CTU Journal of Innovation and Sustainable Development
ISSN 2588-1418 | e-ISSN 2815-6412

DOI:10.22144/ctujoisd.2025.005

Digital mapping of soil electrical conductivity for paddy field

Ho The Anh[1,2], Luu Trong Hieu[1], and Nguyen Chi Ngon[1*]

[1]*Faculty of Automation Engineering, College of Engineering, Can Tho University, Viet Nam*

[2]*Faculty of Mechanical Engineering, Can Tho University of Technology, Viet Nam*

**Corresponding author (ncngon@ctu.edu.vn)*

Article info.

Received 1 Aug 2024
Revised 5 Sep 2024
Accepted 26 Feb 2025

Keywords

CNN, digital soil mapping, KNN, Kriging, IDW, soil electrical conductivity

ABSTRACT

The digital soil electrical conductivity (EC) map has been widely applied in agriculture globally due to its ability to explain various soil characteristics. However, the Mekong Delta lacks comprehensive data on soil EC. This study aims to address this gap by using the common interpolation method —K-Nearest Neighbors (KNN), Inverse Distance Weighting (IDW), Kriging interpolation, and Convolutional Neural Networks (CNN)—to map soil EC over an area of approximately 1.4 hectares. Using 228 data samples, the study found that the Gaussian model within Kriging was the most effective for interpolating soil EC, achieving the highest R-squared values (0.79 with test data and 0.96 with full data) and the lowest RMSE values (0.049 with test data and 0.022 with full data). Additionally, GPS data collection using the U-blox ZED-F9P-01B GPS module, paired with the U-blox ANN-MB-00 antenna, yielded better accuracy and reliability under rice field conditions (Q=1) compared to the performance in orchard settings. This research provides valuable insights into soil management and agricultural practices in the Mekong Delta.

1. INTRODUCTION

Digital soil maps (DSM) have become a popular tool in modern agriculture and can be applied on both local and global scales (Wadoux et al., 2020). Field mapping to monitor temporal variability is essential for various soil assessments, such as planning soil sampling to evaluate soil quality, and site-specific soil management and remediation strategies (Corwin & Scudiero, 2020). Soil conductivity is an important indicator in agriculture, related to many soil attributes, such as salt content, moisture, and soil texture (Corwin & Plant, 2005). In non-saline conditions, EC primarily reflects soil properties such as fertilizer content, nutrients, and chemical characteristics, providing valuable insights for soil and crop management (Corwin & Lesch, 2005; Rhoades & Corwin, 1990). The EC value in soil is not completely stable over time, as it can be influenced by several factors, including fertilization, irrigation, and weather conditions (Corwin & Plant, 2005). There are various methods for measuring soil conductivity in the field, with the use of electrodes being common due to its simplicity and high reliability (Corwin & Scudiero, 2020; Olaojo & Oladunjoye, 2022). Numerous studies have focused on mapping soil conductivity and its applications in modern agriculture.

Inverse Distance Weighting (IDW) is a spatial interpolation technique widely used to estimate values at unknown locations based on known sample points, under the assumption that values closer to the estimated point have greater influence than those farther away (Ismain et al., 2023). This method has been applied in various research contexts, such as mapping particulate matter in megacities like Delhi (Shukla et al., 2020) and the

Klang Valley (Ismain et al., 2023), as well as in characterizing soil pollution, including the spatial distribution of chromium in soil (Chen et al., 2023) and lead in urban soil (Magno & Budianta, 2023). Studies suggest that IDW is particularly suitable for research with a limited number of samples. However, most studies are conducted on large areas and have not demonstrated effectiveness in smaller areas, such as fields in the Mekong Delta.

Kriging interpolation techniques, a widely used geostatistical method in spatial data analysis, play a crucial role in estimating values at unsampled locations based on known data points. In environmental science and agronomy, Kriging has been extensively applied to map soil nutrients, predict soil properties, and assess spatial variability. For example, Sharma et al. (2022) used Kriging interpolation to map soil nutrients in central India, optimizing prediction errors through cross-validation iterations. Similarly, studies by Yuan et al. (2022) demonstrated the utility of ordinary Kriging in interpolating spatial variations of soil contaminants at polluted sites, showing its effectiveness in environmental monitoring and soil pollution research. Although this method is effective, it is necessary to choose an appropriate variogram model for each different subject.

K-Nearest Neighbors (KNN) is a widely used algorithm in various fields, such as materials science, optics, chemistry, remote sensing, and computer science. This method involves finding the K-nearest data points in feature space to make predictions or classifications. KNN has been used to map wildfire areas and burned regions, demonstrating its effectiveness in environmental monitoring applications (Pacheco et al., 2021). Furthermore, KNN has been employed to assess tree cover density in Sri Lanka, although the Support Vector Machine (SVM) method using Sentinel-2 satellite imagery showed promising results with high accuracy (Premakantha et al., 2023). Similar to the two methods presented above, most studies are conducted in large-area conditions but have not demonstrated effectiveness in small-scale area conditions.

In various fields, Convolutional Neural Networks (CNNs) have become a powerful tool because of their ability to efficiently extract features from data. In remote sensing, One-dimension (1D) CNNs have been successfully used to assess flood susceptibility, showing high accuracy in predicting such events, as demonstrated in a study conducted on the island of Euboea, Greece (Tsangaratos et al., 2023). This study confirmed that 1D CNNs provide higher accuracy (0.924) compared to LR (0.904) and DLNN (0.899) (Tsangaratos et al., 2023). Although convolutional neural networks have proven to be highly effective under large data conditions, it is necessary to conduct research on small-scale field plots to compare their effectiveness with traditional models.

