# Quantifying Uncertainty in the Prediction of Soil Properties Using Mid-Infrared Spectra

Osayande Pascal Omondiagbe[a,*], Pierre Roudier[a], Linda Lilburne[a], Yuxin Ma [a] and Stephen McNeill [a]

[a]*Manaaki Whenua - Landcare Research, New Zealand*

ARTICLE INFO

*Keywords*:

Soil spectroscopy

Uncertainty

Mid-infrared prediction

ABSTRACT

Soilś pivotal role in environmental and agricultural processes underscores the importance of accurate soil property predictions for informed decisions and sustainable land management. Spectroscopic techniques, particularly mid-infrared (MIR) spectroscopy, have emerged as rapid and non-destructive tools for soil analysis. Despite advancements in predicting soil properties using spectroscopy, quantifying prediction uncertainties has often been overlooked. Precise uncertainty quantification aids risk assessment and decision-making processes. This study compares three methods—Bootstrapped Partial Least-Squares regression (PLS-BS), Generalized Additive Models (GAM), and Bayesian Convolutional Neural Networks (BCNNs)—for quantifying uncertainty in six soil property predictions (clay, sand, silt, pH, p retention and carbon) based on MIR spectroscopy. In terms of predictive performance and quality of prediction, our evaluation indicated that both GAMs and BCNNs outperformed PLS-BS for all six soil properties. The ability of GAMs and BCNNs to capture non-linear relationships in the data allowed for better fitting to the underlying patterns. BCNNs, in particular, demonstrated superior performance by combining accurate predictions with robust uncertainty quantification. The evaluation of models should extend beyond standard validation metrics, which typically assess prediction accuracy, to include the capability to measure prediction uncertainty. By analysing their performance and estimation capabilities, we enhance our understanding of these approaches and their applicability to soil analysis and land management.

## 1. Introduction

Soil plays a vital role in various environmental and agricultural processes, making accurate predictions of soil properties crucial for informed decision-making and sustainable land management. In recent years, spectroscopic techniques, particularly mid-infrared (MIR) spectroscopy, have gained prominence as a rapid and non-destructive method for soil analysis (dos Santos et al., 2020; Knox et al., 2015). MIR spectra capture essential chemical information that can be leveraged to predict various soil properties, such as organic carbon content, pH levels, and nutrient concentrations Sanderman et al. (2020).

Despite those advances in spectroscopic predictions of soil properties, quantifying the uncertainties of those predictions has largely be ignored (Heuvelink and Webster, 2001). Accurate quantification of uncertainty enables decision-makers to assess the reliability and limitations of soil property predictions, aiding in risk assessment and informed decision-making processes (Carré et al., 2007). Various statistical and machine learning methods have been applied to quantify prediction uncertainty, each with its

*Corresponding author

omondiagbpe@landcareresearch.co.nz (O.P. Omondiagbe); RoudierP@landcareresearch.co.nz (P. Roudier); lilburnel@landcareresearch.co.nz (L. Lilburne); MaY@landcareresearch.co.nz (Y.M. ); mcneills@landcareresearch.co.nz (S.M. )

strengths and limitations (Takoutsing and Heuvelink, 2022; Psaros et al., 2023; Frenklach et al., 2016).

In this paper, we compare three methods for quantifying uncertainty in the prediction of soil properties using MIR spectroscopy: Bootstrapped Partial Least-Squares (PLS) regression, Generalized Additive Models (GAMs), and Bayesian Convolutional Neural Networks (BCNNs). Each of these methods offers distinct approaches to modelling and uncertainty estimation, providing a diverse range of tools for soil prediction.

Partial Least-Squares regression is a powerful multivariate statistical method that can handle collinear and high-dimensional data. It utilizes the correlation structure between the predictor variables (MIR spectra) and the response variables (soil properties) to create predictive models. However, the estimation of prediction uncertainty with PLS regression can be challenging, and various approaches have been proposed to address this issue (Martens and Naes, 1992)

Generalized Additive Models offer a flexible and interpretable framework for modeling non-linear relationships between predictors and response variables. By incorporating smooth functions of the predictors, GAMs can capture complex relationships in the MIR spectra and provide estimates of prediction uncertainty through methods such as bootstrapping or Monte Carlo simulations (Wood, 2017; Hastie, 2017).

Bayesian Convolutional Neural Networks, on the other hand, leverage deep learning techniques to model the MIR spectra and predict soil properties. These neural networks offer the ability to capture intricate patterns in the data (Omondiagbe et al., 2023a), but more importantly, they provide a natural framework for uncertainty estimation through Bayesian inference. Bayesian CNNs can provide not only point predictions but also probability distributions over the predicted soil properties, enabling the quantification of uncertainty (Gal and Ghahramani, 2016; Kendall and Gal, 2017).

In this study, we compare and evaluate these three methods in terms of their predictive performance and their ability to estimate prediction uncertainty in soil property modelling using MIR spectra. We utilise a comprehensive dataset of soil samples with corresponding MIR spectra and reference measurements to conduct our comparative analysis.

The findings of this research will shed light on the strengths and limitations of each method in capturing the inherent uncertainty in soil property predictions. This knowledge will provide valuable insights for researchers, soil scientists, and practitioners involved in soil analysis and land management, aiding them in selecting appropriate methods for quantifying uncertainty in soil property predictions using MIR spectroscopy.

Overall, this study aims to contribute to the advancement of soil analysis by comparing and evaluating the performance of different methods for uncertainty quantification in the prediction of soil properties using MIR spectra. By improving our understanding of the strengths and weaknesses of these approaches, we can enhance the accuracy and reliability of soil property predictions, thereby supporting sustainable land management practices and decision-making processes.

# 2. Material and methods

## 2.1. Benchmark Datasets

### *2.1.1. Data splitting and validation*

The raw spectral data were split into calibration (70 %) and testing (30 %) sets. The BCNN models' calibration dataset was further split into calibration (70 %) and validation (30 %) sets.

### *2.1.2. Data pre-processing methods*

Due to the presence of noise, wavelengths greater than 4000 nm were removed. For each modelling approach, relevant spectral treatments were applied. In the PLS and GAMs approach, pseudo-absorbance values were derived from the spectra, followed by the application of the first-order Savitzky-Golay filter (McCarty et al., 2002), using a window width of 7 points and a polynomial order of 2. Additionally, standard normal variate transformation was performed (Rinnan et al., 2009).

For the BCNNs method, no preprocessing was conducted. This is because CNN has the capability to capture intricate patterns in the data (Omondiagbe et al., 2023b). To reduce the number of variables in our BCNNs model, spectral data was downsampled by a factor of 1:10, retaining one value every 10 nm, approximating a slower-paced sampling sequence. Finally, feature standardisation was performed by removing the mean and scaling to unit variance using the StandardScaler module from Scikit-learn (Bisong and Bisong, 2019).

## 2.2. Uncertainty Assessment

Uncertainty in soil predictions arises due to measurement errors, model assumptions, and variations in the soil spectroscopy data (Heuvelink and Webster, 2001). Assessing uncertainty allows for a more comprehensive understanding of the reliability of the predictions. In this paper, we introduce three model methods in the sub-sections below that can be used for uncertainty assessment. By employing these three model methods, we aim to compare the quality of uncertainty when predicting soil properties, considering various sources of uncertainty in the soil spectroscopy data and modelling process.

### *2.2.1. Partial Least-Squares regression Approach*

PLSR is a multivariate regression technique widely used in chemometrics to understand the relationship between highly collinear multi-dimensional predictor variables and a response variable (Tobias et al., 1995). PLS regression is employed to model the complex relationship between predictor variables (spectral data) and response variables (soil properties). Rossel et al. (2006) demonstrated the effectiveness of PLSR models in capturing the relationships between soil spectral data and soil properties when predicting soil properties such as organic carbon, clay content, and pH.

In PLS, the non-parametric bootstrapping technique (Kuhn et al., 2013) can be used to estimate uncertainty and construct prediction intervals. This allows for generating a distribution of model estimates and predictions, thereby enabling the estimation of uncertainty without relying on specific assumptions about the underlying data distribution. By repeatedly resampling from the original data, it captures the inherent variability and uncertainty in the model estimates.

