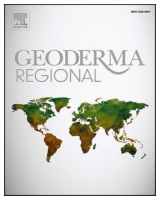
Geoderma Regional 44 (2026) e01048

Contents lists available at ScienceDirect

Geoderma Regional

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



Developing a pedotransfer function for the prediction of nitrogen mineralization in the agricultural soils of Quebec

Chedzer-Clark Clement [a], R. Deragon [b], B. Heung [a], J. Dessureault-Rompré [b], M.O. Gasser [c], J.-B. Mathieu [c], D.L. Burton [a,*]

[a] Center for Sustainable Soil Management, Dalhousie University, Truro, NS B2N 5E3, Canada
[b] Soil and Agri-Food Engineering Department, Université Laval, Quebec City, QC G1V 0A6, Canada
[c] Research and Development Institute for the Agri-environment, Quebec City, QC G1P 3W8, Canada

ARTICLE INFO

Keywords:
Nitrogen mineralization
Potentially mineralizable nitrogen
Pedotransfer function
Random Forest
Machine learning

ABSTRACT

Effective nitrogen (N) management is crucial for maximizing crop yields while minimizing environmental impacts. In the agricultural systems of Quebec, soil organic matter mineralization supplies a significant portion of crop N demand, but direct quantification is challenging and costly. This study utilized zero-N trial data to evaluate a two-pool zero-plus first-order kinetic model for predicting growing season N mineralization (GSNM) based on total N (TN) and potentially mineralizable N (PMN). Additionally, machine learning-based pedotransfer functions (PTFs) were developed to predict TN, PMN, and GSNM from easily measurable soil properties using a large soil health dataset (n = 3117). The kinetic model showed strong agreement between predicted and observed soil N supply, especially with the inclusion of deeper soil layers and early-season mineralization estimates. Recursive feature elimination identified total carbon (TC) and clay as the best predictors for TN, yielding a Lin's concordance correlation coefficient (CCC) of 0.93 and a coefficient of determination (R^2) of 0.86, while soil respiration (SR) and pH best predicted PMN (CCC = 0.89, R^2 = 0.80). For GSNM, SR, TC, and pH were the top predictors (CCC = 0.91, R^2 = 0.83). The developed PTFs provide a practical framework for estimating soil N-pools where direct data is unavailable, ultimately improving site-specific N management decisions. These tools support more efficient fertilizer use and minimize environmental losses. Future research should focus on integrating soil management practices into the development of PTFs and considering spatial and landscape variability through digital soil mapping.

1. Introduction

Nitrogen (N) is an essential nutrient for crop growth and yield; however, managing it effectively remains a significant challenge. Optimal N application can enhance plant productivity, while deficiencies can greatly diminish both yield and crop quality (Zebarth and Rosen, 2007). Conversely, excessive N inputs may increase the vulnerability of the crops to disease and pest infestations (Huber et al., 2012) and lead to environmental degradation through NO_3^- leaching and nitrous oxide emissions (Burton et al., 2008; Clément et al., 2021; Zebarth and Rosen, 2007). Therefore, sustainable N management requires accurate quantification of the soil's capacity to supply N through organic matter (OM) mineralization, allowing for appropriate adjustments in fertilizer applications.

Effectively predicting the amount of N that will be mineralized during the growing season remains a challenging task. Building on the foundational work of Stanford and Smith (1972), Dessureault-Rompré et al. (2015) developed a measurement-based predictive function to estimate growing season N mineralization (GSNM). This model was first evaluated in potato soils in New Brunswick (Dessureault-Rompré et al., 2015) and later tested using soils from various regions across Canada (Dessureault-Rompré et al., 2016). In this model, the stable (i.e., slowly mineralizing) N pool, derived from more recalcitrant soil OM (Dessureault-Rompré et al., 2016), is assumed to mineralize at a constant rate following a zero-order kinetic relationship and can be parameterized using total soil N (TN) analysis (Dessureault-Rompré et al., 2013). In contrast, the labile N pool, which represents the more reactive OM fractions, is described by a first-order kinetic relationship and is typically quantified through short-term incubations, such as the two-week aerobic incubation method known as the N flush (Pool I)

* Corresponding author.
E-mail address: dburton@dal.ca (D.L. Burton).

https://doi.org/10.1016/j.geodrs.2025.e01048
Received 7 July 2025; Received in revised form 17 December 2025; Accepted 23 December 2025
Available online 25 December 2025

(Curtin and Campbell, 2007; Sharifi et al., 2007; Stanford and Smith, 1972). While the model shows promising applicability across different crops and soil types, its parameters, especially the mineralization rate constants, may need to be adjusted to account for regional variations in soil properties, climate, and management practices.

A key limitation to the widespread adoption of these measurement-based models lies in the limited availability of direct measures of TN and labile N fractions. Although TN measurement is relatively simple, it is not typically included in standard soil analytical packages. Moreover, measuring potentially mineralizable N (PMN) through incubation is both time-consuming and costly, which limits its practical application. Developing inexpensive and rapid proxies for labile N would facilitate broader integration of N mineralization estimates into management decisions. One promising strategy involves using predictive modeling that relies on routinely measured soil properties to estimate labile N pools (Laurence et al., 2023; Laurance et al., 2024; Laurence et al., 2025).

Pedotransfer functions (PTFs) provide an efficient way to estimate complex or costly soil parameters by using empirical relationships with easily measured variables (McBratney et al., 2002). PTFs have been widely applied to predict soil organic carbon (Benke et al., 2020; Xiao et al., 2022), bulk density (Arbor et al., 2023, 2024; Ramcharan et al., 2017), hydraulic conductivity (Mozaffari et al., 2022), and available water-holding capacity (Dobarco et al., 2019). While early PTFs primarily used linear regression, recent advancements increasingly incorporate machine learning algorithms, such as random forest (RF), Cubist, support vector machines, and artificial neural networks, that can more effectively capture the complex, hierarchical, and nonlinear relationships among soil properties (Amanabadi et al., 2019; Arbor et al., 2024; Heung et al., 2016; Van Looy et al., 2017; de Castro Moreira da Silva et al., 2023). Machine learning-based PTFs have shown superior accuracy and robustness compared to traditional models, performing consistently well across diverse soils and management systems (Benke et al., 2020; Xiao et al., 2022). Several studies have explored PTFs for estimating soil N pools, including total, stable, and labile fractions (Glendining et al., 2011; Heumann et al., 2003, 2011). However, the transferability of these models is limited because their performance is heavily influenced by regional pedo-climatic conditions and management practices (McBratney et al., 2002; Van Looy et al., 2017; Arbor et al., 2023). Consequently, there is a recognized need for regional calibration to ensure an accurate representation of soil N dynamics.

The pedo-climatic context of Quebec provides a particularly relevant opportunity for developing region-specific models of N mineralization. The province features a diverse range of soil types, from the clay-rich soils of the St. Lawrence Lowlands to the sandy loams of the upland regions, within a humid continental climate characterized by a short growing season (Gasser et al., 2023). These conditions favor substantial N mineralization, with rates reported as high as 150 kg N ha^{-1} (Clément et al., 2021). Despite this potential, mineralized N is rarely accounted for in fertilizer recommendations, primarily because of uncertainties regarding its temporal dynamics and regional variability. As a result, fertilizer decisions often rely on generalized guidelines that may overlook the actual N-supplying capacity of local soils. This gap highlights the need for predictive tools that can more accurately quantify N mineralization under the specific soil and climate conditions of Quebec.

This study evaluated the performance of a two-pool zero-plus first-order kinetic model for estimating GSNM in agricultural soils of Quebec. It also aimed to develop machine learning–based PTFs that can predict stable and labile N pools from easily measurable soil properties, thereby allowing for the estimation of GSNM using accessible proxies for TN and PMN. By combining established concepts of N mineralization kinetics with modern predictive modeling techniques, this study intends to provide accurate, cost-effective, and transferable tools for improving N management. Although the methodology was developed for Quebec, its framework and application can be extended to other temperate regions facing similar challenges in optimizing crop productivity while ensuring environmental sustainability.

2. Materials and methods

This study was conducted in two phases. First, the suitability of the zero-plus first-order kinetic model described in Eq. (1) was evaluated under the soil and climate conditions of Quebec. This evaluation used zero-N trial data from published studies on corn (Cambouris et al., 2016b; Gagnon and Ziadi, 2010; Nyiraneza et al., 2010), spring wheat (Nyiraneza et al., 2012), and potatoes (Cambouris et al., 2016a; Clément et al., 2019). In the second phase, data from the *Étude sur l'État de Santé des Sols Agricoles du Québec* (EESSAQ), conducted between 2017 and 2022 (Gasser et al., 2023), were used to develop PTFs for TN, PMN, and GSNM derived from the zero-plus first-order kinetic equation.

2.1. Study area

The study was conducted in the primary agricultural regions of Quebec. In 2021, cultivated land covered approximately 1.75 million hectares, along with an additional 500,000 ha designated as pasture and newly cleared land. This total represents less than 2 % of province's overall area (Statistics Canada, 2022). The agricultural zones of Quebec are shaped by distinct climatic gradients. The southwestern region, which include Montérégie, the Montreal Plain, and the St. Lawrence Valley, experiences warmer temperatures and a longer growing season. In contrast, the northern and eastern areas, such as Abitibi, Saguenay–Lac-Saint-Jean, and Gaspésie, experience cooler conditions and shorter seasons (Gasser et al., 2023). Precipitation levels increase from west to east, with the highest rainfall recorded in eastern Quebec. The relationship between precipitation and potential evapotranspiration is crucial for determining soil moisture, erosion risk, and crop productivity. These factors, in turn, drive regional variations in cropping systems whereby perennial forages are common in the cooler, wetter areas, while annual crops are more dominant in the warmer regions (Gasser et al., 2023).

The soils in the province reflect a complex history of glacial and alluvial processes, leading to significant variations in texture, drainage, and fertility. The main parent materials include glacial tills, marine and lacustrine clays, alluvial sediments, fluvial sands, and organic deposits (Angers et al., 2022). In the St. Lawrence lowlands, fertile clay soils are predominant; however, they are susceptible to compaction and drainage issues. In contrast, the loamy and sandy soils found in the Appalachian and Laurentian uplands tend to be more acidic and less fertile. Organic soils, typically situated in depressions and reclaimed wetlands, are well-suited for growing forage or specialty crops (Angers et al., 2022).

Cropping systems are closely aligned with the climatic and edaphic conditions. In the southern plains, annual row crops—primarily maize, soybeans, and cereals—are predominant. These crops are often managed under intensive tillage practices, which increase the risk of soil erosion. In contrast, in the cooler and wetter regions, such as Saguenay–Lac-Saint-Jean, Abitibi, and the Appalachians, predominantly feature perennial forages like hay and pasture; and when combined with poor drainage, higher soil OM levels are expected. Specialized cropping systems are also found in the region, including horticultural production near Montréal, potato cultivation on sandy soils, and mixed rotations that integrate annual and perennial crops. (Gasser et al., 2023).

2.2. Estimation of growing season nitrogen mineralization

Growing season N mineralization was estimated using the zero-plus first-order kinetic model (Eq. 1) (Dessureault-Rompré et al., 2015). This function utilizes a two-pool zero-plus first-order equation of the form:

$$N_{min} = K_S t + N_L \left[1 - e^{-k_L t}\right] \quad (1)$$

where K_S is the rate constant of a non-depleting zero-order pool ("stable

N pool") where N is mineralized at a constant rate, and N_L is a depleting first-order pool ("labile N pool") with a mineralization rate constant k_L. The stable, non-depleting zero-order pool is quantified as a function of TN, while the N mineralization measured through incubation methods quantifies the labile pool.

Estimation of growing season N mineralization was made for a 130-day growing season and expressed as cumulative N mineralization (N_{min}) over 130 days based on climate data for the agricultural zones of Quebec. The non-depleting (zero-order) stable fraction ($K_S t$), where N mineralization occurs continuously at a constant rate (K_S), was calculated by estimating K_S (d^{-1}) from the relationship in Eq. 2 using total N (%w w^{-1}) and the amount of N mineralized during 14-day aerobic incubation in mg N kg^{-1} soil as *Pool I* (Dessureault-Rompré et al., 2015).

$$K_S = 0.123 \times Total\,N + 0.00312 \times Pool\,I + 0.0685 \quad (2)$$

The depleting (first-order) labile N-pool, as $N_L(1 - e^{-k_L t})$ was calculated using PMN from a 14-day aerobic incubation as *Pool I* (Eq. 3), and a single fixed k_L value of 0.074 d^{-1} obtained from Dessureault-Rompré et al. (2015) as it was demonstrated that variations of k_L due to soil types are minor and have little effect on predicted values of GSNM when used in a prediction function. The value of N_L was estimated as suggested by Dessureault-Rompré et al. (2015):

$$N_L = 1.16 \times Pool\,I + 3.44 \quad (3)$$

Corrections were made to K_S and k_L to reflect the environmental conditions in which mineralization occurs in Quebec. The values of K_S and k_L were modified daily for soil temperature using a Logistic temperature function (Dessureault-Rompré et al., 2010):

$$\frac{k}{k_0} = \frac{M}{\left\{1 + C \times e^{-r\left[\frac{T-273.16}{T_0-273.16}\right]}\right\}} \quad (4)$$

where *M*, *C,* and *r* are fitting parameters previously estimated by Dessureault-Rompré et al. (2010); *T* is the daily soil temperature in degrees Celsius; and T_0 is the temperature at which the soil was incubated in degrees Celsius. Daily soil temperature was predicted from air temperature normals (1990–2020) using the model described by Kätterer and Andrén (2009). K_S and k_L were also corrected for soil water content using the biophysical water function described by Dessureault-Rompré et al., 2011

$$f(x) = \lambda\frac{(1 - e^{-bx})}{(1 - e^{-b})} + 2(1 - \lambda)\frac{\left[e^{g} - e^{-g(x-1)}\right]}{(e^{g} - 1)[1 + e^{-g(x-1)}]} \quad (5)$$

where λ, *b*, and *g* are fitting parameters (Dessureault-Rompré et al., 2011); and *x* is the proportion of the available water range defined as:

$$x = \frac{(W - W_0)}{(W_{max} - W_0)} \quad (6)$$

where *W* is gravimetric water content; W_0 is the water content at which N mineralization should be negligible, and W_{max} is the water content at the maximum rate of N mineralization. Here, water content was expressed as water-filled pore space (WFPS) to more accurately characterize the effects of water content on surface wetting and pore filling, given that WFPS reflects the quantity of water in relation to soil porosity (Dessureault-Rompré et al., 2011). The scaled proportion of the available water range ($WFPS_s$) was thus defined as:

$$WFPS_s = \frac{(WFPS - WFPS_0)}{(WFPS_{max} - WFPS_0)} \quad (7)$$

where *WFPS* is the actual water content expressed in units of %*WFPS*; $WPFS_0$ is set to a water content of 5 % *WFPS*; and $WFPS_{max}$ is the water content at the maximum rate of N mineralization and was set to the soil water content at field capacity.