The purpose of this study is to apply KNN, IDW, Kriging, and CNN methods in mapping soil EC on a small scale under cultivation conditions in the Mekong Delta. Each method will be implemented with various configurations to select the best model for each method. Consequently, the most effective solution for mapping soil EC under these conditions will be proposed. Additionally, GPS data collection will be tested and applied using the U-blox ZED-F9P-01B GPS module and the U-blox ANN-MB-00 antenna under real conditions.

2. MATERIALS AND METHOD

2.1. Observation area and devices

The study was conducted in January 2024 in Thoi Lai district (Can Tho city) (Figure 1-a) on a rice cultivation area of 35 x 400 *m*. Soil conductivity was measured using the Wenner method with an electrode spacing of 10 cm (Ho et al., 2024). GPS coordinates were collected based on the Cors station reference method using the U-blox ZED-F9P-01B-00 GPS sensor module, U-blox ANN-MB-00-00 antenna (Figures 1-c), and the RTKLIB version “Demo 5 b34” library package, which has been tested in previous studies (Tran et al., 2023). In this study, the grid sampling method (Carter & Gregorich, 2007) was implemented as follows:

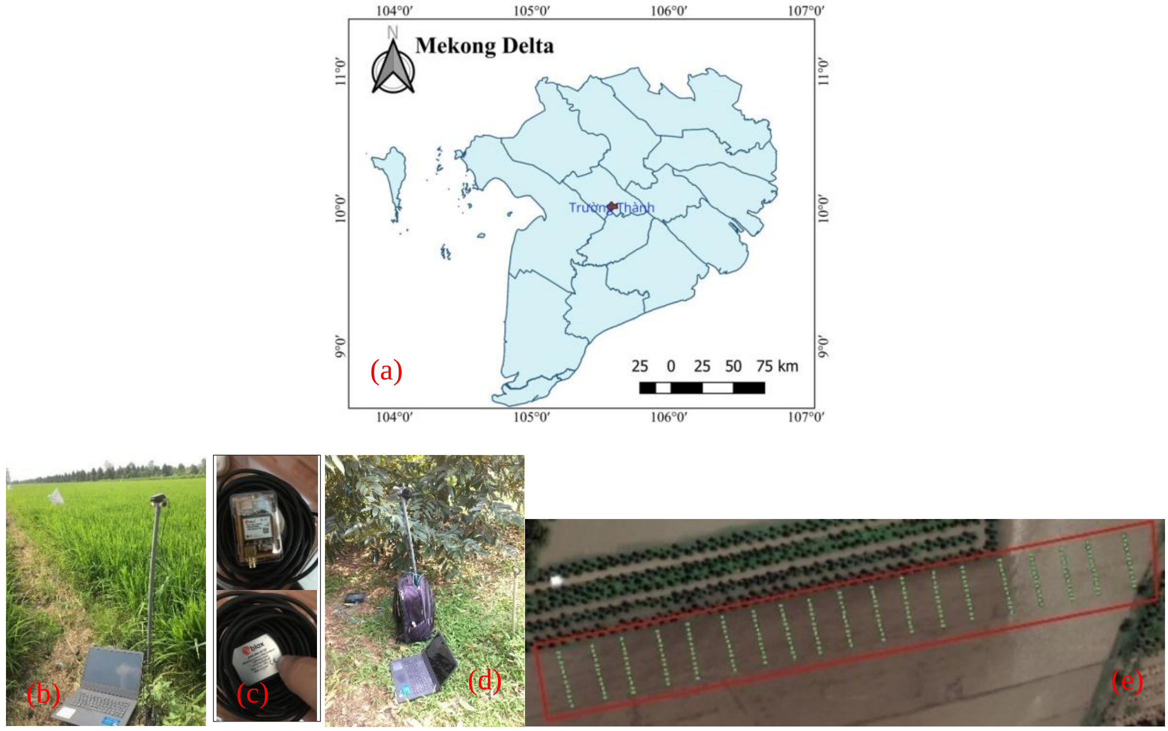

Figure 1. a) – Study area. b) – Collecting GPS data in a rice field. c) – U-blox ZED-F9P-01B GPS module and U-blox ANN-MB-00 antenna. d) - Collecting GPS data in a fruit orchard. e) - Sampling point

– GPS data were collected and accuracy assessed under rice field and fruit orchard conditions with an antenna height of 1.5 m (Figures 1-b, 1-d).

– Fixed points were arranged along both sides of the field to create horizontal rows (Figure 1-e), and GPS coordinates were collected, with 19 horizontal rows established.

– GPS coordinates were corrected so that the horizontal rows had the same latitude.

– Each horizontal row was divided into 12 equal points to determine the location of 228 sampling points. Coordinates at these points were determined based on the two outer points.

– Sampling was conducted at the determined points, and the results were recorded.

– Python software was used to perform calculations (Virtanen et al., 2020) (SciPy and NumPy for numerical calculations) and create charts with four algorithms: KNN in Scikit-learn library (Pedregosa et al., 2012), IDW, Kriging in PyKrige library (Oliver & Webster, 2015), and CNN using TensorFlow library (Chollet, 2021).