### *2.2.2. Generalised Additive Models Approach*

Generalized Additive Models (GAMs) can be used to capture nonlinear relationships between the spectral data and soil properties (Wood, 2017). Through the incorporation of smooth functions and spatially-varying terms, GAMs provide a framework for quantifying uncertainty (Wood, 2017). This can be achieved through different approaches (Chavez-Demoulin and Davison, 2005; Wang and Wahba, 1995; Wood, 2017), such as:

1. Bootstrap Resampling: According to Wang and Wahba (1995), it has been suggested that employing bootstrapping techniques with smoothing splines can provide reasonable pointwise and overall confidence

intervals for the underlying mean curve. This approach takes advantage of the flexibility of smoothing splines to capture complex relationships in the data and leverages bootstrapping to estimate uncertainty (Wood, 2017; Wang and Wahba, 1995; Wahba, 1990). By resampling the data and fitting smoothing splines to each bootstrap sample, confidence intervals can be obtained, allowing for a comprehensive assessment of the uncertainty surrounding the mean curve.

2. Bayesian GAMs: By combining the flexibility of GAMs with Bayesian inference, Bayesian GAMs provide a robust framework for modelling complex relationships between predictor variables and response variables while quantifying uncertainty (Fahrmeir and Lang, 2001). In this case, prior distributions are specified for the model parameters, and posterior distributions are obtained through Bayesian inference techniques such as Markov Chain Monte Carlo sampling (Hastings, 1970). This allows for uncertainty quantification by characterising the variability in the parameter estimates and predictions.
3. Model Comparison and Selection: Assessing uncertainty in GAMs can also involve comparing multiple candidate models and evaluating the variability in model performance (Lubke et al., 2017). Techniques such as cross-validation and information criteria (e.g., AIC, BIC) can be used to select the best-fitting model (Konishi and Kitagawa, 2008) and gauge uncertainty in the model selection process.

#### *2.2.3. Bayesian convolutional neural network*

Bayesian Convolutional Neural Networks (BCNNs) have gained significant attention in recent years due to their ability to estimate uncertainty in predictions. Unlike traditional convolutional neural networks (CNNs), BCNNs incorporate Bayesian principles, allowing for uncertainty quantification in model predictions. The derivation of uncertainty estimates in BCNNs is based on Bayesian inference, which provides a probabilistic framework for modelling and reasoning about uncertainty (Gal and Ghahramani, 2016).

One common approach in BCNNs is to employ Bayesian inference techniques such as variational inference or Markov chain Monte Carlo (MCMC) sampling. These methods aim to approximate the posterior distribution over the network parameters, which captures the uncertainty in the parameter estimates. By sampling from the posterior distribution, BCNNs can produce a distribution of predictions instead of a

single-point estimate, allowing for uncertainty quantification (Neal, 2012).

Other approaches for uncertainty quantification in BCNNs can also be used. Lakshminarayanan et al. (2017) introduced Deep Ensembles, which involves training multiple independently initialised networks and averaging their predictions to obtain reliable uncertainty estimates. Another approach called Monte Carlo Dropout can also be used to approximate the posterior distribution by applying dropout during test-time inference (Kendall and Gal, 2017).

In summary, Bayesian Convolutional Neural Networks provide a principled framework for uncertainty quantification in deep learning models, enabling more robust and informative predictions by capturing and representing uncertainty in a probabilistic manner.

## 2.3. Model Training

In this section, we focus on mid-infrared spectra (MIR) which cover a wavelength range from 600 to 4000 nm with a resolution of 0.5 nm. The spectra were used to predict sand, silt, clay, carbon and P retention by using the three modelling approaches introduced earlier in Section 2.2.

### *2.3.1. Partial Least-Squares regression*

PLS regression can be extended to estimate and quantify uncertainty in soil property predictions. We first developed a PLS model for each soil property and the respective optimal number of the latent variable (LV) was chosen using repeated cross-validation.

Next, we employed the non-parametric bootstrapping approach to derive the prediction intervals from the PLS model. Bootstrapping involves randomly sampling the original dataset with replacement to create multiple bootstrap replicates. Each PLS model was then fitted with 1000 replicates, resulting in a collection of models that capture the variability in the data. The predictions from the bootstrap replicates were then used to construct prediction intervals, which provide a range of values that are likely to contain the true response variable. The width of the prediction interval represents the estimated uncertainty in the soil property prediction.

### *2.3.2. Generalised Additive Model*

We adopt the Bayesian GAMs approach by extracting the latent variable for each variable in our PLS model to train a GAMs model for each soil property. The "mgcv" library implemented in R (Wood, 2001) was used to achieve this.

For six soil properties tested (carbon, clay, sand, pH, silt and P retention), beta regression was used since those variables are bounded between 0% and 100%.

In deriving the 90% prediction interval we use simulation from the posterior distribution of the trained GAMs. For each variable, 1000 simulations were used to construct prediction intervals, which provide a range of values that are likely to contain the true response variable.

### *2.3.3. Architectural Design of Bayesian CNN*

In this section, we present the architectural design of our Bayesian Convolutional Neural Network (CNN) model. Our primary goal is to address the challenge of quantifying uncertainty while maintaining accurate predictions. To achieve this, we leverage the principles of Bayesian deep learning to capture parameter uncertainty, providing valuable insights into model predictions. We use a moderately deep Bayesian CNN network architecture with multiple fully connected layers to predict six different soil properties individually, except in the case of soil texture (sand, silt, and clay), where we use the same network to simultaneously predict these three properties. The number of hidden layers in our BCNN is considered a hyperparameter to be tuned and was restricted to a range between 2 and 7 due to the amount of data available. We optimised our network architecture through a hyperparameter tuning process, selecting the configuration that maximized predictive accuracy for our specific task and dataset. Below, we provide a comprehensive overview of the key components, design choices, and the rationale behind our Bayesian CNN architecture.

**Input and Connected Layers:** The input layer of our Bayesian CNN consists of 340 neurons, accommodating the spectral signals. To reduce the dimensionality of the data, we excluded the first two spectral signals and selected every tenth signal. This initial layer serves as the entry point for our model. Our optimized fully connected layers comprise a series of fully connected (linear) layers, as detailed in Table 1. These layers introduce non-linearity, enabling our model to capture complex patterns in the data.

**Weight and Bias Parameters:** To account for uncertainty in our model, we treat the weights and biases as random variables. The following parameters are employed:

- **Weight Mean:** This represents the mean of the weight distribution.
- **Weight Standard Deviation:** This represents the standard deviation of the weight distribution, initialized with `weight_std` (default 0.001).

- **Bias Mean:** This represents the mean of the bias distribution.
- **Bias Standard Deviation:** This corresponds to the standard deviation of the bias distribution, initialized with `bias_std` (default 1.5).

These parameters allow us to capture the uncertainty associated with the model's parameters, enhancing its robustness through variational inference techniques. This technique optimises these random variables (weight and bias parameters) to find the best approximation to the true posterior distribution. The process involves two key steps: optimizing the Evidence Lower Bound (ELBO), which acts as a measuring stick to gauge the effectiveness of our approximation, and employing optimisation algorithms such as stochastic gradient descent (SGD) or its variants, including Adaptive Moment Estimation (Adam), to update the neural network weights(Hoffman et al., 2013).

In simpler terms, ELBO and variational inference are tools that helps to estimate the parameters of a regression model by finding the best approximation to the underlying data distribution. This process is analogous to finding the best-fitting line or curve through data points while considering the uncertainty in our estimates. The variation inference techniques were performed using the reset parameter and forward method described below.

**Reset parameters:** We introduced a `reset parameters` method responsible for initializing the models weights and biases. We utilize normal distributions with specific means and standard deviations to set the initial parameter values. This initialization strategy plays a crucial role in the model's training and convergence.

- **Weight Mean and Standard Deviation:** These are initialized with small values to provide a reasonable starting point for optimisation.
- **Bias Mean and Standard Deviation:** Similarly, these are initialized with small values to mitigate potential bias.

**Forward Pass:** In the forward pass of our Bayesian CNN model, the uncertain model parameters (weights and biases) are sampled from probability distributions. This process results in a distribution of predictions, offering a probabilistic view of model outputs. This approach enables us to quantify the uncertainty in predictions, which is particularly valuable in scenarios where robustness and uncertainty estimation are essential.

The forward pass of our Bayesian CNN involves the following steps:

1. **Input Data Flattening:** The input data is flattened to prepare it for fully connected layers.
2. **Activation Functions:** We apply Rectified ReLU (RReLU) activation functions after each fully connected layer, introducing non-linearity into the model.
3. **Reparameterization Method:** We employ a reparameterization method to sample weights and biases, incorporating random noise from standard normal distributions. This enables us to estimate model predictions and uncertainty simultaneously.
4. **Non-Negativity Constraint:** To ensure model stability, we enforce non-negativity on the sampled weights and biases using the `torch.clamp` function in PyTorch.
5. **Linear Layer Operation:** The final linear layer computes the output, producing the predicted soil property.