The WFPS was calculated from the gravimetric water content as (Hanks and Ashcroft, 1980):

$$WFPS = 100 \times \frac{SWC}{1 - \frac{BD}{PD}} \quad (8)$$

where *SWC* is the gravimetric water content (g g^{-1}), *BD* is the soil bulk density (Mg m^{-3}), and *PD* is the particle density (2.65 Mg m^{-3}) (Linn and Doran, 1984). Daily *SWC* was calculated using the soil water balance, based on a modified daily Thornthwaite equation, which requires fewer climate variables (Pereira and Pruitt, 2004).

2.3. Evaluation of the zero-plus first-order kinetic model

2.3.1. Validation of the model under Quebec conditions

The performance of the two-pool zero-plus first-order equation in estimating GSNM under the agricultural conditions of Quebec was evaluated by comparing GSNM to field-based estimates of *SNS* in zero-N trials (Eq. 10). The prediction function used to estimate GSNM was of the form:

$$GSNM = M_s\left(K_S t + N_L\left[1 - e^{-k_L t}\right]\right) \quad (9)$$

$$SNS_{field-based} = Total\,N\,uptake + RSN - PPSN \quad (10)$$

where RSN is the residual soil inorganic N at the end of the growing season; *PPSN* is pre-plant soil inorganic N (kg ha^{-1}), representing early-season N mineralization and RSN from the previous growing season that was not lost over winter (carried-over N); and *Ms* is soil mass (kg).

2.3.2. Zero-N trial data

Data from published zero-N fertilizer field trials were used to evaluate the kinetic model. These data were obtained from previously published studies conducted over multiple years and crops. Potato field trials were carried out between 2008 and 2015 (Cambouris et al., 2016a; Clément et al., 2019), whereas corn (Cambouris et al., 2016b; Gagnon and Ziadi, 2010; Nyiraneza et al., 2010) and spring wheat (Nyiraneza et al., 2012) trials were conducted between 2004 and 2010. Plant N uptake was determined by measuring N accumulation in various plant components at physiological maturity.

Whole plants were harvested from representative areas within each plot by cutting aboveground biomass at the soil surface. For potatoes, vines were collected just before desiccation, and tubers were mechanically harvested at maturity. For corn and wheat, harvested plants were mechanically shredded, and subsamples of approximately 500 g were oven-dried at 55 °C until they reach a constant mass. Grain subsamples were obtained after wet shelling for corn or threshing for wheat, then dried and ground for analysis. All plant materials were ground to pass through a 1–2 mm sieve prior to chemical analysis. For potato, N concentration was determined using an Elementar CN analyzer (Varian Macro CN, Elementar, Hanau, Germany), whereas for corn and wheat, ground samples were digested with a mixture of sulfuric and selenious acids, and N concentrations were measured using a Lachat QuikChem 8000 autoanalyzer (Zellweger Analytics Inc., Lachat Instruments Division, Milwaukee, WI, USA). Total N uptake for each crop was calculated by summing the N accumulation for the individual components, obtained by multiplying the dry matter yield of each component by its corresponding tissue N concentrations.

Preplant soil N and RSN were quantified from composite soil samples collected before planting and in late October to early November, respectively. Four soil cores were taken from each experimental unit at three depth intervals: 0–0.30, 0.30–0.60, and 0.60–0.90 m. Cores from each depth interval were combined to form a composite sample per plot and depth. All samples were kept frozen until analysis. Soil mineral N was extracted using 1 M KCl at a soil-to-extractant ratio of 1:10 (Maynard and Kalra, 1993). Soil NH_4-N concentration was determined

colorimetrically by reaction with nitroprusside (Nkonge and Balance, 1982), while soil NO_3-N concentration was determined at 210 nm by liquid chromatography on a Dionex 4000i equipped with a VDM-II UV-V detector (Dionex Corporation, Sunnyvale, CA, USA). Nitrogen concentrations from the three depth intervals were combined and expressed on a 0–0.90 m depth basis. Values were converted to kg N ha^{-1} by multiplying concentrations by the corresponding soil bulk density.

The zero-N trial data did not include any measurements of PMN. To address this, a PTF was trained using the EESSAQ dataset and then applied to the zero-N trial data to estimate PMN, defined as mineralized N over a 14-day aerobic incubation period. The input parameters included TN, TC, soil pH, BD, sand, and clay contents (Table 1). These parameters were selected because they were the only relevant variables common to both the EESSAQ dataset and the zero-N trial data.

For the corn dataset, PPSN levels ranged from 0.4 to 84.0 kg N ha^{-1} in the 0–30 cm soil layer and from 0.6 to 151.0 kg N ha^{-1}, in the 0–90 cm soil layer. Over the growing season, corn plants absorbed an average of 71.0 kg N ha^{-1}, with values ranging from 10.0 to 179.0 kg N ha^{-1}. RSN varied between 0.5 and 64.0 kg N ha^{-1} in the 0–30 cm layer and between 1.0 and 100.0 kg N ha^{-1} in the 0–90 cm layer. GSNM ranged from 33.0 to 196.0 kg N ha^{-1}, with a mean of 84.0 kg N ha^{-1} (Table 1).

For spring wheat, PPSN levels ranged from 1.0 to 99.8 kg N ha^{-1} in the 0–30 cm soil layer and from 1.6 to 123.7 kg N ha^{-1} in the 0–90 cm soil layer. Wheat plants absorbed between 15.5 and 81.2 kg N ha^{-1} during the growing season, with RSN ranging from 0.13 to 330 kg N ha^{-1} in the 0–30 cm layer and from 0.23 to 354.7 kg N ha^{-1} in the 0–90 cm layer. GSNM varied from 44.8 to 97.9 kg N ha^{-1}, with an average of 72.4 kg N ha^{-1} (Table 1).

Pre-plant soil N levels for potatoes ranged from 4.6 to 84.2 kg N ha^{-1} in the 0–30 cm soil layer and from 6.5 to 302.3 kg N ha^{-1} in the 0–90 cm soil layer. Over the growing season, potato plants absorbed between 15.8 and 94.4 kg N ha^{-1}, while GSNM varied from 58.6 to 128.5 kg N ha^{-1}, with an average of 99.7 kg N ha^{-1}. Residual soil N ranged from 3.5 to 80.6 kg N ha^{-1} in the 0–30 cm layer and from 6.5 to 197.9 kg N ha^{-1} in the 0–90 cm layer (Table 1).

2.4. EESSAQ dataset

Data from the EESSAQ dataset were used to develop PTFs for TN, PMN, and GSNM, with GSNM values derived according to the procedures described in Section 2.1. The EESSAQ data consisteted of soil samples from across 431 agricultural fields in Quebec, capturing a wide range of geographical locations, soil types, cropping systems, and management practices throughout the province. The sampling was strategically designed to represent the full variety of pedoclimatic conditions in Quebec, with sites primarily located in the major southern plains, specifically the Montreal Plain and the St. Lawrence Valley. In addition to these regions, the study also included samples acquired from more peripheral and diverse landscapes, such as the Laurentian Plateau, the Abitibi lowlands, the Eastern and Western Appalachian foothills, the Monts Notre-Dame, and the Lake Saint-Jean plain, to reflect the diversity of parent materials and climatic conditions that characterize Quebec's agricultural landscapes. For this study, organic soils were excluded from the dataset while the mineral soils were generally classified at the great-group level as gleysols, humic gleysols, humo-ferric podzols, dystric brunisols, sombric brunisols, or melanic brunisols in accordance with the Canadian System of Soil Classification (3rd Ed.; Soil Classification Working Group, 1998).

At each sampling location, a soil pit approximately 60 to100 cm wide and 40 to 60 cm deep was excavated and georeferenced using a differential GPS with a positional accuracy of ±60 cm. Four soil samples were collected per site, with each sample taken from one of four predetermined points selected by the soil survey teams to represent the soil variability at each location. The main pedological horizons, namely Ap (or Ap1 and Ap2) and B, were identified in the field, based on the Canadian System of Soil Classification. The Ap1 horizon was sampled from

Table 1
Summary statistics of parameters from the Zero-N Trial Dataset: sample size (n), minimum value (Min), first quartile (Q_1), mean, standard deviation (SD), third quartile (Q_3) and maximum (Max).

aw		Corn							Spring wheat							Potato						
	Units	n	Min	Q_1	Mean	SD	Q_3	Max	n	Min	Q_1	Mean	SD	Q_3	Max	n	Min	Q_1	Mean	SD	Q_3	Max
Sand	%	121	0.0	18.0	40.1	25.8	65.3	86.0	48	7.9	32.1	57.3	29.3	84.2	89.1	44	65.4	70.6	75.5	4.6	78.7	79.4
Clay	%	121	5.5	15.7	27.5	15.7	39.0	68.0	48	5.0	6.4	16.7	12.4	23.5	46.0	44	5.5	5.9	6.7	1.3	8.0	9.0
Silt	%	121	6.0	24.0	32.5	13.5	43.0	55.0	48	5.8	9.0	26.1	17.8	42.5	52.0	44	14.7	15.8	17.8	3.4	21.4	25.6
pH		121	5.10	6.20	6.65	0.65	7.1	7.8	48	5.9	6.25	6.57	0.47	7.00	7.30	44	4.7	4.90	5.14	0.25	5.3	5.5
BD	$g.cm^{-3}$	121	0.85	1.26	1.34	0.12	1.43	1.53	48	1.17	1.27	1.36	0.11	1.42	1.60	44	1.22	1.24	1.25	0.03	1.24	1.30
TN	%	121	0.06	0.12	0.16	0.06	0.18	0.38	48	0.06	0.10	0.13	0.03	0.15	0.18	44	0.11	0.18	0.20	0.04	0.23	0.27
TC	%	121	0.64	1.35	1.98	0.84	2.50	4.72	48	1.05	1.37	1.69	0.46	2.03	2.5	44	1.5	2.40	2.67	0.49	3.00	3.34
PMN	$mg.kg^{-1}$	121	7.6	20.5	25.6	9.8	30.3	67.1	48	12.7	16.5	21.6	6.5	27.3	31.5	44	17.5	27.9	32.9	6.9	37.9	43.5
$PPSN_{0-30\ cm}$	$kg.ha^{-1}$	51	0.4	2.7	24.2	34.3	27.0	84.0	48	1.0	6.6	21.5	21.1	26.9	99.8	44	4.6	16.0	27.2	18.9	31.8	84.2
$PPSN_{0-90\ cm}$	$kg.ha^{-1}$	48	0.6	5.8	36.6	45.2	42.7	151.0	48	1.6	10.6	34.7	30.5	47.0	123.7	44	6.5	34.1	69.8	65.7	79.9	302.3
GSNM	$kg.ha^{-1}$	121	33.0	73.7	84.0	26.9	99.5	196.0	48	44.8	63.0	72.4	17.4	92.8	97.9	44	58.6	88.3	99.7	18.5	114.0	128.5
N uptake	$kg.ha^{-1}$	113	10.0	50.0	71.0	33.4	86.5	179.0	48	15.5	27.2	40.1	16.2	50.6	81.2	44	15.8	36.3	51.9	21.4	66.0	94.4
$RSN_{0-30\ cm}$	$kg.ha^{-1}$	51	0.5	1.6	15.5	23.5	13.6	64.0	48	0.13	3.64	25.5	51.1	21.6	330.0	44	3.5	8.6	33.5	22.0	49.6	80.6
$RSN_{0-90\ cm}$	$kg.ha^{-1}$	47	1.0	4.6	18.5	25.1	18.7	100.0	48	0.23	7.9	35.6	59.8	36.7	354.7	44	6.5	46.7	83.1	46.2	118.4	197.9

*BD, Bulk Density; TN, Total Nitrogen; TC, Total Organic Carbon; PMN, Potentially Mineralizable Nitrogen (14-day aerobic incubation); GSNM, estimated Growing Season Nitrogen Mineralization; $PPSN_{0-30\ cm}$, Preplant Soil Nitrogen at 0–30 cm soil depth; $PPSN_{0-90\ cm}$, Preplant Soil Nitrogen at 0–90 cm soil depth, $RSN_{0-30\ cm}$, Residual Soil Nitrogen at 0–30 cm soil depth; $RSN_{0-90\ cm}$, Residual Soil Nitrogen at 0–90 cm soil depth.

the upper 10 cm of the soil surface. The collection of the Ap2 horizon sample was based on the total thickness of the Ap horizon. If the Ap horizon was less than 17 cm thick, no Ap2 sample was collected. For Ap horizons ranging from 17 to 30 cm, the Ap sample was taken from 10 cm to the bottom of the Ap horizon, with a maximum depth of 25 cm. When the Ap horizon exceeded 30 cm in thickness, sampling was conducted between 15 and 30 cm. The B horizon was sampled from the first 15 cm immediately below the lower boundary of the Ap horizon. Detailed methods of soil sampling and analysis are described in Gasser et al. (2023). Table 2 provides a summary of the basic soil properties and parameters of the soils included in the EESSAQ survey data set.

In terms of the soil chemical properties, TC and TN were measured by the dry combustion method (Skjemstad and Baldock, 2007) using a LECO CN828 analyzer at a temperature of 950 °C (LECO Corp., St. Joseph, MI). TC and TN contents ranged from 0.8 to 529 g kg^{-1} and from 0.01 to 24.9 g kg^{-1}, respectively (Gasser et al., 2023). Soil pH was measured using a 1:1 soil-to-water ratio (Hendershot et al., 2007). Active carbon, also known as Permanganate Oxidizable Carbon (POXC), was measured through colorimetric methods from the CASH framework, as adapted from Weil et al. (2003). The process starts by air-drying the soil, followed by mixing it with a potassium permanganate solution. The mixture is then shaken, extracted, and finally measured spectroscopically for POXC (Marshall et al., 2021).

Potentially mineralizable N was estimated through a 14-day aerobic incubation at room temperature (approximately 21 °C) and at 45 % water-filled space as the difference in the mineral N content before and after incubation (Campbell and Curtin, 2007; Parfitt et al., 2005; Scott et al., 1998). The mineral N content ($NH_4^+ + NO_3^-$) was determined colorimetrically using a two-channel Seal Analytical AA3 HR Nutrient Autoanalyzer. Soil respiration (SR) was measured by determining the cumulative CO_2 flush after rewetting air-dried soils incubated for 3 days at room temperature (approximately 21 °C) and at 45 % water-filled pore space (Franzluebbers et al., 2000; Picone et al., 2002). Soil samples were air-dried, weighed, and placed in sealed jars with a vial containing a solution of 0.2 M NaOH to trap the CO_2 evolved, and a vial with water to maintain humidity. After 3 days of incubation, the alkali vials were removed, and the amount of CO_2 was determined by back-titration with 1 M HCl.

