustainable Phosphorus Platform* https://www.phosphorusplatform.eu/scope-in-print/news/2090-eu-zero-pollution-action-plan-published (2025).

44. Heuser, D. I. Soil Governance in current European Union Law and in the European Green Deal. *Soil Secur.* **6**, 100053 (2022).

45. European Parliament. Soil monitoring: deal with Council on new EU law with better support for farmers. *European Parliament News* https://www.europarl.europa.eu/news/en/press-room/20250407IPR27692/soil-monitoring-deal-with-council-on-new-eu-law-with-better-support-for-farmers (2025).

46. European Union. European Union (Good Agricultural Practice for the Protection of Waters) Regulation 2022. https://www.irishstatutebook.ie/eli/2022/si/113/made/en/pdf (2022).

47. Department of Agriculture Food & the Marine. NITRATES EXPLANATORY HANDBOOK for Good Agricultural Practice for the Protection of Waters Regulations 2022. https://www.owenodriscoll.com/contentfiles/Nitrates Explanatory Handbook 2018.pdf (2022).

48. Teagasc. What happens to your soil sample when it comes into a lab? *Teagasc Daily* https://www.teagasc.ie/publications/2020/what-happens-to-your-soil-sample-when-it-comes-into-a-lab.php (2020).

49. Teagasc. Soil Index System. *Teagasc Agriculture and Food Development Authority* https://www.teagasc.ie/crops/soil--soil-fertility/soil-analysis/soil-index-system/ (2017).

50. Teagasc. National Soil Fertility Trends 2021. https://www.teagasc.ie/about/farm-advisory/advisory-regions/cork-east/farm-advice/national-soil-fertility-trends-2021/ (2021).

51. Sharpley, A. N., McDowell, R. W. & Kleinman, P. J. A. Phosphorus loss from land to water: integrating agricultural and environmental management. *Plant Soil* **237**, 287–307 (2001).

52. Hart, M. R., Quin, B. F. & Nguyen, M. L. Phosphorus Runoff from Agricultural Land and Direct Fertilizer Effects: A Review. *J. Environ. Qual.* **33**, 1954–1972 (2004).

53. Ding, N., Tao, F. & Chen, Y. Effects of climate change, crop planting structure, and agricultural management on runoff, sediment, nitrogen and phosphorus losses in the Hai-River Basin since the 1980s. *J. Clean. Prod.* **359**, 132066 (2022).

54. Ramos, M. C., Lizaga, I., Gaspar, L., Quijano, L. & Navas, A. Effects of rainfall intensity and slope on sediment, nitrogen and phosphorous losses in soils with different use and soil hydrological properties. *Agric. Water Manag.* **226**, 105789 (2019).

55. Ockenden, M. C. *et al.* Major agricultural changes required to mitigate phosphorus losses under climate change. *Nat. Commun.* **8**, 161 (2017).

56. Hou, X., Zhang, S., Ruan, Q. & Tang, C. Synergetic impact of climate and vegetation cover on runoff, sediment, and nitrogen and phosphorus losses in the Jialing River Basin, China. *J. Clean. Prod.* **361**, 132141 (2022).

57. Johnes, P. J. Uncertainties in annual riverine phosphorus load estimation: Impact of load estimation methodology, sampling frequency, baseflow index and catchment population density. *J. Hydrol.* **332**, 241–258 (2007).

58. Huang, C.-L. *et al.* Effects of urbanization on phosphorus metabolism in a typical agricultural area. *J. Clean. Prod.* **214**, 803–815 (2019).

59. European Commission. Intense rainfall could wash away 13-23% of EU & UK agri-soils by 2050. *The Joint Research Centre: EU Science Hub* https://joint-research-centre.ec.europa.eu/jrc-news-and-updates/soil-erosion-rainfall-2021-08-27_en (2021).

60. Sharpley, A. N. Soil phosphorus dynamics: agronomic and environmental impacts. *Ecol. Eng.* **5**, 261–279 (1995).

61. Menezes-Blackburn, D. *et al.* A Holistic Approach to Understanding the Desorption of

Phosphorus in Soils. *Environ. Sci. Technol.* **50**, 3371–3381 (2016).

62. Dunne, K. S., Holden, N. M. & Daly, K. A management framework for phosphorus use on agricultural soils using sorption criteria and soil test P. *J Env. Manag.* **299**, 113665 (2021).

63. Długosz, J., Piotrowska-Długosz, A. & Kalisz, B. Vertical changes in P-acquiring enzyme activities and microbial biomass in Luvisols – The effect of different types of agricultural land use and soil-forming processes. *Geoderma* **432**, 116406 (2023).

64. Ruttenberg, K. C. Phosphorus Cycle*. in (ed. Steele, J. H. B. T.-E. of O. S. (Second E.) 401–412 (Academic Press, Oxford, 2001). doi:https://doi.org/10.1016/B978-012374473-9.00277-0.

65. Plassard, C. & Dell, B. Phosphorus nutrition of mycorrhizal trees. *Tree Physiol.* **30**, 1129–1139 (2010).

66. Sims, J. T. & Sharpley, A. N. *Phosphorus: Agriculture and the Environment. Phosphorus: agriculture and the environment* (American Society of Agronomy, 2005).

67. Strawn, D. G. Sorption Mechanisms of Chemicals in Soils. *Soil Systems* vol. 5 at https://doi.org/10.3390/soilsystems5010013 (2021).

68. Aaron, T. & Keith, G. Introduction to the Sorption of Chemical Constituents in Soils. *Nat. Educ. Knowl.* **4**, 7 (2012).

69. Penn, C. J. & Camberato., J. J. A Critical Review on Soil Chemical Processes that Control How Soil pH Affects Phosphorus Availability to Plants. *Agriculture* **9**, 120 (2019).

70. Liao, W., Xu, R. & Stone, A. T. Adsorption of Phosphorus Oxyanions at the FeOOH(goethite)/Water Interface: The Importance of Basicity. *Environ. Sci. Technol.* **55**, 14389–14396 (2021).

71. McGechan, M. B. & Lewis, D. R. SW—Soil and Water: Sorption of Phosphorus by Soil, Part 1: Principles, Equations and Models. *Biosyst. Eng.* **82**, 1–24 (2002).

72. Oyebanjo, O., Ekosse, G.-I. & Odiyo, J. Phosphorus Sorption in Soils and Clay Fractions

Developed from Different Parent Rocks in Limpopo Province, South Africa. *Sustainability* vol. 14 at https://doi.org/10.3390/su14148528 (2022).

73. Yusran, F. The Relationship between Phosphate Adsorption and Soil Organic Carbon from Organic Matter Addition. *J. Tanah Trop.* **15**, (2010).

74. Dubus, I. G. & Becquer, T. Phosphorus sorption and desorption in oxide-rich Ferralsols of New Caledonia. *Soil Res.* **39**, 403–414 (2001).

75. Zhang, S., Huffman, T., Zhang, X., Liu, W. & Liu, Z. Spatial distribution of soil nutrient at depth in black soil of Northeast China: a case study of soil available phosphorus and total phosphorus. *J. Soils Sediments* **14**, 1775–1789 (2014).

76. Xu, G. *et al.* Phosphorus sorption capacity in soils from freshwater restored coastal wetlands increased with restoration age. *Geoderma* **422**, 115926 (2022).

77. Daly, K., Jeffrey, D. & Tunney, H. The effect of soil type on phosphorus sorption capacity and desorption dynamics in Irish grassland soils. *Soil Use Manag.* **17**, 12–20 (2006).

78. Amarh, F. *et al.* Effects of soil depth and characteristics on phosphorus adsorption isotherms of different land utilization types: Phosphorus adsorption isotherms of soil. *Soil Tillage Res.* **213**, 105139 (2021).

79. Hanyabui, E. *et al.* Phosphorus sorption in tropical soils. *AIMS Agric. Food* **5**, 599–616 (2020).

80. Ch'ng, H. Y., Ahmed, O. H. & Majid, N. M. A. Improving Phosphorus Availability in an Acid Soil Using Organic Amendments Produced from Agroindustrial Wastes. *Sci. World J.* **2014**, 506356 (2014).

81. Sparling, G. P., McLay, C. D. A., Tang, C. & Raphael, C. Effect of short-term legume residue decomposition on soil acidity. *Soil Res.* **37**, 561–574 (1999).

82. Wong, M. T. F., Nortcliff, S. & Swift, R. S. Method for determining the acid ameliorating capacity of plant residue compost, urban waste compost, farmyard manure, and peat applied to tropical

soils. *Commun. Soil Sci. Plant Anal.* **29**, 2927–2937 (1998).

83. Panagos, P. *et al.* Improving the phosphorus budget of European agricultural soils. *Sci. Total Environ.* **853**, 158706 (2022).

84. Teagasc. Soil pH & Liming. *Teagasc Agriculture and Food Development Authority* https://www.teagasc.ie/crops/soil--soil-fertility/soil-ph--liming/ (2019).

85. Hart, E. Soil test results: 34% of samples at optimum P, K, and pH on Footprint Farms. *Irish Farmers Journal* (2024).

86. Mead, J. A. A comparison of the Langmuir, Freundlich and Temkin equations to describe phosphate adsorption properties of soils. *Soil Res.* **19**, 333–342 (1981).

87. Ayenew, B., Tadesse, A. M., Kibret, K. & Melese, A. Phosphorous status and adsorption characteristics of acid soils from Cheha and Dinsho districts, southern highlands of Ethiopia. *Environ. Syst. Res.* **7**, 17 (2018).

88. Yang, X., Chen, X. & Yang, X. Effect of organic matter on phosphorus adsorption and desorption in a black soil from Northeast China. *Soil Tillage Res.* **187**, 85–91 (2019).

89. Hussain, A., Ghafoor, A. & Murtaza, D. G. Use of Models for Phosphorus Adsorption on Some Sodic Soils of Punjab. *Int. J. Agric. Biol.* **2**, (2006).

90. Dunne, K. S., Holden, N. M. & Daly, K. Predicting phosphorus sorption isotherm parameters in soil using data of routine laboratory tests. *Pedosphere* **31**, 694–704 (2021).

91. Kusuma, H. S. *et al.* Theoretical perspectives and recent advances in palm-based adsorbents for sustainable heavy metal removal from aqueous systems. *Desalin. Water Treat.* **323**, 101315 (2025).

92. Kleinman, P. J. A. The Persistent Environmental Relevance of Soil Phosphorus Sorption Saturation. *Curr. Pollut. Reports* **3**, 141–150 (2017).

93. Wang, Y. T., Zhang, T. Q., O'Halloran, I. P., Tan, C. S. & Hu, Q. C. A phosphorus sorption index

and its use to estimate leaching of dissolved phosphorus from agricultural soils in Ontario. *Geoderma* **274**, 79–87 (2016).