It is important to note that treating weights and biases as random variables introduces the concept of parameter uncertainty. This allows our model to not only make predictions but also estimate the degree of confidence associated with those predictions.

To determine the best parameters for our architecture, we conducted hyperparameter searches by defining a set of hyperparameters, this includes the learning rate, momentum, optimizer choice, batch size, epochs, the dimension and number of the hidden layers, weight standard deviation, and bias standard deviation. We employed the Population-Based Bandit (PB2) algorithm for this purpose (Parker-Holder et al., 2020). PB2 is a potent method for hyperparameter optimisation in machine learning, which extends the concept of population-based training (PBT) and combines it with Bayesian optimisation. This approach allows for an efficient exploration of hyperparameter configurations that yield optimal results for a given machine-learning model. This choice was influenced by previous research, which demonstrated that a variation of PBT outperformed traditional Bayesian optimisation when designing CNNs for soil properties (Omondiagbe et al., 2023b). The list of hyperparameters is provided in Table 2

The steps involved in our BCNN training include the following:

**Table 1**
Optimal Bayesian Convolutional Neural Network Architectures

| Model | Layer | Output Size/Shape | Activation Function |
|---|---|---|---|
| Texture | Input | Input Size : 340 | - |
| | Fully Connected (Linear) | 512 | ReLU |
| | Fully Connected (Linear) | 28 | ReLU |
| | Fully Connected (Linear) | 64 | ReLU |
| | Output | 3 | Linear |
| | Variational Parameters | - | - |
| | Reparameterization Trick | - | - |
| | Non-Negativity Constraint | - | - |
| P-retention | Input | Input Size : 340 | - |
| | Fully Connected (Linear) | 32 | ReLU |
| | Fully Connected (Linear) | 32 | ReLU |
| | Output | 1 | Linear |
| | Variational Parameters | - | - |
| | Reparameterization Trick | - | - |
| | Non-Negativity Constraint | - | - |
| pH | Input | Input Size : 340 | - |
| | Fully Connected (Linear) | 512 | ReLU |
| | Fully Connected (Linear) | 64 | ReLU |
| | Fully Connected (Linear) | 128 | ReLU |
| | Output | 1 | Linear |
| | Variational Parameters | - | - |
| | Reparameterization Trick | - | - |
| Carbon | Input | Input Size : 340 | - |
| | Fully Connected (Linear) | 8 | ReLU |
| | Fully Connected (Linear) | 32 | ReLU |
| | Fully Connected (Linear) | 32 | ReLU |
| | Fully Connected (Linear) | 256 | ReLU |
| | Fully Connected (Linear) | 128 | ReLU |
| | Fully Connected (Linear) | 8 | ReLU |
| | Output | 1 | Linear |
| | Variational Parameters | - | - |
| | Reparameterization Trick | - | - |

**Table 2**
Hyperparameter Settings

| Hyperparameter | Value Range |
|---|---|
| Learning Rate (lr) | [0.00001, 0.1] |
| Momentum | [0.009, 0.01] |
| Weight Standard Deviation (weight_std) | [0.0001, 10.02] |
| Bias Standard Deviation (bias_std) | [0.01, 17.7] |
| Batch Size | [2, 128] |
| Epochs | [2, 1000] |
| Optimizer | [Adams, SGD,Nadam] |

- **Model Definition:** We define the structure of the Bayesian CNN model as shown in Table 1 This involves specifying the layers, activation functions, and the prior distributions for the weights.
- **Data Preparation:** Prepare the dataset for training and testing. This involves loading the data, applying any necessary transformations, and splitting the initial training dataset into training (80%) and validation (20%) sets. We also standardize the data, ensuring that

all features share the same scale. This process entails learning the mean and standard deviation from the training data and applying the same transformation to our validation and test data.

- **Model Training/Inference:** We define a loss function and optimizer for training the Bayesian CNN. The loss function used in this paper is the mean squared error (MSE) loss. Then, we employ an optimisation method and learning rate derived through our hyperparameter tuning approach to update the model parameters. Inference was performed using the reset parameter and forward method described above. During hyperparameter tuning, we define our objective function to jointly optimize both the coefficient of determination and the prediction interval specifically on the validation set.

- **Uncertainty Quantification:** To estimate uncertainty, we generate 1000 predictions by sampling sets of weights from the posterior distribution. Then we computed the variance across the multiple predictions. Once we had the variance values for each sample, we calculated the 90% prediction interval.

The optimal parameters for the Clay, Carbon, P-rentention, and soil texture (sand, silt, and clay) models are shown in Table 3.

## 2.4. Model evaluation

We computed the following metrics: model efficiency coefficient (MEC) or coefficient of determination (Janssen and Heuberger, 1995), Root Mean Squared Error (RMSE), Lin's concordance correlation coefficient (CCC) (Lawrence and Lin, 1989), the ratio of performance to interquartile range (RPIQ) (Bellon-Maurel and McBratney, 2011), Relative Percent Difference (RPD) , and Standard Error (SE)

RPIQ is a measure that assesses the relative performance of a system or model compared to the spread or variability of the data. This is calculated by dividing the interquartile (IQR) range of the observed values by the RMSE estimate, as shown in Equation 1.

$$RPIQ = \frac{IQR}{RMSE}. \quad (1)$$

MEC is a metric used to assess the quality of model fit. It quantifies the goodness of fit by comparing the predicted

**Table 3**
Optimal Model Parameters

| Model | learning rate | momentum | weight_std | bias_std | batch_size | epochs | optimiser |
|---|---|---|---|---|---|---|---|
| pH | $4.70 \times 10^{-4}$ | 0.0096 | 1.299 | 0.2769 | 58 | 799 | Adams |
| P-rentention | $5.557 \times 10^{-4}$ | 0.0097 | 0.0111 | 0.142 | 115 | 789 | Adams |
| Carbon | $4.05 \times 10^{-4}$ | 0.0091 | 0.0016 | 0.1327 | 86 | 221 | Adams |
| Soil Texture (sand, silt and clay) | $5.89 \times 10^{-5}$ | 0.0090 | 4.7621 | 16.738 | 66 | 325 | Nadam |

values from the model with the observed values. The calculation of MEC 2 is given below.

$$MEC = 1 - \frac{\sum_i (y_i - \hat{y_i})^2}{\sum_i (y_i - \bar{y_i})^2} \tag{2}$$

where $y_i$ is the dependent variable, $\hat{y}_i$ is the output generated from our simulated model, both at index i and $\bar{y_i}$ is the mean of the measurements.

CCC is a statistical measure used to assess the agreement between two continuous variables. It takes into account both the precision (closeness to the best-fit line) and accuracy (deviation from the best-fit line) of the agreement. It is calculated as in equation 3 below:

$$CCC = \frac{2rs_x s_y}{{s_x}^2 + {s_y}^2 + (\bar{x_i} - \bar{y_i})} \tag{3}$$

where $\bar{x_i}$ and $\bar{y_i}$ are the mean of the observed and predicted values; r is the correlation coefficient between the observed and predicted values, $s_x$ and $s_y$ are the variance of observed and predicted values.

RMSE is s a statistical metric used to measure the average magnitude of errors between predicted and observed values. It provides a way to quantify the accuracy of a regression model's predictions. RMSE is calculated as shown in equation 4 below:

$$RMSE(y) = \frac{1}{N^2}\sqrt{\sum_i (y_i - \hat{y}_i)^2} \tag{4}$$

RPD is a metric used to assess the ratio of the standard deviation to the mean in a dataset. It provides a measure of the variability relative to the central tendency of the data. RPD is calculated by dividing the standard deviation of the data by the mean, and then multiplying the result by 100 to express it as a percentage. It is calculated as shown in equation 5 below:

$$RPD = \left(\frac{|y_i - \hat{y}_i|}{(y_i + \hat{y}_i)/2}\right) \times 100 \tag{5}$$

where $y_i$ is the dependent variable and $\hat{y}_i$ is the output generated from the model.

SE is a statistical metric that measures the precision or variability of an estimate or statistic. It quantifies the average amount of deviation between the sample mean or regression coefficient and the true population mean or coefficient. It is calculated as shown in equation 6 below:

$$SE = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n-1}} \tag{6}$$

where $y_i$ is the dependent variable, $\hat{y}_i$ is the output generated from the model and "n" represents the total number of observations.

