Soil Security 16 (2024) 100161

Contents lists available at ScienceDirect

Soil Security

journal homepage: www.sciencedirect.com/journal/soil-security

Rapid soil attribute evaluation for soil security assessments in data-poor environments in the Pacific region

J.P. Moloney [a,*], Y. Ma [b], U. Stockmann [a], V.T. Manu [c], V. Minoneti [c], S.T. Hui [c], S. M. Halavatau [d], S. Patolo [e], T. Tukia [e], S. Foliaki [f], T. Carter [a], B.C.T. Macdonald [a], J. Barringer [b], P. Roudier [b]

[a] CSIRO Agriculture and Food, Australia
[b] Manaaki Whenua Landcare Research, New Zealand
[c] Ministry of Agriculture, Forests and Food, Kingdom of Tonga
[d] Independant consultant to Manaaki Whenua Landcare Research, New Zealand
[e] Mainstreaming of Rural Development Innovation Tonga Trust, Kingdom of Tonga
[f] Pacific Seeds for Life, Kingdom of Tonga

ARTICLE INFO

Keywords:
The Kingdom of Tonga
PICTs
Vis-NIR spectroscopy
Memory-based learning
Soil Organic Carbon
Soil pH
Total Nitrogen

ABSTRACT

Many global environments face increasing pressures on soil resources, and effective, scalable methods for assessment of soil condition and capital are essential to respond to tangible soil threats. This situation is common across Pacific Island Countries and Territories (PICTs), where high throughput soil analysis laboratories are limited, and issues such as soil organic carbon decline, acidification and fertility declines are present. Soil spectral inference presents an opportunity in such regions to provide rapid insights into soil capital and condition, though the need for robust calibration libraries remains a limiting factor. This work investigates the utility of a regionally appropriate spectral library, the New Zealand Soil Spectral Library (NZSSL) to support the development of soil spectral inference in data-poor environments, such as PICTs, through a case study on the island of Tongatapu in The Kingdom of Tonga. We contrast the performance of existing partial least squares regression (PLSR) models developed for New Zealand soils on soils from Tongatapu and explore the opportunities for enhancement of predictions formed through memory-based learning (MBL) supplemented with local data. Our work shows the potential for cost-effective and timely soil monitoring through soil spectral inference in PICTs. The work further underscores the importance of regional cooperation and data-sharing for addressing soil security.

Introduction

In a global context of increased pressures on soil resources, sustaining the ecosystem services provided by soils (such as the mitigation of climate risk, the securing of water resources, the protection of biodiversity and human health, and the support of the production of food and delivery) is a central part of the soil security challenge (Evangelista et al., 2023; McBratney et al., 2014). It is of particular importance within Pacific Island Countries and Territories (PICTs), which face a unique combination of factors, including changing precipitation patterns, sea level rise, and rapid shifts towards intensive commercial farming systems (Antille et al., 2022). These challenges exacerbate existing threats to soils across PICTs including declining soil fertility, seen in loss of soil organic carbon (SOC), declining soil nutrients, increasing soil acidity, salinisation, soil sodicity and erosion (McKenzie et al., 2015; Naidu et al., 1991).

Tongatapu, the largest island within The Kingdom of Tonga is not immune to these common threats across PICTs. The intensification of agricultural production systems and tillage regimes, particularly for squash and taro crops, has led to declining SOC stocks across Tongatapu (Manu et al., 2014). This SOC decline, paired with other issues common across the PICTs, such as nutrient mismanagement, and a general trend towards acidic soils due to high humidity and weathering of soil material, compromises soil security (McKenzie et al., 2015). Furthermore, soil wet chemistry laboratories across the PICTs are limited, with The Kingdom of Tonga currently lacking a functioning soil chemistry

* Corresponding author.
E-mail address: james.moloney@csiro.au (J.P. Moloney).

https://doi.org/10.1016/j.soisec.2024.100161
Received 22 May 2024; Received in revised form 18 July 2024; Accepted 21 July 2024
Available online 7 August 2024

laboratory. Thus, intervention into declining soil security across Tongatapu will benefit greatly from scalable methods and analysis techniques with direct involvement by stakeholders.

Visible Near-Infrared (Vis-NIR) proximal sensors are a widely adopted tool for the inference and prediction of several soil properties. Vis-NIR reflectance spectroscopy has proven to be valuable in accurately predicting SOC and total nitrogen (TN) contents of soils and presents a useful tool for identifying trends in soil pH (Nocita et al., 2015; Stenberg et al., 2010). However, Vis-NIR proximal sensing relies on the availability of a spectral library to establish the relationship between spectra and soil analyte values, which must be representative of the soils on which inference is to be performed. The construction of new and comprehensive spectral libraries within PICTs would be a slow and costly investment, requiring a large amount of samples to be scanned and analysed in the laboratory. As mentioned, in many PICTs, high throughput laboratories are not established, further delaying the construction of new soil spectral libraries due to long sample processing times or requiring the expatriation of soil samples to laboratories in other countries, with all the associated freight and quarantine induced time lags, and additional expenses incurred.

Existing soil spectral libraries may be a useful tool for soil spectral inference in PICTs, but direct usage of existing soil spectral resources in new environments has historically proved difficult. Regression or decision tree based models trained on large datasets or local subsets from these libraries often struggle to maintain prediction accuracy in new areas due to limitations in representing soil attributes and spectral data (Grunwald et al., 2018; Wetterlind and Stenberg, 2010). These issues are exacerbated in PICTs due to their small geographic extents, spatial discontinuity of land area, and diversity of soil forming factors making each new island a unique prediction challenge. More recently, deep learning techniques, such as convolutional neural networks have shown potential in predicting soil spectra in new domains (Padarian et al., 2019b; Wang et al., 2019). And a focus on transfer learning approaches, wherein a neural network trained on a large soil spectral library is adapted to local situation through re training of the final layers, has shown success in predicting soil properties at country or regional scales (Padarian et al., 2019a; Viscarra Rossel et al., 2024). Such methods may hold promise for use in PICTs as locally collected datasets grow but remain impractical for implementation currently without well distributed and representative local soil libraries.

An alternative, more accessible approach for PICTs may exist in using library constraining approaches, such as memory-based learning. Rather than producing a generalised model for all unknown spectra, memory based learning produces individual weighted averaged partial least squares regression (WAPLSR) models for each unknown sample, consisting of only the most similar samples from a reference library in the spectral domain (Ramirez-Lopez et al., 2013). Thus making any spectral inference more representative of the relationship between soil spectra and soil analytical values seen in only the most similar samples.

As a large spectral library within the region, with existing Partial Least Squares Regression (PLSR) models of many soil properties, the New Zealand Soil Spectral Library (NZSSL, Ma et al., 2023) presents a promising candidate resource for predicting soil properties for monitoring soil security within the soils of Tongatapu. Beyond the benefits of regional proximity, the soils of New Zealand hold many potential sites with properties like those found within Tongatapu. Developed on a raised atoll, Tongatapu's soils have been formed from two layers of fine-grained andesitic volcanic ash, deposited 5000 and 20,000 years ago, respectively, sitting atop a raised coral limestone (Cowie, 1980; Gibbs, 1976). This depositional environment may hold analogues in soils classified within the Allophanic and Recent orders of the New Zealand Soil Classification (Hewitt et al., 2021) predominantly located in the North Island (Taranaki, Waikato, Bay of Plenty regions in particular) (Parfitt, 1990). The presence of these potentially analogous sites makes the NZSSL a more attractive option for use in Tongatapu than other large soil spectral libraries in the region, such as the CSIRO Australian Vis-NIR Soil Spectral Library (Moloney et al., 2023), or globally available libraries (Shepherd et al., 2022).

This work aims to assess the applicability of the NZSSL (Ma et al., 2023) for predicting three soil properties for monitoring soil security within Tongatapu, namely SOC, total nitrogen, and soil pH. To this end, this work will:

- Compare soil spectra collected from two surveys of Tongatapu island conducted in 2018 and 2022, representing periods pre and post the 2022 Hunga Tonga-Hunga Ha'apai volcanic eruption and tsunami to those soil spectra within the NZSSL,
- Evaluate the performance of the national scale NZSSL Vis-NIR PLSR models for pH, total nitrogen, and SOC on the soils of Tongatapu,
- Highlight the importance of robust spectral libraries in the production of PLSR models for PICTs, through building models of pH, total nitrogen and SOC using subsets of varying size of the spectra collected in the 2018 and 2022 surveys of Tongatapu,
- Investigate methods for the use of the most appropriate spectra within the NZSSL for spectral inference of Tongatapu soils, through the use of memory-based learning, and
- Evaluate the relative improvement in memory-based learning predictions when supplementing the NZSSL with local data from Tongatapu.