2.2. Data processing

2.2.1. Inverse Distance Weighting (IDW)

IDW is a type of deterministic method for multivariate interpolation with a known scattered set of points. The basic idea is that the influence of a known data point decreases with distance. Assume a set of data points (x_i, y_i, EC_i) with *i = 1, 2, 3, …, n*, where (x_i, y_i) are the sampling coordinates and EC_i is the conductivity value at the sampling point. The value $EC(u)$ at location u is estimated using the IDW method as follows (Dooley, 1976)(Liu et al., 2021):

$$EC(u) = \frac{\sum_{i=1}^{n} \frac{EC_i}{d_i^p}}{\sum_{i=1}^{n} \frac{1}{d_i^p}} \quad (1)$$

– d_i is the distance from the interpolation point u to the data point (x_i, y_i).

$$d_i = \sqrt{(x - x_i)^2 + (y - y_i)^2} \quad (2)$$

– p is an exponent (often called the power parameter), which indicates the influence of distance, usually chosen as $p = 2$

– $\sum_{i=1}^{n} \frac{EC_i}{d_i^p}$ is the sum of the EC values of the data points, each divided by the distance from the interpolation point to the data point raised to the power of p. Points closer to the interpolation point will have a larger weight.

– $\sum_{i=1}^{n} \frac{1}{d_i^p}$ is the sum of the inverse distance weights, used to normalize the sum of the values in the numerator.

2.2.2. K-Nearest Neighbors (KNN)

In the context of spatial interpolation, KNN is applied to predict values at unknown locations based on the values of the nearest known points (Peterson, 2009). Assume a set of data points (x_i, y_i, EC_i) with *i = 1, 2, 3, …, n,* where (x_i, y_i) are the sampling coordinates and EC_i is the conductivity value at the sampling point. The value $EC(u)$ at location u is estimated using the KNN method as follows:

$$EC(u) = \frac{1}{k} \sum_{j=1}^{k} EC_j \tag{3}$$

– k is the number of nearest neighbors chosen to estimate the value.

– $z_{(i_j)}$ is the value of the j-th data point among the k nearest neighbors to the location u.

2.2.3. Kriging interpolation

This method allows for the interpolation of values at unsampled locations based on existing data, improving the accuracy of predictions and spatial data analysis (Cressie, 2015)(Chiles & Delfiner, 2012)(Loonis & de Bellefon, 2018). Assume a set of data points (x_i, y_i, EC_i) with *i = 1, 2, 3, …, n,* where (x_i, y_i) are the sampling coordinates and EC_i is the conductivity value is at the sampling point.

The objective of Kriging is to find the weights λ_i to calculate the interpolated value $Z^*(u)$ at a location u:

$$Z^*(u) = \sum_{i=1}^{n} \lambda_i Z(x_i) \tag{4}$$

There are many possible choices for the weights λ_i. Using the second-order stationarity assumption or intrinsic hypothesis, we have:

$$E\left[Z(u)\right] = m \quad \forall m \in D \tag{5}$$

This means for the linear estimator:

$$E\left[Z^*(u)\right] = \sum_{i=1}^{n} \lambda_i E\left[Z(x_i)\right] = m \tag{6}$$

The weights λ_i must satisfy the condition:

$$\sum_{i=1}^{n} \lambda_i = 1 \tag{8}$$

This is called the unbiasedness condition. The goal of Kriging is to minimize the estimation variance $\sigma^2(u)$ based on the covariance model $C(h)$. The estimation variance can be expressed as a function of the covariance:

$$\sigma^2(u) = C(0) + \sum_{j=1}^{n}\sum_{i=1}^{n} \lambda_i \lambda_j C(x_i - x_j) - 2\sum_{j=1}^{n} \lambda_i C(x_i - u) \tag{9}$$

To minimize $\sigma^2(u)$ under the unbiasedness condition, we use the Lagrange multiplier μ. The weights that minimize $\sigma^2(u)$ are the solution to:

$$\sum_{j=1}^{n} \lambda_i C(x_i - x_j) + \mu = C(x_i - u) \tag{10}$$

$$i = 1, 2, 3, \ldots, n$$

This system of equations is called the Kriging system, and the weights λ_i are called Kriging weights. The minimum estimation variance $\sigma_K^2(u)$ can be achieved by substituting the Kriging weights λ_i on equation (9):

$$\sigma_K^2(u) = C(0) + \sum_{j=1}^{n}\sum_{i=1}^{n} \lambda_i \lambda_j C(x_i - x_j) - 2\sum_{j=1}^{n} \lambda_i C(x_i - u) \tag{11}$$

In the specific case of soil conductivity, the Kriging method will be used to interpolate EC values at unsampled locations.

The variogram model describes the spatial correlation of a random variable, such as soil conductivity, and indicates the degree of correlation between data points based on their distance. The variogram is used to understand and model spatial variation patterns in the data.

Some common variogram models include:

Linear: The simplest model, where the semivariance increases linearly with distance.

Exponential: A model where the semivariance increases rapidly and reaches a threshold.

Spherical: A model where the semivariance increases rapidly and then stabilizes at a threshold.

Gaussian: A model where the semivariance increases quadratically and reaches a threshold.

2.2.4. Convolutional Neural Network (CNN)

CNN utilizes convolutional layers and pooling layers to reduce the number of parameters and computations while capturing essential features from the input data (Nielsen, 2015)(Book, 2017).

Convolutional Layer:

In this layer, filters are applied to the input to create feature maps. This process can be described by the following formula:

$$Z_{i,j,k} = \sum_{m=0}^{M-1}\sum_{n=0}^{N-1} X_{i+m,j+n}.\mathrm{W}_{m,n,k} + b_k \tag{12}$$

– $Z_{i,j,k}$ is the value at position *(i, j)* of the k-th channel in the feature map.

– . $Z_{i,j}$ .is the value at position *(i, j)* in the input.

– $\mathrm{W}_{m,n,k}$ is the value at position *(m, n)* of the *k-th* filter.

– b_k is the bias of the *k-th* filter.

– *M* and *N* are the dimensions of the filter.