94. Daly, K. *Eutrophication from Agricultural Sources : Phosphorus Chemistry of Mineral and Peat Soils in Ireland (2000-LS-2.1.1b-M2) : Final Report*. *Phosphorus chemistry of mineral and peat soils in Ireland* (Johnstown Castle, Co. Wexford : Environmental Protection Agency, Johnstown Castle, Co. Wexford, 2005).

95. Amrani, M., D. G., W. & and Moughli, L. Improving the soil test phosphorus calibration for calcareous soils. *Commun. Soil Sci. Plant Anal.* **30**, 129–144 (1999).

96. Yang, Y.-F. *et al.* Characteristics of phosphorus adsorption and desorption of soils from wetlands recovered from farmlands in Caizi Lake. *Chinese J. Appl. Ecol.* **25**, (2014).

97. Ahmed, W. *et al.* Impacts of long-term inorganic and organic fertilization on phosphorus adsorption and desorption characteristics in red paddies in southern China. *PLoS One* **16**, e0246428 (2021).

98. Xu, M. *et al.* Biochar addition to soil highly increases P retention and decreases the risk of phosphate contamination of waters. *Environ. Chem. Lett.* **17**, 533–541 (2019).

99. Bache, B. W. & Williams, E. G. A phosphate sorption index for soils. *J. Soil Sci.* **22**, 289–301 (1971).

100. Pautler, M. C. & Sims, J. T. Relationships Between Soil Test Phosphorus, Soluble Phosphorus, and Phosphorus Saturation in Delaware Soils. *Soil Sci. Soc. Am. J.* **64**, 765–773 (2000).

101. Renneson, M. *et al.* Degree of phosphorus saturation in agricultural loamy soils with a near-neutral pH. *Eur. J. Soil Sci.* **66**, 33–41 (2015).

102. Xue, Q. *et al.* Deriving sorption indices for the prediction of potential phosphorus loss from calcareous soils. *Environ. Sci. Pollut. Res.* **21**, 1564–1571 (2014).

103. McKeague, J. A. & Day, J. H. Dithionite- and Oxalate-Extractable Fe and Al as Aids in

Differentiating Various Classes of Soils. *Can. J. Soil Sci.* **46**, 13– 22 (1966).

104. Beek, J. Phosphate retention by soil in relation to waste disposal. *Wageningen Univ. Res.* (1979).

105. Maguire, R. O. & Sims, J. T. Measuring Agronomic and Environmental Soil Phosphorus Saturation and Predicting Phosphorus Leaching with Mehlich 3. *Soil Sci. Soc. Am. J.* **66**, 2033–2039 (2002).

106. Ige, D. V, Akinremi, O. O., Flaten, D. N., Ajiboye, B. & Kashem, M. A. Phosphorus sorption capacity of alkaline Manitoba soils and its relationship to soil properties. *Can. J. Soil Sci.* **85**, 417–426 (2005).

107. Van der Zee, S. E. A. T. M. & Van Riemsdijk, W. H. Van der Zee, S.E.A.T.M. & Van Riemsdijk, W.H. 1988 Model for long-term phosphate reaction kinetics in soil. *J. Environ. Qual.* **17**, 35– 41 (1988).

108. Maguire, R. O., Foy, R. H., Bailey, J. S. & Sims, J. T. Estimation of the phosphorus sorption capacity of acidic soils in Ireland. *Eur. J. Soil Sci.* **52**, 479–487 (2001).

109. Burkitt, L. L., Moody, P. W., Gourley, C. J. P. & Hannah, M. C. A simple phosphorus buffering index for Australian soils. *Aust. J. Soil Res.* **40**, 497–513 (2002).

110. Burkitt, L., Sale, P. W. G. & Gourley, C. J. P. Soil phosphorus buffering measures should not be adjusted for current phosphorus fertility. *Aust. J. Soil Res.* **46**, (2008).

111. Peverill, K. I., Sparrow, L. A. & Reuter, D. J. *Soil Analysis: An Interpretation Manual.* (CSIRO, 1999).

112. Moody, P. W. Interpretation of a single-point P buffering index for adjusting critical levels of the Colwell soil P test. *Soil Res.* **45**, 55–62 (2007).

113. Hall, R. *et al.* Geospatial modelling of soil phosphorus fractions and sorption indicators from heterogeneous landscapes in Ireland. *Geoderma Reg.* **35**, e00729 (2023).

114. Vandecar, K. L., Lawrence, D. & Clark, D. Phosphorus Sorption Dynamics of Anion Exchange Resin Membranes in Tropical Rain Forest Soils. *Soil Sci. Soc. Am. J.* **75**, 1520–1529 (2011).

115. Antonangelo, J. A. *et al.* Phosphorus speciation by P-XANES in an Oxisol under long-term no-till cultivation. *Geoderma* **377**, 114580 (2020).

116. Soriano-Disla, J. M., Janik, L. J., Rossel, R. A. V., MacDonald, L. M. & McLaughlin, M. J. The Performance of Visible, Near-, and Mid-Infrared Reflectance Spectroscopy for Prediction of Soil Physical, Chemical, and Biological Properties. *Appl. Spectrosc. Rev.* **49**, 139–186 (2014).

117. Janik, L., Merry, R. H. & Skjemstad, J. O. Can mid infrared diffuse reflectance analysis replace soil extractions? *Aust. J. Exp. Agric.* **38**, 681–696 (1998).

118. Ng, W., Minasny, B., Jeon, S. H. & McBratney, A. Mid-infrared spectroscopy for accurate measurement of an extensive set of soil properties for assessing soil functions. *Soil Secur.* **6**, 100043 (2022).

119. Malik, A. K., Kumar, R. & Heena. Spectroscopy: Types. in (eds. Caballero, B., Finglas, P. M. & Toldrá, F. B. T.-E. of F. and H.) 64–72 (Academic Press, Oxford, 2016). doi:https://doi.org/10.1016/B978-0-12-384947-2.00643-7.

120. Alpert, N. L., Keiser, W. E. & Szymanski, H. A. *IR: Theory and Practice of Infrared Spectroscopy*. (Springer US, 2012).

121. Śliwińska, A., Smolinski, A. & Kucharski, P. Simultaneous Analysis of Heavy Metal Concentration in Soil Samples. *Applied Sciences* vol. 9 at https://doi.org/10.3390/app9214705 (2019).

122. Durig, J. R., Panikar, S. S. & Klaassen, J. J. Far Infrared Spectroscopy Applications. in *Reference Module in Chemistry, Molecular Sciences and Chemical Engineering* (2013). doi:10.1016/B978-0-12-409547-2.05309-9.

123. Chapter 3 - The Current Situation in Ultra-Precision Technology – Silicon Single Crystals as an Example. in (eds. Doi, T., Marinescu, I. D. & Kurokawa, S. B. T.-A. in C. M. P. P. T.) 15–111 (William Andrew Publishing, Oxford, 2012). doi:https://doi.org/10.1016/B978-1-4377-7859-5.00003-X.

124. Stenberg, B., Viscarra Rossel, R. A., Mouazen, A. M. & Wetterlind, J. Chapter Five - Visible and Near Infrared Spectroscopy in Soil Science. in (ed. Sparks, D. L. B. T.-A. in A.) vol. 107 163–215 (Academic Press, 2010).

125. Zhou, X., Deng, P. & Xu, K. *波谱解析 Introduction to Spectroscopy*. (化学工业出版社, 2015).

126. Raphael, L. Application of FTIR spectroscopy to agricultural soils analysis. *Fourier Transform. Anal. Approaches FTIR Strateg.* 385–404 (2011).

127. Viscarra Rossel, R. A., Walvoort, D. J. J., McBratney, A. B., Janik, L. J. & Skjemstad, J. O. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma* **131**, 59–75 (2006).

128. Dalal, R. C. & Henry, R. J. Simultaneous Determination of Moisture, Organic Carbon, and Total Nitrogen by Near Infrared Reflectance Spectrophotometry. *Soil Sci. Soc. Am. J.* **50**, 120–123 (1986).

129. Margenot, A. J., Calderón, F. J., Goyne, K. W., Dmukome, F. N. & Parikh, S. J. IR spectroscopy, soil analysis applications. in *Encyclopedia of spectroscopy and spectrometry* 448–454 (Elsevier, 2016).

130. Volkov, D. S., Rogova, O. B. & Proskurnin, M. A. Organic Matter and Mineral Composition of Silicate Soils: FTIR Comparison Study by Photoacoustic, Diffuse Reflectance, and Attenuated Total Reflection Modalities. *Agronomy* vol. 11 at https://doi.org/10.3390/agronomy11091879 (2021).

131. Robertson, J., Shand, C. & Perez-Fernandez, E. The application of Fourier transform infrared, near infrared and X-ray fluorescence spectroscopy to soil analysis. **28**, 9–13 (2016).

132. Xing, Z. *et al.* Agricultural soil characterization by FTIR spectroscopy at micrometer scales: Depth profiling by photoacoustic spectroscopy. *Geoderma* **335**, 94–103 (2019).

133. Tinti, A., Tugnoli, V., Bonora, S. & Francioso, O. Recent applications of vibrational mid-Infrared

(IR) spectroscopy for studying soil components: a review. *J. Cent. Eur. Agric.* **16**, 1–22 (2015).

134. Stuart, B. H. *Infrared Spectroscopy: Fundamentals and Applications*. (John Wiley & Sons, 2004).

135. Nguyen, T. T., Janik, L. J. & Raupach, M. Diffuse reflectance infrared Fourier transform (DRIFT) spectroscopy in soil studies. *Soil Res.* **29**, 49–67 (1991).

136. Nath, D. *et al.* Strategies to admittance soil quality using mid-infrared (mid-IR) spectroscopy an alternate tool for conventional lab analysis: A global perspective. *Environ. Challenges* **7**, 100469 (2022).

137. Soriano-Disla, J. M. *et al.* The use of diffuse reflectance mid-infrared spectroscopy for the prediction of the concentration of chemical elements estimated by X-ray fluorescence in agricultural and grazing European soils. *Appl. Geochemistry* **29**, 135–143 (2013).

138. Stenberg, B. & Rossel, R. A. V. Diffuse reflectance spectroscopy for high-resolution soil sensing. *Prox. soil Sens.* 29–47 (2010).

139. Singh, K. *et al.* Near infrared diffuse reflectance spectroscopy for rapid and comprehensive soil condition assessment in smallholder cacao farming systems of Papua New Guinea. *Catena* **183**, 104185 (2019).