Materials and methods

Datasets

Vis-NIR soil spectral library of the Kingdom of Tonga – Tongatapu

Soil survey and soil data collection. A total of 194 soil samples were collected across Tongatapu island, in the Kingdom of Tonga, during two soil surveys: the first survey was conducted in 2018 to assess the baseline state of the soils across the island, and the second survey conducted in 2022, after the Hunga Tonga-Hunga Ha'apai volcanic eruption and subsequent tsunami, to assess the impact of these events on soils. 23 soil sampling locations across the island were chosen at agricultural sites and sites deemed 'native' or uncultivated with existing soil legacy information (Cowie et al., 1991; Potter, 1986), allowing for comparison of the impact of management practices on the soil's status (Stockmann et al., under revision; Stockmann, U., Farrell, M., Carter, T., Krawitz, S., Wong, V.N.L., Minoneti, V., Finau, D., Sharma, A., Barringer, J., Manu, V.T., Karunaratne, S., Macdonald, B.C.T. "State and trend of surface soil organic carbon stocks and associated nutrients in Tongatapu, Kingdom of Tonga". Soil Research). In 2022, additional sampling was conducted in the western region of the island, examining fifteen new sites which experienced the most impact from the January 2022 tsunami wave (Fig. 1).

For the 2018 and 2022 surveys, soil core samples were taken from the centre of 1 ha plots at each site using a hand auger, representative of the top- and subsoil (i.e., 0–15, 15–30 and 30–60 cm). Subsequently, the soil samples were shipped to the CSIRO Analytical Services Unit Quarantine Laboratory in Adelaide, South Australia. Soils were processed to air-dry 40 °C condition, ground, and sieved to obtain fine earth material with a diameter of <2 mm. A series of laboratory analyses were performed to acquire various soil property information, including total soil organic carbon % by dry combustion (Leco) using the method 6B2 from Rayment & Lyons (2011), soil total N % (method 7A8), and pH in 1:5 soil to water (method 4A1) and $CaCl_2$ (method 4B4) solution ratio. In addition, the 2018 soil samples were scanned with an ASD FieldSpec 3 (Malvern Panalytical, Worcestershire, UK), while the 2022 soil samples were scanned using an ASD Terraspec 4 (Malvern Panalytical, Worcestershire, UK) Vis-NIR spectrometer, with an ASD high intensity contact probe.

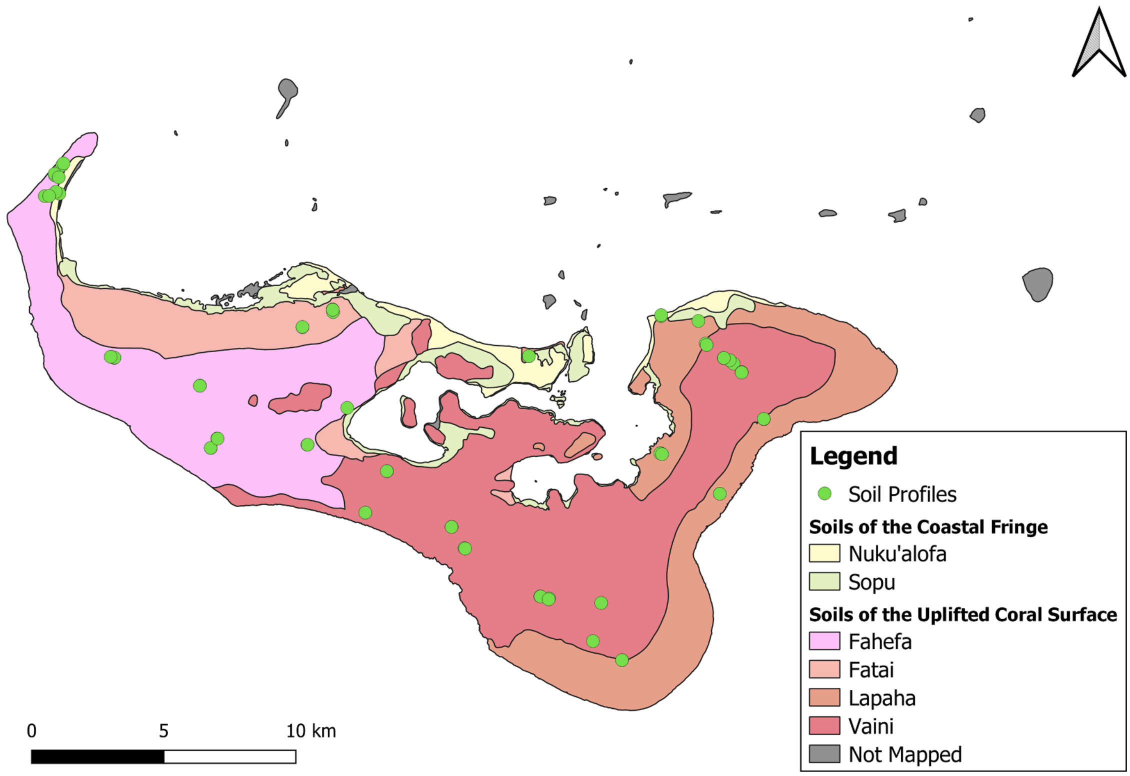

Fig. 1. Map of the soil types of the Island of Tongatapu, displaying sampling locations from the 2018 and 2022 Surveys.

Soils of Tongatapu. Dominant soils on Tongatapu are allophanic soils, originating from volcanic ash, with younger reddish-brown tephra overlying older browner and finer textured tephra deposits. These generally fertile soils are well drained with clayey textures and deep dark-coloured A horizons. Smaller areas also have occurrence of soils formed from coastal coral sands. These soils tend to be nutrient deficient with high pH and elevated calcium content, which makes them less suitable for cropping (Cowie et al., 1991).

Vis-NIR soil spectral library of New Zealand. The New Zealand Vis-NIR soil spectral library (NZSSL) is comprised of 11,246 paired Vis-NIR spectra and corresponding soil analytical values from a diverse range of soils of New Zealand, collected since 1971. The collection includes observations on soils classified as Allophanic and Recent orders in the NZSC, corresponding to Andisols and Entisols or Inceptisols of the World Reference Base classification system, respectively (Hewitt et al., 2021). Because of the similarity in parent material type and age, these soil orders are the most likely to be similar to the soils of Tongatapu. All samples were scanned in air-dry condition, following grinding and sieving for fine earth material (<2 mm). Samples were then scanned with an ASD FieldSpec 3 (Malvern Panalytical, Cambridge, UK) with an ASD high intensity contact probe, which covers the same spectral range of the instrument used to scan the Tongan soils. Within the spectral library, a total of 10,164 total carbon observations and spectra are recorded, alongside 10,110 total Nitrogen observations, and 2833 observations of pH in 1:5 soil:water solution. A full description of the library is available in Ma et al. (2023).

Pre-Processing. Spectra from both the NZSSL and those collected from Tongatapu soils underwent the same pre-processing procedures used to calibrate the national PLSR models from the Vis-NIR data of the NZSL (Ma et al., 2023). Using the spectacles package (Roudier, 2023) within the R statistical computing language (R Core Team, 2020), spectra were first converted to absorbance, trimmed to wavelengths between 450 and 2450 nm, and treated with a Savitzky Golay filter with a window size of 9, using a second order polynomial, to remove high frequency noise, and then treated with a standard normal variate (SNV) transformation.

To compare the spectra of the NZSSL with those recorded from the soils of Tongatapu, spectra from both collections were projected to the principal component space of the NZSSL. Pairwise comparisons of the distribution of spectra from both collections across the first three principal components were plotted, as well as frequency distributions and correlations.