Activation Function:

After convolution, an activation function like ReLU (Rectified Linear Unit) is often applied to introduce non-linearity:

$$A_{i,j,k} = ReLU(Z_{i,j,k}) = max(0, Z_{i,j,k}) \tag{13}$$

Pooling Layer:

The pooling layer (typically max pooling) reduces the spatial dimensions of the feature map:

$$P_{i,j,k} = \max_{(m,n)\in pool} A_{i+m,j+n,k} \tag{14}$$

– $P_{i,j,k}$ is the value at position *(i, j)* of the k-th channel in the pooled feature map.

The max function is applied over a defined window (pool) size.

Fully Connected Layer:

The output from the previous layers is flattened and fed into one or more fully connected layers:

$$O_j = \sigma\left(\sum_i A_i.W_{ij} + b_j\right) \tag{15}$$

– O_j is the output of the j-th node.

– A_i is the input value from the previous layer.

– W_{ij} is the weight connecting the i-th input to the j-th node.

– b_j is the bias of the j-th node.

– σ is the activation function (typically softmax for the final layer in classification problems).

With the collected data, consisting of 228 samples, they are divided into two subsets:

– The first subset, comprising 70% (about 163 data samples), is used for calculations or model training.

– The second subset, comprising 30% (about 65 data samples), is used for testing.

The evaluation criteria include R-squared, MAE, and RMSE.

R-squared: This metric evaluates the explanatory power of the model. A higher value indicates a better model.

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \tag{16}$$

MAE (Mean Absolute Error): This metric measures the average absolute error between the predicted and the actual values. A lower value indicates a better model.

$$MSE = \frac{1}{n}\sum(y_i - \hat{y}_i)^2 \tag{17}$$

RMSE (Root Mean Squared Error): This metric measures the average squared error between the predicted and the actual values. A lower value indicates a better model.

$$RMSE = \sqrt{\frac{1}{n}\sum(y_i - \hat{y}_i)^2} \tag{18}$$

$\sum(y_i - \hat{y}_i)^2$: This represents the Sum of Squared Errors (SSE), which is the total sum of the squared differences between the observed actual values (y_i) and the predicted values (y_i).

$\sum(y_i - \bar{y})^2$: which is the total sum of the squared differences between the observed actual values (yi) and the mean of the observed values $\left(\bar{y}\right)$.

3. RESULTS AND DISCUSSION

3.1. GPS Data Collection

The image window from the RTKNAVI software demo version 5 b34f.1, RTKPLOT, is a tool for positioning and navigation. In the image, a black point in the center represents the current position. Yellow Path represents the movement path of the positioning device, showing a short recorded route. The X and Y axes are divided into units, with each small grid square corresponding to 1 square meter:

– X-axis usually represents the East-West direction, with positive values indicating eastward displacement and negative values indicating westward displacement.

– Y-axis usually represents the North-South direction, with positive values indicating northward displacement and negative values indicating southward displacement.

The bottom of the screen displays the time and coordinates of the current point: January 18, 2024, at 07:43:36.000 GPST, with coordinates at 10.02447528^{0}N, 105.58124326^{0}E, and an altitude of -3.8267 meters (Figures 2-a). The bottom right corner includes a scale bar indicating 1 meter and a Q=1 indicator, showing that the signal quality is FIX, meaning the highest accuracy, or FLOAT (Q=2), indicating lower accuracy (Figures 2-b,2-c).

Results from rice field experiments showed accurate and consistent GPS coordinates with good signal quality (Q=1), ensuring precise and reliable data. In contrast, orchard experiments showed slight fluctuations in GPS coordinates, likely due to environmental factors affecting the signal. The signal quality only reached FLOAT level (Q=2), indicating lower accuracy, possibly due to tree canopies obstructing the signal compared to the open rice fields.

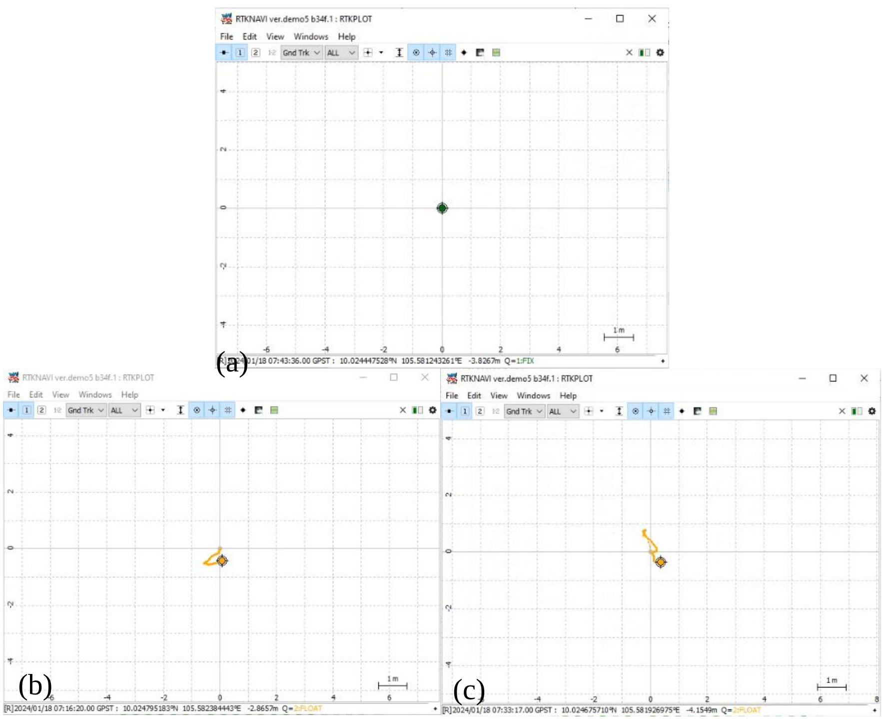

Figure 2. a – GPS Data for Rice Fields; b - GPS in the gap between two tree canopies, c – GPS directly under the tree canopy

3.2. Model Evaluation

3.2.1. KNN method

Figure 3 illustrates the soil electrical conductivity (EC) map generated using the KNN method. The points on the map represent EC values according to geographic coordinates, with different colors indicating different EC levels. Blue represents low EC values, while red indicates high EC values.