140. Siebielec, G., McCarty, G. W., Stuczynski, T. I. & Reeves III, J. B. Near-and mid-infrared diffuse reflectance spectroscopy for measuring soil metal content. *J. Environ. Qual.* **33**, 2056–2069 (2004).

141. Reeves III, J. B. & Smith, D. B. The potential of mid-and near-infrared diffuse reflectance spectroscopy for determining major-and trace-element concentrations in soils from a geochemical survey of North America. *Appl. Geochemistry* **24**, 1472–1481 (2009).

142. Reeves, J. B., McCarty, G. W. & Reeves, V. B. Mid-infrared diffuse reflectance spectroscopy for the quantitative analysis of agricultural soils. *J. Agric. Food Chem.* **49**, 766–772 (2001).

143. Ge, Y., Wadoux, A. & Peng, Y. *A Primer on Soil Analysis Using Visible and Near-Infrared (Vis-NIR)*

and Mid-Infrared (MIR) Spectroscopy. (FAO, Rome, 2022). doi:https://doi.org/10.4060/cb9005en.

144. Leue, M., Ellerbrock, R. H., Bänninger, D. & Gerke, H. H. Impact of Soil Microstructure Geometry on DRIFT Spectra: Comparisons with Beam Trace Modeling. *Soil Sci. Soc. Am. J.* **74**, 1976–1986 (2010).

145. Bachion de Santana, F. & Daly, K. A comparative study of MIR and NIR spectral models using ball-milled and sieved soil for the prediction of a range soil physical and chemical parameters. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **279**, 121441 (2022).

146. Breure, T. S. *et al.* Comparing the effect of different sample conditions and spectral libraries on the prediction accuracy of soil properties from near- and mid-infrared spectra at the field-scale. *Soil Tillage Res.* **215**, 105196 (2022).

147. Dunne, K. S., Holden, N. M., O'Rourke, S. M., Fenelon, A. & Daly, K. Prediction of phosphorus sorption indices and isotherm parameters in agricultural soils using mid-infrared spectroscopy. *Geoderma* **358**, (2020).

148. Barthès, B. *et al.* Studying the Physical Protection of Soil Carbon with Quantitative Infrared Spectroscopy. *J. Near Infrared Spectrosc.* **24**, 199 (2016).

149. Baumann, P. *et al.* Developing the Swiss mid-infrared soil spectral library for local estimation and monitoring. *SOIL* **7**, 525–546 (2021).

150. Terra, F. S., Demattê, J. A. M. & Viscarra Rossel, R. A. Spectral libraries for quantitative analyses of tropical Brazilian soils: Comparing vis–NIR and mid-IR reflectance data. *Geoderma* **255**–**256**, 81–93 (2015).

151. Hutengs, C., Ludwig, B., Jung, A., Eisele, A. & Vohland, M. Comparison of portable and bench-top spectrometers for mid-infrared diffuse reflectance measurements of soils. *Sensors* **18**, 993 (2018).

152. Deiss, L., Margenot, A. J., Culman, S. W. & Demyan, M. S. Optimizing acquisition parameters in diffuse reflectance infrared Fourier transform spectroscopy of soils. *Soil Sci. Soc. Am. J.* **84**, 930–948 (2020).

153. Wang, X., Sanderman, J. & Yoo, K. Climate-dependent topographic effects on pyrogenic soil carbon in southeastern Australia. *Geoderma* **322**, 121–130 (2018).

154. Shi, L., O'Rourke, S., de Santana, F. B. & Daly, K. Prediction of soil bulk density in agricultural soils using mid-infrared spectroscopy. *Geoderma* **434**, 116487 (2023).

155. Mohamed, E. S., Saleh, A. M., Belal, A. B. & Gad, A. Application of near-infrared reflectance for quantitative assessment of soil properties. *Egypt. J. Remote Sens. Sp. Sci.* **21**, 1–14 (2018).

156. Stenberg, B. Effects of soil sample pretreatments and standardised rewetting as interacted with sand classes on Vis-NIR predictions of clay and soil organic carbon. *Geoderma* **158**, 15–22 (2010).

157. Bellon-Maurel, V., Fernandez-Ahumada, E., Palagos, B., Roger, J.-M. & McBratney, A. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. *TrAC Trends Anal. Chem.* **29**, 1073–1081 (2010).

158. Wang, L. & Wang, R. Determination of soil pH from Vis-NIR spectroscopy by extreme learning machine and variable selection: A case study in lime concretion black soil. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **283**, 121707 (2022).

159. Fang, Q. *et al.* Visible and Near-Infrared Reflectance Spectroscopy for Investigating Soil Mineralogy: A Review. *J. Spectrosc.* **2018**, 3168974 (2018).

160. Tümsavaş, Z., Tekin, Y., Ulusoy, Y. & Mouazen, A. M. Prediction and mapping of soil clay and sand contents using visible and near-infrared spectroscopy. *Biosyst. Eng.* **177**, 90–100 (2019).

161. Islam, K., Singh, B. & McBratney, A. Simultaneous estimation of several soil properties by ultra-violet, visible, and near-infrared reflectance spectroscopy. *Soil Res.* **41**, 1101–1114 (2003).

162. Chang, C.-W., Laird, D. A., Mausbach, M. J. & Hurburgh, C. R. Near-Infrared Reflectance Spectroscopy–Principal Components Regression Analyses of Soil Properties. *Soil Sci. Soc. Am. J.* **65**, 480–490 (2001).

163. Rossel, R. A. V. & Webster, R. Predicting soil properties from the Australian soil visible–near infrared spectroscopic database. *Eur. J. Soil Sci.* **63**, 848–860 (2012).

164. Krischenko, V. P., Samokhvalov, S. G., Fomina, L. G. & Novikova, G. A. Use of infrared spectroscopy for the determination of some properties of soil. *Interagrotech, Moscow, USSR* 239–249 (1991).

165. Thomasson, J. A., Sui, R., Cox, M. S. & Al–Rajehy, A. Soil reflectance sensing for determining soil properties in precision agriculture. *Trans. ASAE* **44**, 1445 (2001).

166. He, Y., Song, H., Pereira, A. G. & Gómez, A. H. A New Approach to Predict N, P, K and OM Content in a Loamy Mixed Soil by Using Near Infrared Reflectance Spectroscopy BT - Advances in Intelligent Computing. in (eds. Huang, D.-S., Zhang, X.-P. & Huang, G.-B.) 859–867 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2005).

167. Daniel, K. W., Tripathi, N. K. & Honda, K. Artificial neural network analysis of laboratory and <emph type="2">in situ </emph>spectra for the estimation of macronutrients in soils of Lop Buri (Thailand). *Soil Res.* **41**, 47–59 (2003).

168. Bogrekci, I., Suk Lee, W. & Herrera, J. Spectral Signatures for the Lake Okeechobee Soils Using UV-VIS-NIR Spectroscopy and Predicting Phosphorus Concentrations. at https://doi.org/https://doi.org/10.13031/2013.16148 (2004).

169. Bogrekci, I. & S. Lee, W. SPECTRAL MEASUREMENT OF COMMON SOIL PHOSPHATES. *Trans. ASAE* **48**, 2371–2378 (2005).

170. Mouazen, A. M., Kuang, B., De Baerdemaeker, J. & Ramon, H. Comparison among principal component, partial least squares and back propagation neural network analyses for accuracy

of measurement of selected soil properties with visible and near infrared spectroscopy. *Geoderma* **158**, 23–31 (2010).

171. Reda, R. *et al.* Predicting soil phosphorus and studying the effect of texture on the prediction accuracy using machine learning combined with near-infrared spectroscopy. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **242**, 118736 (2020).

172. Serrano, J. *et al.* Evaluation of Near Infrared Spectroscopy (NIRS) for Estimating Soil Organic Matter and Phosphorus in Mediterranean Montado Ecosystem. *Sustainability* vol. 13 at https://doi.org/10.3390/su13052734 (2021).

173. Abdi, D., Tremblay, G. F., Ziadi, N., Bélanger, G. & Parent, L.-É. Predicting Soil Phosphorus-Related Properties Using Near-Infrared Reflectance Spectroscopy. *Soil Sci. Soc. Am. J.* **76**, 2318–2326 (2012).

174. Mouazen, A. M. & Kuang, B. On-line visible and near infrared spectroscopy for in-field phosphorous management. *Soil Tillage Res.* **155**, 471–477 (2016).

175. Recena, R., Fernández-Cabanás, V. M. & Delgado, A. Soil fertility assessment by Vis-NIR spectroscopy: Predicting soil functioning rather than availability indices. *Geoderma* **337**, 368–374 (2019).

176. Saidi, S. *et al.* Use of Vis-NIR reflectance spectroscopy for estimating soil phosphorus sorption parameters at the watershed scale. *Soil Tillage Res.* **248**, 106460 (2025).

177. Chang, C.-W., Laird, D. A. & Hurburgh, C. R. J. INFLUENCE OF SOIL MOISTURE ON NEAR-INFRARED REFLECTANCE SPECTROSCOPIC MEASUREMENT OF SOIL PROPERTIES. *Soil Sci.* **170**, (2005).

178. Zornoza, R. *et al.* Near infrared spectroscopy for determination of various physical, chemical and biochemical properties in Mediterranean soils. *Soil Biol. Biochem.* **40**, 1923–1930 (2008).

179. Xu, D. *et al.* Assessment of important soil properties related to Chinese Soil Taxonomy based

on vis–NIR reflectance spectroscopy. *Comput. Electron. Agric.* **144**, 1–8 (2018).

180. Yang, M., Mouazen, A., Zhao, X. & Guo, X. Assessment of a soil fertility index using visible and near-infrared spectroscopy in the rice paddy region of southern China. *Eur. J. Soil Sci.* **71**, 615–626 (2020).

181. Qi, J. *et al.* A Novel Model for Soil Organic Matter and Total Nitrogen Detection Based on Visible/Shortwave Near-Infrared Spectroscopy. *Land* vol. 14 at https://doi.org/10.3390/land14020329 (2025).

182. Yu, B., Yan, C., Yuan, J., Ding, N. & Chen, Z. Prediction of soil properties based on characteristic wavelengths with optimal spectral resolution by using Vis-NIR spectroscopy. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **293**, 122452 (2023).

183. Davari, M., Karimi, S. A., Bahrami, H. A., Taher Hossaini, S. M. & Fahmideh, S. Simultaneous prediction of several soil properties related to engineering uses based on laboratory Vis-NIR reflectance spectroscopy. *CATENA* **197**, 104987 (2021).

184. Reeves, J. B. & McCarty, G. W. Quantitative analysis of agricultural soils using near infrared reflectance spectroscopy and a fibre-optic probe. *J. Near Infrared Spectrosc.* **9**, 25–34 (2001).