In terms of the soil physical properties, particle size distribution in soil was determined using the modified Bouyoucos method that involved using a hydrometer after removing organic matter and carbonate with H_2O_2 and acetic acid, respectively (Kroetsch and Wang, 2007). Here, the soils varied in texture, with sand and clay contents ranging from 2 to 974 g kg^{-1} and from 2 to 828 g kg^{-1}, respectively. Soil bulk density and porosity were measured using soil cores following the method described by Hao et al. (2007).

Aggregate stability was determined on an undisturbed soil block using the wet-sieving method described by Angers et al. (2008), with minor adaptations to improve reproducibility and particle recovery. The undisturbed soil block measuring 8 × 8 × 8 cm was collected from the surface horizon (Ap1, 1–10 cm depth) using a stainless-steel cutter to minimize structural disturbance. Each block was immediately placed in a pre-labeled airtight plastic bag, sealed to preserve its natural moisture, and transported in a rigid protective container to prevent structural alteration during transport. Approximately 100 g of moist soil, pre-sieved to 8 mm, was placed on the top sieve of a nested set with mesh sizes of 4, 2, 1, 0.5, and 0.25 mm (w_i). The sieve stack was immersed in distilled water for 5 min to ensure full wetting and then oscillated vertically for 10 min using a mechanical sieve shaker to simulate water-stable aggregate breakdown. After sieving, the material retained on each sieve was transferred into labeled Erlenmeyer flasks and oven-dried at 105 °C to determine the dry mass of aggregates (w_{2i}). To remove sand and organic debris, 5 mL of a 5 g L^{-1} sodium hexametaphosphate dispersing solution was added to each flask and diluted to 50 mL with deionized water. The suspension was mechanically shaken for 45 min, after which the remaining coarse fraction was washed back onto its respective sieve, dried, and weighed. The mass of coarse fragments and organic residues (w3i) was subtracted from the corresponding aggregate fraction (w2i).

The proportion of water-stable aggregates (WSA_i) in each of the size fractions was calculated using the method described by Angers et al. (2008):

$$WSA_i = \frac{w_{2i} - w_{3i}}{\frac{w_1}{1+e_g} - \sum_{i=1}^{n} w_{3i}}$$

where w_1 is the mass of the moist soil sample (g); e_g is the gravimetric water content of the sample (g g^{-1}); w_{2i} is the mass of aggregates retained on sieve i after drying (g); w_{3i} is the mass of coarse fragments and debris in fraction I (g); and n is the total number of sieves.

The mean weight diameter (MWD) was calculated to express the size distribution of soil aggregates:

$$MWD = \sum_{i=1}^{n} x_i \times WSA_i$$

where i corresponds to each fraction collected, x_i is the mean diameter of each size fraction, and WSA_i is the proportion of water-stable aggregates in each size fraction. Larger MWD values indicate a higher proportion of stable macroaggregates, reflecting greater structural stability and resistance to slaking.

Table 2
Summary Statistics of Soil Parameters of Ap1 and Ap2 Horizons from the Soil Health Database (EESSAQ): sample size (n), minimum value (Min), first quartile (Q_1), mean, standard deviation (SD), third quartile (Q_3) and maximum (Max).

Parameter*		Horizon Ap1 (0–10 cm)							Horizon Ap2 (10–25 or 30 cm)						
	Units	n	Min	Q_1	Mean	SD	Q_3	Max	n	Min	Q_1	Mean	SD	Q_3	Max
Sand	%	1627	0.0	11.5	36.5	26.0	55.2	97.4	1479	0.0	11.9	36.9	26.4	56.3	94.7
Clay	%	1627	2.5	14.5	29.0	19.9	37.8	96.5	1479	2.9	13.8	28.8	19.9	38.1	96.1
Silt	%	1627	0.0	24.0	34.5	16.0	45.0	79.5	1479	0.1	23.7	34.3	16.2	44.8	79.8
pH		1634	4.3	5.7	6.1	0.6	6.6	8.0	1482	4.5	5.7	6.2	0.6	6.6	8.1
BD	$g.cm^{-3}$	1634	0.47	1.04	1.18	0.20	1.32	1.73	1483	0.44	1.19	1.32	0.19	1.45	1.90
TN	%	1634	0.03	0.16	0.25	0.12	0.31	0.99	1482	0.04	0.13	0.19	0.09	0.24	0.88
TC	%	1634	0.57	2.00	3.13	1.78	3.86	21.46	1482	0.37	1.64	2.50	1.38	3.02	19.76
POXC	$mg.kg^{-1}$	1634	111.0	468.9	637.0	252.1	764.3	2735.2	1482	104.5	362.6	498.3	214.1	598.4	2475.8
PMN	$mg.kg^{-1}$	1630	1.36	36.8	67.9	44.0	89.9	253.9	1478	3.8	25.7	42.6	23.4	55.4	145.9
SR	$mg.kg^{-1}$	1634	34.0	166.4	291.8	180.1	371.1	1759.8	1482	33.1	125.0	184.1	89.1	228.3	771.6
GSNM	$kg.ha^{-1}$	1539	48.6	161.7	258.1	136.0	324.4	844.0	1395	41.1	125.8	177.5	74.2	216.0	546.6
AS	mm	1633	0.03	2.1	2.8	1.02	3.5	5.4							

* BD, Bulk Density; TN, Total Nitrogen; TC, Total Carbon; POXC, Permanganate Oxidable Carbon (Active Carbon); PMN, Potentially Mineralizable Nitrogen (14-day aerobic incubation); SR, Soil Respiration (3-day CO_2-burst); GSNM, estimated Growing Season Nitrogen Mineralization; AS, Aggregate Stability (expressed as Mean Weight Diameter).

2.5. Development of pedotransfer functions

The modeling framework follows that of Kasraei et al. (2024), which applies a feature selection by first removing highly correlated predictors using variance inflation factor (VIF) analysis; and secondly, by removing irrelevant predictors using recursive feature elimination (RFE).

2.5.1. Variance inflation factor analysis

The feature selection process began with an assessment of multicollinearity and examination of correlations among the predictors using VIF analysis (Marquardt, 1970). Here, a VIF value is calculated for all predictors, whereby the predictor with the highest VIF was removed. Following an iterative (re)calculation of VIF and removal of the highest correlated variable, the process was terminated once all remaining predictors had VIF < 5 (Kasraei et al., 2024). The remaining predictors were retained for predictive modeling purposes.

2.5.2. Predictive modeling

To develop the PTFs, multiple linear regression (MLR) and RF were employed. RF is a machine learning technique that constructs an ensemble of tree predictors, each generated from a random subset of the data and predictor variables (Breiman, 2001). By aggregating the predictions from multiple trees, RF improves model stability and predictive accuracy. This algorithm is particularly well suited to soil datasets, which often include variables of different types, such as numerical, ordinal, and categorical, and complex, nonlinear relationships among them (Cutler et al., 2007). RF has been successfully applied in numerous soil and environmental studies, including the development of PTFs, predictive digital soil mapping, water resource management, and spatial imaging (Deragon et al., 2023; Heung et al., 2016; Paul et al., 2023; Tyralis et al., 2019).

When using RF, m_{try}, which represents the number of randomly selected covariates at each node (Kuhn et al., 2020), is the main parameter that needs to be tuned. The m_{try} value can range from 1 to the total number of predictors. In the initial modeling phase, the best m_{try} value for each model was determined using a leave-one-site-out cross-validation, with soil profiles from the same sampling site grouped together. This approach ensured that observations from the same site, likely to be spatially autocorrelated, were not included in both training and test sets, thereby minimizing the risk of data leakage and overfitting (John et al., 2025).

2.5.3. Feature selection and model simplification

Recursive feature elimination (RFE) was performed with the *rfe* function in the *caret* package (Kuhn et al., 2020) to reduce the number of predictors for training the final models. Unlike VIF, which is conducted independently of the response variable to address multicollinearity, RFE eliminates irrelevant predictors based on model performance in predicting the response variable (Deragon et al., 2023). The RFE process begins with building a model on the entire set of predictors and assigning an importance score to each predictor. The predictors with the least importance are then removed, and the model is rebuilt with the remaining predictors. This process is repeated until model performance metrics are optimized. To investigate the effect of each covariate on the predictions, an accumulated local effects (ALE) plot was generated using the *iml* package. Each plot illustrates the first-order effect between the predicted variable and one of the covariates in the model.

2.5.4. Accuracy metrics

The performance metrics were obtained via a 5-fold leave profile out cross-validation created during the initial modeling phase. To improve the stability and reliability of model accuracy estimates, the cross-validation process was repeated 5 times.

Final model predictions were evaluated using three key metrics: normalized root mean square error (NRMSE), the coefficient of determination (R^2), and Lin's concordance correlation coefficient (CCC). NRMSE expresses root mean square error relative to the mean of observed values, thus providing a unitless criterion:

$$NRMSE = \frac{RMSE}{Mean}$$

Model performance was assessed according to the classification criteria for R^2 values proposed by Li et al. (2016). According to these criteria, an R^2 value of less than 0.50 indicates poor predictions, a value between 0.50 and 0.75 represents acceptable predictions, and an R^2 value of 0.75 or higher denotes good predictive performance. The CCC measures the reproducibility of predictions by evaluating the agreement between observed and predicted values. It assesses both precision and accuracy by analyzing how closely data points align with the 45° line through the origin (Lawrence and Lin, 1989). CCC values range from −1 to 1, with values near 0 indicating a lack of concordance between predicted and actual values.

2.5.5. Uncertainty estimation

To quantify the uncertainty associated with the predictions of N pools, a quantile regression (QR) approach was applied. Originally introduced by Koenker and Bassett Jr. (1978) and later adapted for digital soil mapping applications by Kasraei et al. (2021), QR offers several advantages, including its robustness to outliers (Staffa et al., 2019) and the ability to provide information about the effect of predictors on selected quantiles rather than just the mean (Das et al., 2019). Here, QR was implemented to estimate uncertainty for both the RF and MLR models developed after RFE. For each model, quantile levels $\tau = 0.05$, and 0.95 were used to compute the lower, and upper conditional quantiles of predicted TN, PMN, and GSNM. All QR analyses were conducted using the *quantreg* package (Koenker, 2025) in R statistical software. For each response variable, the model residuals and corresponding predictions were utilized to estimate the conditional quantiles. From the QR results, the 5 % lower and 95 % upper prediction limits were derived, and the difference between these limits defined the 90 % prediction interval used to visualize model uncertainty.

3. Results

3.1. Evaluation of the zero-plus first-order prediction function

A highly significant relationship ($p < 0.0001$) was observed between the predicted GSNM and the field-based estimates of SNS in zero-N trials for both corn and spring wheat (Figs. 1 and 2). The R^2 values ranged from 0.69 to 0.79 for corn and 0.35 to 0.64 for spring wheat. The relationship improved further when the SNS estimates accounted for nitrate concentrations in deeper soil layers. Specifically, using SNS estimates from the 0–90 cm depth resulted in higher R^2 values compared to estimates based solely on the 0–30 cm layer (Figs. 1 and 2). Moreover, the strongest model performance was obtained when GSNM was predicted for the first 60 days of the growing season. A significant relationship between predicted GSNM and field-based SNS was also found for potato crops when SNS measurements were based on the 0–90 cm soil layer (Fig. 3). In contrast, when only the initial 30 cm soil layer was considered (PPSN and RSN at the 30 cm depth), the relationship was notably weak, with an R^2 value of 0.05 (data not shown).

3.2. Development of pedotransfer functions

3.2.1. Model performance and predictive accuracy of control models

Random forest-based PTFs were implemented in R using the 12 soil parameters from the database (Table 2). Because GSNM was calculated from TN and PMN, these variables were excluded from the PTFs for GSNM. To address concerns related to multicollinearity among predictors (Menard, 1995), a VIF analysis was conducted, resulting in the reduction of PTFs to 8 parameters for each variable of interest. Three RF models, generated using these parameters, were used as controls to

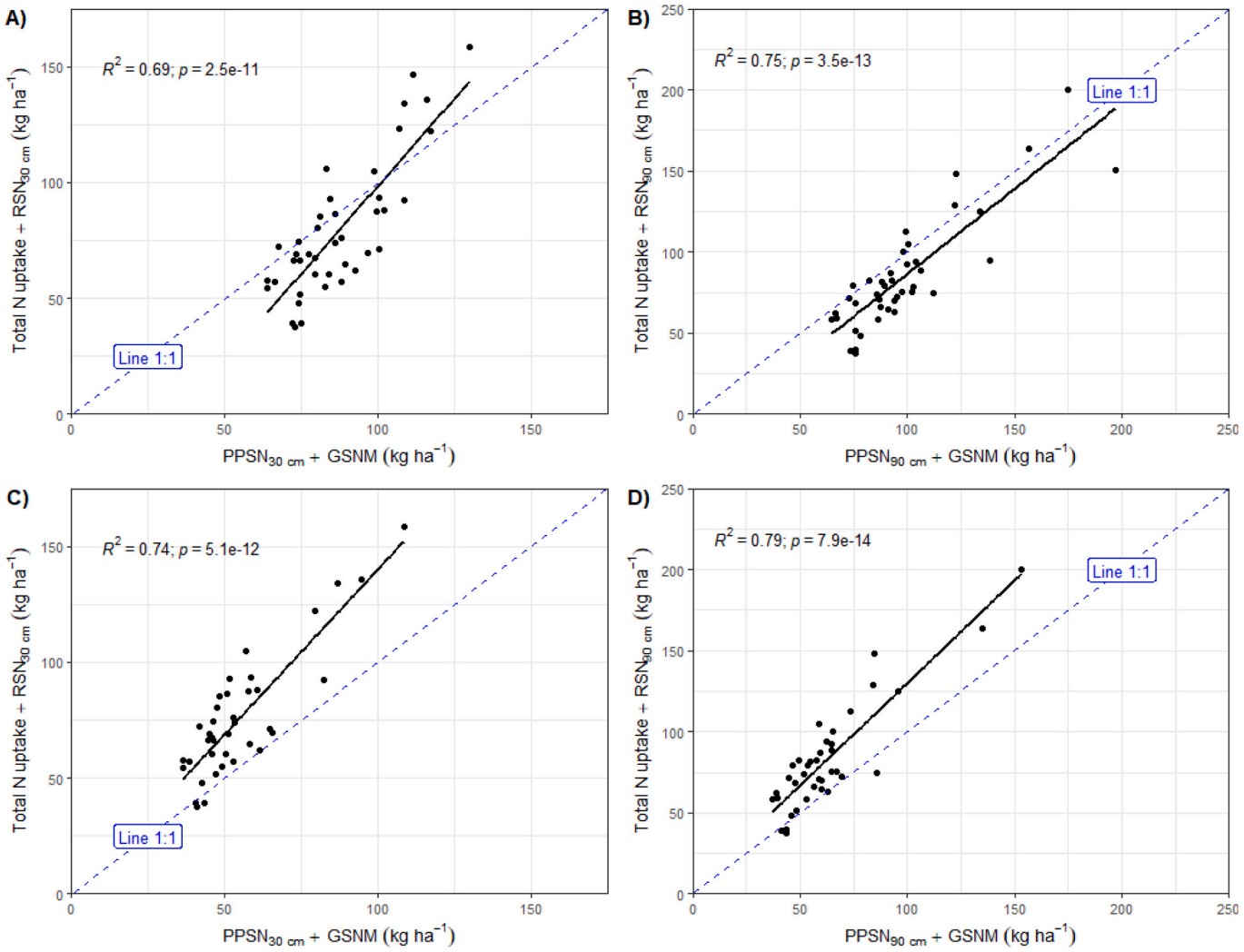


Fig. 1. Relationship between predicted soil N supply [(pre-plant soil mineral N (PPSN) + growing season N mineralization at 30 cm depth (GSNM)] and field-based soil N supply in zero-N trials [Total N uptake + residual soil N at the end of the growing season (RSN)] for corn. Panels show comparisons using: A) 130-day GSNM and N measurements from 0 to 30 cm soil depth; B) 130-day GSNM and N measurements from 0 to 90 cm soil depth; C) 60-day GSNM and N measurements from 0 to 30 cm soil depth; D) 60-day GSNM and N measurements from 0 to 90 cm soil depth.

compare against the PTFs with a reduced number of parameters, following the RFE procedure (Table 3).