With K = 6, the R-squared value reaches its highest on the test set (0.75).

With K = 5, the MAE reaches its lowest on the test set (0.042).

With K = 6, the RMSE reaches its lowest on the test set (0.054).

However, when considering the performance on the full dataset, K = 4 shows better performance with lower R-squared, MAE, and RMSE compared to K = 6 and K = 7.

K = 5 appears to be the best choice as it has the highest R-squared and RMSE on the test set and the lowest MAE on the test set.

K = 6 is also a good choice as it has the highest R-squared and RMSE, but its MAE is not the lowest on the test set.

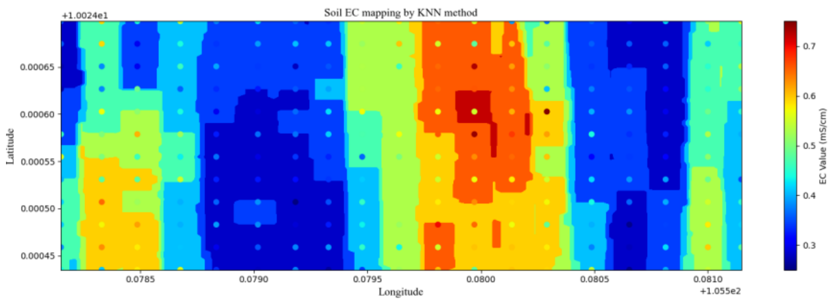

Figure 3. Soil EC map by KNN method

Based on these metrics, we can conclude that K = 5 is the most effective K value for the KNN model, as it provides the best balance between the evaluation metrics on both the test set and the full dataset.

Table 1. Performance Metrics of KNN Model with Different Numbers of Neighbors

Neighbors	Test Data			Full Data		
	R – squared	MAE	RMSE	R – squared	MAE	RMSE
K = 4	0.70	0.044	0.059	0.75	0.043	0.056
K = 5	0.75	0.042	0.054	0.75	0.044	0.057
K = 6	0.75	0.043	0.054	0.72	0.046	0.060
K = 7	0.73	0.046	0.056	0.70	0.048	0.062

3.2.2. IDW method

Figure 4 shows the soil EC map generated via the IDW method. The map uses a color gradient to represent EC values, with blue indicating lower conductivity and red representing higher conductivity. The model's predictive accuracy is reflected in the test set results, where an RMSE of 0.051 demonstrates the model's ability to closely approximate actual values.

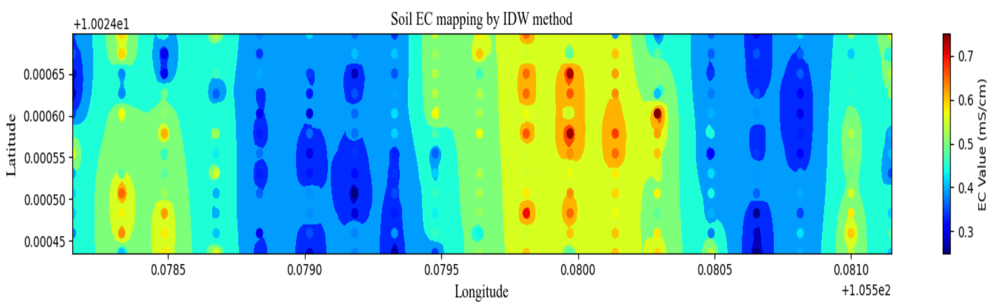

Figure 4. Soil EC map by IDW method

On the full dataset, the RMSE is slightly higher (0.067), indicating a slight difference in prediction capability between the test set and the full dataset. Testing on the full dataset provides an opportunity to assess the model's performance across the entire data range, rather than just a limited subset (the test set). This helps determine whether the model is consistently accurate across all data points.

On the test set, the R-squared value reached 0.77, indicating that the IDW model can explain 77.17% of the data variability. This is a fairly good indicator, showing that the model performs well on the test set.

On the full dataset, the R-squared value decreased to 0.65, indicating that the model explains about 64.53% of the data variability when considering the full dataset. This may suggest that the model might not be entirely suitable for the full dataset or there might be differences in data structure between the test set and the full dataset.

On the test set, the MAE reached a value of 0.041, indicating that the mean absolute error between the predicted values and the actual values is small, demonstrating the high accuracy of the IDW model on the test set.

On the full dataset, the MAE is slightly higher (0.052), indicating a slightly larger mean absolute error when considering the full dataset.

3.2.3. Kriging method

The four variogram models (Figure 5) (Linear, Spherical, Exponential, Gaussian) were applied and compared with the experimental variogram to determine the most appropriate model for capturing the variability in soil EC. The findings revealed that the Linear model was inadequate for this dataset, as it displayed only a linear increase in semivariance and failed to capture the complexity of EC variation over longer distances. In contrast, the Spherical, Exponential, and Gaussian models provided better fits to the experimental data at shorter distances (within 100 meters). Nonetheless, at greater distances, none of the models accurately represented the semivariance variation, particularly the decline after its peak.

