A generative deep neural network as an alternative to co-kriging

Herbert Rakotonirina, Paul Honeine, Olivier Atteia, Antonin van Exem

A generative deep neural network as an alternative to co-kriging

Herbert RAKOTONIRINA[a,c], Paul HONEINE[a], Olivier ATTEIA[b], Antonin VAN EXEM[c]

[a] *Univ Rouen Normandie, INSA Rouen Normandie, Université Le Havre Normandie, Normandie Univ, LITIS UR 4108, F-76000 Rouen, France*
[b] *Univ Bordeaux, CNRS, Bordeaux INP, EPOC, UMR 5805, F-33600 Pessac, France*
[c] *Tellux, 11 Rue du Moulin À Poudre, Maromme, 76150, France*

Abstract

In geosciences, kriging is leading spatial interpolation, and co-kriging is the most commonly used method for accomplishing spatial interpolation of a target variable by incorporating information from a secondary variable. Co-kriging relies on the assumption of spatial stationarity, which may not hold true in all geospatial contexts, leading to potential inaccuracies in interpolation. The effectiveness of co-kriging can be compromised in areas with sparse data, impacting the reliability of interpolated results. Moreover, it can be resource-intensive when used for interpolation with a substantial volume of data, especially in the case of 3D interpolation. In this paper, we introduce a new method for spatial interpolation that considers two variables using a generative deep neural network. This approach utilizes a convolutional neural network with an encoder-decoder architecture, featuring a single encoder and two decoders to handle the two variables. Additionally, we introduce a loss function that facilitates the control over the relationships between the two variables. Traditional Deep Learning methods require prior training and labeled data, whereas the proposed approach eliminates this requirement and simplifies the interpolation process. In order to assess the performance of our method, we use two real-world datasets. The first one is a 2D dataset of total soil organic carbon combined with the Normalized Difference Vegetation Index. The second one is a 3D dataset that combines concentrations of Hydrocarbon and Fluoride obtained from hyperspectral analysis of soil cores with very limited number of boreholes. The experimental results demonstrate that the proposed method outperforms ordinary kriging and co-kriging, showing a significant improvement when both variables are used. We also demonstrate how the inclusion of the auxiliary variable serves as a means to mitigate the overfitting of the model.

Keywords: Co-kriging, Kriging, Deep Learning, Environmental data, Soil pollution, Soil organic carbon, Hydrocarbon, Fluoride

Article published in Applied Computing and Geosciences, Volume 24, December 2024.

1. Introduction

Spatial interpolation is central in geosciences, aiming to estimate and map a variable of interest based on discrete data points. This process is fundamental to creating comprehensive and accurate representations of spatial phenomena. By converting point data into a continuous surface, spatial interpolation allows us to visualize and understand the distribution of a variable across a geographical area. This is particularly crucial in geosciences, where such understanding can inform everything from resource management to environmental conservation efforts. There are various spatial interpolation methods, but kriging remains particularly prominent in geosciences, especially for soil mapping (Bangroo et al., 2020; Gia Pham et al., 2019; Ouabo et al., 2020). Co-kriging is a variant of kriging that takes into account an auxiliary variable to improve the interpolation of the main variable (Matheron, 1963; Journel and Huijbregts, 1976). This auxiliary variable contributes valuable additional spatial information, improving the overall accuracy of predictions in geostatistical analyses. Although the use of co-kriging does not require the presence of the second variable at all prediction locations, it does rely on a correlation between the main variable and the auxiliary variable. For a recent survey on co-kriging, see (Dowd and Pardo-Igúzquiza, 2023).

Co-kriging can be employed for characterizing hydrocarbon reservoir deposits in rocks (Ochie and Rotimi, 2018). It is also used for mapping the distribution of soil chemical and physical properties such as arsenic, lead, heavy metals, pH, hydraulic acidity, porosity, and permeability (Kim et al., 2019; Leenaers et al., 2020; Milillo et al., 2012; Tziachris et al., 2017). The scope of co-kriging applications extends beyond geosciences. For instance, Liu et al. (2022) use it to optimize ship hull forms, Elsayed (2015) for optimizing the geometric separator of a cyclone, and Koziel et al. (2014) for the rapid optimization of multi-objective antennas.