185. Ulusoy, Y., Tekin, Y., Tümsavaş, Z. & Mouazen, A. M. Prediction of soil cation exchange capacity using visible and near infrared spectroscopy. *Biosyst. Eng.* **152**, 79–93 (2016).

186. Lu, P., Wang, L., Niu, Z., Li, L. & Zhang, W. Prediction of soil properties using laboratory VIS–NIR spectroscopy and Hyperion imagery. *J. Geochemical Explor.* **132**, 26–33 (2013).

187. Tigga, P. *et al.* Potential of Soil Spectroscopy as an Alternative to Soil Testing. in 81–104 (2022).

188. Wijewardane, N. K., Ge, Y., Wills, S. & Libohova, Z. Predicting Physical and Chemical Properties of US Soils with a Mid-Infrared Reflectance Spectral Library. *Soil Sci. Soc. Am. J.* **82**, 722–731 (2018).

189. Barra, I., Haefele, S. M., Sakrabani, R. & Kebede, F. Soil spectroscopy with the use of

chemometrics, machine learning and pre-processing techniques in soil diagnosis: Recent advances–A review. *TrAC Trends Anal. Chem.* **135**, 116166 (2021).

190. Ma, F., Du, C. W., Zhou, J. M. & Shen, Y. Z. Investigation of soil properties using different techniques of mid-infrared spectroscopy. *Eur. J. Soil Sci.* **70**, 96–106 (2019).

191. Minasny, B., Tranter, G., McBratney, A. B., Brough, D. M. & Murphy, B. W. Regional transferability of mid-infrared diffuse reflectance spectroscopic prediction for soil chemical properties. *Geoderma* **153**, 155–162 (2009).

192. Ramaroson, H. V., Becquer, T., Rabeharisoa, H. R. A. R. L. & Rakotondrazafy, A. F. M. Prediction of Phosphorus Sorption Index and Availability by NIR and MIR in Soils in Madagascar. *Land* **12**, 196 (2023).

193. Janik, L. J., Forrester, S. T. & Rawson, A. The prediction of soil chemical and physical properties from mid-infrared spectroscopy and combined partial least-squares regression and neural networks (PLS-NN) analysis. *Chemom. Intell. Lab. Syst.* **97**, 179–188 (2009).

194. Burkitt, L. L., Gourley, C. J. P., Hannah, M. C. & Sale, P. W. G. Assessing alternative approaches to predicting soil phosphorus sorption. *Soil Use Manag.* **22**, 325–333 (2006).

195. Forrester, S. *et al. Prediction of a Range of Phosphorus Buffering Techniques Using Mid-Infrared Spectroscopy.* (2003).

196. Janik, L. J., Merry, R. H. & Skjemstad, J. O. Can mid infrared diffuse reflectance analysis replace soil extractions? *Aust. J. Exp. Agric.* **38**, 681–696 (1998).

197. Forrester, S. T. *et al.* Use of handheld mid-infrared spectroscopy and partial least-squares regression for the prediction of the phosphorus buffering index in Australian soils. *Soil Res.* **53**, 67–80 (2015).

198. Barra, I., Elmoatassem, T. & Kebede, F. Soil Particle Size Thresholds in Soil Spectroscopy and Its Effect on the Multivariate Models for the Analysis of Soil Properties. *Sensors* **23**, (2023).

199. Knox, N. M. *et al.* Modelling soil carbon fractions with visible near-infrared (VNIR) and mid-infrared (MIR) spectroscopy. *Geoderma* **239**–**240**, 229–239 (2015).

200. McCarty, G. W., Reeves, J. B., Reeves, V. B., Follett, R. F. & Kimble, J. M. Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil carbon measurement. *Soil Sci. Soc. Am. J.* **66**, 640–646 (2002).

201. Hati, K. M. *et al.* Mid-Infrared Reflectance Spectroscopy for Estimation of Soil Properties of Alfisols from Eastern India. *Sustainability* vol. 14 at https://doi.org/10.3390/su14094883 (2022).

202. de Santana, F. B., Grunsky, E. C., Fitzsimons, M. M., Gallagher, V. & Daly, K. Diffuse reflectance mid infra-red spectroscopy combined with machine learning algorithms can differentiate spectral signatures in shallow and deeper soils for the prediction of pH and organic matter content. *CATENA* **218**, 106552 (2022).

203. Viscarra Rossel, R. A., Adamchuk, V. I., Sudduth, K. A., McKenzie, N. J. & Lobsey, C. Chapter Five - Proximal Soil Sensing: An Effective Approach for Soil Measurements in Space and Time. in *Advances in Agronomy* (ed. Sparks, D. L. B. T.-A. in A.) vol. 113 243–291 (Academic Press, 2011).

204. Metzger, K., Zhang, C., Ward, M. & Daly, K. Mid-infrared spectroscopy as an alternative to laboratory extraction for the determination of lime requirement in tillage soils. *Geoderma* **364**, 114171 (2020).

205. Hong, T., Yin, J.-Y., Nie, S.-P. & Xie, M.-Y. Applications of infrared spectroscopy in polysaccharide structural analysis: Progress, challenge and perspective. *Food Chem. X* **12**, 100168 (2021).

206. Beć, K. B., Grabska, J. & Huck, C. W. Principles and Applications of Miniaturized Near-Infrared (NIR) Spectrometers. *Chem. – A Eur. J.* **27**, 1514–1532 (2021).

207. Metrohm USA Inc. A Technology Comparison of Near-Infrared Spectroscopy. *AZoM*

https://www.azom.com/article.aspx?ArticleID=14371 (2017).

208. AZoM. FT-NIR Analyzers vs. Dispersion Spectrometers: A Comparison. *Process Insights – Optical Absorption Spectroscopy* https://www.azom.com/article.aspx?ArticleID=17314 (2018).

209. Ozaki, Y., Genkawa, T. & Futami, Y. Near-Infrared Spectroscopy. in (eds. Lindon, J. C., Tranter, G. E. & Koppenaal, D. W. B. T.-E. of S. and S. (Third E.) 40–49 (Academic Press, Oxford, 2017). doi:https://doi.org/10.1016/B978-0-12-409547-2.12164-X.

210. Process Insights. Dispersive NIR Spectrometer Advantages in Process Spectroscopy. *Process Insights, Inc.* https://www.process-insights.com/products-3/products-industrial/nir-uv-vis-process-and-lab-analyzers/dispersive-nir-spectrometers/ (2025).

211. Jeon, S. H. *et al.* Predicting soil properties for fertiliser recommendation in South Korea using MIR spectroscopy. *Geoderma Reg.* **39**, e00901 (2024).

212. Dangal, S. R. S., Sanderman, J., Wills, S. & Ramirez-Lopez, L. Accurate and Precise Prediction of Soil Properties from a Large Mid-Infrared Spectral Library. *Soil Systems* vol. 3 at https://doi.org/10.3390/soilsystems3010011 (2019).

213. Shi, Z. *et al.* Comparison of Depth-Specific Prediction of Soil Properties: MIR vs. Vis-NIR Spectroscopy. *Sensors* vol. 23 at https://doi.org/10.3390/s23135967 (2023).

214. Mammadov, E., Denk, M., Mamedov, A. & Cornelia, G. Predicting Soil Properties for Agricultural Land in the Caucasus Mountains Using Mid-Infrared Spectroscopy. (2024).

215. Hutengs, C., Seidel, M., Oertel, F., Ludwig, B. & Vohland, M. In situ and laboratory soil spectroscopy with portable visible-to-near-infrared and mid-infrared instruments for the assessment of organic carbon in soils. *Geoderma* **355**, 113900 (2019).

216. Metzger, K., Zhang, C. & Daly, K. From benchtop to handheld MIR for soil analysis: Predicting lime requirement and organic matter in agricultural soils. *Biosyst. Eng.* **204**, 257–269 (2021).

217. Esquerre, C. A. *et al.* Use of an NIR MEMS spectrophotometer and visible/NIR hyperspectral

imaging systems to predict quality parameters of treated ground peppercorns. *LWT* **131**, 109761 (2020).

218. Sharififar, A., Singh, K., Jones, E., Ginting, F. I. & Minasny, B. Evaluating a low-cost portable NIR spectrometer for the prediction of soil organic and total carbon using different calibration models. *Soil Use Manag.* **35**, 607–616 (2019).

219. Ng, W. *et al.* Developing a soil spectral library using a low-cost NIR spectrometer for precision fertilization in Indonesia. *Geoderma Reg.* **22**, e00319 (2020).

220. Tang, Y., Jones, E. & Minasny, B. Evaluating low-cost portable near infrared sensors for rapid analysis of soils from South Eastern Australia. *Geoderma Reg.* **20**, e00240 (2020).

221. Dhawale, N. M., Adamchuk, V. I., Prasher, S. O., Rossel, R. A. V. & Ismail, A. A. Evaluation of Two Portable Hyperspectral-Sensor-Based Instruments to Predict Key Soil Properties in Canadian Soils. *Sensors* vol. 22 at https://doi.org/10.3390/s22072556 (2022).

222. Priori, S., Mzid, N., Pascucci, S., Pignatti, S. & Casa, R. Performance of a Portable FT-NIR MEMS Spectrometer to Predict Soil Features. *Soil Systems* vol. 6 at https://doi.org/10.3390/soilsystems6030066 (2022).

223. Metzger, K. *et al.* The use of visible and near-infrared spectroscopy for in-situ characterization of agricultural soil fertility: A proposition of best practice by comparing scanning positions and spectrometers. *Soil Use Manag.* **40**, e12952 (2024).

224. Izaurralde, R. C. *et al.* Evaluation of Three Field-Based Methods for Quantifying Soil Carbon. *PLoS One* **8**, e55560 (2013).

225. Kuang, B. & Mouazen, A. M. Calibration of visible and near infrared spectroscopy for soil analysis at the field scale on three European farms. *Eur. J. Soil Sci.* **62**, 629–636 (2011).

226. Wenjun, J., Zhou, S., Jingyi, H. & Shuo, L. In Situ Measurement of Some Soil Properties in Paddy Soil Using Visible and Near-Infrared Spectroscopy. *PLoS One* **9**, e105708 (2014).

227. Loria, N., Lal, R. & Chandra, R. Handheld In Situ Methods for Soil Organic Carbon Assessment. *Sustainability* vol. 16 at https://doi.org/10.3390/su16135592 (2024).

228. Viscarra Rossel, R. A., Cattle, S. R., Ortega, A. & Fouad, Y. In situ measurements of soil colour, mineral composition and clay content by vis–NIR spectroscopy. *Geoderma* **150**, 253–266 (2009).