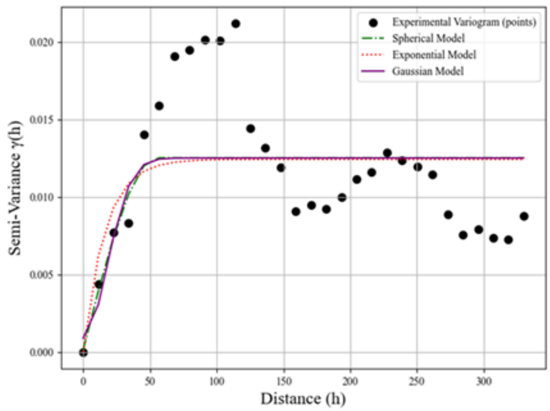

Figure 5. Experimental variograms of soil EC

The soil EC distribution across the study area is depicted in Figure 6, created using the Kriging method. The gradient from blue to red illustrates varying EC levels, with blue denoting lower values and red higher ones. This map underscores the effectiveness of the Gaussian method within the Kriging approach, which achieved superior performance based on R-squared and RMSE metrics.

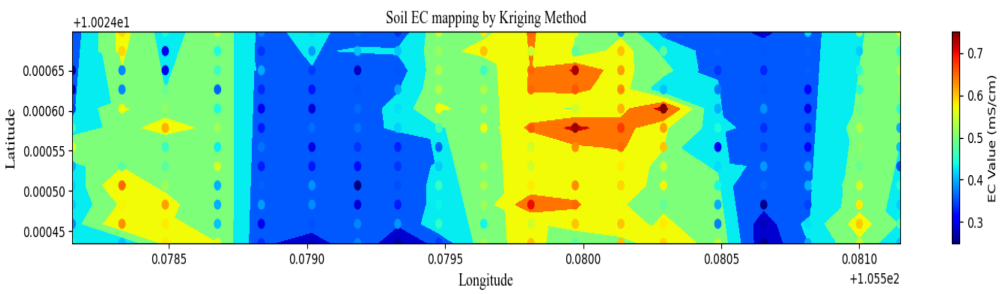

Figure 6. Soil EC map by Kriging method

The Linear method shows a negative R-squared (-0.014) when tested with the test data, indicating that this model is not suitable for soil EC interpolation.

The Exponential, Spherical, and Gaussian methods all show much better results with both test data and full data. Among them, the Gaussian method has the highest R-squared and the lowest RMSE, indicating it is the best model for EC interpolation. The highest R-squared (0.79 with test data and 0.96 with full data) and the lowest RMSE (0.049 with test data and 0.022 with full data) are achieved with this method.

Table 2. Performance metrics of different variogram models

Variogram	Test Data			Full Data		
Model	**R – squared**	**MAE**	**RMSE**	**R – squared**	**MAE**	**RMSE**
Linear	-0.014	0.092	0.11	0.82	0.019	0.049
Exponential	0.77	0.040	0.051	0.96	0.0081	0.023
Spherical	0.79	0.042	0.049	0.96	0.0086	0.022
Gaussian	0.79	0.040	0.049	0.96	0.0080	0.022

3.2.4. CNN method

Figure 7 depicts the spatial variation of soil EC generated by the CNN method. A color gradient from blue to red is used to represent different levels of EC, where blue corresponds to lower values and red to higher ones. The CNN model's performance is particularly validated at 140 epochs where the balance between R-squared, MAE, and RMSE is optimal. CNN model is designed to predict soil EC values using latitude and longitude as inputs. The model uses a convolutional layer (Conv2D) with 64 filters and a kernel size of (1, 2). After convolution, the output is flattened and passed through two dense layers, where the final layer has one output neuron for regression.

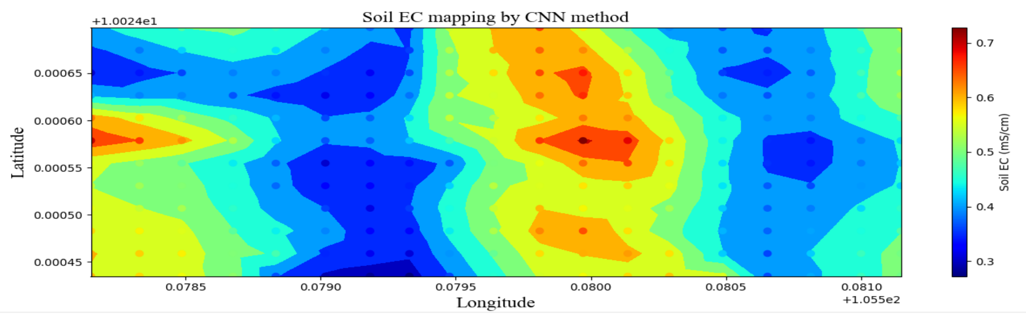

Figure 7. Soil EC map by CNN method

Test Data: The highest R-squared value is achieved at 130 epochs (0.39) with MAE and RMSE of 0.56 and 0.74, respectively. However, at 140 epochs, R-squared slightly decreases but remains relatively high (0.37), while MAE and RMSE do not differ much from those at 130 epochs. At 150 epochs, R-squared significantly drops (0.21), indicating that increasing the number of epochs may lead to overfitting.

Full Data: At 140 epochs, R-squared reaches the highest value (0.70) with MAE and RMSE of 0.36 and 0.55, respectively. This is the best result compared to other epoch numbers, indicating that the model performs best when trained with 140 epochs.