# 3. Results

## 3.1. Prediction performance

The optimal number of independent latent variables (LVs) used for deriving the GAM model is shown in Figure 1 below. The LVs were used for beta regression. To visualize how the Bootstrapped PLS (PLS-BS), GAM and BCNN perform, we plot the measured value against the predicted value for all six properties in Figs 2, 3, and 4 respectively. The performance of the original PLS, PLS-BS, GAM, and BCNN models is shown in Table 4. Given that GAM and Bayesian CNN could capture nonlinear relationships between the spectral data and soil properties, an increase in performance was noticed for five of the soil properties compared to when using the PLS-BS model. Also, the Bootstrapped PLS was seen to outperform the original PLS in all six properties.

## 3.2. Quality of the prediction intervals

We derived a 90% prediction interval for each model using the validation datasets. This is shown in Fig 5, 6 and 7. The quality was assessed using the Prediction Interval Coverage Probability (PICP, %), which represents the number of observations from the validation set falling within the predicted prediction interval. The result of this is shown in Table 5 and Fig 8. When predicting soil texture using the GAM model, all PICP values were very close to the target value of 90%. This differed in the case of the Bayesian CNN, where we observed that all PICP values were above the target value of 90%. The PICP for all six properties with the Bootstrap PICP was below the target value of 90%.