The RF models demonstrated strong predictive performance across all variables. For TN, the model achieved an R^2 of 0.92, an NRMSE of 0.13, and a CCC of 0.95, based on eight predictors and 3097 observations (Table 3). The PMN-PTF showed slightly lower, yet still highly accurate, predictive performance with an R^2 of 0.84, an NRMSE of 0.27, and a CCC of 0.91. In comparison, the GSNM-model outperformed its PMN counterpart, achieving an R^2 of 0.86, an NRMSE of 0.20, and a CCC of 0.92, indicating a strong agreement between predicted and observed values.

3.2.2. *Simplified pedotransfer functions*

Recursive feature elimination combined with repeated cross-validation was applied to systematically identify the most relevant parameters for developing simplified RF-based PTFs for TN (RF-TN_{RFE}), PMN (RF-PMN_{RFE}), and GSNM (RF-$GSNM_{RFE}$). To enhance interpretability and create straightforward predictive equations, the selected covariates were then used to train MLR models. It should be noted that even though MLR and RF are highly contrasting in model structure, the use of RF for selecting relevant features for MLR was still deemed appropriate given the similarity in accuracy metrics. This approach enabled the estimation of each target variable based on the chosen predictors and their associated coefficients.

3.2.2.1. Total nitrogen. The RFE process identified TC and clay content as the primary predictors of TN. This result is consistent with the correlation analysis, which revealed a strong positive association between TC and TN ($r = 0.92$; Fig. 4). The ALE plots further illustrate the relative influence of TC and clay content on the predicted TN levels (Fig. 5A). TC had a strong, nearly linear positive effect on TN, with the relationship becoming even more pronounced at TC levels above 2 %. The reliability of this trend was reinforced by a high density of observations between 1 % and 8 % TC. In contrast, the ALE curve for clay content remained relatively flat, showing minimal variation across the 0–100 % range. Model performance metrics confirmed the robustness of the RFE-selected predictors. With a dataset of 3097 observations, the RF-TN_{RFE} model achieved an R^2 of 0.86, a NRMSE of 0.18, and a CCC of 0.93. Similarly, the MLR-based TN_{RFE} model showed comparable performance with an R^2 of 0.86, NRMSE of 0.18, and CCC of 0.92 (Fig. 6A; Table 3).

3.2.2.2. Potentially mineralizable nitrogen. Soil respiration and pH were identified as the key predictors in predicting PMN through the RFE process. In line with this finding, correlation analysis showed a strong

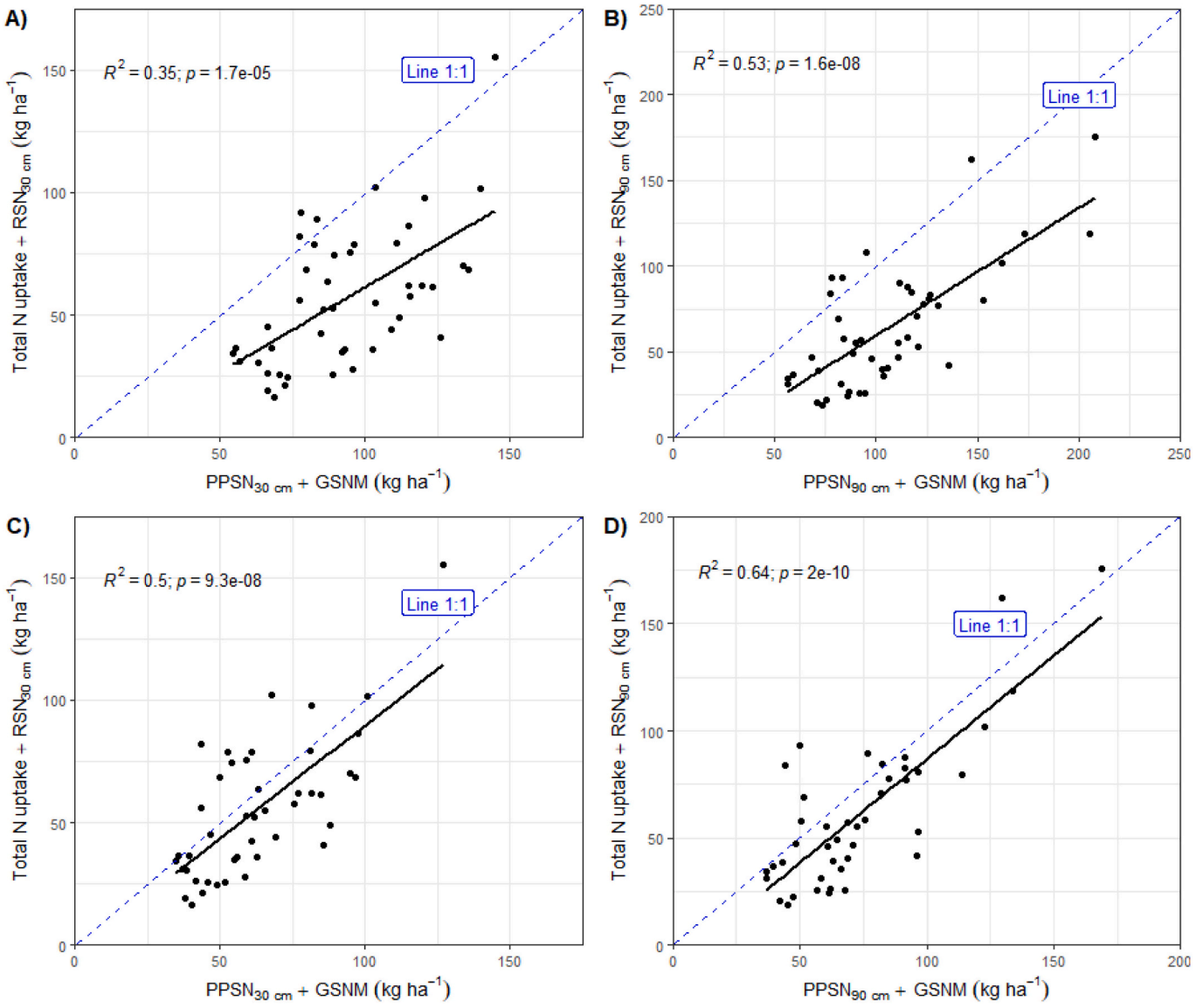


Fig. 2. Relationship between predicted soil N supply [(pre-plant soil mineral N (PPSN) + growing season N mineralization at 30 cm depth (GSNM)] and field-based soil N supply in zero-N trials [Total N uptake + residual soil N at the end of the growing season (RSN)] for spring wheat. Panels show comparisons using: A) 130-day GSNM and N measurements from 0 to 30 cm soil depth; B) 130-day GSNM and N measurements from 0 to 90 cm soil depth; C) 60-day GSNM and N measurements from 0 to 30 cm soil depth; D) 60-day GSNM and N measurements from 0 to 90 cm soil depth.

relationship between SR and PMN (r = 0.90; Fig. 4). The ALE plot illustrated that PMN was lowest when SR was below 200 mg kg^{-1}, increased rapidly between 200 and 600 mg kg^{-1}, and continued to rise more gradually beyond this range (Fig. 5B). For pH, the ALE plot demonstrated a unimodal relationship, with PMN reaching its peak at approximately pH 5.5. Interestingly, although TC and TN were significantly correlated with PMN (r = 0.65 and r = 0.76, respectively), they were not identified as top predictors in the RFE process. Both the RF-PMN_{RFE} and MLR-PMN_{RFE} models showed comparable predictive performance using the same set of predictors. The RF model achieved an R^2 of 0.80, an NRMSE of 0.3, and a CCC of 0.89 (Table 3). Similarly, the MLR model produced nearly identical results, with an R^2 of 0.80, an NRMSE of 0.3, and a CCC of 0.89 (Fig. 6B; Table 3).

3.2.2.3. Growing season nitrogen mineralization. Based on 2923 observations, the RF-$GSNM_{RFE}$ model demonstrated a strong predictive performance, achieving an R^2 of 0.83, an NRMSE of 0.22, and a CCC of 0.91 (Table 3). The MLR-based $GSNM_{RFE}$ model performed similarly, with an R^2 of 0.82, an NRMSE of 0.23, and a CCC of 0.90 (Fig. 6C; Table 3. |The RFE process identified SR, TC, and pH as the most significant predictors for GSNM, with correlation coefficients of 0.90, 0.65, and 0.24, respectively (Fig. 4). The ALE plot in Fig. 5C further illustrates the individual contributions of each variable to the model. SR demonstrated the strongest and near-linear positive effect on GSNM. TC also had a positive influence, increasing GSNM by approximately 3–4 %, after which the response plateaued. In contrast, pH had a non-linear relationship, with: GSNM peaking around pH 5.5 before declining at higher pH values.

3.2.2.4. Prediction uncertainty. Quantile regression was used to estimate and visualize the predictive uncertainty of the RF and MLR models for TN, PMN, and GSNM (Figs. 7–9). The width of the shaded interval, which indicates prediction uncertainty, was similar for both the RF and MLR models across all variables. However, both models consistently underestimated values at the upper end of TN, PMN, and GSNM, as their fitted blue regression lines fell below the 1:1 reference line. Additionally, a clear heteroscedastic pattern was observed in all cases, where the prediction intervals widened as the observed values increased. Points

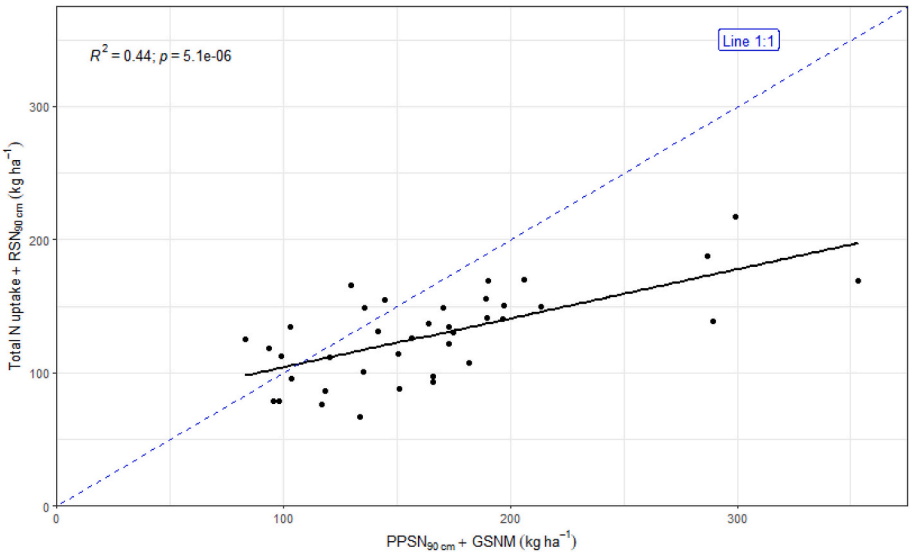


Fig. 3. Relationship between predicted soil N supply [(pre-plant soil mineral N at 90 cm depth ($PPSN_{90\ cm}$) + growing season N mineralization at 30 cm depth (GSNM)] estimated using a first-order prediction function, and field-based soil nitrogen supply in zero-N trials [(Total N uptake + residual soil N at 90 cm depth ($RSN_{90\ cm}$)] for potato.

Table 3
Results of Pedotransfer Functions (PTFs) for Total Nitrogen (TN), Potentially Mineralizable Nitrogen (PMN), and Growing Season Nitrogen Mineralization (GSNM).

Parameter*	Method*	Learner*	Conceptual Models*	Predictors	n	R^2	RMSE	CCC
TN	Control	RF	pH + TC + PMN + POXC + AS + Silt + Clay + BD	8	3097	0.92	0.13	0.95
PMN	Control	RF	SR + pH + TC + AS + Silt+ Clay + POXC + BD	8	3097	0.84	0.27	0.91
GSNM	Control	RF	SR + pH + TC + POXC + AS + Silt + Clay+ BD	8	2923	0.86	0.2	0.92
TN	RFE	RF	TC + Clay	2	3097	0.86	0.18	0.93
	RFE	MLR	TC + Clay	2	3097	0.86	0.18	0.92
PMN	RFE	RF	SR + pH	2	3097	0.80	0.3	0.89
	RFE	MLR	SR + pH	2	3097	0.80	0.3	0.89
GSNM	RFE	RF	SR + TC + pH	3	2923	0.83	0.22	0.91
	RFE	MLR	SR + TC + pH	3	2923	0.82	0.23	0.90

* BD, Bulk Density; TN, Total Nitrogen; TC, Total Carbon; POXC, Permanganate Oxidable Carbon (Active Carbon); PMN, Potentially Mineralizable Nitrogen (14-day aerobic incubation); SR, Soil Respiration (3-day CO_2-burst); GSNM, Growing Season Nitrogen Mineralization; AS, Aggregate Stability (expressed as Mean Weight Diameter); RFE, Recursive Feature Elimination; RF, Random Forest; MLR, Multiple Linear Regression.

outside the shaded 90 % prediction interval denote extreme prediction errors, where observed values lay beyond the models' expected confidence range. The frequency and distribution of these outliers were similar for both models, indicating comparable overall predictive performance across the three variables.