229. Syafinaz, A., Chia, K. S. & Mohd. Fuzi, S. F. Z. A Portable in-situ Near-infrared LEDs-based Soil Nitrogen Sensor Using Artificial Neural Network. *Int. J. Integr. Eng.* **10**, 81–87 (2018).

230. Yang, S., White, B., de Santana, F. B., Hall, R. L. & Daly, K. Comparing the potential of benchtop and handheld mid-infrared spectrometers for predicting soil phosphorus (P) sorption capacity and evaluating the influence of sample preparation. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **322**, 124856 (2024).

231. Reeves, J. B., McCarty, G. W. & Hively, W. D. Mid- Versus Near-Infrared Spectroscopy for On-Site Analysis of Soil BT - Proximal Soil Sensing. in (eds. Viscarra Rossel, R. A., McBratney, A. B. & Minasny, B.) 133–142 (Springer Netherlands, Dordrecht, 2010). doi:10.1007/978-90-481-8859-8_11.

232. Soriano-Disla, J. M., Janik, L. J., Allen, D. J. & McLaughlin, M. J. Evaluation of the performance of portable visible-infrared instruments for the prediction of soil properties. *Biosyst. Eng.* **161**, 24–36 (2017).

233. Martínez-España, R., Bueno-Crespo, A., Soto, J., Janik, L. J. & Soriano-Disla, J. M. Developing an intelligent system for the prediction of soil properties with a portable mid-infrared instrument. *Biosyst. Eng.* **177**, 101–108 (2019).

234. Soriano-Disla, J. M. *et al.* The use of mid-infrared diffuse reflectance spectroscopy for acid sulfate soil analysis. *Sci. Total Environ.* **646**, 1489–1502 (2019).

235. Metzger, K. Development of mid-infrared spectroscopic methods for the determination of soil chemical quality in tillage fields. (NUI Galway, 2021). doi:https://doi.org/10.13025/17109.

236. Janik, L. J., Soriano-Disla, J. M., Forrester, S. T. & McLaughlin, M. J. Moisture effects on diffuse reflection infrared spectra of contrasting minerals and soils: A mechanistic interpretation. *Vib. Spectrosc.* **86**, 244–252 (2016).

237. Seidel, M. *et al.* Soil moisture effects on predictive VNIR and MIR modeling of soil organic carbon and clay content. *Geoderma* **427**, 116103 (2022).

238. Dai, L. *et al.* In-situ prediction of soil organic carbon contents in wheat-rice rotation fields via visible near-infrared spectroscopy. *Soil Environ. Heal.* **2**, 100113 (2024).

239. Reeves, J. B. Near- versus mid-infrared diffuse reflectance spectroscopy for soil analysis emphasizing carbon and laboratory versus on-site analysis: Where are we and what needs to be done? *Geoderma* **158**, 3–14 (2010).

240. Yin, J. *et al.* Prediction of Soil Properties in a Field in Typical Black Soil Areas Using in situ MIR Spectra and Its Comparison with vis-NIR Spectra. *Remote Sensing* vol. 15 at https://doi.org/10.3390/rs15082053 (2023).

241. Dhawale, N. M. *et al.* Proximal soil sensing of soil texture and organic matter with a prototype portable mid-infrared spectrometer. *Eur. J. Soil Sci.* **66**, 661–669 (2015).

242. Ji, W. *et al.* Assessment of soil properties in situ using a prototype portable MIR spectrometer in two agricultural fields. *Biosyst. Eng.* **152**, 14–27 (2016).

243. Greenberg, I., Seidel, M., Vohland, M., Koch, H.-J. & Ludwig, B. Performance of in situ vs laboratory mid-infrared soil spectroscopy using local and regional calibration strategies. *Geoderma* **409**, 115614 (2022).

244. Courtney, R. & Harrington, T. Assessment of plant-available phosphorus in a fine textured sodic substrate. *Ecol. Eng.* **36**, 542–547 (2010).

245. Williams, W. L. *et al.* The Evolution of Data-Driven Modeling in Organic Chemistry. *ACS Cent. Sci.* **7**, 1622–1637 (2021).

246. Sjögren, R. Synergies between Chemometrics and Machine Learning TT - Synergier mellan kemometri och maskininlärning (swe). (Umeå Universitet, Department of Chemistry, Faculty of Science and Technology, Umeå University, 2021).

247. Wadoux, A. M. J.-C., Malone, B., Minasny, B., Fajardo, M. & McBratney, A. B. Introduction BT - Soil Spectral Inference with R: Analysing Digital Soil Spectra using the R Programming Environment. in (eds. Wadoux, A. M. J.-C., Malone, B., Minasny, B., Fajardo, M. & McBratney, A. B.) 1–10 (Springer International Publishing, Cham, 2021). doi:10.1007/978-3-030-64896-1_1.

248. Mokari, A., Guo, S. & Bocklitz, T. Exploring the Steps of Infrared (IR) Spectral Analysis: Pre-Processing, (Classical) Data Modelling, and Deep Learning. *Molecules* vol. 28 at https://doi.org/10.3390/molecules28196886 (2023).

249. Gautam, R., Vanga, S., Ariese, F. & Umapathy, S. Review of multidimensional data processing approaches for Raman and infrared spectroscopy. *EPJ Tech. Instrum.* **2**, 1–38 (2015).

250. Wadoux, A. M. J. C., Odeh, I. O. & McBratney, A. B. Overview of Pedometrics, in: Reference Module in Earth Systems and Environmental Sciences. *Elsevier* (2021).

251. Vestergaard, R.-J. *et al.* Evaluation of Optimized Preprocessing and Modeling Algorithms for Prediction of Soil Properties Using VIS-NIR Spectroscopy. *Sensors* vol. 21 at https://doi.org/10.3390/s21206745 (2021).

252. Barra, I., Haefele, S., Sakrabani, R. & Kebede, F. Soil spectroscopy with the use of chemometrics, machine learning and pre-processing techniques in soil diagnosis: Recent advances -A review. *TrAC Trends Anal. Chem.* **135**, (2020).

253. Helin, R., Indahl, U. G., Tomic, O. & Liland, K. H. On the possible benefits of deep learning for spectral preprocessing. *J. Chemom.* **36**, e3374 (2022).

254. Jiao, Y., Li, Z., Chen, X. & Fei, S. Preprocessing methods for near-infrared spectrum calibration. *J. Chemom.* **34**, (2020).

255. Rinnan, Å., Berg, F. van den & Engelsen, S. B. Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends Anal. Chem.* **28**, 1201–1222 (2009).

256. Singh, K. *et al.* Optimization of spectral pre-processing for estimating soil condition on small farms. *Soil Use Manag.* **38**, 150–163 (2022).

257. Wu, D., Pengcheng, N., Yong, H., Zhiping, W. & and Wu, H. Spectral Multivariable Selection and Calibration in Visible-Shortwave Near-Infrared Spectroscopy for Non-Destructive Protein Assessment of Spirulina Microalga Powder. *Int. J. Food Prop.* **16**, 1002–1015 (2013).

258. Xue, J. *et al.* The validity domain of sensor fusion in sensing soil quality indicators. *Geoderma* **438**, 116657 (2023).

259. Barra, I., Briak, H. & Kebede, F. The application of statistical preprocessing on spectral data does not always guarantee the improvement of the predictive quality of multivariate models: Case of soil spectroscopy applied to Moroccan soils. *Vib. Spectrosc.* **121**, 103409 (2022).

260. Lazim, S. S. R., Nawi, N., Chen, G., Jensen, T. & Md Rasli, A. M. Influence of different pre-processing methods in predicting sugarcane quality from near-infrared (NIR) spectral data. **23**, S231–S236 (2016).

261. Witteveen, M. *et al.* Comparison of preprocessing techniques to reduce nontissue-related variations in hyperspectral reflectance imaging. *J. Biomed. Opt.* **27**, 106003 (2022).

262. Salazar, O. *et al.* Evaluation of a miniaturized portable NIR spectrometer for the prediction of soil properties in Mediterranean central Chile. *Geoderma Reg.* **34**, e00675 (2023).

263. Gao, X., Lv, B., He, K. & Tang, X. Effects of different pretreatment and modeling methods on soil moisture content detected by near infrared spectroscopy . *2021 ASABE Annual International Virtual Meeting* 1 at https://doi.org/https://doi.org/10.13031/aim.202100335 (2021).

264. Ji, W., Viscarra Rossel, R. A. & Shi, Z. Accounting for the effects of water and the environment

on proximally sensed vis–NIR soil spectra and their calibrations. *Eur. J. Soil Sci.* **66**, 555–565 (2015).

265. Tan, Y., Jiang, Q., Yu, L., Liu, H. & Zhang, B. Reducing the Moisture Effect and Improving the Prediction of Soil Organic Matter With VIS-NIR Spectroscopy in Black Soil Area. *IEEE Access* **9**, 5895–5905 (2021).

266. de Santana, F. B., de Giuseppe, L. O., de Souza, A. M. & Poppi, R. J. Removing the moisture effect in soil organic matter determination using NIR spectroscopy and PLSR with external parameter orthogonalization. *Microchem. J.* **145**, 1094–1101 (2019).

267. Nawar, S., Abdul Munnaf, M. & Mouazen, A. M. Machine Learning Based On-Line Prediction of Soil Organic Carbon after Removal of Soil Moisture Effect. *Remote Sensing* vol. 12 at https://doi.org/10.3390/rs12081308 (2020).

268. Munnaf, M. A. & Mouazen, A. M. Removal of external influences from on-line vis-NIR spectra for predicting soil organic carbon using machine learning. *CATENA* **211**, 106015 (2022).

269. Roger, J.-M., Chauchard, F. & Bellon-Maurel, V. EPO–PLS external parameter orthogonalisation of PLS application to temperature-independent measurement of sugar content of intact fruits. *Chemom. Intell. Lab. Syst.* **66**, 191–204 (2003).

270. Metzger, K., Liebisch, F., Herrera, J. M., Guillaume, T. & Bragazza, L. Prediction Accuracy of Soil Chemical Parameters by Field- and Laboratory-Obtained vis-NIR Spectra after External Parameter Orthogonalization. *Sensors* vol. 24 at https://doi.org/10.3390/s24113556 (2024).

271. Minasny, B. *et al.* Removing the effect of soil moisture from NIR diffuse reflectance spectra for the prediction of soil organic carbon. *Geoderma* **167**–**168**, 118–124 (2011).

272. Wijewardane, N. K., Ge, Y. & Morgan, C. L. S. Moisture insensitive prediction of soil properties from VNIR reflectance spectra based on external parameter orthogonalization. *Geoderma* **267**, 92–101 (2016).