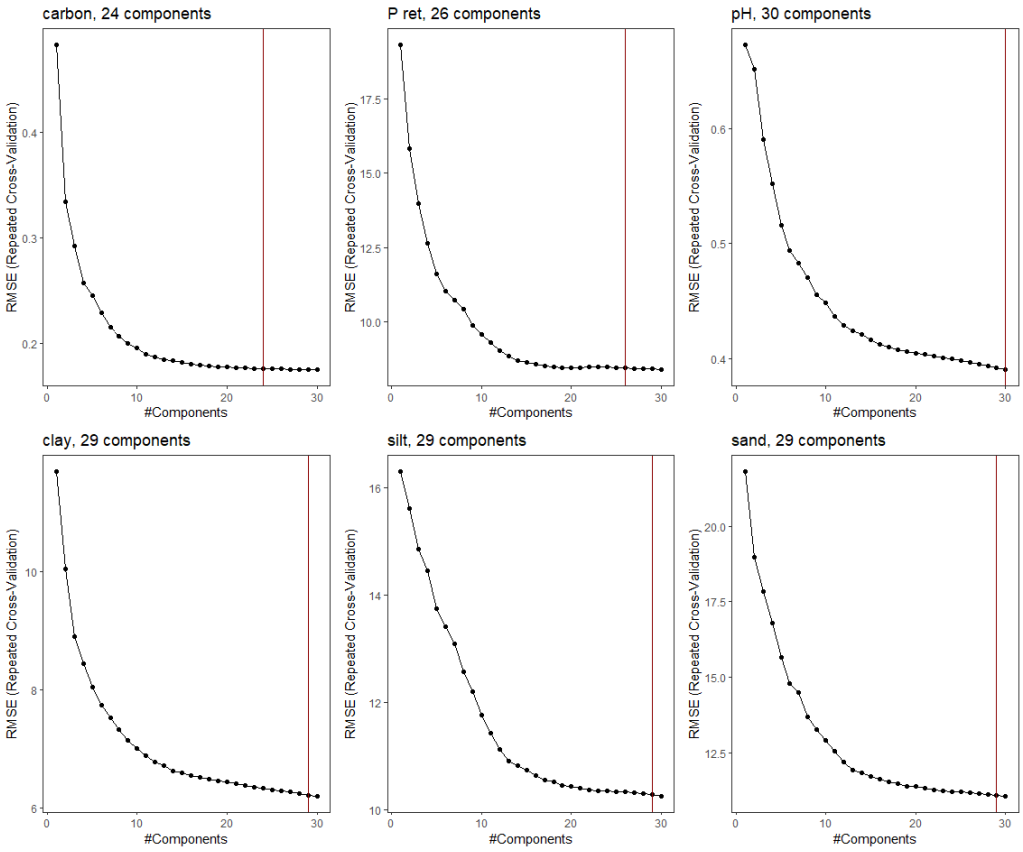


**Figure 1:** Number of LV retained for each PLS model

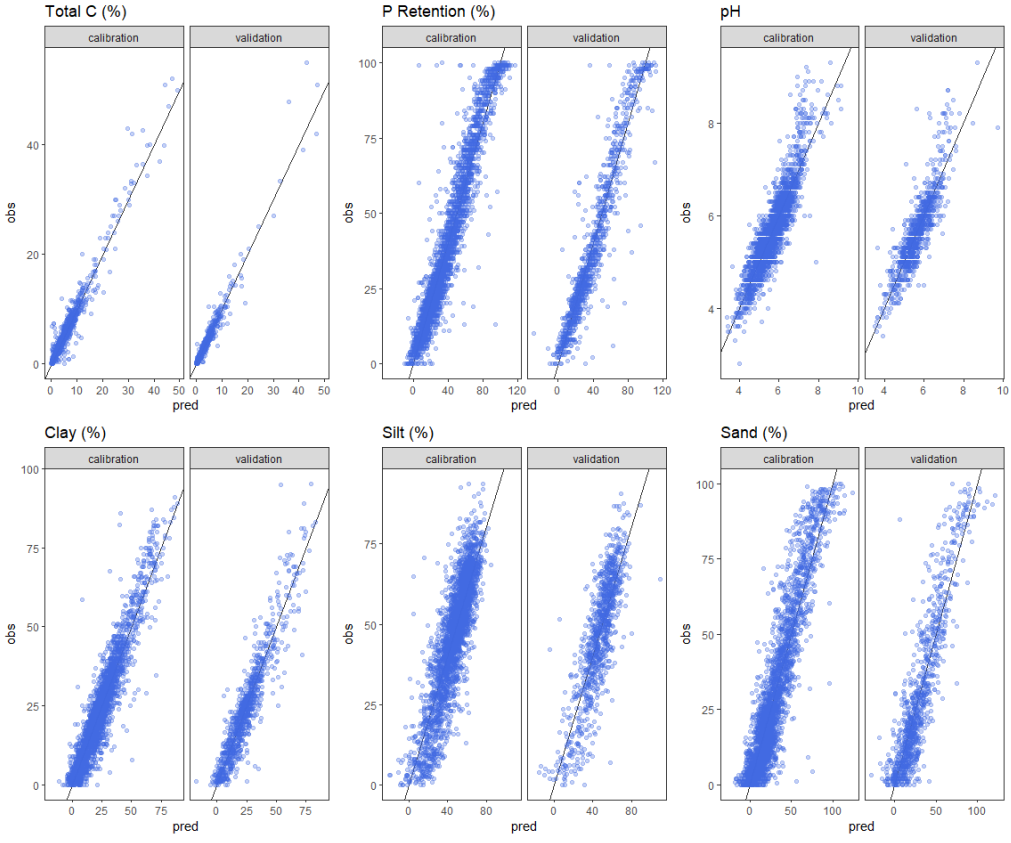


**Figure 2:** PLS model performance

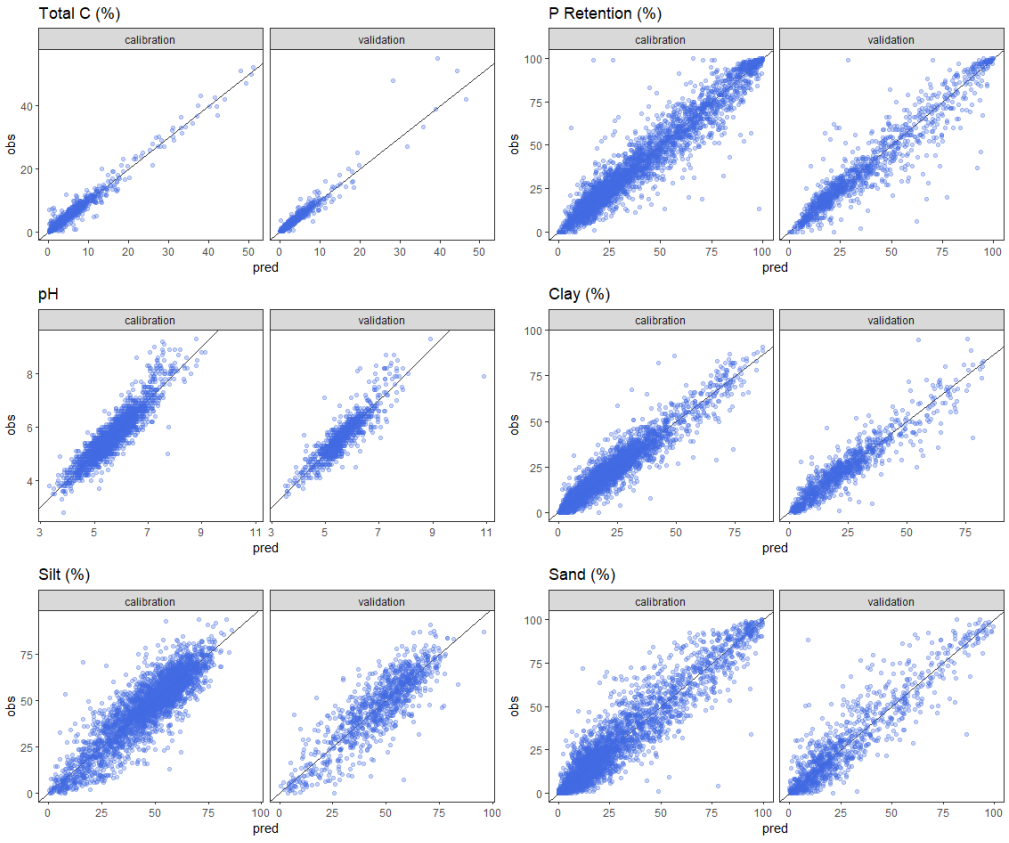


**Figure 3:** GAM model performance

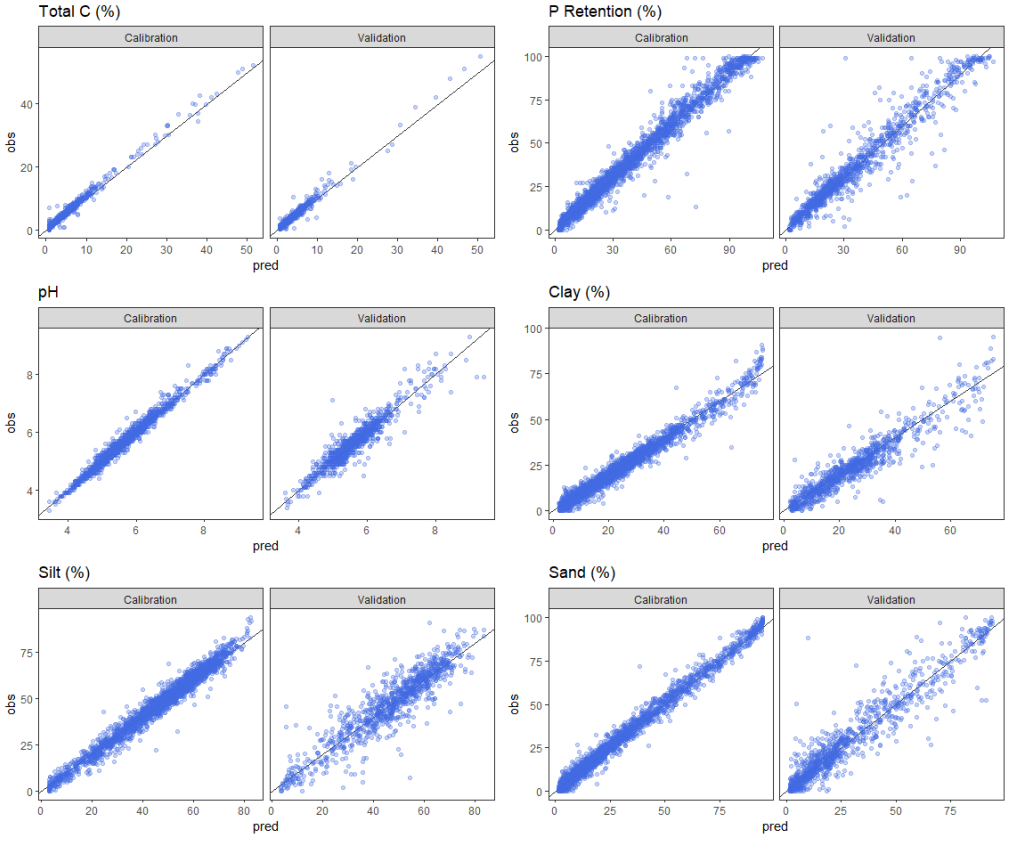


**Figure 4:** Bayesian CNN model performance

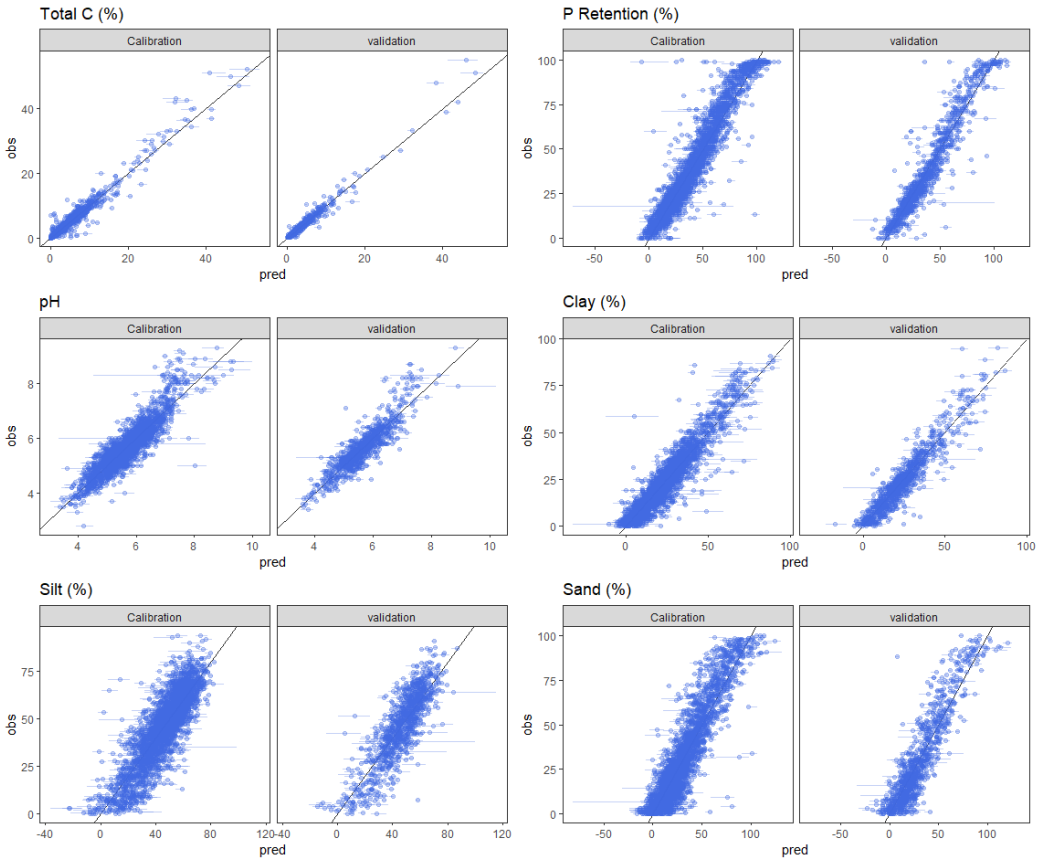


**Figure 5:** PLS-BS prediction interval with 1000 replicates for each model

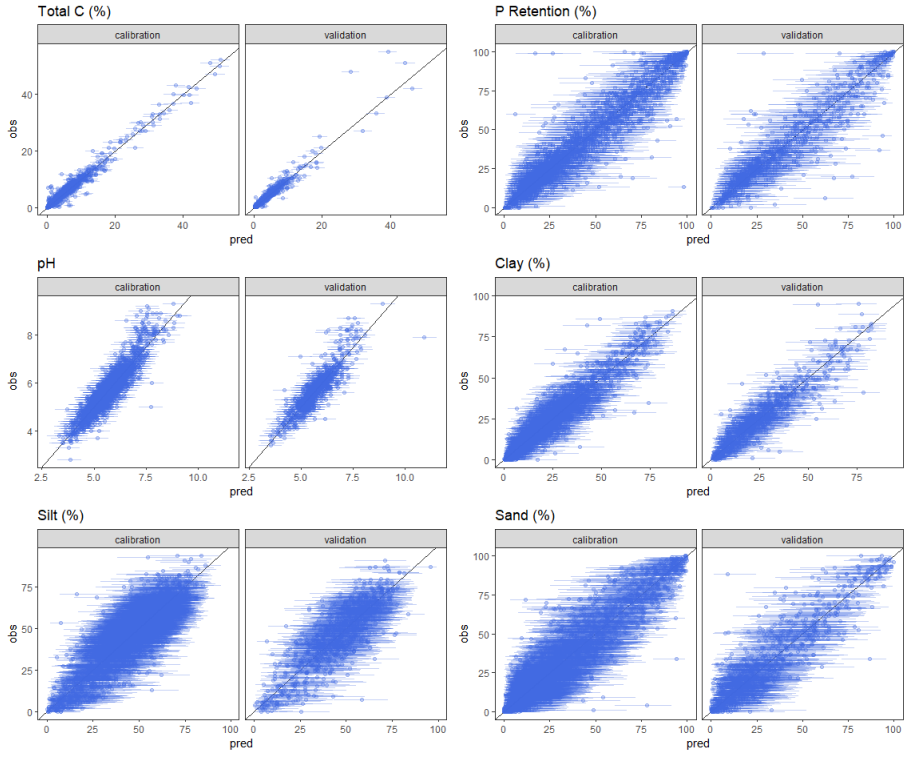


**Figure 6:** GAM prediction interval

| Variable | Method | RMSE | Rsquared | RPD | RPIQ | CCC | Bias | SE |
|---|---|---|---|---|---|---|---|---|
| carbon | Bayesian CNN | 0.77 | 0.98 | 5.05 | 3.24 | 0.98 | -0.13 | 0.77 |
| carbon | GAM | 0.95 | 0.94 | 4.08 | 2.62 | 0.97 | -0.04 | 0.95 |
| Carbon | PLS | 0.79 | 0.96 | 5.04 | 3.24 | 0.98 | 0.00 | 0.77 |
| carbon | PLS-BS | 0.73 | 0.97 | 5.34 | 3.44 | 0.98 | -0.03 | 0.73 |
| P retention | Bayesian CNN | 7.99 | 0.91 | 3.32 | 4.63 | 0.95 | 0.22 | 8.00 |
| P retention | GAM | 8.76 | 0.89 | 3.03 | 4.22 | 0.95 | 0.64 | 8.76 |
| P retention | PLS | 8.63 | 0.89 | 3.07 | 4.29 | 0.94 | 0.57 | 8.63 |