4. Discussion

4.1. Evaluation of the zero-plus first-order prediction function

The strong correlations between the predicted GSNM and field-based SNS for corn and spring wheat indicated that the zero-plus first-order model effectively captured N mineralization dynamics under field conditions in Quebec. Notably, the relationship improved when the field-based SNS was measured to a depth of 0–90 cm, which included both residual and leached nitrate pools. This finding was consistent with the well-documented mobility of nitrate in soil. Due to its high solubility, nitrate readily leaches below the surface layer as water moves through the soil (Zebarth and Rosen, 2007; Zvomuya et al., 2003). By incorporating nitrate from deeper soil layers, the 0–90 cm measurements provided a more accurate representation of the N mineralized and available to crops throughout the growing season.

The strongest relationship was achieved when GSNM was predicted for the first 60 days of the growing season, which aligned with the N uptake patterns of the studied crops. For corn, approximately 63 % of total N uptake occurs by the flowering stage (R1), which is around 63 to 68 days after emergence (Abendroth et al., 2011). In contrast, spring wheat has a more gradual N uptake, with about 30 % occurring before the first node stage, 40 % between the node and flag leaf stages, and the remaining 20 % between the flag leaf and flowering stages (Agriculture and Horticulture Development Board (AHDB), 2018). These patterns suggest that early-season N mineralization is critical for meeting crop N requirements and should be a focal point for management practices and modeling efforts.

For corn, the model performance was consistently strong across all timeframes and soil depths (R^2 = 0.69–0.79; Fig. 1), indicating that the model is well-suited for high N-demand crops characterized by a clear early-season uptake pattern. In contrast, R^2 values for spring wheat were moderately lower (0.35–0.64; Fig. 2), likely reflecting its more extended and variable N uptake pattern. This variability, combined with potential N losses beyond the root zone during the growing season, makes accurate modeling more challenging for wheat.

For potatoes, strong relationships between predicted GSNM and field-based SNS were observed only when SNS was estimated to a depth of 0–90 cm, whereas relationships based on the 0–30 cm layer were weak. This weaker performance observed in the 30 cm soil layer may be due to the unique physiological characteristics and soil interactions associated with potato systems. Potato plants typically have a shallow

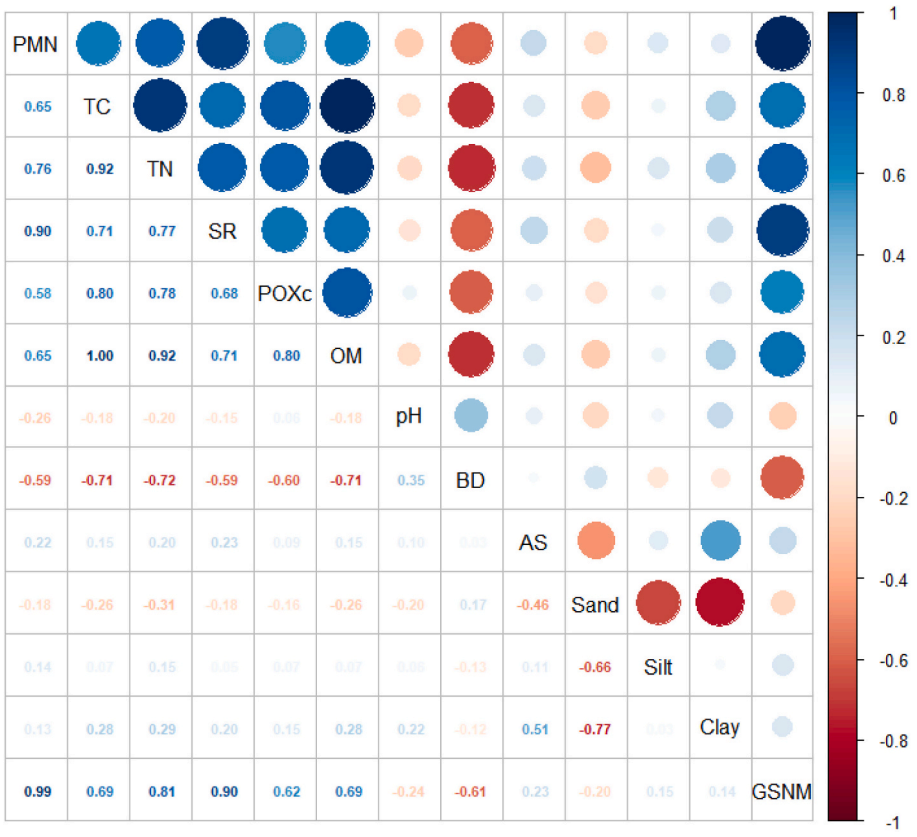


Fig. 4. Correlations between soil health database variables (BD, Bulk Density; TN, Total Nitrogen; TC, Total Organic Carbon; OM, Organic Matter; POXC, Permanganate Oxidable Carbon (Active Carbon); PMN, Potentially Mineralizable Nitrogen (14-day aerobic incubation); SR, Soil Respiration (3-day CO_2-burst); GSNM, Growing Season Nitrogen Mineralization; AS, Aggregate Stability (expressed as Mean Weight Diameter).

root system, with most roots concentrated in the upper 30–60 cm of soil (Zebarth and Rosen, 2007). This root structure leads to lower N uptake efficiency and a greater susceptibility to N losses through leaching. As a result, higher levels of RSN often remain at the end of the growing season (Errebhi et al., 1998; Zvomuya et al., 2003).

Gasser et al. (2002) also observed an inverse relationship between RSN and the sum of plant N uptake and nitrate leaching during the growing season. Interestingly, Dessureault-Rompré et al. (2015) reported stronger model performance (R^2 = 0.38–0.41) within the upper 15–30 cm of soil. The differences between their results and those of the present study may be attributed to variations in soil texture. While Dessureault-Rompré et al. (2015) examined a wide range of soil types, from clay to sandy, most zero-N trial sites used in this study were located on sandy soils. Such soils promote rapid nitrate leaching and reduce the retention of mineral N, thereby diminishing relationships at shallower depths (Zvomuya et al., 2003).

Overall, the zero-plus first-order model showed strong performance across various cropping systems, although there was some variability depending on the specific crop. The results highlight that rooting depth, soil texture, and leaching potential are key factors affecting the accuracy of GSNM predictions. For cereal crops such as corn and wheat, the model effectively captures early-season dynamics. However, for potatoes, which are typically grown on sandy soils, it is crucial to pay close attention to leaching behavior and deeper soil N pools to ensure accurate assessments.

4.2. Pedotransfer functions

4.2.1. Total nitrogen

The strong predictive performance of the control RF model highlights the potential of machine learning approaches to accurately estimate TN under various soil and environmental conditions. Compared to existing PTFs, this model outperformed the leading PTF developed by Laurence et al. (2023), which achieved an R^2 of 0.69 and a CCC of 0.81, along with a NRMSE of 0.18 based on 1984 observations and nine predictors (AS, PMN, POXC, SR, Clay, Sand, Silt, OM, pH). Similarly, Mesele and Ajiboye (2020) developed soil-property-based PTFs using 312 observations from tropical soils, where TOC, texture, and pH explained 77 to 85 % of TN variability (R^2 = 0.77–0.85). Collectively, these results confirm that soil attributes such as TOC, particle size distribution, and pH remain strong predictors of TN. The superior agreement between predicted and observed values in this study likely reflects the broader and more representative training dataset, which allowed the RF model to better capture the inherent non-linear relationships among soil properties governing N dynamics.

The identification of TC and clay content by the RFE process as the most influential predictors of TN was consistent with the strong positive correlation observed between TC and TN (r = 0.92). This relationship reflects the central role of OM, which serves as the primary reservoir for both C and N, with their availability being jointly governed by decomposition and mineralization processes (Schimel and Bennett, 2004). Comparable trends have been documented across various ecosystems, including montane forests and agricultural soils, where increases in TOC are consistently associated with proportional increases in TN (Shedayi et al., 2016; Mesele and Ajiboye, 2020). This consistent linear relationship indicates that TC is a dominant factor influencing N stocks, reflecting the simultaneous accumulation of C and N during humus formation and stabilization. Therefore, management practices that enhance organic matter inputs, such as residue retention and reduced tillage, are likely to improve both soil C sequestration and N fertility concurrently. In contrast, clay content showed little to no direct effect, suggesting that its effect on TN is either weak or mediated through its

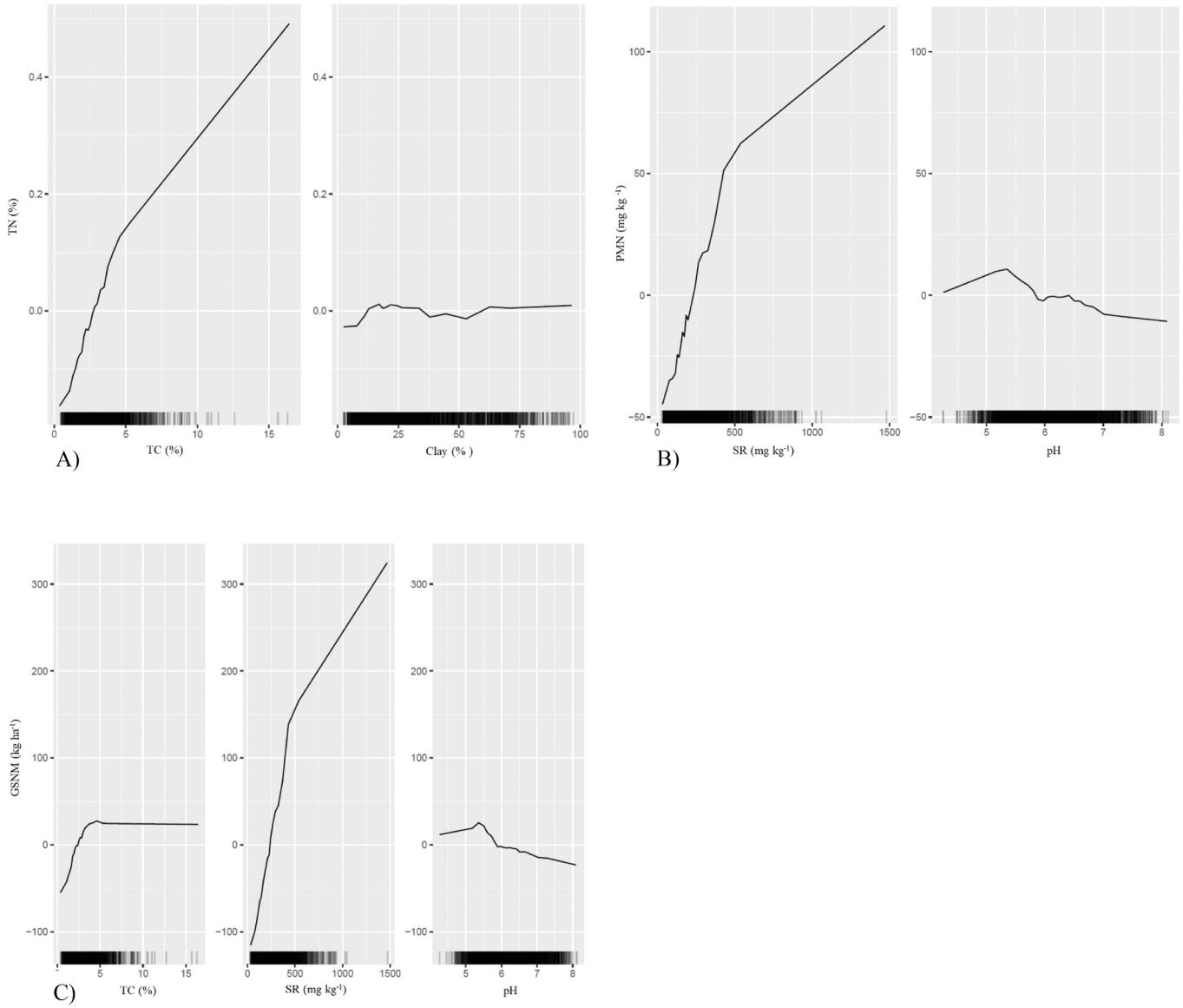


Fig. 5. Accumulated local effects (ALE) plot of A) total nitrogen (TN), B) potentially mineralizable nitrogen (PMN), C) growing season nitrogen mineralization (GSNM). TC, total carbon; SR, soil respiration. Each hatch mark at the bottom of each plot, represents an observed data point at a corresponding covariate value. A higher density of hatch marks indicates greater reliability of ALE results at those covariate values.

interactions with other soil properties.

The comparable performance of the RF-TN_{RFE} and MLR-TN_{RFE} models ($R^2 = 0.86$, CCC ≈ 0.92) confirms that the relationship between TN and its predictors was predominantly linear, allowing for the development of simpler and more interpretable predictive models without sacrificing accuracy. The results demonstrated that TN can be reliably estimated from fundamental soil properties, particularly TC, and that more complex machine learning approaches provide limited additional benefit for estimating this parameter.

When compared to other published PTFs for TN, the TN_{RFE} models developed in this study outperformed the cubist-based TN_{RFE} model by Laurence et al. (2023), which used PMN, POXC, sand, and soil OM as predictors, reporting an R^2 of 0.67, NRMSE of 0.18, and CCC of 0.80. Additionally, previous studies by Rashidi and Seilsepour (2009) and Mesele and Ajiboye (2020) employed linear regression analysis using TOC as the sole predictor, achieving R^2 values of 0.83; however, neither the NRMSE nor CCC were reported in those studies.

4.2.2. *Potentially mineralizable nitrogen*

The control PMN model achieved better accuracy metrics compared to those reported by Laurence et al. (2023), who found an R^2 of 0.66, an NRMSE of 0.30 and a CCC of 0.80. The improved performance in this study may be attributed to a more diverse and representative training dataset, as well as methodological differences. Specifically, PMN was measured using a 14-day aerobic incubation without a leaching step, which contrasts with the approach used by Laurence et al. (2023).