273. Yu, W., Hong, Y., Chen, S., Chen, Y. & Zhou, L. Comparing Two Different Development Methods of External Parameter Orthogonalization for Estimating Organic Carbon from Field-Moist Intact Soils by Reflectance Spectroscopy. *Remote Sensing* vol. 14 at https://doi.org/10.3390/rs14061303 (2022).

274. Mirzaei, S. *et al.* Minimising the effect of moisture on soil property prediction accuracy using external parameter orthogonalization. *Soil Tillage Res.* **215**, 105225 (2022).

275. Yang, P. *et al.* Predicting soil organic carbon content using simulated insitu spectra and moisture correction algorithms in southern Xinjiang, China. *Geoderma Reg.* **37**, e00783 (2024).

276. S. Veum, K., A. Parker, P., A. Sudduth, K. & H. Holan, S. Predicting Profile Soil Properties with Reflectance Spectra via Bayesian Covariate-Assisted External Parameter Orthogonalization. *Sensors* vol. 18 at https://doi.org/10.3390/s18113869 (2018).

277. Ackerson, J. P., Demattê, J. A. M. & Morgan, C. L. S. Predicting clay content on field-moist intact tropical soils using a dried, ground VisNIR library with external parameter orthogonalization. *Geoderma* **259**–**260**, 196–204 (2015).

278. Ge, Y., Morgan, C. L. S. & Ackerson, J. P. VisNIR spectra of dried ground soils predict properties of soils scanned moist and intact. *Geoderma* **221**, 61–69 (2014).

279. Wadoux, A. M. soilspec: functions and data for the book Soil Spectral Inference with R. *R Packag.* (2020).

280. Héberger, K. Chapter 7 - Chemoinformatics—multivariate mathematical–statistical methods for data evaluation. in (eds. Vékey, K., Telekes, A. & Vertes, A. B. T.-M. A. of M. S.) 141–169 (Elsevier, Amsterdam, 2008). doi:https://doi.org/10.1016/B978-044451980-1.50009-4.

281. Sansana, J. *et al.* Recent trends on hybrid modeling for Industry 4.0. *Comput. Chem. Eng.* **151**, 107365 (2021).

282. Kahsay, A., Haile, M., Gebresamuel, G. & Mohammed, M. Developing soil quality indices to

investigate degradation impacts of different land use types in Northern Ethiopia. *Heliyon* **11**, e41185 (2025).

283. Abdel-Fattah, M. K. *et al.* Quantitative Evaluation of Soil Quality Using Principal Component Analysis: The Case Study of El-Fayoum Depression Egypt. *Sustainability* vol. 13 at https://doi.org/10.3390/su13041824 (2021).

284. Rossel, R. A. V. & Behrens, T. Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma* **158**, 46–54 (2010).

285. Rossel, R. A. V. *et al.* A global spectral library to characterize the world's soil. *Earth-Science Rev.* **155**, 198–230 (2016).

286. Wold, H. Systems Analysis by Partial Least Squares. in *Measuring the Unmeasurable* 221–251 (Springer Netherlands, Dordrecht, 1983). doi:10.1007/978-94-009-5079-5_11.

287. Dunn, K. *Process Improvement Using Data*. (Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0), 2025).

288. Janik, L., Merry, R., Forrester, S., Lanyon, D. & Rawson, A. Rapid Prediction of Soil Water Retention using Mid Infrared Spectroscopy. *Soil Sci. Soc. Am. J.* **Vol. 71**, 507–514 (2007).

289. Deiss, L., Margenot, A. J., Culman, S. W. & Demyan, M. S. Tuning support vector machines regression models improves prediction accuracy of soil properties in MIR spectroscopy. *Geoderma* **365**, 114227 (2020).

290. Jia, X. *et al.* Organic carbon prediction in soil cores using VNIR and MIR techniques in an alpine landscape. *Sci. Rep.* **7**, 2144 (2017).

291. Kuang, B., Tekin, Y. & Mouazen, A. M. Comparison between artificial neural network and partial least squares for on-line visible and near infrared spectroscopy measurement of soil organic carbon, pH and clay content. *Soil Tillage Res.* **146**, 243–252 (2015).

292. Zhu, J., Jin, Y., Zhu, W. & Lee, D. K. VIS-NIR spectroscopy and environmental factors coupled

with PLSR models to predict soil organic carbon and nitrogen. *Int. Soil Water Conserv. Res.* **12**, 844–854 (2024).

293. Chang, C.-W. & Laird, D. A. Near-infrared reflectance spectroscopic analysis of soil C and N. *Soil Sci.* **167**, 110–116 (2002).

294. Kawamura, K. *et al.* Vis-NIR Spectroscopy and PLS Regression with Waveband Selection for Estimating the Total C and N of Paddy Soils in Madagascar. *Remote Sensing* vol. 9 at https://doi.org/10.3390/rs9101081 (2017).

295. Reeves, J. B., McCarty, G. W. & Meisinger, J. J. Near Infrared Reflectance Spectroscopy for the Analysis of Agricultural Soils. *J. Near Infrared Spectrosc.* **7**, 179–193 (1999).

296. Singha, C., Swain, K. C., Sahoo, S. & Govind, A. Prediction of soil nutrients through PLSR and SVMR models by VIs-NIR reflectance spectroscopy. *Egypt. J. Remote Sens. Sp. Sci.* **26**, 901–918 (2023).

297. Janik, L. J. & Skjemstad, J. O. Characterization and analysis of soils using mid-infrared partial least-squares. 2. Correlations with some laboratory data. *Soil Res.* **33**, 637–650 (1995).

298. de Santana, F. B., Otani, S. K., de Souza, A. M. & Poppi, R. J. Comparison of PLS and SVM models for soil organic matter and particle size using vis-NIR spectral libraries. *Geoderma Reg.* **27**, e00436 (2021).

299. Masserschmidt, I. *et al.* Determination of organic matter in soils by FTIR/diffuse reflectance and multivariate calibration. *J. Chemom.* **13**, 265–273 (1999).

300. Walvoort, D. J. J. & McBratney, A. B. Diffuse reflectance spectrometry as a proximal sensing tool for precision agriculture. in *ECPA* 503–507 (2001).

301. Cozzolino, D. & Moron, A. The potential of near-infrared reflectance spectroscopy to analyse soil chemical and physical characteristics. *J. Agric. Sci.* **140**, 65–71 (2003).

302. Marmette, M.-C., Adamchuk, V., Nault, J., Tabatabai, S. & Cocciardi, R. *Comparison of the*

Performance of Two Vis-NIR Spectrometers in the Prediction of Various Soil Properties. (2018).

303. Kuhn, M. & Quinlan, R. Cubist: Rule- And Instance-Based Regression Modeling. (2024).

304. Kuhn, M., Weston, S., Keefer, C. & Coulter, N. Cubist models for regression. *R Packag. Vignette R Packag. version 0.0* **18**, 480 (2012).

305. Chen, J., He, Y., Liang, Y., Wang, W. & Duan, X. Estimation of gross calorific value of coal based on the cubist regression model. *Sci. Rep.* **14**, (2024).

306. Minasny, B., Mcbratney, A., Stockmann, U. & Hong, S.-Y. *Cubist, a Regression Rule Approach for Use in Calibration of NIR Spectra*. (2013).

307. Minasny, B., McBratney, A. B., Tranter, G. & Murphy, B. W. Using soil knowledge for the evaluation of mid-infrared diffuse reflectance spectroscopy for predicting soil physical and mechanical properties. *Eur. J. Soil Sci.* **59**, 960–971 (2008).

308. Haghi, R. K., Pérez-Fernández, E. & Robertson, A. H. J. Prediction of various soil properties for a national spatial dataset of Scottish soils based on four different chemometric approaches: A comparison of near infrared and mid-infrared spectroscopy. *Geoderma* **396**, 115071 (2021).

309. Stevens, A., Nocita, M., Tóth, G., Montanarella, L. & van Wesemael, B. Prediction of Soil Organic Carbon at the European Scale by Visible and Near InfraRed Reflectance Spectroscopy. *PLoS One* **8**, e66409 (2013).

310. Liu, S. *et al.* Estimating forest soil organic carbon content using vis-NIR spectroscopy: Implications for large-scale soil carbon spectroscopic assessment. *Geoderma* **348**, 37–44 (2019).

311. Xu, S., Zhao, Y. & Wang, Y. Optimizing machine learning models for predicting soil pH and total P in intact soil profiles with visible and near-infrared reflectance (VNIR) spectroscopy. *Comput. Electron. Agric.* **218**, 108643 (2024).

312. Breiman, L. Random Forests. *Mach. Learn.* **45**, 5–32 (2001).

313. Padarian, J., Minasny, B. & McBratney, A. B. Machine learning and soil sciences: A review aided

by machine learning tools. *Soil* **6**, 35–52 (2020).

314. de Santana, F. B., Mazivila, S. J., Gontijo, L. C., Neto, W. B. & Poppi, R. J. Rapid Discrimination Between Authentic and Adulterated Andiroba Oil Using FTIR-HATR Spectroscopy and Random Forest. *Food Anal. Methods* **11**, 1927–1935 (2018).

315. Ramírez, P. B., Calderón, F. J., Jastrow, J. D., Ping, C.-L. & Matamala, R. Applying NIR and MIR spectroscopy for C and soil property prediction in northern cold-region ecosystems. Which approach works better? *Geoderma Reg.* **32**, e00617 (2023).

316. de Santana, F. B., de Souza, A. M. & Poppi, R. J. Visible and near infrared spectroscopy coupled to random forest to quantify some soil quality parameters. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **191**, 454–462 (2018).

317. Canero, F. M., Rodriguez-Galiano, V. & Aragones, D. Machine Learning and Feature Selection for soil spectroscopy. An evaluation of Random Forest wrappers to predict soil organic matter, clay, and carbonates. *Heliyon* **10**, e30228 (2024).

318. Katuwal, S. *et al.* Predicting the dry bulk density of soils across Denmark: Comparison of single-parameter, multi-parameter, and vis–NIR based models. *Geoderma* **361**, 114080 (2020).

319. Iyer, R. Demystifying Support Vector Machines: Kernel Machines. *MLDemystified* https://mldemystified.com/posts/basics-of-ml/support-vector-machines/svm-kernel-machines/ (2024).

320. Jain, A. SVM kernels and its type. *Medium* https://medium.com/%40abhishekjainindore24/svm-kernels-and-its-type-dfc3d5f2dcd8 (2024).

321. Sui, X. *et al.* A review of non-probabilistic machine learning-based state of health estimation techniques for Lithium-ion battery. *Appl. Energy* **300**, 117346 (2021).