| P retention | PLS-BS | 7.71 | 0.92 | 3.45 | 4.93 | 0.96 | 0.01 | 7.71 |
| pH | Bayesian CNN | 0.29 | 0.88 | 2.84 | 3.44 | 0.93 | 0.02 | 0.29 |
| pH | GAM | 0.39 | 0.78 | 2.14 | 2.59 | 0.88 | -0.01 | 0.39 |
| pH | PLS | 0.36 | 0.81 | 2.31 | 2.79 | 0.90 | 0.00 | 0.36 |
| pH | PLS-BS | 0.36 | 0.81 | 2.31 | 2.79 | 0.90 | 0.00 | 0.36 |
| clay | Bayesian CNN | 5.29 | 0.90 | 3.10 | 3.59 | 0.94 | 0.36 | 5.30 |
| clay | GAM | 6.01 | 0.87 | 2.73 | 3.16 | 0.93 | -0.08 | 6.02 |
| Clay | PLS | 5.64 | 0.88 | 2.88 | 3.37 | 0.94 | 0.00 | 5.64 |
| clay | PLS-BS | 5.64 | 0.88 | 2.89 | 3.37 | 0.94 | 0.01 | 5.64 |
| sand | Bayesian CNN | 8.02 | 0.90 | 3.16 | 4.40 | 0.94 | -0.44 | 8.03 |
| sand | GAM | 9.89 | 0.85 | 2.56 | 3.57 | 0.92 | -0.83 | 9.90 |
| Sand | PLS | 10.90 | 0.82 | 2.33 | 3.24 | 0.90 | -0.25 | 10.90 |
| sand | PLS-BS | 10.17 | 0.84 | 2.50 | 3.31 | 0.91 | -0.01 | 10.17 |
| silt | Bayesian CNN | 7.62 | 0.83 | 2.43 | 3.24 | 0.91 | 0.08 | 7.63 |
| silt | GAM | 9.49 | 0.74 | 1.95 | 2.60 | 0.85 | -0.36 | 9.49 |
| Silt | PLS | 10.14 | 0.70 | 1.83 | 2.43 | 0.83 | 0.01 | 10.14 |
| silt | PLS-BS | 9.42 | 0.74 | 1.96 | 2.61 | 0.85 | 0.00 | 9.42 |