Soil respiration emerged as the key determinant of PMN because it reflects the metabolic activity of soil microorganisms that are responsible for OM decomposition and nutrient cycling (Booth et al., 2005). As an integrative measure, SR captures the combined effects of substrate availability, microbial biomass, and environmental factors such as moisture, temperature, and aeration, which regulate N mineralization and the supply of plant-available N (St. Luce et al., 2011). Although TC and TN were significantly correlated with PMN ($r = 0.65$ and 0.76, respectively), they were not top predictors. This is likely because TC and TN represent the total stable pools of these elements rather than their biological availability or turnover rates. While TC and TN provide

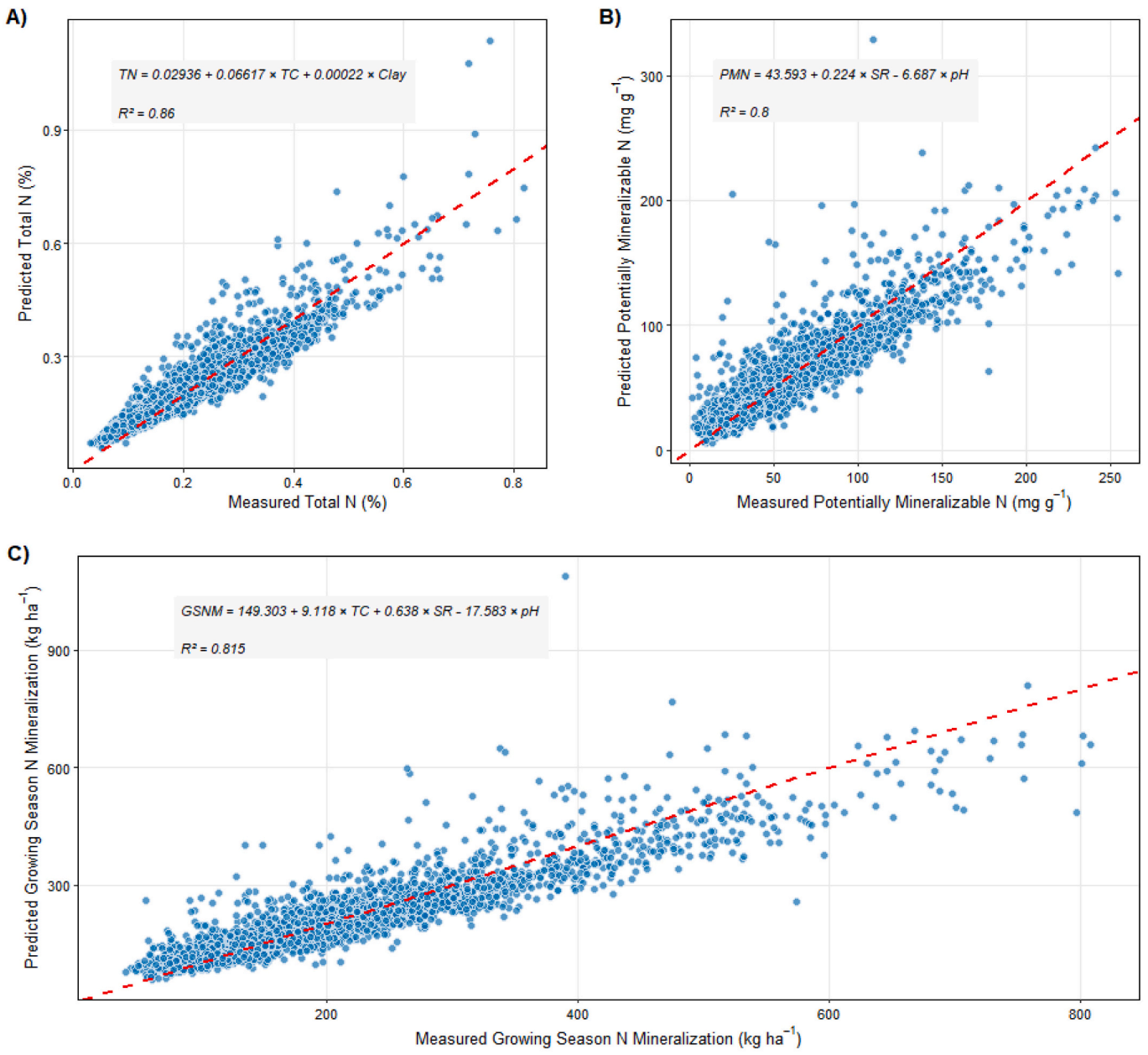


Fig. 6. Multiple Linear Regression Models Following Recursive Feature Elimination for A) Total Nitrogen, B) Potentially Mineralizable Nitrogen, and C) Growing Season Nitrogen Mineralization (TC, total carbon; SR, soil respiration). The red dashed line represents a perfect 1:1 relationship between observed and predicted values.

insight into the potential supply of OM and N, they do not accurately reflect how actively these resources are being processed by the soil microbial community. In contrast, SR directly indicates soil microbial activity that drives decomposition and N mineralization, providing a more immediate measure of the soil's capacity to release mineralizable N (Booth et al., 2005). Moreover, SR responds dynamically to management practices, organic inputs, and environmental changes, whereas TC and TN are relatively static and do not capture short-term fluctuations in mineralization potential.

Although RF is a non-parametric algorithm capable of capturing complex, nonlinear interactions and higher-order relationships among predictors (Li et al., 2014), its performance advantage over MLR was negligible. This similarity suggests that any nonlinear relationships within the dataset may have been either weak or statistically insignificant. As a result, the additional computational complexity and reduced interpretability of RF did not provide any practical benefits. Therefore, the simpler, more transparent MLR model is preferable, especially in applications where interpretability and ease of use are prioritized (Levy and O'Malley, 2020; Li et al., 2014; Smith et al., 2013).

When compared with previously published PTFs for PMN, the performance of the current models falls within or exceeds the ranges reported in the literature. Rasiah (1995) developed PTFs for N mineralization parameters using one- and two- pool models, predicting PMN from TN, residue removal factor, pH, and clay content, achieving an R^2 of 0.94 (NRMSE and CCC not reported). Similarly, Heumann et al. (2011) found that PTFs for the labile N pool performed better when grouped by former land use, using clay content, humus class, and mean fall temperature as predictors, yielding R^2 values of 0.34–0.42. More recently, Laurence et al. (2023) developed a PTF using AS, POXC, SR, and TN as predictors, achieving an R^2 of 0.64, NRMSE of 0.3, and a CCC of 0.78.

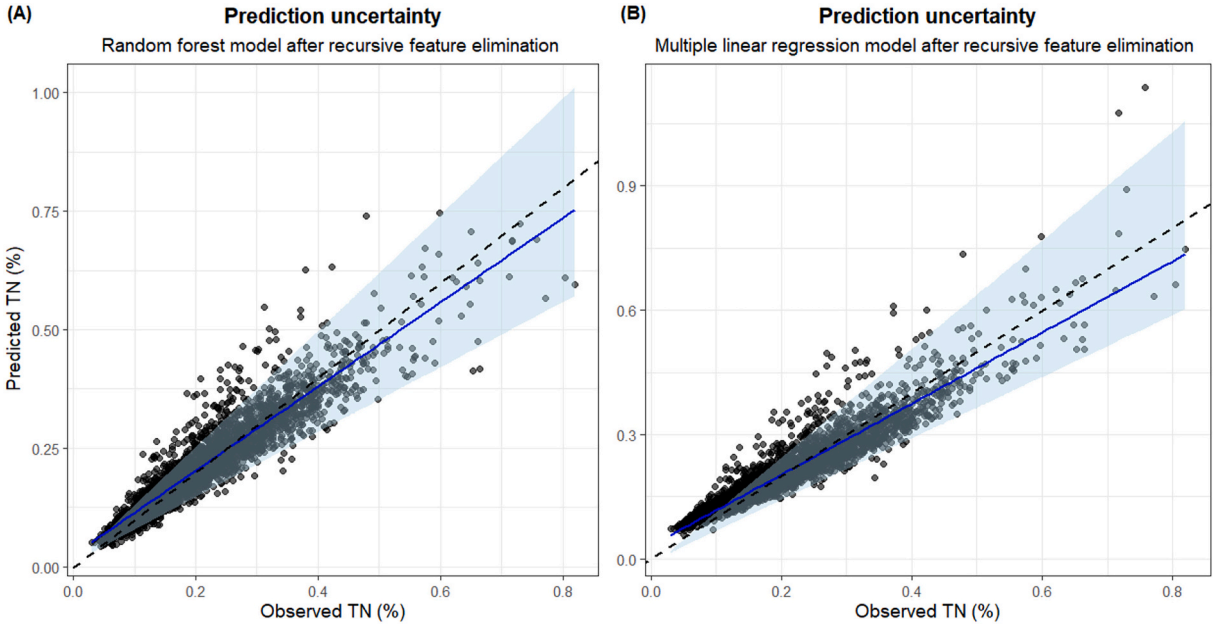


Fig. 7. Observed versus predicted value plot for total nitrogen (TN) models. The solid blue line represents the fitted regression between observed and predicted TN values, and the shaded area indicates the 95 % prediction interval. The dashed black line represents the 1:1 reference line.

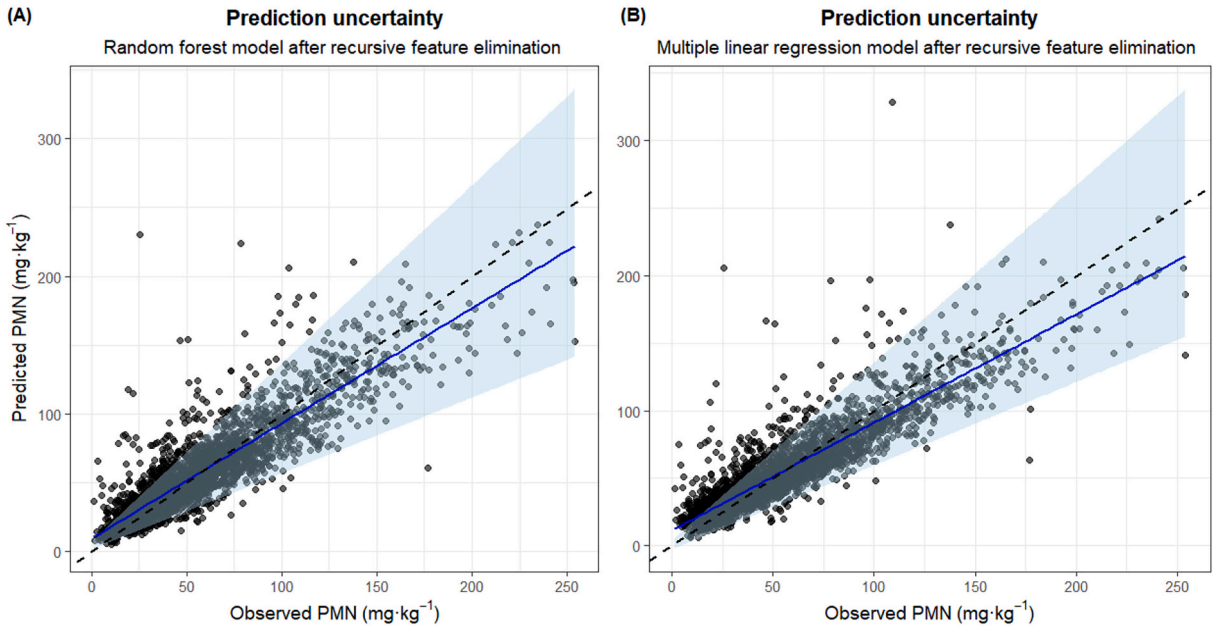


Fig. 8. Observed versus predicted value plot for potentially mineralizable nitrogen (PMN) models. The solid blue line represents the fitted regression between observed and predicted TN values, and the shaded area indicates the 95 % prediction interval. The dashed black line represents the 1:1 reference line.

4.2.3. Growing season nitrogen mineralization

The similar performance of the RF-GSNM$_{RFE}$ and MLR-GSNM$_{RFE}$ models indicates that the relationship between GSNM and its key predictors is largely linear, a pattern also observed for PMN. This consistency suggests that nonlinear interactions among variables were minimal in both instances. As a result, increasing model complexity does not offer significant additional benefits beyond what can be achieved with a linear approach.

Soil respiration emerged as the most significant predictor of GSNM, underscoring the crucial role of microbial activity in regulating N transformations during the growing season. Laurence et al. (2023) similarly identified SR as a primary predictor of GSNM in a PTF developed for soils in Prince Edward Island, confirming its broader relevance across various soil types and climatic conditions. This strong correlation aligns with the well-established relationship between microbial activity and N mineralization, as heterotrophic microbes break down OM, releasing CO_2 and mineral N (Booth et al., 2005). Research has shown that short-term CO_2 emissions from soil can explain a large proportion of the variation in longer-term N mineralization, reinforcing SR as a strong indicator of microbial metabolic activity driving N transformations (Haney et al., 2008).

The positive relationship between TC and GSNM indicates that the increased C availability enhances microbial activity and mineralization at lower concentrations, but this relationship reaches a saturation point

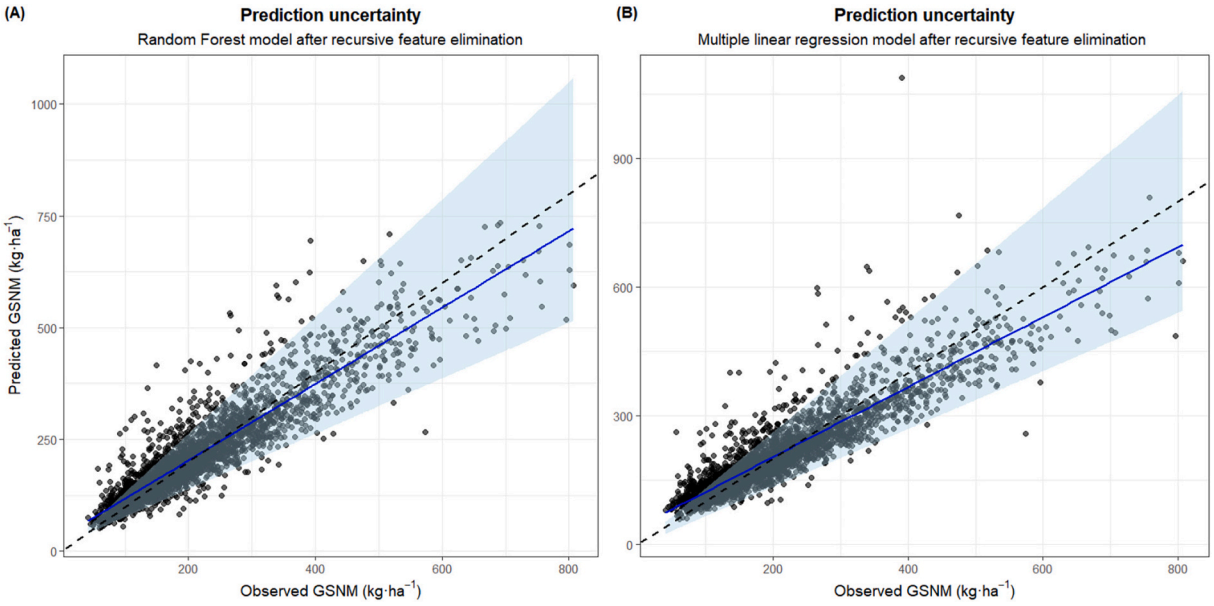


Fig. 9. Observed versus predicted value plot for growing season nitrogen mineralization (GSNM) models. The solid blue line represents the fitted regression between observed and predicted TN values, and the shaded area indicates the 95 % prediction interval. The dashed black line represents the 1:1 reference line.

at around 3–4 %, as illustrated by the ALE plot. This pattern suggests that once there is enough organic substrate, C is no longer the limiting factor for microbial processes. Instead, other factors such as nutrient balance and oxygen availability become more significant influences (Manzoni et al., 2012; Schimel and Schaeffer, 2012). The non-linear response of GSNM to pH highlights the complex influence of soil acidity on microbial communities and enzymatic activity (Aciego Pietri and Brookes, 2008; Kemmitt et al., 2006). Optimal mineralization occurs under slightly acidic to near-neutral conditions, which promote microbial diversity and enzyme efficiency. In contrast, highly acidic environments tend to constrain microbial activity due to aluminum toxicity and reduced enzymatic function (Breugem et al., 2024).