Calibration data split. In order to establish a realistic assessment of model quality under different modelling techniques and data scenarios, the Tongan spectra were split into calibration and validation sets through selection of the spectral space. A subset (40 %) of the spectra was selected using conditioned Latin hypercube sampling (Minasny and McBratney, 2006) using the clhs package in R (Roudier, 2011) of the first 5 principal components of the pre-processed spectra, and held out of the modelling workflow as a validation set. All co-located spectra to those selected in this manner were also added to the test group, to improve the independence of the two populations. All goodness of fit metrics under the different chemometric modelling approaches explored are reported on this validation set, to ensure a consistent point of comparison. We reported five goodness of fit metrics against our validation set. The coefficient of determination (R^2) and Lin's concordance correlation coefficient (LCCC) to identify the agreement between our observed and predicted soil property values. The root mean squared error (RMSE) and model bias (Bias) to assess the size and trends in our residuals, and the ratio of prediction to the interquartile range (RPIQ) was used as a general modelling performance indicator which is less impacted by skewed datasets, as is common in soil properties.

Modelling

Existing New Zealand NZSSL models

The current operational models from the NZSSL were applied to the pre-processed spectra from Tongatapu for our three soil properties of interest. These models, outlined further in section 2.4 of Ma et al. (2023), are partial least squares regression (PLSR) models, where the observed spectra are transformed into a latent variable space that both reduces the dimensionality of the predictor variables (spectra) for use in modelling, and maximises the covariance between the soil properties and predictor variables. The number of latent variables used in the final model was optimised through 50 repeated 10-fold cross validation runs. A summary of their performance on New Zealand soils is shown in Table 1. As inorganic carbon content within the soils of New Zealand is generally considered to be negligible (Metson et al., 1979; Sparling and Schipper, 2002), we consider the national NZSSL models for soil total carbon, to act as an appropriate proxy for soil organic carbon in this work.

Local Tongatapu PLSR models

To highlight the importance of robust soil spectral libraries in the construction of PLSR models for soil spectral inference, and emulate model construction under data poor conditions, local PLSR models were created from the Tongatapu samples for each of the three soil properties of interest. Twenty subsets of data were evaluated from amongst those Tongatapu samples within the calibration set, ranging from 5 to 100 samples. The PLSR models were constructed using each of these sets with the caret package (Kuhn, 2022) in the R statistical programming language. LCCC, Bias and RMSE values for each of these models was reported and compared to those of the existing NZSSL PLSR models when applied to Tongatapu spectra.

Memory-based learning

Memory-based learning was used to produce predictions for Tongatapu soils using the most spectrally similar candidates within the NZSSL for soil organic carbon, total nitrogen, and pH in a 1:5 soil to water suspension. Memory-based learning predictions were first generated on the test set of Tongatapu data, using exclusively spectra and analytical values from the NZSSL. To avoid the influence of organic horizons with large organic carbon values performing poorly on mineral soils, all horizons with TOC greater than 18 % defined as peaty horizons in the New Zealand Soils Classification (Hewitt, 2010) were eliminated.

Using the resemble package (Ramirez-Lopez et al., 2016) in the R statistical language, for each of the three indicators of soil security assessed, unique weighted average partial least squares regression (WAPLSR) models (Cleveland and Devlin, 1988; Shenk et al., 1997). These WAPLS predictions were produced for each Tongatapu sample within the test set, by constructing a weighted PLS calibration based on a limited number of nearest neighbours in principal component space to our unknown sample from the NZSSL. Weights were determined by a tri-cubic function of the Mahalanobis distance to our unknown sample in the principal component space of the spectra. We contrasted different ranges of nearest neighbours from which to construct the WAPLS models ranging from a population of 50 to 300 samples in increments of fifty.

Once a value of nearest neighbours was identified which produced predictions with a good balance of goodness of fit metrics, we evaluated the relative improvements in performance offered by supplementing the population from which the nearest neighbours could be drawn with additional spectra from Tongatapu. To this end the twenty representative subsets of Tongatapu soils used for local PLSR model calibration, of sizes ranging from 5 to 100 samples, were added to the NZSSL, and the goodness of fit metrics of memory-based learning predictions were recorded using the previously identified optimal number of nearest neighbours.

Results

Soil analytical values

Observed soil organic carbon, total nitrogen, and pH values within Tongatapu were quite varied, representing a variety of soil conditions, with different degrees of consistency to the distributions of soil properties within the NZSSL (Table 2, Fig. 2). Organic carbon across the Tongatapu soils ranged from 0.58 % to 7.19 %, with a median of 2.79 %. Comparatively, values within the NZSSL held a much larger range, with some organic soil horizons reaching as high as 62.2 %, though the median value of 2.14 % was similar to the Tongatapu soils. Total nitrogen in the Tongatapu soils was again centred similarly to the distribution seen in the NZSSL, at 0.23 % and 0.18 % respectively, though the upper range of observations within the New Zealand data were higher than that seen in Tongatapu. The pH range in Tongatapu was in the higher range of the samples within the NZSSL, with Tongatapu soils ranging from a pH of 5.57 to 9.6. The lower bound of the Tongatapu pH values was close to the median of those within the NZSSL, at 5.7, with no overlap between the central two quartiles of pH in each soil population (Fig. 2). The agreement between the distributions of total organic carbon and total nitrogen across the Tongatapu and NZSSL soils is a positive indicator towards the utility of shared soil information benefiting soil spectral inference, though the difference in the centres of pH between the two regions is not as promising.

Spectral space comparison Tongatapu - NZSSL

Initial comparison of the spectra collected from the two Tongatapu surveys indicated that they were well positioned within the spectral space of the NZSSL (Fig. 3). The Tongatapu samples were well centred within the first two principal components of the NZSSL, and largely overlap with the third. This supports our hypothesis that there are indeed similar soils to the Tongatapu soils within the NZSSL. Within the Tongan datasets, each of the first three principal component projections show a decent amount of correlation, with PC1 and PC3 holding a correlation of 0.592 in contrast to their orthogonal, uncorrelated nature in the New Zealand dataset from which they were generated.

Table 2
Summary statistics of three soil properties from the sampled soils of Tongatapu contrasted against those within the NZSSL.

Library	Soil Property	Mean	Median	Range	Min	Max
Tongatapu	Total Organic Carbon (%)	2.76	2.69	6.61	0.58	7.19
NZSSL	Total Organic Carbon (%)	4.25	2.14	62.14	0.06	62.20
Tongatapu	Total Nitrogen (%)	0.24	0.23	0.65	0.04	0.69
NZSSL	Total Nitrogen (%)	0.26	0.18	3.09	0.01	3.10
Tongatapu	pH 1:5 Soil in Water	7.04	6.60	4.03	5.57	9.60
NZSSL	pH 1:5 Soil in Water	5.74	5.70	6.50	2.80	9.30

Table 1
A summary table displaying the validation goodness of fit measures of the NZSSL national partial least squares regression models for our three soil properties, adapted from Ma et al. (2023, Table 4).

Indicator	Transformation	Latent Variables	R^2	LCCC	RMSE	Bias	RPIQ
Total C (%)	Square Root	29	0.92	0.96	2.2	−0.05	1.44
Total N (%)	Square Root	28	0.85	0.91	0.12	−0.01	2.09
pH	–	28	0.58	0.74	0.49	−0.00	1.80

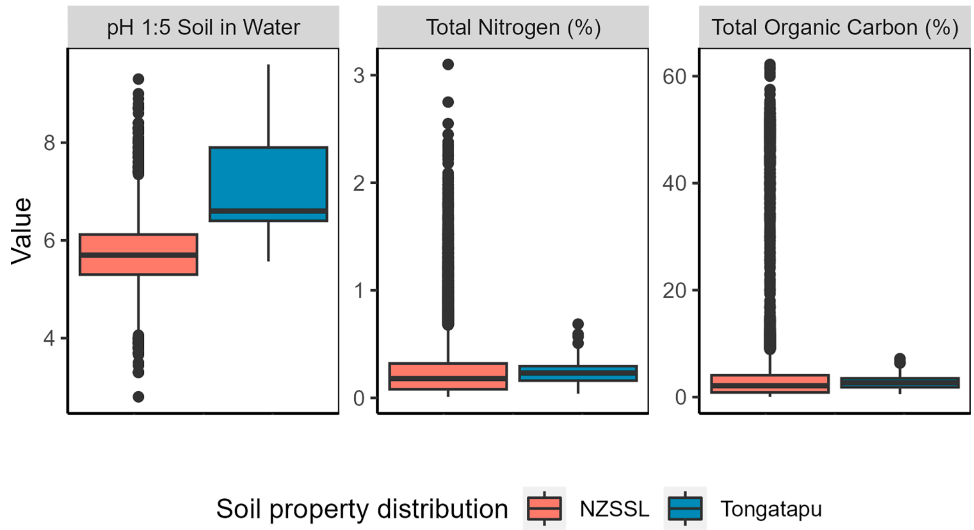

Fig. 2. Boxplots showing the distribution of pH, total nitrogen, and total carbon in the sampled soils of tongatapu contrasted against the ditributions within the NZSSL.