**Table 4**
Bayesian CNN, GAM, PLS and PLS-BS performance on the validation datasets

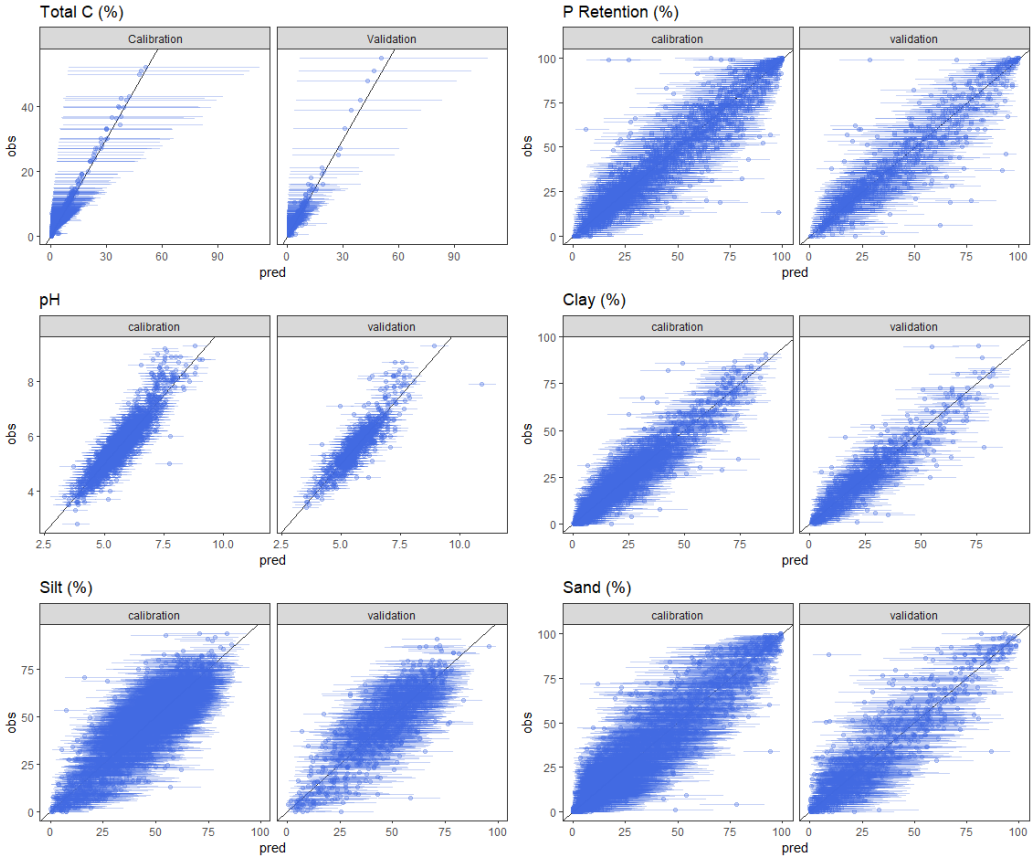


**Figure 7:** CNN prediction interval

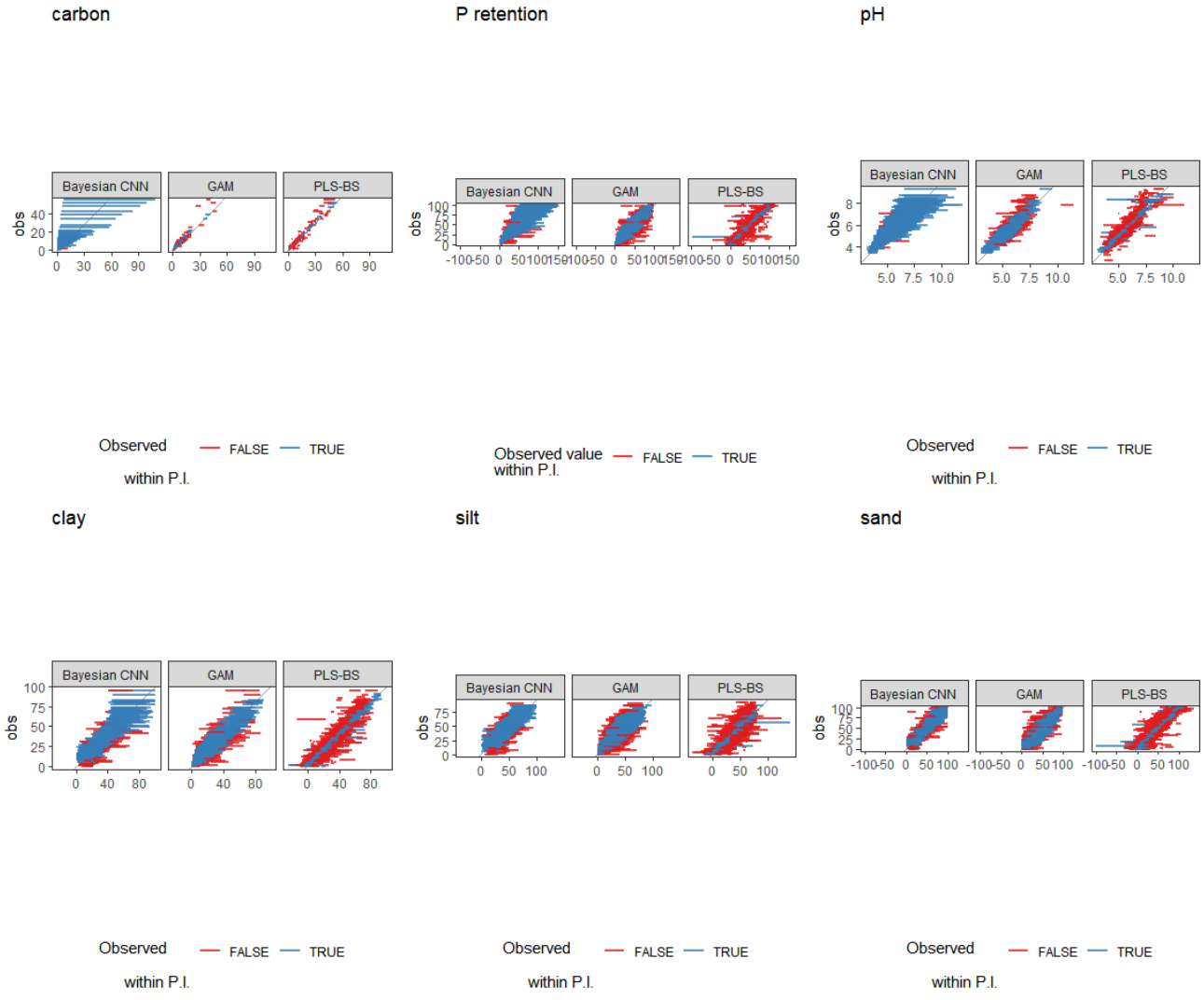


**Figure 8:** Quality of the prediction intervals

**Table 5**
Prediction Interval Coverage Probability (PCIP, %) for the different variables modelled, on the validation set. The target is a 90% prediction interval

| Method | P Retention | Total Carbon | Clay | pH | Sand | Silt |
|---|---|---|---|---|---|---|
| GAM | 91.5 | 92.8 | 91.2 | 88.8 | 88.6 | 87.7 |
| Boostraped PLS | 32 | 33.5 | 30.8 | 28.8 | 28.8 | 26.5 |
| Bayesian CNN | 92.1 | 91.7 | 93.2 | 97.2 | 91.8 | 92.3 |

## 4. Discussion

In this paper, we systematically assessed the predictive precision of two statistical models (PLS-BS and GAM) and a deep learning model (BCNN) with the aim of quantifying uncertainty. Our investigation into uncertainty quantification methods revealed distinct strengths and limitations for each approach, aligning with findings from previous studies using different methods (Padarian et al., 2022; Takoutsing and Heuvelink, 2022). In general, the predictive performance of

the PLS-BS and GAM methods were very similar in all of the six soil properties with a slight, non-significant, higher error for the PLS-BS method as shown in Table 2. While PLS has demonstrated its effectiveness in capturing relationships between predictor variables and response variables in soil property prediction (Nanni et al., 2018), it faces challenges in accurately estimating uncertainty. In contrast, this study showcased the flexibility and effectiveness of GAM in modelling non-linear relationships. The success of GAMs in providing reliable estimates of prediction uncertainty, achieved through beta regression approach has been investigated in other areas (Liu et al., 2021; Clark and Wells, 2023). In observing the prediction interval coverage probability (PCIP, %) for the six soil properties in Table 8, we could see the GAM model outperformed the PLS-BS for all six properties.

Bayesian Convolutional Neural Networks (BCNNs), as demonstrated in recent studies (Zhu and Zabaras, 2018; Gawlikowski et al., 2023), have emerged as powerful tools for accurate predictions and principled uncertainty estimation. By leveraging deep learning techniques and Bayesian principles, BCNNs showcase their ability to capture complex patterns and generate probabilistic predictions with quantified uncertainty. Across all six properties, BCNNs consistently outperformed GAM and PLS-BS regression in terms of predictive performance and prediction interval coverage probability (PCIP), except for carbon, where the carbon PCIP was slightly higher than that of the BCNN. Additionally, the uncertainty quantification capabilities (PCIP) were much closer to that matched of BCNNs across all six properties. From these results, it is clear that the BCNNs and GAMs model are much more certain about what it does not know. In all six properties, we were able to achieeve a 90% prediction interval across all six properties.

### 4.1. Risks to Experimental Validity

While our study aimed to provide valuable insights into the comparison of PLS regression, GAMs, and BCNNs for uncertainty quantification in predicting soil properties using mid-infrared (MIR) spectra, there are certain risks to the experimental validity that should be considered.

1. **Selection Bias:** The selection of the soil samples and the specific MIR spectra used in the study may introduce bias. It is essential to ensure a representative and diverse sample set to obtain reliable and generalisable results.
2. **Model Assumptions:** Each method, including PLS regression, GAMs, and BCNNs, relies on specific

assumptions. Violations of these assumptions may impact the performance and validity of the results. Careful consideration and assessment of these assumptions are crucial.

3. **Hyperparameter Tuning:** The performance of machine learning models, such as GAMs and BCNNs, heavily relies on the selection and tuning of hyperparameters. Improper hyperparameter settings could affect the results and lead to suboptimal performance.
4. **Overfitting:** Overfitting occurs when a model learns the training data too well, resulting in poor generalization to unseen data. It is important to employ appropriate regularisation techniques and cross-validation methods to mitigate the risk of overfitting and ensure the robustness of the models.
5. **Data Quality and Preprocessing:** The quality of the MIR spectra data, including potential noise, missing values, and outliers, can impact the performance and validity of the models. Careful data preprocessing steps, such as data cleaning, normalization, and feature selection, should be applied to ensure data quality and reduce bias.
6. **Generalisability:** The findings of this study may be specific to the dataset and experimental setup used. Although we modelled six different soil types, it is still important to consider the generalisability of the results to other soil types, geographic regions, or environmental conditions.

Addressing and mitigating these risks to experimental validity is crucial for ensuring the reliability and credibility of the study's conclusions. By acknowledging these potential limitations and taking appropriate measures, future research can further enhance the understanding and application of uncertainty quantification methods for soil property prediction using MIR spectroscopy.

## 4.2. Future work

While this study provides valuable insights into the comparison of PLS regression, GAMs, and BCNNs for uncertainty quantification in predicting soil properties using mid-infrared (MIR) spectra, there are several avenues for future research to explore:

1. **Hybrid Approaches:** Future research can investigate the development of hybrid approaches that leverage the strengths of different methods. For example, combining PLS regression with GAMs or BCNNs may

yield improved accuracy and uncertainty estimation capabilities.

2. **Alternative Techniques:** Exploring alternative techniques beyond PLS regression, GAMs, and BCNNs can provide additional insights. Other machine learning algorithms, such as support vector machines or random forests, could be evaluated for their performance in uncertainty quantification.
3. **Ensemble Methods:** Ensemble methods, which combine multiple models to improve predictions, could be explored to further enhance the accuracy and reliability of soil property prediction. Building an ensemble of diverse models, including PLS regression, GAMs, BCNNs, and others, may result in more robust predictions and uncertainty estimates.
4. **External Validation:** Conducting external validation of the models on independent datasets from different geographical regions or soil types is crucial to assess their generalisability. This would provide a more comprehensive evaluation of the methods and their performance across diverse soil contexts.
5. **Model Interpretability:** Developing methods to interpret and explain the predictions and uncertainties generated by the models can enhance their practical applicability. Understanding the factors influencing soil property predictions can assist in decision-making processes related to land management and soil characterization.