Table 3. Performance Metrics of CNN Model with Different Epochs

Epochs	Test Data			Full Data		
	R – squared	**MAE**	**RMSE**	**R – squared**	**MAE**	**RMSE**
120	0.31	0.57	0.79	0.68	0.36	0.57
130	0.39	0.56	0.74	0.69	0.36	0.56
140	0.37	0.56	0.75	0.70	0.36	0.55
150	0.21	0.61	0.84	0.64	0.60	0.37

3.3. Overall evaluation

With the test data, KNN has a fairly good R-squared (0.51) and relatively low MAE and RMSE (0.072 and 0.090). However, with the full data, the R-squared slightly decreases (0.66), but the MAE and RMSE remain acceptable. IDW shows good results with the test data, with R-squared (0.56), MAE (0.067), and RMSE (0.085). However, when using the full data, the R-squared significantly decreases (0.51), and both MAE and RMSE increase (0.070 and 0.090).

Kriging provides the best results among all methods with the full data, having a very high R-squared (0.91) and very low MAE and RMSE (0.013 and 0.038). With the test data, Kriging still gives good results with R-squared (0.57), MAE (0.067), and RMSE (0.084).

CNN performs worse than the other methods with the test data, having a low R-squared (0.37) and high MAE and RMSE (0.56 and 0.75). However, with the full data, CNN significantly improves the R-squared (0.70), but the MAE and RMSE remain higher compared to traditional methods as Kriging.

Table 4. Performance Metrics of Different Methods

Methods	Test Data			Full Data		
	R – squared	MAE	RMSE	R – squared	MAE	RMSE
KNN	0.51	0.072	0.090	0.66	0.058	0.075
IDW	0.56	0.066	0.085	0.51	0.070	0.090
Kriging	0.79	0.040	0.049	0.96	0.0080	0.022
CNN	0.37	0.56	0.75	0.70	0.36	0.55

Kriging is the best method for predicting soil EC. This method provides the best results with the full data, having high R-squared and low MAE and RMSE. Even with the test data, Kriging still delivers more reliable and accurate results compared to other methods.

3.4. Discussion

The results indicate that the surrounding environment significantly affects GPS signal accuracy, with rice fields providing more stable signals than orchards. Overall, the Kriging model is identified as the best method for predicting soil electrical conductivity, exhibiting the highest and most consistent evaluation metrics on both the test set and the full dataset.

The experiments were conducted in specific areas, so the results may not represent all soil types and environmental conditions. Additionally, the models used may be influenced by input data, potentially leading to differences in prediction outcomes.

Future studies should expand the range of soil types and environmental conditions to verify these results. Combining different methods or developing new models could also enhance the accuracy of soil conductivity predictions. Further research should consider factors such as weather and seasonal effects on GPS signal quality and prediction results.

4. CONCLUSION

In this study, we evaluated the performance of various interpolation and prediction methods for estimating soil EC. The methods examined included KNN, IDW, Kriging, and CNN. The Kriging method emerged as the superior model for soil EC interpolation. Even with the test data, Kriging showed reliable performance with an R-squared value of 0.57, MAE of 0.067, and RMSE of 0.084. With the full dataset, Kriging achieved the highest R-squared value (0.91) and the lowest MAE (0.014) and RMSE (0.038). This indicates that Kriging is highly effective in explaining the variability in soil EC and minimizing prediction errors.

The GPS data collected using the U-blox ZED-F9P-01B module and U-blox ANN-MB-00 antenna showed high accuracy and stability in open environments like rice fields (FIX quality). However, in more complex environments like orchards, the data exhibited variation and lower signal quality (FLOAT), likely due to obstacles like tree canopies and terrain. This indicates that while the GPS device performs well in open areas, it requires improvements for more accurate data in complex environments.

CONFLICT OF INTEREST

This research was conducted with self-funded support and does not conflict with any individual or organization's interests.

ACKNOWLEDGMENT

We would like to express our sincere gratitude to Associate Professor Dr. Nguyễn Chánh Nghiệm (Can Tho University) for allowing us to use research equipment during the execution of this project. The support and sharing from Associate Professor Dr. Nguyễn Chánh Nghiệm have not only been a great source of encouragement but have also helped us successfully conduct this research.

REFERENCES

Book, D. L. (2017). *Ian Goodfellow, Yoshua Bengio, Aaron Courville*. MIT Press. http://www.deeplearningbook.org-March.

Carter, M. R., & Gregorich, E. G. (2007). *Soil sampling and methods of analysis*. CRC press.

Chen, J., Han, C., Peng, Y., Wang, M., & Zhao, Y. (2023). Improved three-dimensional mapping of soil chromium pollution with sparse borehole data: Incorporating multisource auxiliary data into IDW-based interpolation. *Soil Use and Management*, *39*(2), 933–947. https://doi.org/https://doi.org/10.1111/sum.12899

Chiles, J.-P., & Delfiner, P. (2012). *Geostatistics: modeling spatial uncertainty* (Vol. 713). John Wiley & Sons.

Chollet, F. (2021). *Deep learning with Python*. Simon and Schuster.

Corwin, D. L., & Lesch, S. M. (2005). Apparent soil electrical conductivity measurements in agriculture. *Computers and Electronics in Agriculture*, *46*(1), 11–43. https://doi.org/https://doi.org/10.1016/j.compag.2004.10.005

Corwin, D. L., & Plant, R. E. (2005). Applications of apparent soil electrical conductivity in precision agriculture. *Computers and Electronics in Agriculture*, *46*(1), 1–10. https://doi.org/https://doi.org/10.1016/j.compag.2004.10.004

Corwin, D. L., & Scudiero, E. (2020). Field-scale apparent soil electrical conductivity. *Soil Science Society of America Journal*, *84*(5), 1405–1441. https://doi.org/https://doi.org/10.1002/saj2.20153

Cressie, N. (2015). *Statistics for spatial data*. John Wiley & Sons.