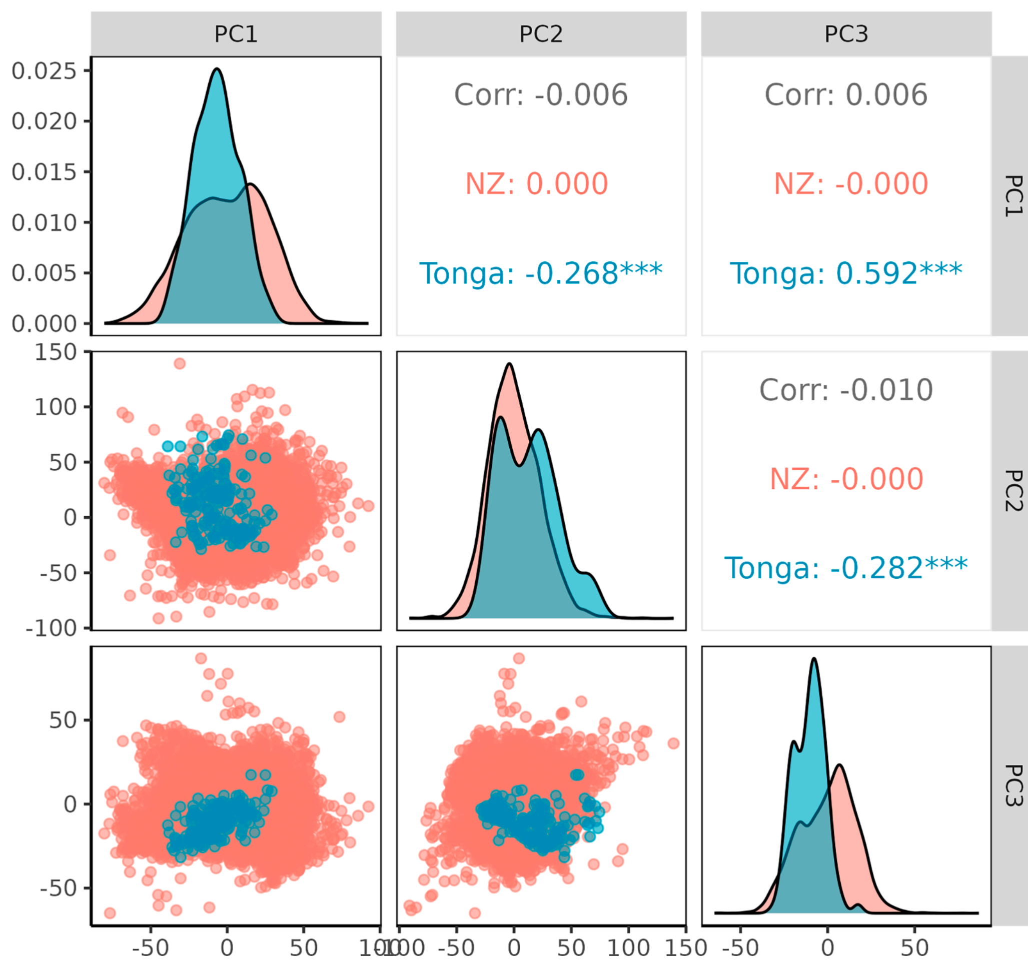

Fig. 3. Pairwise plots of the distribution of samples from the New Zealand (red) and Tongan (blue) soil spectra across the first three principal components of the NZSSL.

Direct application of NZSSL models to Tongatapu spectra

When the NZSSL PLSR models for total carbon, total nitrogen, and pH 1:5 soil in water (Ma et al., 2023) were applied directly to the spectra of the Tongatapu validation set, model quality metrics indicated a low level of performance (Fig. 4, Table 3). The total carbon model, here used as a proxy for organic carbon due to a lack of carbonate presence in New Zealand soils, performed the best, with an LCCC of 0.44, showing a

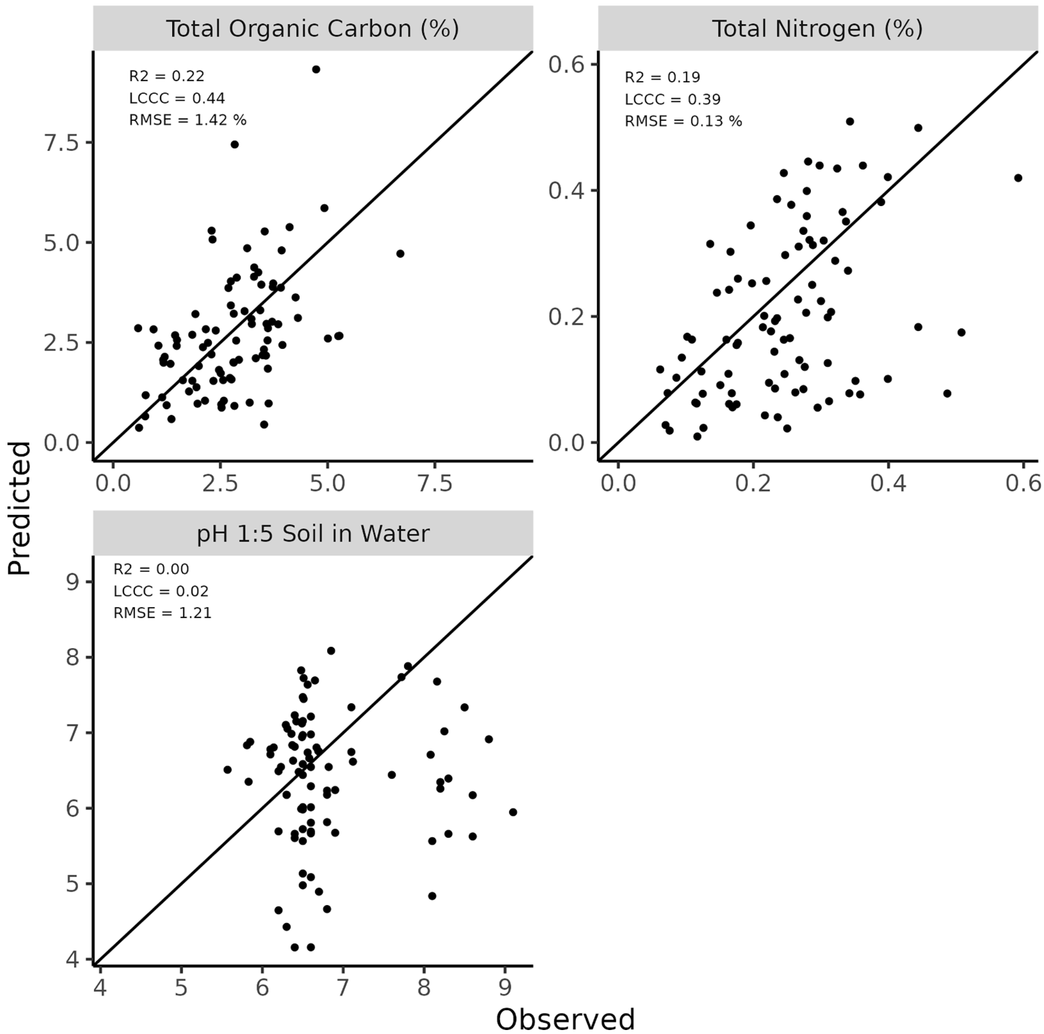

Fig. 4. The observed values of Total Organic Carbon, Total Nitrogen and pH for the test set of Tongatapu soils plotted against the predicted values from the NZSSL PLSR models (Ma et al., 2023).

Table 3
Model quality of fit metrics for the NZSSL PLSR calibration models applied to the Tongatapu validation set.