322. Karatzoglou, A., Smola, A. & Hornik, K. kernlab: Kernel-Based Machine Learning Lab. (2024).

323. Smola, A. J. & Schölkopf, B. A tutorial on support vector regression. *Stat. Comput.* **14**, 199–222 (2004).

324. Liu, Y., Wang, H., Zhang, H. & Liber, K. A comprehensive support vector machine-based classification model for soil quality assessment. *Soil Tillage Res.* **155**, 19–26 (2016).

325. Barman, U. & Choudhury, R. D. Soil texture classification using multi class support vector machine. *Inf. Process. Agric.* **7**, 318–332 (2020).

326. Were, K., Bui, D. T., Dick, Ø. B. & Singh, B. R. A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape. *Ecol. Indic.* **52**, 394–403 (2015).

327. Liao, K., Xu, S., Wu, J., Zhu, Q. & An, L. Using support vector machines to predict cation exchange capacity of different soil horizons in Qingdao City, China. *J. Plant Nutr. Soil Sci.* **177**, (2014).

328. Behrens, T. & Scholten, T. A comparison of data-mining techniques in predictive soil mapping. *Dev. Soil Sci.* **31**, 353–617 (2006).

329. Meliho, M. *et al.* Spatial Prediction of Soil Organic Carbon Stock in the Moroccan High Atlas Using Machine Learning. *Remote Sensing* vol. 15 at https://doi.org/10.3390/rs15102494 (2023).

330. Li, X. *et al.* Can fusion of vis-NIR and MIR spectra at three levels improve the prediction accuracy of soil nutrients? *Geoderma* **441**, 116754 (2024).

331. Shi, T. *et al.* Comparison of multivariate methods for estimating soil total nitrogen with visible/near-infrared spectroscopy. *Plant Soil* **366**, 363–375 (2013).

332. Xu, S., Zhao, Y., Wang, M. & Shi, X. Comparison of multivariate methods for estimating selected soil properties from intact soil cores of paddy fields by Vis–NIR spectroscopy. *Geoderma* **310**, 29–43 (2018).

333. Gao, Yin *et al.* Estimating Soil Organic Carbon Content with Visible–Near-Infrared (Vis-NIR)

Spectroscopy. *Appl. Spectrosc.* **68**, 712–722 (2014).

334. Li, S. *et al.* In situ measurements of organic carbon in soil profiles using vis-NIR spectroscopy on the Qinghai–Tibet plateau. *Environ. Sci. Technol.* **49**, 4980–4987 (2015).

335. Morellos, A. *et al.* Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using VIS-NIR spectroscopy. *Biosyst. Eng.* **152**, 104–116 (2016).

336. Vohland, M., Besold, J., Hill, J. & Fründ, H.-C. Comparing different multivariate calibration methods for the determination of soil organic carbon pools with visible to near infrared spectroscopy. *Geoderma* **166**, 198–205 (2011).

337. Araújo, S. R., Wetterlind, J., Demattê, J. A. M. & Stenberg, B. Improving the prediction performance of a large tropical vis-NIR spectroscopic soil library from B razil by clustering into smaller subsets or use of data mining calibration techniques. *Eur. J. Soil Sci.* **65**, 718–729 (2014).

338. Nawar, S., Buddenbaum, H., Hill, J., Kozak, J. & Mouazen, A. M. Estimating the soil clay content and organic matter by means of different calibration methods of vis-NIR diffuse reflectance spectroscopy. *Soil Tillage Res.* **155**, 510–522 (2016).

339. Iznaga, A. C. *et al.* Vis/NIR spectroscopic measurement of selected soil fertility parameters of Cuban agricultural Cambisols. *Biosyst. Eng.* **125**, 105–121 (2014).

340. Ji, W., Shi, Z., Huang, J. & Li, S. Correction: In Situ Measurement of Some Soil Properties in Paddy Soil Using Visible and Near-Infrared Spectroscopy. *PLoS One* **11**, e0159785 (2016).

341. Cohen, M., Paris, J. & Clark, M. P-sorption capacity estimation in Southeastern USA wetland soils using visible/near infrared (VNIR) reflectance spectroscopy. *Wetlands* **27**, 1098–1111 (2007).

342. Niederberger, J. *et al.* Use of near-infrared spectroscopy to assess phosphorus fractions of different plant availability in forest soils. *Biogeosciences* **12**, 3415–3428 (2015).

343. Dunne, K. S. *The Application of Spectroscopic Techniques for the Prediction of Phosphorus*

Dynamics in Agricultural Soils. (University College Dublin. School of Biosystems and Food Engineering, 2021).

344. Carstensen, A., Herdean, A., Birkelund, S., Schmidt, Anurag Sharma, Cornelia Spetea, M. P. & Husted, S. The impacts of phosphorus deficiency on the photosynthetic electron transport chain. *Plant Physiol.* **177**, 271–284 (2018).

345. Hall, R. *et al.* A machine learning approach to predicting plant available phosphorus that accounts for soil heterogeneity and regional variability. *J. Soils Sediments* **24**, 1–12 (2023).

346. Kovar, J. & Pierzynski, G. Methods of Phosphorus Analysis for Soils, Sediments, Residuals, and Waters Second Edition. *South. Coop. Ser. Bull. 408 South. Ext. Res. Act.* (2009).

347. Hu, Y. *et al.* Soil acidification suppresses phosphorus supply through enhancing organomineral association. *Sci. Total Environ.* **905**, 167105 (2023).

348. Jindo, K. *et al.* Biotic and abiotic effects of soil organic matter on the phytoavailable phosphorus in soils: a review. *Chem. Biol. Technol. Agric.* **10**, 29 (2023).

349. Møller, A. B. *et al.* Mapping the phosphorus sorption capacity of Danish soils in four depths with quantile regression forests and uncertainty propagation. *Geoderma* **430**, 116316 (2023).

350. Kumaragamage, D. *et al.* Degree of Phosphorus Saturation as a Predictor of Redox-Induced Phosphorus Release from Flooded Soils to Floodwater. *J. Environ. Qual.* **48**, 1817–1825 (2019).

351. van der Zee, S. E. A. T. M., Nederlof, M. M., Riemsdijk, W. H. van & Haan, F. A. M. de. Spatial Variability of Phosphate Adsorption Parameters. *J. Environ. Qual.* **17**, 682–688 (1988).

352. Lynch, J., Donnellan, T., Finn, J. A., Dillon, E. & Ryan, M. Potential development of Irish agricultural sustainability indicators for current and future policy evaluation needs. *J. Environ. Manage.* **230**, 434–445 (2019).

353. Wall, D. P. & Plunkett, M. *Major and Micro Nutrient Advice for Productive Agricultural Crops (5th Edition)*. (Teagasc Johnstown Castle, Wexford, 2020).

354. GSI. *Tellus Geochemical Survey: Deeper Topsoil Data from Outer Dublin, the Border, North, and West of Ireland*. https://gsi.geodata.gov.ie/downloads/Geochemistry/Reports/GSI_Tellus_Geochemistry_G1_G3_G6_S_QC_Report.pdf (2020).

355. Browne, M. A., Szpak, M. & V., G. *Tellus Geochemical Survey: Deeper Topsoil Multi-Element Maps for the Northern Half of Ireland*. https://gsi.geodata.gov.ie/downloads/Geochemistry/Reports/Tellus_S_Geochemistry_Soils_ICPar_50percent_Map_Booklet.pdf (2021).

356. GSI. Data and maps. https://www.gsi.ie/en-ie/data-and-maps/Pages/default.aspx (2025).

357. GSI. Terra soil. https://www.gsi.ie/en-ie/programmes-and-projects/tellus/projects/terra-soil/Pages/default.aspx#:~:text=The Terra Soil project is,identify farming and environmental signals (2025).

358. de Santana, F. B. *et al.* A systematic approach to predicting and mapping soil particle size distribution from unknown samples using large mid-infrared spectral libraries covering large-scale heterogeneous areas. *Geoderma* **434**, 116491 (2023).

359. Ture, M. D. *Tellus Merge-2019B Airborne Geophysical Technical Report (Operations, Logistics and Data Merging).* (2020).

360. Nair, P. S. *et al.* Interlaboratory Comparison of a Standardized Phosphorus Adsorption Procedure. *J. Environ. Qual.* **13**, 591–595 (1984).

361. Pautler, M. C. & Sims, J. T. Relationships Between Soil Test Phosphorus, Soluble Phosphorus, and Phosphorus Saturation in Delaware Soils. *Soil Sci. Soc. Am. J.* **64**, 765–773 (2000).

362. Dytham, C. *Choosing and Using Statistics: A Biologist's Guide*. (John Wiley & Sons, Ltd., Publication, 2017).

363. Muggeo, V. M. R. segmented: An R Package to FitRegression Models with Broken-

LineRelationships. *R News* **8**, 20–25 (2008).

364. Muggeo, V. M. R. Testing with a nuisance parameter present only under the alternative: a score-based approach with application to segmented modelling. *J. Stat. Comput. Simul.* **86**, 3059–3067 (2016).

365. Xia, Y. Chapter Eleven - Correlation and association analyses in microbiome study integrating multiomics in health and disease. in *The Microbiome in Health and Disease* (ed. Sun, J. B. T.-P. in M. B. and T. S.) vol. 171 309–491 (Academic Press, 2020).

366. Wang, L. & Liang, T. Effects of exogenous rare earth elements on phosphorus adsorption and desorption in different types of soils. *Chemosphere* **103**, 148–155 (2014).

367. Goyne, eith W., Jun, H.-J., Anderson, S. H. & Motavalli, P. P. Phosphorus and Nitrogen Sorption to Soils in the Presence of Poultry Litter-Derived Dissolved Organic Matter. *J. Environ. Qual.* **37**, (2018).

368. Zhang, T., Wang, Y., Tan, C. S. & Welacky, T. An 11-Year Agronomic, Economic, and Phosphorus Loss Potential Evaluation of Legacy Phosphorus Utilization in a Clay Loam Soil of the Lake Erie Basin. *Front. Earth Sci.* **8**, (2020).

369. Gatiboni, L., Brunetto, G., Pavinato, P. S. & George, T. S. Editorial: Legacy Phosphorus in Agriculture: Role of Past Management and Perspectives for the Future. *Front. Earth Sci.* **8**, 619935 (2020).

370. McDowell, R., Dodd, R., Pletnyakov, P. & Noble, A. The Ability to Reduce Soil Legacy Phosphorus at a Country Scale. *Front. Environ. Sci.* **8**, (2020).

371. Johnston, Fixen & Poulton. The Efficient Use of Phosphorus in Agriculture. *Better Crops* 22–24 (2014).

372. McCarty, G. W. & Reeves, J. B. Comparison of Near Infrared and Mid Infrared Diffuse Reflectance Spectroscopy for Field-Scale Measurement of Soil Fertility Parameters. *Soil Sci.* **171**,

(2006).