The GSNM, calculated from TN and PMN, maintained similar predictor variables, which aligns with the findings of Laurence et al. (2023). However, the Cubist-based model for predicting GSNM developed by Laurence et al. (2023) utilized AS, POXC, SR, organic matter, and pH as predictors to achieve an R^2 of 0.69, an NRMSE of 0.19, and a CCC of 0.82. The higher NRMSE values in this study may be attributed to the broader range of soil types and cropping systems examined compared to the study by Laurence et al. (2023).

4.3. Prediction uncertainty

The quantile regression analysis showed that both RF and MLR models had similar predictive uncertainty across TN, PMN, and GSNM. The comparable widths of the 90 % prediction intervals suggest that neither modeling approach provided a clear advantage in terms of overall confidence in predictions. This similarity implies that the additional flexibility of the RF model did not result in a meaningful reduction in uncertainty compared to the simpler MLR model. This may be due to limited nonlinearity in the predictor–response relationships or constraints related to the data distribution.

For all three variables, both models consistently underestimated values at the upper extremes. This downward bias highlights the challenge these models face in accurately capturing high-value observations, which is a common limitation when target variables have skewed or heavy-tailed distributions (Vaysse and Lagacherie, 2017). This underestimation may also arise from the relatively small number of high-value observations available for model training, which may reduce the predictive capability in those regions (Padarian et al., 2020).

The observed heteroscedastic pattern, where prediction uncertainty increased with higher observed values, underscores the difficulty in modeling variability in N dynamics (Sun et al., 2017). This pattern suggests that model performance may decline under conditions of elevated N availability or microbial activity, where the processes governing N mineralization can become more complex and spatially variable (Masunga et al., 2016). Incorporating additional explanatory variables related to environmental conditions and land use may help to reduce this scale-dependent uncertainty.

Overall, these results indicate that while both models effectively capture the general trends in soil N indicators, their ability to represent extreme values and heteroscedastic variability is still limited.

4.4. Implications and future work

The results of this study help address existing limitations by showing that a small set of easily measurable soil properties can effectively predict key N indices across various crops and soil types. The comparable performance of the MLR and RF models suggests that interpretable, low-complexity approaches can be both scientifically robust and practically applicable. However, as with all PTFs, the applicability of the developed relationships depends on the soil types and conditions represented in the training dataset (McBratney et al., 2002; Van Looy et al., 2017; Arbor et al., 2023; Mozaffari et al., 2024). The MLR-PTFs performed well within the range of soils examined here, but their use in other regions or contrasting soil environments should be approached cautiously until validated with independent datasets. Although the PTFs may also perform well in areas with similar soil characteristics, further testing is needed to confirm their transferability and ensure reliable predictions across broader contexts.

To enhance the generalizability of the model, it is essential to extend its calibration to cover a broader range of soils, climates, and cropping systems. Developing hierarchical or stratified PTFs based on factors such as soil texture, land use, and topography, could help capture soil-specific relationships and minimize regional biases. Additionally, future research should evaluate the two-pool kinetic model using additional measurements. Most studies conducted so far have focused on 14-day incubation methods; therefore, exploring shorter incubation periods or alternative proxies, like the 7-day aerobic mineralizable N and the CO_2 burst, could provide valuable insights into the model's robustness across

various methodological timeframes.

In Canada, each province uses different methods or proxies to measure PMN, primarily based on the analytical services offered by local commercial laboratories (Clément et al., 2025). This lack of uniformity in methodology limits the comparability of results and restricts transferability of models across regions. Evaluating kinetic and statistical models in relation to these province-specific approaches, including short-term incubations or labile C and N fractions, would help standardize interpretations and identify practical indicators that are appropriate for different regional contexts. Moreover, as some provinces lack laboratory analyses for labile N pools, developing PTFs by leveraging spatial predictors could offer a scalable alternative to predict GSNM. Incorporating these functions into national or provincial soil databases would provide consistent, spatially explicit estimates of N mineralization, ultimately improving nutrient management strategies across Canada.

5. Conclusion

This study demonstrated an effective integration of mechanistic modeling and data-driven approaches to enhance the estimation of GSNM in agricultural soils in Quebec. A refined two-pool zero-plus first-order kinetic model reliably estimated GSNM, with validation against field data from various cropping systems. Machine learning methods, specifically RF and MLR, were employed to develop PTFs that accurately predicted TN, PMN, and GSNM based on accessible soil properties. The comparable performance of RF and MLR indicates largely linear relationships between predictors and N pools, which supports the use of simpler, interpretable, and computationally efficient models. The MLR-based PTFs provide a practical and scalable tool for estimating GSNM in areas with limited data, allowing for more informed, site-specific N management. For farmers, these predictive tools represent a significant opportunity to optimize N use, improve economic returns, and mitigate environmental impacts. Future research should prioritize integrating these models into farm-level decision support tools. Additionally, improving PTFs by incorporating soil management practices and considering spatial and landscape variability through digital soil mapping will enhance their accuracy and applicability.

CRediT authorship contribution statement

Chedzer-Clark Clement: Resources, Data curation, Writing – review & editing. **R. Deragon:** Methodology, Investigation, Formal analysis, Conceptualization, Writing – review & editing. **B. Heung:** Software, Methodology, Formal analysis, Conceptualization, Writing – review & editing. **J. Dessureault-Rompré:** Methodology, Conceptualization, Writing – review & editing. **M.O. Gasser:** Methodology, Data curation. **J.-B. Mathieu:** Methodology, Data curation. **D.L. Burton:** Supervision, Resources, Project administration, Funding acquisition, Conceptualization, Writing – review & editing.

Declaration of competing interest

The financial support of this research provided by the Weston Family Foundation Soil Health Imitative (DB) and the Natural Sciences and Engineering Research Council of Canada (NSERC) through a Discovery Grant (DB). Soil health data was provided by Institut de recherche et de développement en agroenvironnement (IRDA). The authors have no competing interests to declare that are relevant to the content of this article. The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Acknowledgements

We would like to acknowledge the contributions of all scientists who generated and made the two datasets used in this study available. This research was funded by The Weston Family Foundation, with support from Dalhousie University, Laval University, and the Quebec Research and Development Institute for the Agri-environment.

Data availability

The authors do not have permission to share data.

References

Abendroth, L., Elmore, R., Boyer, M., Marlay, S., 2011. Corn Growth and Development (PMR 1009). Iowa State University Extension.

Aciego Pietri, J.C., Brookes, P.C., 2008. Relationships between soil pH and microbial properties in a UK arable soil. Soil Biol. Biochem. 40 (7), 1856–1861.

Agriculture and Horticulture Development Board (AHDB), 2018. Wheat Growth Guide. AHDB Cereals & Oilseeds, Kenilworth, UK.

Amanabadi, S., Vazirinia, M., Vereecken, H., Vakilian, K.A., Mohammadi, M.H., 2019. Comparative study of statistical, numerical and machine learning-based Pedotransfer functions of water retention curve with particle size distribution data. Eurasian Soil Sci. 52 (12), 1555–1571.

Angers, D.A., Bullock, M., Mehuys, G., 2008. Aggregate Stability to Water. Soil Sampling and Methods of Analysis, 2, pp. 811–819.

Angers, D., Ouimet, R., Roy-Léveillé, P., Garneau, M., 2022. Priorities for management and protection of Québec soils. Geoderma Reg. 29, e00523. https://doi.org/10.1016/j.geodrs.2022.e00523.

Arbor, A., Schmidt, M., Saurette, D., Zhang, J., Bulmer, C., Filatow, D., Kasraei, B., Smukler, S., Heung, B., 2023. A framework for recalibrating pedotransfer functions using nonlinear least squares and estimating uncertainty using quantile regression. Geoderma 439, 116674.

Arbor, A., Schmidt, M., Zhang, J., Bulmer, C., Filatow, D., Kasraei, B., Smukler, S., Heung, B., 2024. Filling the gaps in soil data: a multi-model framework for addressing data gaps using pedotransfer functions and machine-learning with uncertainty estimates to estimate bulk density. Catena 245, 108310.

Benke, K.K., Norng, S., Robinson, N.J., Chia, K., Rees, D.B., Hopley, J., 2020. Development of pedotransfer functions by machine learning for prediction of soil electrical conductivity and organic carbon content. Geoderma 366, 114210.

Booth, M.S., Stark, J.M., Rastetter, E., 2005. Controls on nitrogen cycling in terrestrial ecosystems: a synthetic analysis of literature data. Ecol. Monogr. 75 (2), 139–157.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.

Breugem, A., Kros, H., de Vries, W., 2024. Impacts of pH on Mechanisms and Rates of Carbon and Nitrogen Mineralisation: A Review.

Burton, D.L., Zebarth, B.J., Gillam, K.M., MacLeod, J.A., 2008. Effect of split application of fertilizer nitrogen on N2O emissions from potatoes. Can. J. Soil Sci. 88 (2), 229–239.

Cambouris, A.N., St. Luce, M., Zebarth, B.J., Ziadi, N., Grant, C.A., Perron, I., 2016a. Potato response to nitrogen sources and rates in an irrigated Sandy soil. Agron. J. 108 (1), 391–401.

Cambouris, A.N., Ziadi, N., Perron, I., Alotaibi, K.D., Luce, M.S., Tremblay, N., 2016b. Corn yield components response to nitrogen fertilizer as a function of soil texture. Can. J. Soil Sci. 96 (4), 386–399.

Campbell, C.A., Curtin, D., 2007. Mineralizable Nitrogen, Soil Sampling and Methods of Analysis, Second edition. CRC Press.

Clément, C.-C., Cambouris, A.N., Ziadi, N., Zebarth, B.J., Karam, A., 2019. Nitrogen source and rate effects on residual soil nitrate and overwinter NO3-N losses for irrigated potatoes on sandy soils. Can. J. Soil Sci. 100 (1), 44–57.

Clément, C.-C., Cambouris, A.N., Ziadi, N., Zebarth, B.J., Karam, A., 2021. Potato yield response and seasonal nitrate leaching as influenced by nitrogen management. Agronomy 11 (10), 2055. https://doi.org/10.3390/agronomy11102055.

Clément, C.C., Burton, D.L., Laurence, L., Fehr, P.A., Congreves, K.A., Dessureault-Rompré, J., 2025. Nitrogen mineralization in Canadian agricultural soils: a review of methods for quantifying soil nitrogen mineralization potential and estimating growing season nitrogen mineralization. Can. J. Soil Sci. 105, 1–19. https://doi.org/10.1139/cjss-2025-0019.

Curtin, D., Campbell, C.A., 2007. Mineralizable Nitrogen. In: Carter, M.R., Gregorich, E. G. (Eds.), Soil and Methods of Analysis, Second Edition. CRC Press, Boca Raton, pp. 599–606.

Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., Lawler, J.J., 2007. Random forests for classification in ecology. Ecology 88 (11), 2783–2792.

Das, K., Krzywinski, M., Altman, N., 2019. Quantile regression. Nat. Methods 16, 451–452.

de Castro Moreira da Silva, L., Amorim, R.S.S., Fernandes Filho, E.I., Bocuti, E.D., da Silva, D.D., 2023. Pedotransfer functions and machine learning: advancements and challenges in tropical soils. Geoderma Reg. 35, e00720.

Deragon, R., Saurette, D.D., Heung, B., Caron, J., 2023. Mapping the maximum peat thickness of cultivated organic soils in the southwest plain of Montreal. Can. J. Soil Sci. 103 (1), 103–120.

Dessureault-Rompré, J., Zebarth, B.J., Georgallas, A., Burton, D.L., Grant, C.A., Drury, C. F., 2010. Temperature dependence of soil nitrogen mineralization rate: comparison of mathematical models, reference temperatures and origin of the soils. Geoderma 157 (3), 97–108.