Property	R^2	LCCC	RMSE	Bias	RPIQ
Total Organic Carbon	0.22	0.44	1.42%	−0.06%	1.10
Total Nitrogen	0.19	0.39	0.13%	−0.04%	1.02
pH 1:5 Soil in Water	0.00	0.02	1.21	−0.42	0.37

moderate fit of the observed and predicted carbon values, but with a few large outliers of overprediction. The RMSE of 1.42 % demonstrates a large error, influenced by those overpredicted values, however the Bias demonstrates a minor, general underprediction at −0.06 %. Total nitrogen showed a similar quality of prediction, with an LCCC of 0.39, though the Bias of −0.04 % should be viewed more critically due to the lower concentration of Nitrogen within the Tongatapu soils. In contrast, predictions of pH 1:5 soil in water were extremely poor, with an R^2 approaching 0, and a LCCC of just 0.02 highlighting little to no trend between observed and predicted values.

Local Tongatapu PLSR models

Local PLSR models generated on subsets of the calibration samples from Tongatapu demonstrated enhanced model quality with increasing amounts of calibration data (Fig. 5). Local SOC models exceeded the prediction qualities of the exogenic NZSSL PLSR for both LCCC and RMSE at around 25 calibration samples, and stabilised at an LCCC between 0.72 and 0.8, and an RMSE of <0.8 %. Similarly, the performance of total nitrogen PLSR models from Tongatapu soils also exceeded the NZSSL PLSR model at 25 samples, with an LCCC of 0.65, and an RMSE of 0.07 %. results for total nitrogen were however less consistent, with LCCC values returning to values below 0.6 until 50 Tongatapu calibration samples were used, eventually peaking at 0.76 with 85 samples. Conversely pH models with Tongatapu spectra exceeded NZSSL PLSR models with as few as five spectra, eventually plateauing with LCC values around 0.72, with 14 or more samples, and RMSE values nearing 0.5. In our investigation, local Tongatapu PLSR models generally exceeded the performance of the large, exogeneous models calibrated on spectra from New Zealand, with calibration sets of 25 samples or more. Model qualities from Tongatapu PLSR models with small numbers of calibration samples were largely inconsistent, until a point of plateauing around a calibration set size of 50–60 samples where model metrics stabilised.

Memory-based learning

Memory-based learning predictions using exclusively soils of the NZSSL as a reference library for inference on soils of Tongatapu generally produced predictions of generally poor quality (Fig. 6). For soil total organic carbon, all nearest neighbour numbers assessed in the WAPLSR

Fig. 5. RMSE and LCCC trends under increasing calibration sample size for Tongatapu PLSR models for total organic carbon, total nitrogen, and pH, contrasted against the performance of a PLSR model trained exclusively on soils from the New Zealand Soil Spectral Library.

models produced predictions with both lower LCCC values, and higher RMSE values than the NZSSL PLSR models, with a general trend of decreasing model metric quality as nearest neighbours increased. In predictions of total nitrogen, model metrics were stable across different numbers of nearest neighbours, with LCCC values ranging from 0.26 to 0.28, again lower than the NZSSL PLSR model of the same property. RMSE values for total nitrogen was comparable to predictions from the NZSSL PLSR total nitrogen model, at 0.13 %. Conversely, the MBL predictions of pH 1:5 soil in water produced predictions of greater quality than the NZSSL PLSR model, though still poor. LCCC reached a maximum and RMSE a minimum at 250 nearest neighbours, at 0.2, and 0.91 respectively. In all, predictions of soils of Tongatapu using the NZSSL as a reference library for memory-based learning demonstrated no substantial improvement over the use of PLSR models generated on the NZSSL.

We continued to investigate the influence of introducing up to 100 Tongatapu soils to the reference library for WAPLSR construction, using 150 nearest neighbours for SOC and total nitrogen, so as to not represent purely local models when Tongatapu data were the closest in Mahalanobis distance, and 250 for pH, where we observed the maxima in LCCC and minima in RMSE with New Zealand soils. Supplementing the NZSSL reference library with Tongatapu soil samples significantly enhanced the memory-based learning prediction accuracy for all soil properties investigated. With as few as ten supplementary samples, all predictions showed LCCC values greater than those in the NZSSL PLSR models, and RMSE values lower across all three soil properties. For SOC, prediction quality increased steadily with more Tongatapu soils, until 45 supplemental samples were added to the reference library, where LCCC stabilised above 0.75, and RMSE below 0.8 %. A similar pattern was observed in total nitrogen, where quality stabilised at 55 Tongatapu samples in the reference library, with LCCC values reaching only 0.6, and RMSE values of 0.09 or lower. for pH, all predictions with 35 or more supplemental samples held an RMSE of 0.61 or lower and LCCC values remained above 0.6 once there were 30 or more Tongatapu samples included in the reference library. As with Local PLSR models, adding Tongatapu soil samples to the NZSSL reference library significantly improved MBL prediction accuracy for all soil properties when compared to the exogenous NZSSL PLSR models.

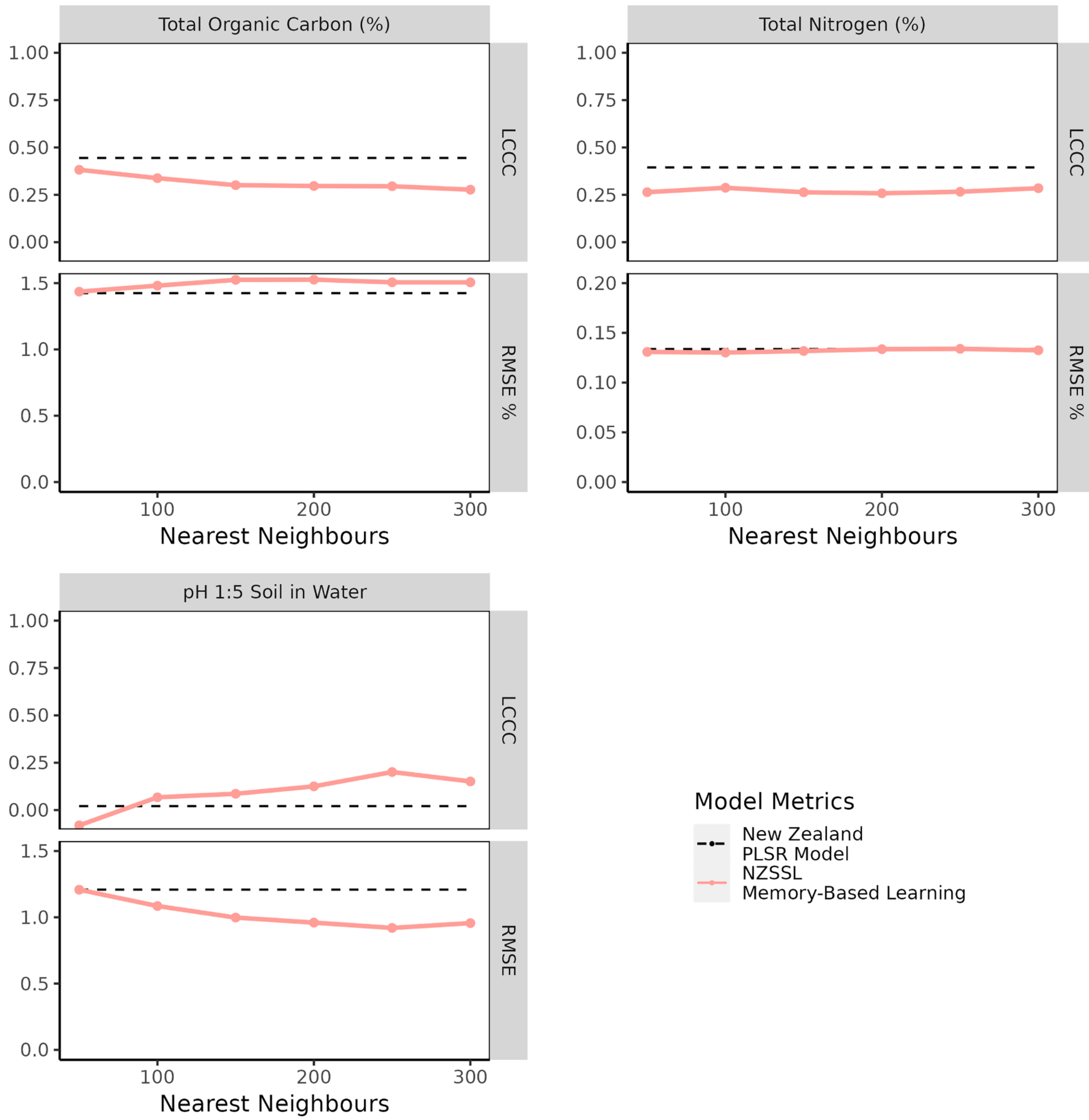

Fig. 6. RMSE and LCCC values of MBL predictions using different nearest neighbours for total organic carbon, total nitrogen and pH, predictions of Tongatapu soils, using only soils from the New Zealand Soil Spectral Library (NZSSL), contrasted against the performance of a PLSR model trained exclusively on soils from the NZSSL.