Co-kriging is built upon certain assumptions that may pose challenges in real-world applications, particularly in contexts such as soil pollution assessment.The main assumptions are spatial stationarity (consistent statistical properties across the study area), isotropy (directional independence of these properties), and linearity (linear relationship between variables). These assumptions may not always hold true in real-world environmental contexts, potentially leading to inaccuracies in spatial interpolation, particularly in areas with varying spatial characteristics or nonlinear relationships. Similar to traditional kriging, co-kriging requires variographic analysis to model value evolution based on distance. This process requires defining prior parameters, which requires domain expertise and may not be universally applicable across different environmental datasets. Co-kriging may face challenges in scenarios with sparse data, such as soil pollution assessment, where data samples are limited and may not sufficiently support the assumptions and modeling requirements of co-kriging.

The scope of Deep Learning applications for environmental data has expanded significantly in recent years. For instance, it includes forest zone classification as mentioned in (Haq et al., 2021) and weed classification in (Haq, 2022). Additionally, Yuan et al. (2020) conducted a review of the potential uses of remote sensing with Deep Learning in environmental science.New Deep Learning methods have been proposed to address the limitations of co-kriging. Kirkwood et al. (2022) use a Bayesian neural network to perform spatial interpolation, while offering the possibility of adding auxiliary information even outside the interpolation grid. Wadoux et al. (2019) utilize a Convolutional Neural Network (CNN) for digital soil mapping, considering uncertain measurements of the soil property. However, these methods use a CNN to handle the auxiliary variable, which is defined across the entire grid. Other hybrid methods exist where kriging is combined with other existing techniques. For instance, Tao et al. (2019) suggest combining backpropagation neural networks, the Nemerow pollution index, and 3D kriging for the delineation of soil contaminant plumes. However, since the method utilizes kriging, it cannot entirely free itself from the constraints associated with its use. In certain cases, such as soil pollution, chemical analyses involve significant costs, resulting in a limited number of observed data points. This constraint restricts the use of supervised machine learning methods for spatial interpolation.

In this paper, we present a novel Deep Learning architecture for spatial interpolation involving an auxiliary variable. This architecture builds upon our prior work on 2D spatial interpolation and map generation, denoted DIP-SI (Rakotonirina et al., 2024), where we adapted the inpainting approach of Deep Image Prior (Ulyanov et al., 2018). This method entails reconstructing an image from the remaining pixels without prior training or labeled data. The core idea of the method proposed in the present paper involves employing a CNN of an encoder-decoder type, generating from a random vector two maps: one for the variable of interest and another one for the auxiliary variable. The neural network is fine-tuned based on the difference between the output maps and the observed values for both variables. Diverging from conventional encoder-decoder architectures, our approach incorporates two decoders to accommodate both variables. Recognizing that the correlation links between the main and secondary variables may vary in importance, we introduce a parameter in the loss function to control the impact of the auxiliary variable on the main variable. The method we propose does not require prior assumptions like kriging and co-kriging methods, nor does it need substantial training data as with Deep Learning generation methods. This offers a new perspective in spatial interpolation with an auxiliary variable when the available data are very limited or when spatial variabilities are significant.

In order to demonstrate the effectiveness of the proposed method, we evaluate the model on two datasets. The first dataset is a 2D dataset of Soil Organic Carbon (SOC) with 469 surface soil samples collected in the United States (Ahmed et al., 2018). For this dataset, the auxiliary variable we use is the Normalized Difference Vegetation Index (NDVI). We compare our method to ordinary kriging, co-kriging and also DIP-SI with one variable. The second dataset

is a 3D dataset from a site contaminated with hydrocarbons and Fluoride in France. The data are soil core samples obtained from hyperspectral analysis using the Tellux method. This method provides the concentrations at every point in the soil cores, by inferring the hyperspectral model trained on some samples with ground-truth being laboratory chemical analysis (Dhaini et al., 2021; Exem et al., 2023; Feray et al., 2023). To assess the proposed method and the influence of the auxiliary variable, we operate a validation protocol as follows: we exclude portions of certain boreholes and utilize these sections for evaluation.

The main contributions of this study can be summarized as follows:

1. **Introduction of a Novel Generative Neural Network Architecture:**
   We propose a new architecture for generative neural networks that does not require prior training. This architecture is designed to handle two continuous variables, showcasing its versatility in both 2D and 3D contexts. In our case, we use the architecture to incorporate additional statistical and spatial information into the interpolation of the main variable using the secondary variable.
2. **Loss Function Parameter to Manage Auxiliary Variable:**
   We propose a parameter, denoted as θ, to balance the influence between the primary and auxiliary variables within the loss function. This parameter is strategically included to accommodate the variability observed in spatial and statistical correlation.
3. **Extension from 2D to 3D Interpolation:**
   By nullifying the parameter θ, we get a 3D interpolation method, which is an extension of the 2D interpolation of our previous work (Rakotonirina et al., 2024). Experimental results on the 3D dataset confirms its capability to consider spatial information comprehensively, including both vertical and horizontal dimensions.
4. **Mitigation of Overfitting with Second Variable:**
   Our observations indicate that the inclusion of the second variable plays a crucial role in mitigating overfitting issues commonly associated with DIP methods as corroborated for instance in (Wang et al., 2023).