Dessureault-Rompré, J., Zebarth, B.J., Georgallas, A., Burton, D.L., Grant, C.A., 2011. A biophysical water function to predict the response of soil nitrogen mineralization to soil water content. Geoderma 167-168, 214–227.
Dessureault-Rompré, J., Zebarth, B.J., Burton, D.L., Gregorich, E.G., Goyer, C., Georgallas, A., Grant, C.A., 2013. Are soil mineralizable nitrogen pools replenished during the growing season in agricultural soils? Soil Sci. Soc. Am. J. 77 (2), 512–524.
Dessureault-Rompré, J., Zebarth, B.J., Burton, D.L., Georgallas, A., 2015. Predicting soil nitrogen supply from soil properties. Can. J. Soil Sci. 95 (1), 63–75.
Dessureault-Rompré, J., Zebarth, B.J., Burton, D.L., Grant, C.A., 2016. Depth distribution of mineralizable nitrogen pools in contrasting soils in a semi-arid climate. Can. J. Soil Sci. 96 (1), 1–11.
Dobarco, R.M., Cousin, I., Le Bas, C., Martin, M.P., 2019. Pedotransfer functions for predicting available water capacity in French soils, their applicability domain and associated uncertainty. Geoderma 336, 81–95.
Errebhi, M., Rosen, C.J., Gupta, S.C., Birong, D.E., 1998. Potato yield response and nitrate leaching as influenced by nitrogen management. Agron. J. 90 (1), 10–15.
Franzluebbers, A.J., Haney, R.L., Honeycutt, C.W., Schomberg, H.H., Hons, F.M., 2000. Flush of carbon dioxide following rewetting of dried soil relates to active organic pools. Soil Sci. Soc. Am. J. 64 (2), 613–623.
Gagnon, B., Ziadi, N., 2010. Grain corn and soil nitrogen responses to Sidedress nitrogen sources and applications. Agron. J. 102 (3), 1014–1022.
Gasser, M.O., Laverdière, M.R., Lagacé, R., Caron, J., 2002. Impact of potato-cereal rotations and slurry applications on nitrate leaching and nitrogen balance in sandy soils. Can. J. Soil Sci. 82 (4), 469–479.
Gasser, M.-O., Bossé, C., Clément, C.-C., Bernard, C., Grenon, L., Mathieu, J.-B., Tremblay, M.-E., 2023. Étude sur l'état de la santé des sols du Québec, rapport no. 1: État de santé des principales séries de sols cultivés. IRDA, Québec, Québec (CA), p. 1.
Glendining, M.J., Dailey, A.G., Powlson, D.S., Richter, G.M., Catt, J.A., Whitmore, A.P., 2011. Pedotransfer functions for estimating total soil nitrogen up to the global scale. Eur. J. Soil Sci. 62 (1), 13–22.
Haney, R.L., Brinton, W.H., Evans, E., 2008. Estimating soil carbon, nitrogen, and phosphorus mineralization from short-term carbon dioxide respiration. Commun. Soil Sci. Plant Anal. 39 (17–18), 2706–2720.
Hanks, R.J., Ashcroft, G.L., 1980. Water Quantities, Applied Soil Physics: Soil Water and Temperature Applications. Springer US, New York, NY, pp. 1–19.
Hao, X., Ball, B.C., Culley, J.L.B., Carter, M.R., Parkin, G.W., 2007. Soil density and porosity. In: Carter, M.R., Gregorich, E.G. (Eds.), Soil Sampling and Methods of Analysis, Second edition. CRC Press, Boca Raton, Fl, USA, pp. 743–760.
Hendershot, W.H., Lalande, H., Duquette, M., 2007. Soil Reaction and Exchangeable Acidity, Soil Sampling and Methods of Analysis, Second edition. CRC Press.
Heumann, S., Böttcher, J., Springob, G., 2003. Pedotransfer functions for the pool size of slowly mineralizable organic N in sandy arable soils. J. Plant Nutr. Soil Sci. 166 (3), 308–318.
Heumann, S., Ringe, H., Böttcher, J., 2011. Field-specific simulations of net N mineralization based on digitally available soil and weather data: II. Pedotransfer functions for the pool sizes. Nutr. Cycl. Agroecosyst. 91 (3), 339–350.
Heung, B., Ho, H.C., Zhang, J., Knudby, A., Bulmer, C.E., Schmidt, M.G., 2016. An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. Geoderma 265, 62–77.
Huber, D., Römheld, V., Weinmann, M., 2012. Chapter 10 - relationship between nutrition, plant diseases and pests. In: Marschner, P. (Ed.), Marschner's Mineral Nutrition of Higher Plants, Third edition. Academic Press, San Diego, pp. 283–298.
John, K., Saurette, D.D., Heung, B., 2025. The problematic case of data leakage: A case for leave-profile-out cro-validation in 3-dimensional digital soil mapping. Geoderma 455, 117223.
Kasraei, B., Heung, B., Saurette, D.D., Schmidt, M.G., Bulmer, C.E., Bethel, W., 2021. Quantile regression as a generic approach for estimating uncertainty of digital soil maps produced from machine-learning. Environ. Model Softw. 144, 105139.
Kasraei, B., Schmidt, M.G., Zhang, J., Bulmer, C.E., Filatow, D.S., Arbor, A., Pennell, T., Heung, B., 2024. A framework for optimizing environmental covariates to support model interpretability in digital soil mapping. Geoderma 445, 116873.
Kätterer, T., Andrén, O., 2009. Predicting daily soil temperature profiles in arable soils in cold temperate regions from air temperature and leaf area index. Acta Agric. Scandinavica Sect. B — Soil & Plant Sci. 59 (1), 77–86.
Kemmitt, S.J., Wright, D., Goulding, K.W.T., Jones, D.L., 2006. pH regulation of carbon and nitrogen dynamics in two agricultural soils. Soil Biol. Biochem. 38 (5), 898–911.
Koenker, R., 2025. quantreg:Quantile Regression. R Package Version 6.1. https://CRAN.R-project.org/package=quantreg.
Koenker, R., Bassett Jr., G., 1978. Regression Quantiles. Econometrica 46 (1), 33–50.
Kroetsch, D., Wang, C., 2007. Particle Size Distribution, Soil Sampling and Methods of Analysis, Second edition. CRC Press.
Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., 2020. Caret: Classification and Regression Training. R Package Version 6.0–86, p. 2020. http://CRAN.R-project.org/package=caret.
Laurance, L., Heung, B., Zhang, J., Pennell, T., Nyiraneza, Strom, H., Stiles, K., and Burton, D., 2024. Integrating multi-year crop inventories as a proxy for soil management within a digital soil mapping framework for predicting nitrogen indices. Geoderma 448, 116944.
Laurence, L., Heung, B., Strom, H., Stiles, K., Burton, D., 2023. Towards a cost-effective framework for estimating soil nitrogen pools using pedotransfer functions and machine learning. Geoderma 440, 116692.
Laurence, L., Burton, D.L., Heung, B., MacDonald, E., Nyiraneza, J., Stiles, K., 2025. Applying provincially derived pedotransfer function and digital soil map predictions of nitrogen indices at the field scale. Can. J. Soil Sci. 105, 1–19.
Lawrence, I., Lin, K., 1989. A concordance correlation coefficient to evaluate reproducibility. Biometrics 45 (1), 255–268.
Levy, J.J., O'Malley, A.J., 2020. Don't dismiss logistic regression: the case for sensible extraction of interactions in the era of machine learning. BMC Med. Res. Methodol. 20 (1), 171.
Li, H., Leung, K.S., Wong, M.H., Ballester, P.J., 2014. Substituting random forest for multiple linear regression improves binding affinity prediction of scoring functions: Cyscore as a case study. BMC Bioinforma. 15 (1), 291.
Li, L., Lu, J., Wang, S., Ma, Y., Wei, Q., Li, X., Cong, R., Ren, T., 2016. Methods for estimating leaf nitrogen concentration of winter oilseed rape (Brassica napus L.) using in situ leaf spectroscopy. Ind. Crop. Prod. 91, 194–204.
Linn, D.M., Doran, J.W., 1984. Effect of water-filled pore space on carbon dioxide and nitrous oxide production in tilled and nontilled soils. Soil Sci. Soc. Am. J. 48 (6), 1267–1272.
Manzoni, S., Taylor, P., Richter, A., Porporato, A., Ågren, G.I., 2012. Environmental and stoichiometric controls on microbial carbon-use efficiency in soils. New Phytol. 196 (1), 79–91.
Marquardt, D.W., 1970. Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. Technometrics 12 (3), 591–612.
Marshall, C.B., Burton, D.L., Heung, B., Lynch, D.H., 2021. Influence of cropping system and soil type on soil health. Can. J. Soil Sci. 101 (4), 626–640.
Masunga, R.H., Uzokwe, V.N., Mlay, P.D., Odeh, I., Singh, A., Buchan, D., De Neve, S., 2016. Nitrogen mineralization dynamics of different valuable organic amendments commonly used in agriculture. Appl. Soil Ecol. 101, 185–193.
Maynard, D.G., Kalra, Y.P., 1993. Nitrate and exchangeable ammonium nitrogen. In: Carter, M.R. (Ed.), Soil Sampling and Methods of Soil Analysis. Lewis Publishers, Boca Raton, FL, pp. 25–38.
McBratney, A.B., Minasny, B., Cattle, S.R., Vervoort, R.W., 2002. From pedotransfer functions to soil inference systems. Geoderma 109 (1), 41–73.
Menard, S., 1995. Applied Logistic Regression Analysis. Sage University Papers: Quantitative Applications in the Social Sciences. Sage Publications, Thousand Oaks, CA.
Mesele, S., Ajiboye, G., 2020. Pedo-transfer functions for predicting total soil nitrogen in different land use types under some tropical environments. Ghana J. Sci. 61 (2), 45–56.
Mozaffari, H., Moosavi, A.A., Ostovari, Y., Cornelis, W., 2022. Comparing visible-near-infrared spectroscopy with classical regression pedotransfer functions for predicting near-saturated and saturated hydraulic conductivity of calcareous soils. J. Hydrol. 613, 128412.
Mozaffari, H., Moosavi, A.A., Ostovari, Y., 2024. Feasibility of proximal sensing for predicting soil loss tolerance. Catena 247, 108503. https://doi.org/10.1016/j.catena.2024.108503.
Nkonge, C., Balance, G.M., 1982. A sensitive colorimetric procedure for nitrogen in micro Kjeldahl digest. J. Agric. Food Chem. 30, 416–420. https://doi.org/10.1021/jf00111a002.
Nyiraneza, J., N'Dayegamiye, A., Gasser, M.O., Giroux, M., Grenier, M., Landry, C., Guertin, S., 2010. Soil and crop parameters related to corn nitrogen response in eastern Canada. Agron. J. 102 (5), 1478–1490.
Nyiraneza, J., Cambouris, A.N., Ziadi, N., Tremblay, N., Nolin, M.C., 2012. Spring wheat yield and quality related to soil texture and nitrogen fertilization. Agron. J. 104 (3), 589–599.
Padarian, J., Minasny, B., McBratney, A.B., 2020. Machine learning and soil sciences: A review aided by machine learning tools. Soil 6, 35–52. https://doi.org/10.5194/soil-6-35-2020.
Parfitt, R.L., Yeates, G.W., Ross, D.J., Mackay, A.D., Budding, P.J., 2005. Relationships between soil biota, nitrogen and phosphorus availability, and pasture growth under organic and conventional management. Appl. Soil Ecol. 28 (1), 1–13.
Paul, S.S., Heung, B., Lynch, D.H., 2023. Modeling of total and active organic carbon dynamics in agricultural soil using digital soil mapping: a case study from Central Nova Scotia. Can. J. Soil Sci. 103 (1), 64–80.
Pereira, A.R., Pruitt, W.O., 2004. Adaptation of the Thornthwaite scheme for estimating daily reference evapotranspiration. Agric. Water Manag. 66 (3), 251–257.
Picone, L.I., Cabrera, M.L., Franzluebbers, A.J., 2002. A rapid method to estimate potentially mineralizable nitrogen in soil. Soil Sci. Soc. Am. J. 66 (6), 1843–1847.
Ramcharan, A., Hengl, T., Beaudette, D., Wills, S., 2017. A soil bulk density Pedotransfer function based on machine learning: A case study with the NCSS soil characterization database. Soil Sci. Soc. Am. J. 81 (6), 1279–1287.
Rashidi, M., Seilsepour, M., 2009. Modeling of soil total nitrogen based on soil organic carbon. ARPN J. Agric. Biol. Sci. 4 (2), 1–5.
Rasiah, V., 1995. Comparison of pedotransfer functions to predict nitrogen-mineralization parameters of one-and two-pool models. Commun. Soil Sci. Plant Anal. 26 (11−12), 1873–1884.
Schimel, J.P., Bennett, J., 2004. Nitrogen mineralization: challenges of a changing paradigm. Ecology 85 (3), 591–602.
Schimel, J.P., Schaeffer, S.M., 2012. Microbial control over carbon cycling in soil. Front. Microbiol. 3, 348.
Scott, N.A., Parfitt, R.L., Ross, D.J., Salt, G.J., 1998. Carbon and nitrogen transformations in New Zealand plantation forest soils from sites with different N status. Can. J. For. Res. 28 (7), 967–976.
Sharifi, M., Zebarth, B.J., Burton, D.L., Grant, C.A., Cooper, J.M., 2007. Evaluation of some indices of potentially mineralizable nitrogen in soil. Soil Sci. Soc. Am. J. 71 (4), 1233–1239.
Shedayi, A.A., Xu, M., Naseer, I., Khan, B., 2016. Altitudinal gradients of soil and vegetation carbon and nitrogen in a high altitude nature reserve of Karakoram ranges. SpringerPlus 5 (1), 320.
Skjemstad, J.O., Baldock, J.A., 2007. Total and Organic Carbon, Soil Sampling and Methods of Analysis, Second edition. CRC Press.

Smith, P.F., Ganesh, S., Liu, P., 2013. A comparison of random forest regression and multiple linear regression for prediction in neuroscience. J. Neurosci. Methods 220 (1), 85–91.
Soil Classification Working Group, 1998. The Canadian System of Soil Classification, 3rd ed. Research Branch, Agriculture and Agri-Food Canada, Ottawa, ON.
St. Luce, M., Whalen, J.K., Ziadi, N., Zebarth, B.J., 2011. Chapter two - nitrogen dynamics and indices to predict soil nitrogen supply in humid temperate soils. In: Sparks, D.L. (Ed.), Advances in Agronomy. Academic Press, pp. 55–102.
Staffa, S.J., Kohane, D.S., Zurakowski, D., 2019. Quantile regression and its applications: a primer for anesthesiologists. Anesth. Analg. 128, 820–830.
Stanford, G., Smith, S.J., 1972. Nitrogen mineralization potentials of soils. Soil Sci. Soc. Am. J. 36 (3), 465–472.
Statistics Canada, 2022. Land use, Census of Agriculture. 2021. Table 32-10-0249-01. Retrieved from. https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=3210024901.
Sun, R., Yuan, H., Liu, X., 2017. Effect of heteroscedasticity treatment in residual error models on model calibration and prediction uncertainty estimation. J. Hydrol. 554, 680–692.
Tyralis, H., Papacharalampous, G., Langousis, A., 2019. A brief review of random forests for water scientists and practitioners and their recent history in water resources. Water 11 (5), 910.
Van Looy, K., Bouma, J., Herbst, M., Koestel, J., Minasny, B., Mishra, U., Montzka, C., Nemes, A., Pachepsky, Y.A., Padarian, J., Schaap, M.G., Tóth, B., Verhoef, A., Vanderborght, J., van der Ploeg, M.J., Weihermüller, L., Zacharias, S., Zhang, Y., Vereecken, H., 2017. Pedotransfer functions in earth system science: challenges and perspectives. Rev. Geophys. 55 (4), 1199–1256.
Vaysse, K., Lagacherie, P., 2017. Using quantile regression forest to estimate uncertainty of digital soil mapping products. Geoderma 291, 55–64. https://doi.org/10.1016/j.geoderma.2016.12.017.
Weil, R.R., Islam, K.R., Stine, M.A., Samson-Liebig, S.E., Gruver, J.B., 2003. Estimating active carbon for soil quality assessment: a simplified method for laboratory and field use. Am. J. Altern. Agric. 18 (1), 3–17.
Xiao, Y., Xue, J., Zhang, X., Wang, N., Hong, Y., Jiang, Y., Zhou, Y., Teng, H., Hu, B., Lugato, E., Richer-de-Forges, A.C., Arrouays, D., Shi, Z., Chen, S., 2022. Improving pedotransfer functions for predicting soil mineral associated organic carbon by ensemble machine learning. Geoderma 428, 116208.
Zebarth, B.J., Rosen, C.J., 2007. Research perspective on nitrogen bmp development for potato. Am. J. Potato Res. 84 (1), 3–18.
Zvomuya, F., Rosen, C.J., Russelle, M.P., Gupta, S.C., 2003. Nitrate leaching and nitrogen recovery following application of polyolefin-coated urea to potato. J. Environ. Qual. 32 (2), 480–489.