Discussion

Initial investigation demonstrated good coverage of the spectral, and soil property variation seen in the soil surveys of Tongatapu by the NZSSL. Observations of pH, total nitrogen and total organic carbon in the soils sampled in Tongatapu were within the range of those properties seen within the NZSSL, accepting the substitution of total carbon for total organic carbon. There are differences in the distribution of pH, with the data from Tongatapu being centred at a pH higher than that of the New Zealand soils, likely due to a higher carbonate presence. The principal component space of the of the NZSSL showed good inclusion of the Tongan dataset (Fig. 3), and the spectra from Tongatapu were well centred across the pairwise investigations of the first three principal components of the New Zealand spectra. Whilst these properties show promise for the usefulness of New Zealand spectra for regional soil spectral inference within Tongatapu and potentially across other PICTs, further research may benefit from analysis of other properties for selection of appropriate soils and spectra, such as soil forming factors, or latent variable space from PLSR models.

The NZSSL models performed inadequately when applied directly to the soils of Tongatapu, with poor performance metrics for total organic carbon, total nitrogen, and soil pH. This is not an unexpected finding, as the soils of New Zealand represent a large and varied range of soil formation factors, not all of which are present across all soils of Tongatapu. Tephric and allophanic features seen in the soils of Tongatapu are present across many regions of the central North Island of New Zealand (Hewitt et al., 2021), but the bulk of the soils of New Zealand are derived from a variety of other parent materials, through aeolian, colluvial and alluvial transport. As the ability of PLSR models trained on large spectral libraries to predict values across regions which represent a limited subset of the total variation seen within the library can be quite poor (Brown et al., 2005; Wetterlind and Stenberg, 2010), this may be no small part of the reason for the poor performance of the NZSSL models in Tongatapu. In addition, some features of the soils of Tongatapu, such as coral-derived, carbonate rich soils are not represented at all within the soils of New Zealand, meaning that some of the soil variation seen within the test set were not represented well within the NZSSL. Thus, the NZSSL models present a poor standard for spectral inference in Tongatapu, with much room for improvement upon, reinforcing the need for locally appropriate spectral models and infrastructure.

The results of the local PLSR models demonstrate a definite potential for the use of Vis-NIR spectroscopy within The Kingdom of Tonga, and likely more broadly across PICTs. The LCCC of 0.6 or greater were seen across all three soil properties assessed, and low RMSE values indicate a model error comparable to those seen on the NZSSL models when applied on New Zealand soils (Ma et al., 2023). Whilst this was demonstrably true for those models produced with 50 or more calibration samples employed from amongst the broader calibration set of soils from Tongatapu, the results of the low-calibration sample models demonstrated inconsistent, poor model performances that do not appear robust to the new data upon which they are predicting. Previous studies of optimal calibration set sizes for soil spectral inference have highlighted the need for calibration set sizes in the range of 60 to 70 samples (Debaene et al., 2014; Dorantes et al., 2022; Lucà et al., 2017), but selection methods to ensure representative calibration sets are employed has been shown to reduce these required sizes somewhat (Ramirez-Lopez et al., 2014). Thus, efforts to employ soil spectral inference in PICTs using partial least squares regression models must be mindful of the dangers of small calibration sample sizes and seek to ensure representative selection of samples for laboratory analysis.

Memory-based learning approaches using only spectra from the NZSSL to produce WAPLSR models for soils of Tongatapu were generally unsuccessful. Prediction goodness of fit metrics for total organic carbon, and total nitrogen were poorer than those of the NZSSL models for all numbers of nearest neighbours explored. Whilst pH predictions made using 100 or greater nearest neighbours did hold greater LCCC values, and lower RMSE values than the NZSSL PLSR models, the performance of MBL using purely New Zealand spectra was still quite poor. When the spectra from Tongatapu were added to the NZSSL library, creating regional spectral libraries of differing sizes, prediction performance

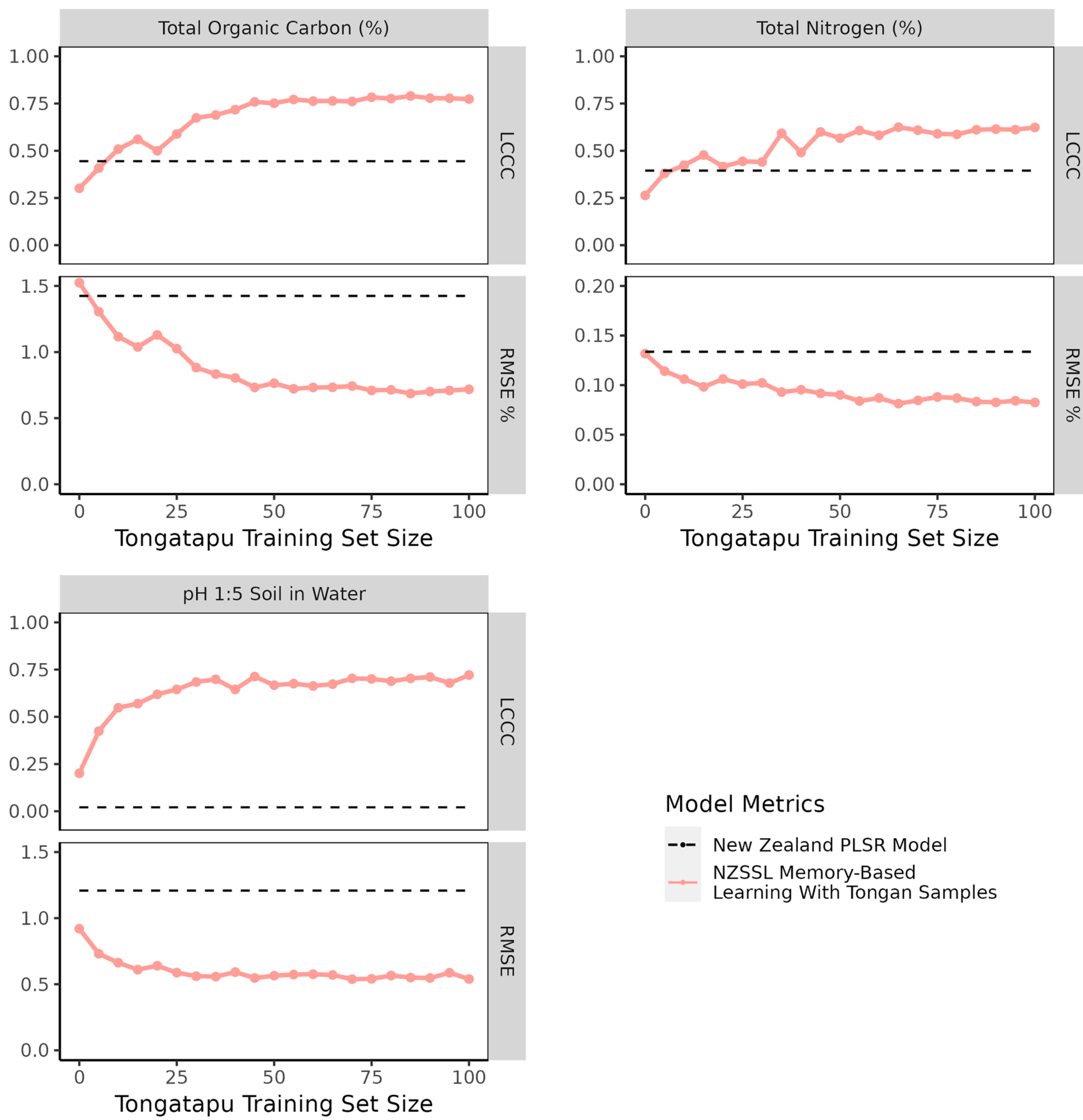

Fig. 7. RMSE and LCCC trends under increasing additions of Tongatapu data to a reference library based on the New Zealand Soil Spectral Library (NZSSL) for memory-based learning predictions for total organic carbon, total nitrogen and pH, contrasted against the performance of a PLSR model trained exclusively on soils from the NZSSL.