The remainder of this paper is organized as follows. The proposed method is introduced in Section 2, which is divided into two subsections: the learning process with the loss function and the architecture of the proposed method. The two datasets are presented in Section 3, and we discuss the experimental results in Section 4. Finally, Section 5 concludes this paper.

2. Proposed Method

In this section, we elaborate in detail on the different components of the proposed method. Firstly, we outline how both variables, namely the variable of interest and the auxiliary variable, are considered in the learning process.

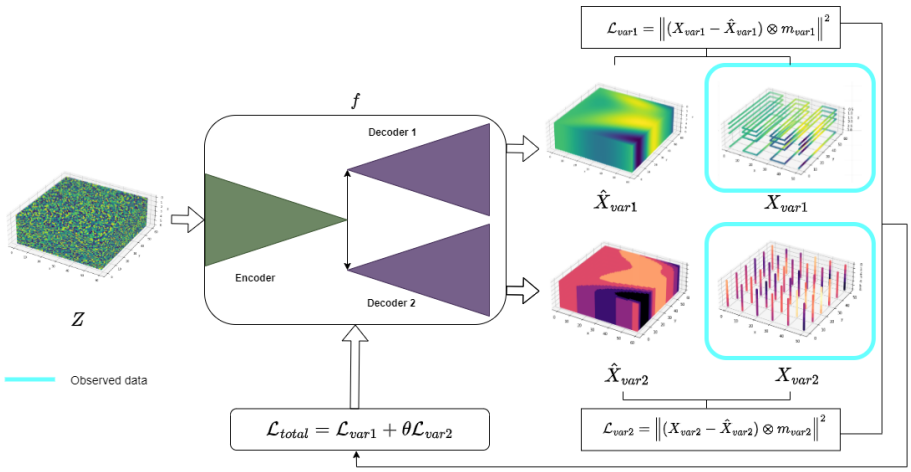

Figure 1: Representation of the proposed method for (3D) spatial interpolation with 2 variables.

We also explain how we can adjust the impact of the auxiliary variable on the main variable by introducing the parameter θ into the loss function. Following that, we provide a detailed exposition of the architecture we propose. This architecture is designed to enable the neural network to learn effectively from both variables.

2.1. Overview of the proposed method

An overview of the proposed method is depicted in Figure 1, and can be briefly described as follows.

As a generative method, we aim to determine a function f that generates, from a randomly generated data Z, the maps of the 2 variables, namely the main and the auxiliary variables. To this end, the function f is defined by a neural network with an architecture of encoder-decoder. Subsequently, the decoder section splits into two parts: Decoder 1 for the main variable and Decoder 2 for the secondary variable. The outputs of these two decoders are the maps $\hat{X}_{var1}$ and $\hat{X}_{var2}$, respectively. To adjust the parameters of the generative neural network f, the output maps of both decoders are multiplied by two binary masks, m_{var1} and m_{var2}. This allows us to compute the mean squared error with the maps of the observed values X_{var1} and X_{var2}. This process yields two loss functions, $\mathcal{L}_{var1}$ and $\mathcal{L}_{var2}$, which are combined with the parameter θ to have the total Loss $\mathcal{L}_{total}$. θ enables us to adjust the influence of the auxiliary variable in the interpolation of the main variable.

2.2. Detail on the Learning Process

The proposed approach builds upon our previous work on DIP for spatial interpolation (DIP-SI) (Rakotonirina et al., 2024), which accomplishes spatial

interpolation using a CNN without pre-training. We aim to revisit this work in order to propose a method that incorporates spatial and statistical information from an auxiliary variable. Thus, this method aims to provide a Deep Learning alternative to co-kriging.

The main part of the process is the function f, with its input Z and outputs $\hat{X}_{\text{var1}}$ and $\hat{X}_{\text{var2}}$, described as follows. The function f operates as a generative neural network, where it is a common practice for the neural network f to take a random vector Z. The randomness of the input can act as a form of regularization during training, as investigated in Generative Adversarial Networks (Goodfellow et al., 2020).