These avenues of investigation have the potential to improve the accuracy, reliability, and interpretability of models, contributing to more informed decision-making and sustainable land management practices.

## 5. Conclusion

In this study, we compared Partial Least-Squares (PLS) regression, Generalized Additive Models (GAMs), and Bayesian Convolutional Neural Networks (BCNNs) for quantifying uncertainty in predicting soil properties using mid-infrared (MIR) spectra. Our findings revealed that PLS regression tends to underestimate uncertainties, while GAMs and BCNNs provide more reliable estimates. BCNNs, in particular, offer accurate predictions and direct estimation of uncertainty through posterior distributions. The choice of method depends on computational resources and the desired balance between efficiency and accuracy. This study contributes to soil analysis techniques, providing insights for researchers and practitioners in land management. In conclusion, by

improving our understanding of the strengths and limitations of different methods for uncertainty quantification, we can enhance the accuracy and reliability of soil property predictions, supporting sustainable land management practices and informed decision-making processes in various domains.

## Declaration of competing interest

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

## References


Bellon-Maurel, V., McBratney, A., 2011. Near-infrared (nir) and mid-infrared (mir) spectroscopic techniques for assessing the amount of carbon stock in soils–critical review and research perspectives. Soil Biology and Biochemistry 43, 1398–1410.

Bisong, E., Bisong, E., 2019. Introduction to scikit-learn. Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners , 215–229.

Carré, F., McBratney, A.B., Mayr, T., Montanarella, L., 2007. Digital soil assessments: Beyond dsm. Geoderma 142, 69–79.

Chavez-Demoulin, V., Davison, A.C., 2005. Generalized additive modelling of sample extremes. Journal of the Royal Statistical Society: Series C (Applied Statistics) 54, 207–222.

Clark, N.J., Wells, K., 2023. Dynamic generalised additive models (dgams) for forecasting discrete ecological time series. Methods in Ecology and Evolution 14, 771–784.

Fahrmeir, L., Lang, S., 2001. Bayesian inference for generalized additive mixed models based on markov random field priors. Journal of the Royal Statistical Society Series C: Applied Statistics 50, 201–220.

Frenklach, M., Packard, A., Garcia-Donato, G., Paulo, R., Sacks, J., 2016. Comparison of statistical and deterministic frameworks of uncertainty quantification. SIAM/ASA Journal on Uncertainty Quantification 4, 875–901.

Gal, Y., Ghahramani, Z., 2016. Dropout as a bayesian approximation: Representing model uncertainty in deep learning, in: international conference on machine learning, PMLR. pp. 1050–1059.

Gawlikowski, J., Tassi, C.R.N., Ali, M., Lee, J., Humt, M., Feng, J., Kruspe, A., Triebel, R., Jung, P., Roscher, R., et al., 2023. A survey of uncertainty in deep neural networks. Artificial Intelligence Review , 1–77.

Hastie, T.J., 2017. Generalized additive models, in: Statistical models in S. Routledge, pp. 249–307.

Hastings, W.K., 1970. Monte carlo sampling methods using markov chains and their applications .

Heuvelink, G., Webster, R., 2001. Modelling soil variation: past, present, and future. Geoderma 100, 269–301.

Hoffman, M.D., Blei, D.M., Wang, C., Paisley, J., 2013. Stochastic variational inference. Journal of Machine Learning Research .

Janssen, P., Heuberger, P., 1995. Calibration of process-oriented models. Ecological Modelling 83, 55–66.

Kendall, A., Gal, Y., 2017. What uncertainties do we need in bayesian deep learning for computer vision? Advances in neural information processing systems 30.

Knox, N.M., Grunwald, S., McDowell, M., Bruland, G.L., Myers, D., Harris, W., 2015. Modelling soil carbon fractions with visible near-infrared (vnir) and mid-infrared (mir) spectroscopy. Geoderma 239, 229–239.

Konishi, S., Kitagawa, G., 2008. Information criteria and statistical modeling .

Kuhn, M., Johnson, K., et al., 2013. Applied predictive modeling. volume 26. Springer.

Lakshminarayanan, B., Pritzel, A., Blundell, C., 2017. Simple and scalable predictive uncertainty estimation using deep ensembles. Advances in neural information processing systems 30.

Lawrence, I., Lin, K., 1989. A concordance correlation coefficient to evaluate reproducibility. Biometrics , 255–268.

Liu, J., Chai, L., Dong, J., Zheng, D., Wigneron, J.P., Liu, S., Zhou, J., Xu, T., Yang, S., Song, Y., et al., 2021. Uncertainty analysis of eleven multisource soil moisture products in the third pole environment based on the three-corned hat method. Remote sensing of environment 255,

112225.

Lubke, G.H., Campbell, I., McArtor, D., Miller, P., Luningham, J., van den Berg, S.M., 2017. Assessing model selection uncertainty using a bootstrap approach: an update. Structural Equation Modeling: A Multidisciplinary Journal 24, 230–245.

Martens, H., Naes, T., 1992. Multivariate calibration. John Wiley & Sons.

McCarty, G., Reeves, J., Reeves, V., Follett, R., Kimble, J., 2002. Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil carbon measurement .

Nanni, M.R., Cezar, E., Silva Junior, C.A.d., Silva, G.F.C., da Silva Gualberto, A.A., 2018. Partial least squares regression (plsr) associated with spectral response to predict soil attributes in transitional lithologies. Archives of Agronomy and Soil Science 64, 682–695.

Neal, R.M., 2012. Bayesian learning for neural networks. volume 118. Springer Science & Business Media.

Omondiagbe, O.P., Lilburne, L., Licorish, S.A., MacDonell, S.G., 2023a. Soil texture prediction with automated deep convolutional neural networks and population-based learning. Geoderma 436, 116521. URL: https://www.sciencedirect.com/science/article/pii/S0016706123001982, doi:https://doi.org/10.1016/j.geoderma.2023.116521.

Omondiagbe, O.P., Lilburne, L., Licorish, S.A., MacDonell, S.G., 2023b. Soil texture prediction with automated deep convolutional neural networks and population-based learning. Geoderma 436, 116521.

Padarian, J., Minasny, B., McBratney, A., 2022. Assessing the uncertainty of deep learning soil spectral models using monte carlo dropout. Geoderma 425, 116063.

Parker-Holder, J., Nguyen, V., Roberts, S.J., 2020. Provably efficient online hyperparameter optimization with population-based bandits. Advances in neural information processing systems 33, 17200–17211.

Psaros, A.F., Meng, X., Zou, Z., Guo, L., Karniadakis, G.E., 2023. Uncertainty quantification in scientific machine learning: Methods, metrics, and comparisons. Journal of Computational Physics , 111902.

Rinnan, Å., Van Den Berg, F., Engelsen, S.B., 2009. Review of the most common pre-processing techniques for near-infrared spectra. TrAC Trends in Analytical Chemistry 28, 1201–1222.

Rossel, R.V., Walvoort, D., McBratney, A., Janik, L.J., Skjemstad, J., 2006. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma 131, 59–75.

Sanderman, J., Savage, K., Dangal, S.R., 2020. Mid-infrared spectroscopy for prediction of soil health indicators in the united states. Soil Science Society of America Journal 84, 251–261.

dos Santos, U.J., de Melo Dematte, J.A., Menezes, R.S.C., Dotto, A.C., Guimarães, C.C.B., Alves, B.J.R., Primo, D.C., Sampaio, E.V.d.S.B., 2020. Predicting carbon and nitrogen by visible near-infrared (vis-nir) and mid-infrared (mir) spectroscopy in soils of northeast brazil. Geoderma Regional 23, e00333.

Takoutsing, B., Heuvelink, G.B., 2022. Comparing the prediction performance, uncertainty quantification and extrapolation potential of regression kriging and random forest while accounting for soil measurement errors. Geoderma 428, 116192.

Tobias, R.D., et al., 1995. An introduction to partial least squares regression, in: Proceedings of the twentieth annual SAS users group international conference, Citeseer.

Wahba, G., 1990. Spline models for observational data. SIAM.

Wang, Y., Wahba, G., 1995. Bootstrap confidence intervals for smoothing splines and their comparison to bayesian confidence intervals. Journal of Statistical Computation and Simulation 51, 263–279.

Wood, S.N., 2001. mgcv: Gams and generalized ridge regression for r. R news 1, 20–25.

Wood, S.N., 2017. Generalized additive models: an introduction with R. CRC press.

Zhu, Y., Zabaras, N., 2018. Bayesian deep convolutional encoder–decoder networks for surrogate modeling and uncertainty quantification. Journal of Computational Physics 366, 415–447.