improved. In cases where only small amounts of Tongatapu reference information was added to the NZSSL, MBL outperformed PLSR models constructed using the same number of samples, generally for 25 or fewer samples in total organic carbon, and total nitrogen, and 15 or fewer samples in the case of pH. Generally prediction qualities were not as varied under low Tongatapu spectral information scenarios under MBL as they were with small sample PLSR models (Fig. 5, Fig. 7), indicating that such an approach of adding calibration samples to a larger regional library may be a more reliable method to make use of spectral inference in PICTs when traditional chemical analysis budgets may be low. As these methods were tested with a dataset consisting of a whole island survey of Tongatapu, and across multiple soil types and parent materials, the prediction qualities and lessons learned may change when applied to field scale studies, as such approaches using spiking of local information in MBL approaches has shown success at field and regional scales in the past (Li et al., 2022; Moloney et al., 2023).

Soil spectral inference offers a pathway to cost-effective assessment of soil condition and capital, which is crucial in insulating PICTs against tangible threats to soil including acidification, carbon loss and declines in fertility. The use of MBL techniques in PICTs are an opportunity to enable the rapid development of an operational spectral inference approach for soil monitoring and evaluation, to support soil security. These spectral inferences approaches will be improved overtime as soil type specific spectral libraries are produced for each PICTs. Investment in regional surveys across the PICTs that target the initial development of spectral inference approaches via MBL will enable the quantification of soil security monitoring and evaluation. These investments in capability building and soil measurement, could be part of a regional monitoring programs coordinated by the FAO Pacific Soil Partnership. For example, historic and current investments by the Australian Centre of International Agricultural Research NIR and MIR capability and soil measurement in Tonga, Vanuatu, Tuvalu, Samoa, Kiribati, and Fiji could be augmented with addition spectral information from New Zealand and other volcanic, (e.g. Hawaii;(McDowell et al., 2012)) and atoll islands databases and surveys to develop a larger regional library.

Conclusions

This work has highlighted the utility of soil spectral inference as a low-cost tool for the assessment of soil properties in Tongatapu, and across PICTs. Large spectral libraries with soils of similar soil forming factors may provide a useful resource in generating robust predictions of soil properties whilst local soil spectral libraries are being established in The Kingdom of Tonga and beyond.

Though small calibration sample sizes can hold limitations in conventional soil spectral inference approaches, the opportunities presented by MBL techniques makes the concept of regional cooperation in the sharing and use of soil spectral information an appealing option for developing soil spectral infrastructure across PICTs.

Further work is needed to evaluate the utility of these approaches for farm scale or system scale applications of soil spectral inference as our investigations represented a diverse range of soil types and properties across an entire island.

CRediT authorship contribution statement

J.P. Moloney: Writing – review & editing, Writing – original draft, Visualization, Investigation, Formal analysis, Data curation, Conceptualization. **Y. Ma:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation. **U. Stockmann:** Writing – review & editing, Investigation, Data curation, Conceptualization. **V.T. Manu:** Writing – review & editing, Resources, Investigation, Data curation. **V. Minoneti:** Writing – review & editing, Resources, Investigation, Data curation. **S.T. Hui:** Resources, Investigation, Data curation. **S.M. Halavatau:** Writing – review & editing, Resources, Investigation, Data curation. **S. Patolo:** Writing – review & editing, Resources, Investigation, Data curation. **T. Tukia:** Writing – review & editing, Resources, Investigation, Data curation. **S. Foliaki:** Writing – review & editing, Resources, Investigation, Data curation. **T. Carter:** Writing – review & editing, Validation, Investigation, Formal analysis. **B.C.T. Macdonald:** Writing – review & editing, Resources, Investigation. **J. Barringer:** Visualization, Investigation, Data curation. **P. Roudier:** Writing – review & editing, Investigation, Data curation, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Funding for this work was provided by ACIAR Grants SLAM/2020/139 and SMCN/2016/111 (2018 field sampling), a CSIRO Agriculture and Food Strategic Investment Project (2018 soil analysis) and the Ministry of Foreign Affairs and Trade New Zealand (post Hunga Tonga-Hunga Ha'apai eruption 2022 field sampling and soil analysis).

Data availability

Data was prepared by CSIRO and Manaaki Whenua Landcare Research. Data is owned by the Government of Tonga. It is not available for download or use without permission from the Government of Tonga.

Acknowledgements

The authors gratefully acknowledge the landowners who allowed the team access to their land, to sample and remove soil for testing. In addition, the authors would like to acknowledge several people who assisted in the field sampling and soil analysis including Brigitte Small, Shaun Kravitz, Derek Finau and Seija Tuomi for the 2018 survey, and Ellen Iramu and Mathilda Simmons for the 2022 survey. Finally, the authors thank Jane Lattimore for her assistance in project coordination and support.

References

Antille, D.L., Field, D.J., Halavatau, S.M., Iramu, E.T., Macdonald, B.C.T., Singh, K., Webb, M.J., 2022. Regional soil priorities creating partnerships with Australia and New Zealand across the Pacific. Geoderma Reg. 29, e00517. https://doi.org/10.1016/j.geodrs.2022.e00517.

Brown, D.J., Bricklemyer, R.S., Miller, P.R., 2005. Validation requirements for diffuse reflectance soil characterization models with a case study of VNIR soil C prediction in Montana. Geoderma 129 (3), 251–267. https://doi.org/10.1016/j.geoderma.2005.01.001.

Cleveland, W.S., Devlin, S.J., 1988. Locally weighted regression: an approach to regression analysis by local fitting. J. Am. Stat. Assoc. 83 (403), 596–610.

R Core Team, 2020. R: A language and environment for statistical computing. R Foundat. Statistic. Comput. https://www.R-project.org/.

Cowie, J.D., Searle, P.L., Widdowson, J.P., Orbell, G.E., 1991. *Soils of Tongatapu, Kingdom of Tonga* (DSIR Land resources scientific report 21; p. 56). New Zealand department of scientific and industrial research. https://doi.org/10.7931/DL1-DLRSR-21.

Cowie, J.D., 1980. Soils from Andesitic Tephra and their variability, Tongatapu, Kingdom of Tonga. Soil Res. 18 (3), 273–284. https://doi.org/10.1071/sr9800273.

Debaene, G., Niedźwiecki, J., Pecio, A., Żurek, A., 2014. Effect of the number of calibration samples on the prediction of several soil properties at the farm-scale. Geoderma 214–215, 114–125. https://doi.org/10.1016/j.geoderma.2013.09.022.

Dorantes, M.J., Fuentes, B.A., Miller, D.M., 2022. Calibration set optimization and library transfer for soil carbon estimation using soil spectroscopy—A review. Soil Sci. Soc. Am. J. 86 (4), 879–903. https://doi.org/10.1002/saj2.20435.

Evangelista, S.J., Field, D.J., McBratney, A.B., Minasny, B., Ng, W., Padarian, J., Román Dobarco, M., Wadoux, A.M.J.-C., 2023. A proposal for the assessment of soil security: soil functions, soil services and threats to soil. Soil Secur. 10, 100086. https://doi.org/10.1016/j.soisec.2023.100086.

Gibbs, H.S., 1976. Soils of Tongatapu Island, Tonga. New Zealand Soil Sur. Report 35, 17. https://doi.org/10.7931/DL1-SSR-35.

Grunwald, S., Yu, C., Xiong, X., 2018. Transferability and scalability of soil total carbon prediction models in Florida, USA. Pedosphere 28 (6), 856–872. https://doi.org/10.1016/S1002-0160(18)60048-7.

Hewitt, A.E., Balks, M.R., Lowe, D.J., 2021. The Soils of Aotearoa New Zealand. Springer.