This helps stabilize learning by introducing some unpredictability and preventing the generator from learning overly rigid representation. The network's output comprises two distinct maps: $\hat{X}_{\text{var1}}$ for the main variable and $\hat{X}_{var2}$ for the auxiliary variable. Therefore, the function can be expressed as $f(Z) = \left(\hat{X}_{\text{var1}}, \hat{X}_{\text{var2}}\right)$. The two maps for the main and auxiliary variables are therefore derived from the same neural network f. While the encoder is similar to the encoder of the single-variable DIP-SI, it is in the decoder part that each map has its own decoder. The proposed deep neural architecture is described in detail in Section 2.3.

For the generative neural networks f to learn from the observed values of both variables, denoted as X_{var1} and X_{var2}, it is essential to quantify the estimation errors for each. To do this, as illustrated in Figure 1, we have two loss functions defined by:

$$\mathcal{L}_{var1} = \left\| (X_{var1} - \hat{X}_{var1}) \otimes m_{var1} \right\|^2, \tag{1}$$

$$\mathcal{L}_{var2} = \left\| (X_{var2} - \hat{X}_{var2}) \otimes m_{var2} \right\|^2. \tag{2}$$

where m_{var1} and m_{var2} are binary masks of both variables, with values 1 when the point is part of the observed points and 0 otherwise. These masks serve to filter the output maps using the Hadamard Product $\otimes$, i.e., element-wise multiplication, ensuring that the loss functions only consider the observed values. In other terms, the loss functions are the squared Euclidean distance between the original data and the reconstructed data, where the contribution of each element is weighted by the corresponding element in the masks. It is worth noting that the m_{var1} and m_{var2} need not to be the same, namely the two variables can be sampled differently. Moreover, Z, X_{var1}, and X_{var2} need not to have the same dimensions in general, even though we set their dimensions to D in this paper.

These two loss functions are specific to each variable. We propose to simultaneously minimize $\mathcal{L}_{var1}$ and $\mathcal{L}_{var2}$ so that the neural network can learn from both variables. To achieve this, we combine the two loss functions and introduce the parameter θ for $\mathcal{L}_{var2}$ to balance the influence of the secondary variable during training. The use of such a parameter is common in many Machine

Learning paradigms where bi-objective optimization may rise (or more generally multi-objective optimization), such as in multimodal models (Ma et al., 2023), in physics-informed models (Heldmann et al., 2023) and in combining physics-induced and data-driven models (Zhu and Honeine, 2016; Zhu et al., 2020). We introduce this parameter because we are aware that the statistical and spatial links between the two variables can vary from case to case. The total loss function is expressed as follows:

$$\mathcal{L}_{total} = \mathcal{L}_{var1} + \theta\, \mathcal{L}_{var2}. \tag{3}$$

Setting the parameter θ to 0 leads to a single-variable DIP-SI. By minimizing $\mathcal{L}_{\text{total}}$, the network f can learn the spatial and statistical configurations of both variables.

In this paper, the value of the parameter θ is fixed. To determine its value, we adopt the optimization method used by Liu et al. (2023), which leverages DIP for electrical impedance tomography reconstruction. This method involves varying the value of θ between 0, where the loss function $\mathcal{L}_{var2}$ has no impact on the total loss function $\mathcal{L}_{total}$, and 1, where the auxiliary variable has the same influence as the primary variable. Following the approach outlined in studies such as (Feurer and Hutter, 2019; Yang and Shami, 2020), we evaluate the performance of each θ value using validation data and select the one that yields the best results. It is important to note that this parameter does not affect the execution time of a training step.

For training, the optimization algorithm operates iteratively a gradient descent that adjusts the parameters of the model f to minimize the value of the loss function $\mathcal{L}_{total}$ considering both the main variable and the auxiliary variable. This process aims to ensure that the predictions closely align with the ground truth values. In order to obtain a model that generalizes well to unobserved data, it is crucial to monitor model performance to prevent overfitting, where the model excessively adjusts to the observed data, capturing even noise or insignificant variations. This issue is common in learning processes like DIP. Recently, Wang et al. (2023) studied four different stopping criteria for DIP: (i) inspect visually, which is a subjective approach, (ii) monitor the learning curve's evolution, (iii) tune the number of iterations, which is limited as the number of iterations varies based on the data, and (iv) use validation data for stopping training. In our case, we opted for the last method, following the work of Yaman et al. (2021) and Ding et al. (2022), ensuring the selection of representative validation samples.