Dooley, J. C. (1976). Two-dimensional interpolation of irregularly spaced data using polynomial splines. *Physics of the Earth and Planetary Interiors*, *12*(2), 180–187. https://doi.org/https://doi.org/10.1016/0031-9201(76)90046-7

Ho, T. A., Bui, V. H., Nguyen, V. K., Nguyen, V. K., & Nguyen, C. N. (2024). An application of soil electrical conductivity measurement by Wenner Method in paddy field. *International Journal of Engineering Trends and Technology*, *72*(2), 58–68. https://doi.org/10.14445/22315381/IJETT-V72I2P107

Ismain, S. H. A., Salleh, S. A., Mohammad Sham, N., Wan Azmi, W. N. F., Zulkiflee, A. L., & Ab Rahman, A. Z. (2023). Spatial distribution of particulate matter (PM2.5) in Klang Valley using Inverse Distance Weighting Interpolation Model. *IOP Conference Series: Earth and Environmental Science*, *1217*(1), 12033. https://doi.org/10.1088/1755-1315/1217/1/012033

Liu, Z.-N., Yu, X.-Y., Jia, L.-F., Wang, Y.-S., Song, Y.-C., & Meng, H.-D. (2021). The influence of distance weight on the inverse distance weighted method for ore-grade estimation. *Scientific Reports*, *11*(1), 2689. https://doi.org/10.1038/s41598-021-82227-y

Loonis, V., & de Bellefon, M. P. (2018). Handbook of Spatial Analysis: Theory and practical application with R. *Insee Méthodes*, *131*(2).

Magno, J. L., & Budianta, W. (2023). Land transportation influence on the spatial distribution of Lead (Pb) in urban soils of Yogyakarta, Indonesia. *IOP Conference Series: Earth and Environmental Science*, *1233*(1), 12038. https://doi.org/10.1088/1755-1315/1233/1/012038

Nielsen, M. A. (2015). *Neural networks and deep learning* (Vol. 25). Determination Press San Francisco, CA, USA.

Olaojo, A. A., & Oladunjoye, M. A. (2022). Field-scale apparent electrical conductivity mapping of soil properties in precision agriculture. *Brazilian Journal of Geophysics*, *40*(3), 1–31. https://doi.org/10.22564/brjg.v40i3.2171

Oliver, M. A., & Webster, R. (2015). *Basic steps in geostatistics: the variogram and kriging*. Springer.

Pacheco, A. D., Junior, J. A., Ruiz-Armenteros, A. M., & Henriques, R. F. (2021). Assessment of k-Nearest Neighbor and Random Forest Classifiers for Mapping Forest Fire Areas in Central Portugal Using Landsat-8, Sentinel-2, and Terra Imagery. In *Remote Sensing* (Vol. 13, Issue 7). https://doi.org/10.3390/rs13071345

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., & Louppe, G. (2012). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*.

Peterson, L. (2009). K-nearest neighbor. *Scholarpedia*, *4*, 1883. https://doi.org/10.4249/scholarpedia.1883

Premakantha, K., Chandani, R., Silva, G., Gunatilaka, R., & Pushpakumara, D. (Gamini). (2023). *Tree Cover Assessment : Assessment of tree cover density of Sri Lanka using visual interpretation of open-source high-resolution imagery and geographic information system interface mapping*. https://doi.org/10.4038/jnsfsr.v51i4.11429

Rhoades, J. D., & Corwin, D. L. (1990). Soil electrical conductivity: Effects of soil properties and application to soil salinity appraisal. *Communications in Soil Science and Plant Analysis*, *21*(11–12), 837–860. https://doi.org/10.1080/00103629009368274

Sharma, R. P., Chattaraj, S., Jangir, A., Tiwari, G., Dash, B., Daripa, A., & Naitam, R. K. (2022). Geospatial variability mapping of soil nutrients for site specific input optimization in a part of central India. *Agronomy Journal*, *114*(2), 1489–1499. https://doi.org/https://doi.org/10.1002/agj2.21025

Shukla, K., Kumar, P., Mann, G. S., & Khare, M. (2020). Mapping spatial distribution of particulate matter using Kriging and Inverse Distance Weighting at supersites of megacity Delhi. *Sustainable Cities and Society*, *54*, 101997. https://doi.org/https://doi.org/10.1016/j.scs.2019.101997

Tran, N.-T., Huynh-Nhu-Y, N., Nguyen, C.-N., & Nguyen, C.-N. (2023). A Combination of Low-cost, Dual-frequency, Multi-GNSS Receiver and CORS Network for Precise Positioning Applications in Vietnam. *IOP Conference Series: Earth and Environmental Science*, *1170*(1), 12012.

Tsangaratos, P., Ilia, I., Chrysafi, A.-A., Matiatos, I., Chen, W., & Hong, H. (2023). Applying a 1D Convolutional Neural Network in Flood Susceptibility Assessments—The Case of the Island of Euboea, Greece. *Remote Sensing*, *15*(14). https://doi.org/10.3390/rs15143471

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., & Bright, J. (2020). Fundamental algorithms for scientific computing in python and SciPy 1.0 contributors. SciPy 1.0. *Nat. Methods*, *17*, 261–272.

Wadoux, A. M. J.-C., Minasny, B., & McBratney, A. B. (2020). Machine learning for digital soil mapping: Applications, challenges and suggested solutions. *Earth-Science Reviews*, *210*, 103359. https://doi.org/https://doi.org/10.1016/j.earscirev.2020.103359

Yuan, Y., Yang, K., Cheng, L., Bai, Y., Wang, Y., Hou, Y., & Ding, A. (2022). Effect of Normalization Methods on Accuracy of Estimating Low- and High-Molecular Weight PAHs Distribution in the Soils of a Coking Plant. *International Journal of Environmental Research and Public Health*, *19*(23). https://doi.org/10.3390/ijerph192315470