Hewitt, A.E., 2010. New Zealand soil classification. Landcare Res. Sci. Ser. 1.

Kuhn, M. (2022). *caret: classification and Regression Training.* https://CRAN.R-project.org/package=caret.

Li, H., Li, Y., Yang, M., Chen, S., Shi, Z., 2022. Strategies for efficient estimation of soil organic content at the local scale based on a national spectral database. Land. Degrad. Dev. 33 (10), 1649–1661. https://doi.org/10.1002/ldr.4223.

Lucà, F., Conforti, M., Castrignanò, A., Matteucci, G., Buttafuoco, G., 2017. Effect of calibration set size on prediction at local scale of soil carbon by Vis-NIR spectroscopy. Geoderma 288, 175–183. https://doi.org/10.1016/j.geoderma.2016.11.015.

Ma, Y., Roudier, P., Kumar, K., Palmada, T., Grealish, G., Carrick, S., Lilburne, L., Triantafilis, J., 2023. A soil spectral library of New Zealand. Geoderma Regional 35, e00726. https://doi.org/10.1016/j.geodrs.2023.e00726.

Manu, V., Whitbread, A., Blair, N., Blair, G., 2014. Carbon status and structural stability of soils from differing land use systems in the Kingdom of Tonga. Soil. Use Manage 30 (4), 517–523. https://doi.org/10.1111/sum.12135.

McBratney, A., Field, D.J., Koch, A., 2014. The dimensions of soil security. Geoderma 213, 203–213. https://doi.org/10.1016/j.geoderma.2013.08.013.

McDowell, M.L., Bruland, G.L., Deenik, J.L., Grunwald, S., Knox, N.M., 2012. Soil total carbon analysis in Hawaiian soils with visible, near-infrared and mid-infrared diffuse reflectance spectroscopy. Geoderma 189–190, 312–320. https://doi.org/10.1016/j.geoderma.2012.06.009.

McKenzie, N., Baldock, J., Balks, M., Arbestain, M.C., Condron, L., Elder-Ratutokarua, M., Grundy, M., Hewitt, A., Kelliher, F., Leys, J., McDowell, R., Morrison, R., Schoknecht, N., 2015. Regional assessment of soil change in the Southwest Pacific. Main Report 476–519.

Metson, A.J., LC, B., & DA, R. (1979). *Methods for the determination of soil organic carbon: a review, and application to New Zealand soils.*

Minasny, B., McBratney, A.B., 2006. A conditioned Latin hypercube method for sampling in the presence of ancillary information. Comput. Geosci. 32 (9). Article 9.

Moloney, J.P., Malone, B.P., Karunaratne, S., Stockmann, U., 2023. Leveraging large soil spectral libraries for sensor-agnostic field condition predictions of several agronomically important soil properties. Geoderma 439, 116651. https://doi.org/10.1016/j.geoderma.2023.116651.

Naidu, R., Haynes, R.J., Gawandar, J.S., Morrison, R.J., Fitzpatrick, R.W., 1991. Chemical and mineralogical properties and soil solution composition of acid soils from the South Pacific Islands. In: Wright, R.J., Baligar, V.C., Murrmann, R.P. (Eds.), *Plant-Soil Interactions at Low pH: Proceedings of the second international symposium on plant-soil interactions at low* pH, 24–29 June 1990. Springer Netherlands, Beckley West Virginia, USA, pp. 43–53. https://doi.org/10.1007/978-94-011-3438-5_5.

Nocita, M., Stevens, A., van Wesemael, B., Aitkenhead, M., Bachmann, M., Barthès, B., Ben Dor, E., Brown, D.J., Clairotte, M., Csorba, A., Dardenne, P., Demattê, J.A.M., Genot, V., Guerrero, C., Knadel, M., Montanarella, L., Noon, C., Ramirez-Lopez, L., Robertson, J., Sakai, H., Soriano-Disla, J.M., Shepherd, K.D., Stenberg, B., Towett, E. K., Vargas, R., Wetterlind, J., 2015. Chapter four - soil spectroscopy: an alternative to wet chemistry for soil monitoring. In: Sparks, D.L. (Ed.), Advances in Agronomy. Academic Press, pp. 139–159. https://doi.org/10.1016/bs.agron.2015.02.002.

Padarian, J., Minasny, B., McBratney, A.B., 2019a. Transfer learning to localise a continental soil vis-NIR calibration model. Geoderma 340, 279–288. https://doi.org/10.1016/j.geoderma.2019.01.009.

Padarian, J., Minasny, B., McBratney, A.B., 2019b. Using deep learning to predict soil properties from regional spectral data. Geoderma Regional 16, e00198. https://doi.org/10.1016/j.geodrs.2018.e00198.

Parfitt, R.L., 1990. Allophane in New Zealand-a review. Soil Research 28 (3), 343–360.

Potter, L.M., 1986. *Tongan soils: site characteristics and management practices* (Occassional Paper 7; South pacific smallholder project). University of New England. https://books.google.com.au/books?id=mLxFAgAACAAJ.

Ramirez-Lopez, L., Behrens, T., Schmidt, K., Stevens, A., Demattê, J.A.M., Scholten, T., 2013. The spectrum-based learner: a new local approach for modeling soil vis–NIR spectra of complex datasets. Geoderma 195, 268–279.

Ramirez-Lopez, L., Schmidt, K., Behrens, T., van Wesemael, B., Demattê, J.A.M., Scholten, T., 2014. Sampling optimal calibration sets in soil infrared spectroscopy. Geoderma 226–227, 140–150. https://doi.org/10.1016/j.geoderma.2014.02.002.

Ramirez-Lopez, L., Stevens, A., Viscarra Rossel, R., Lobsez, C., Wadoux, A., Breure, T., 2016. Resemble: regression and similarity evaluation for memory-based learning in spectral chemometrics. In: R Package Version, 1. Article 2.

Rayment, G.E., Lyons, D.J., 2011. Soil chemical methods: Australasia, Vol. 3. CSIRO publishing.

Roudier, P. (2011). *clhs: A R package for conditioned Latin hypercube sampling.*

Roudier, P. (2023). *Spectacles: storing and manipulating spectroscopy data in R.*

Shenk, J.S., Westerhaus, M.O., Berzaghi, P., 1997. Investigation of a LOCAL calibration procedure for near infrared instruments. J. Near. Infrared. Spectrosc. 5 (4), 223–232.

Shepherd, K.D., Ferguson, R., Hoover, D., Egmond, F.van, Sanderman, J., Ge, Y, 2022. A global soil spectral calibration library and estimation service. Soil Secu. 7, 100061 https://doi.org/10.1016/j.soisec.2022.100061.

Sparling, G.P., Schipper, L.A., 2002. Soil quality at a national scale in New Zealand. J. Environ. Qual. 31 (6), 1848–1857. https://doi.org/10.2134/jeq2002.1848.

Stenberg, B., Viscarra Rossel, R.A., Mouazen, A.M., Wetterlind, J., 2010. Chapter five - visible and near infrared spectroscopy in soil science. In: Sparks, D.L. (Ed.), Advances in Agronomy. Academic Press, pp. 163–215. https://doi.org/10.1016/S0065-2113(10)07005-7.

Viscarra Rossel, R.A., Shen, Z., Ramirez Lopez, L., Behrens, T., Shi, Z., Wetterlind, J., Sudduth, K.A., Stenberg, B., Guerrero, C., Gholizadeh, A., Ben-Dor, E., St Luce, M., Orellano, C., 2024. An imperative for soil spectroscopic modelling is to think global but fit local with transfer learning. Earth. Sci. Rev. 254, 104797 https://doi.org/10.1016/j.earscirev.2024.104797.

Wang, W., Hu, L., Stafford, J.W., Connett, M., Wilson, B.K., Keller, M.D., 2019. Mining spectral libraries with machine learning for soil sensing in low resource settings. In: 2019 IEEE *GloBAL HUMANITARIAN TECHNOLOGY CONFerence (GHTC)*, pp. 1–7. https://doi.org/10.1109/GHTC46095.2019.9033121.

Wetterlind, J., Stenberg, B., 2010. Near-infrared spectroscopy for within-field soil characterization: small local calibrations compared with national libraries spiked with local samples. Eur. J. Soil. Sci. 61, 823–843. https://doi.org/10.1111/j.1365-2389.2010.01283.x.