373. Bellon-Maurel, V. & McBratney, A. Near-infrared (NIR) and mid-infrared (MIR) spectroscopic techniques for assessing the amount of carbon stock in soils – Critical review and research perspectives. *Soil Biol. Biochem.* **43**, 1398–1410 (2011).

374. Kruse, J. *et al.* Innovative methods in soil phosphorus research: A review. *J. Plant Nutr. Soil Sci.* **178**, 43–88 (2015).

375. Gruszczyński, W. & Gruszczyński, S. Assessing the Information Potential of MIR Spectral Signatures for Prediction of Multiple Soil Properties Based on Data from the AfSIS Phase I Project. *Int J Env. Res Public Heal.* **19**, 15210 (2022).

376. Ahmed, M. U. *et al.* A Machine Learning Approach for Biomass Characterization. *Energy Procedia* **158**, 1279–1287 (2019).

377. Engel, J. *et al.* Breaking with trends in pre-processing? *TrAC Trends Anal. Chem.* **50**, 96–106 (2013).

378. Nieuwoudt, H. H., Prior, B. A., Pretorius, I. S., Manley, M. & Bauer, F. F. Principal Component Analysis Applied to Fourier Transform Infrared Spectroscopy for the Design of Calibration Sets for Glycerol Prediction Models in Wine and for the Detection and Classification of Outlier Samples. *J. Agric. Food Chem.* **52**, 3726–3735 (2004).

379. Song, K., Qin, Y., Xu, B., Zhang, N. & Yang, J. Study on outlier detection method of the near infrared spectroscopy analysis by probability metric. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **280**, 121473 (2022).

380. Rodionova, O. Y. & Pomerantsev, A. L. Detection of Outliers in Projection-Based Modeling. *Anal. Chem.* **92**, 2656–2664 (2020).

381. Mehmood, T., Liland, K. H., Snipen, L. & Sæbø, S. A review of variable selection methods in Partial Least Squares Regression. *Chemom. Intell. Lab. Syst.* **118**, 62–69 (2012).

382. Saptoro, A., Tadé, M. O. & Vuthaluru, H. A Modified Kennard-Stone Algorithm for Optimal Division of Data for Developing Artificial Neural Network Models. *Chem. Prod. Process Model.* **7**, (2012).

383. Ferreira, R. de A., Teixeira, G. & Peternelli, L. A. Kennard-Stone method outperforms the Random Sampling in the selection of calibration samples in SNPs and NIR data. *Ciência Rural* **52**, (2022).

384. Nduwamungu, C., Ziadi, N., Parent, L.-E., Tremblay, G. F. & Thurie`s, L. Opportunities for, and limitations of, near infrared reflectance spectroscopy applications in soil analysis: A review. *Can. J. SOIL Sci.* **89**, 531–541 (2009).

385. Ludwig, B. *et al.* Estimation accuracies of near infrared spectroscopy for general soil properties and enzyme activities for two forest sites along three transects. *Geoderma* **288**, 37–46 (2017).

386. Frizzarin, M., O'Callaghan, T. F., Murphy, T. B., Hennessy, D. & Casa, A. Application of machine-learning methods to milk mid-infrared spectra for discrimination of cow milk from pasture or total mixed ration diets. *J. Dairy Sci.* **104**, 12394–12402 (2021).

387. Wu, C., Zheng, Y., Yang, H., Yang, Y. & Wu, Z. Effects of different particle sizes on the spectral prediction of soil organic matter. *CATENA* **196**, 104933 (2021).

388. McHugh, M. L. Interrater reliability: The kappa statistic. *Biochem. Medica* **22**, 276–282 (2012).

389. Subramanian, A. & Rodriguez-Saona, L. Chapter 7 - Fourier Transform Infrared (FTIR) Spectroscopy. in (ed. Sun, D.-W. B. T.-I. S. for F. Q. A. and C.) 145–178 (Academic Press, San Diego, 2009). doi:https://doi.org/10.1016/B978-0-12-374136-3.00007-9.

390. Saidi, S., Ayoubi, S., Shirvani, M., Azizi, K. & Zhao, S. Digital mapping of soil phosphorous sorption parameters (PSPs) using environmental variables and machine learning algorithms. *Int. J. Digit. Earth* **16**, 1752–1769 (2023).

391. Gardiner, M. J. & Radford, T. Soil associations of Ireland and their land use potential:

explanatory bulletin to soil map of Ireland 1980. *An Foras Taluntais* (1980).

392. European Commission (EU). *COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS Pathway to a Healthy Planet for All EU Action Plan: 'Towards Zero Pollution for Air, Water and Soil'. EU*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021DC0400 (2021).

393. United Nations. The UN Sustainable Development Goals: THE 17 GOALS. *United Nations* https://sdgs.un.org/goals (2023).

394. Velten, E. K. *et al. Flagship Report: State of EU Progress to Climate Neutrality. An Indicator-Based Assessment across 13 Building Blocks for a Climate Neutral Future.* https://www.ecologic.eu/19346 (2023).

395. Trodd, W., O'Boyle, S. & Gurrie, M. *Water Quality in Ireland 2016 – 2021*. (ENVIRONMENTAL PROTECTION AGENCY, Wexford,Ireland, 2022).

396. Wang, Y., Li, M., Ji, R., Wang, M. & Zheng, L. Comparison of Soil Total Nitrogen Content Prediction Models Based on Vis-NIR Spectroscopy. *Sensors* vol. 20 at https://doi.org/10.3390/s20247078 (2020).

397. Kühnel, A. & Bogner, C. In-situ prediction of soil organic carbon by vis–NIR spectroscopy: an efficient use of limited field data. *Eur. J. Soil Sci.* **68**, 689–702 (2017).

398. Debaene, G., Bartmiński, P. & Siłuch, M. In Situ VIS-NIR Spectroscopy for a Basic and Rapid Soil Investigation. *Sensors* vol. 23 at https://doi.org/10.3390/s23125495 (2023).

399. Maleki, M. R. *et al.* Phosphorus Sensing for Fresh Soils using Visible and Near Infrared Spectroscopy. *Biosyst. Eng.* **95**, 425–436 (2006).

400. Metzger, K., Zhang, C. & Daly, K. From benchtop to handheld MIR for soil analysis: Predicting lime requirement and organic matter in agricultural soils. *Biosyst. Eng.* **204**, 257–269 (2021).

401. Le Pevelen, D. D. & Tranter, G. E. FT-IR and Raman Spectroscopies, Polymorphism Applications. in (eds. Lindon, J. C., Tranter, G. E. & Koppenaal, D. W. B. T.-E. of S. and S. (Third E.) 750–761 (Academic Press, Oxford, 2017). doi:https://doi.org/10.1016/B978-0-12-409547-2.12161-4.

402. Cevoli, C., Iaccheri, E., Fabbri, A. & Ragni, L. Data fusion of FT-NIR spectroscopy and Vis/NIR hyperspectral imaging to predict quality parameters of yellow flesh “Jintao” kiwifruit. *Biosyst. Eng.* **237**, 157–169 (2024).

403. ThomasTKtungnung. Soil Water Holding Capacity Determination Practical Experiment. *Youtube* https://www.youtube.com/watch?v=ojCRX6ILzRY (2021).

404. Hale, G. M. & Querry, M. R. Optical Constants of Water in the 200-nm to 200-µm Wavelength Region. *Appl. Opt.* **12**, 555–563 (1973).

405. Post, J. L. & Noble, P. N. The Near-Infrared Combination Band Frequencies of Dioctahedral Smectites, Micas, and Illites. *Clays Clay Miner.* **41**, 639–644 (1993).

406. Guillou, F. Le *et al.* How does grinding affect the mid-infrared spectra of soil and their multivariate calibrations to texture and organic carbon? *Soil Res.* **53**, 913–921 (2015).

407. Li, M., HAN, D. & Wang, X. 光谱分析技术及其应用. (科学出版社, Beijing, 2006).

408. Liu, W. *et al.* Relating soil surface moisture to reflectance. *Remote Sens. Environ.* **81**, 238–246 (2002).

409. Hutengs, C. *et al.* Enhanced VNIR and MIR proximal sensing of soil organic matter and PLFA-derived soil microbial properties through machine learning ensembles and external parameter orthogonalization. *Geoderma* **450**, 117037 (2024).

410. Tang, R., Jiang, K., Li, C., Li, X. & Wu, J. Modeling to Correct the Effect of Soil Moisture for Predicting Soil Total Nitrogen by Near-Infrared Spectroscopy. *Electronics* vol. 12 at https://doi.org/10.3390/electronics12061271 (2023).

411. Roudier, P., Hedley, C. B., Lobsey, C. R., Viscarra Rossel, R. A. & Leroux, C. Evaluation of two

methods to eliminate the effect of water from soil vis–NIR spectra for predictions of organic carbon. *Geoderma* **296**, 98–107 (2017).

412. Yang, M. *et al.* Removing the moisture effect on predicting soil organic matter using vis-NIR spectroscopy with external parameter orthogonalization. *Geoderma Reg.* **37**, e00797 (2024).

413. Yang, M. *et al.* Strategies for predicting soil organic matter in the field using the Chinese Vis-NIR soil spectral library. *Geoderma* **433**, 116461 (2023).

414. Gorthi, S., Chakraborty, S., Li, B. & Weindorf, D. C. A field-portable acoustic sensing device to measure soil moisture. *Comput. Electron. Agric.* **174**, 105517 (2020).

415. Ruszczak, B. & Boguszewska-Mańkowska, D. Soil Moisture a Posteriori Measurements Enhancement Using Ensemble Learning. *Sensors* vol. 22 at https://doi.org/10.3390/s22124591 (2022).

416. Liu, Y., Cai, W. & Shao, X. Standardization of near infrared spectra measured on multi-instrument. *Anal. Chim. Acta* **836**, 18–23 (2014).

417. Pu, Y.-Y. *et al.* Calibration Transfer from Micro NIR Spectrometer to Hyperspectral Imaging: a Case Study on Predicting Soluble Solids Content of Bananito Fruit (Musa acuminata). *Food Anal. Methods* **11**, (2018).

418. Pierna, J. A. F., Vermeulen, P., Lecler, B., Baeten, V. & Dardenne, P. Calibration Transfer from Dispersive Instruments to Handheld Spectrometers. *Appl. Spectrosc.* **64**, 644–648 (2010).

419. Parrott, A. *et al.* Calibration model transfer in mid-infrared process analysis with in situ attenuated total reflectance immersion probes. *Anal. Methods* **14**, (2022).