2.3. Proposed Architecture of the Neural Network

The proposed architecture is divided into two distinct parts: the encoder (dimension reduction part) and the decoder (dimension augmentation part). CNNs are used to build this encoder-decoder neural network. The choice of CNNs is corroborated by several studies demonstrating that CNNs are well-suited for capturing spatial configuration; For instance, Jo and Pyrcz (2022) recently proposed a CNN-based method that automates variographic analysis

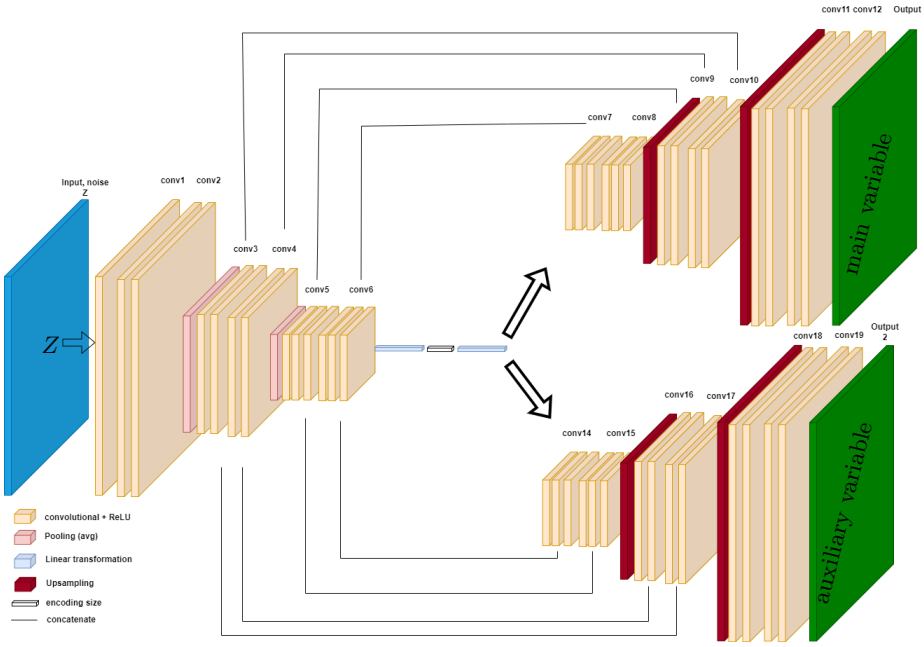

Figure 2: Proposed deep neural network architecture for interpolation with a auxiliary variable.

for capturing spatial configuration. More recently, we have demonstrated the relevance of CNNs for spatial interpolation using DIP in (Rakotonirina et al., 2024). In the current paper, we take advantage of this previous work, by using an encoder similar to the one of the single-variable DIP-SI (Rakotonirina et al., 2024), while each of the two variables has its own decoder, as illustrated in Figure 2.

The model takes a random value map Z of dimension D as input and transforms it into an encoded representation through several convolutional layers. The first two layers, conv1 and conv2, are convolutional layers followed by a rectified linear unit (ReLU) function to introduce non-linearities and enhance the model's learning capacity. From the conv3 layer onwards, pooling functions are introduced to reduce the dimension of the layers down to conv6. We use an average pooling function. The final step of the encoder involves transforming the conv6 layer into an encoding vector.

In the second part, the decoder, the encoded representation is transformed into a 2D or 3D representation, depending on the case that we have. Unlike the encoder, we employ two decoders to generate two maps, one for the main variable and one for the auxiliary variable. This approach of having dual decoders is quite common in Deep Learning literature. It finds applications in image segmentation (Tomar et al., 2021; Bougourzi et al., 2023), as well as in speech recognition and translation (Inaguma et al., 2021; Le et al., 2020). Despite having two decoders,

the operations performed in both are exactly the same. Firstly, the encoded representation is transformed by the first convolutional layer. Then, in contrast to the encoder, we perform upsampling operations. We use the same function as in our previous work Rakotonirina et al. (2024), namely the bicubic function for the 2D case. Concerning the 3D case, the bicubic function is not feasible, and in this context, computational costs can be significant. Therefore, we opted for the Nearest upsampling function. We also retain the skip connections, as shown in Figure 2. These connections help maintain information from both variables since they share the same encoder, and also help preserve information that may have been perturbed by the various nonlinear functions in our model as recently corroborated by Tran and Yang (2022). Wang et al. (2022) demonstrated that using a dual decoder in a U-net with skip connections allows for more detailed features. After all these operations, we reach the model's output with the two maps for the main and auxiliary variables.

3. Datasets and Case Studies

In this section, we present the case studies used to demonstrate the benefits of the proposed method. Firstly, we employ a 2D dataset with Soil Organic Carbon (SOC) as the main variable and NDVI as the auxiliary variable. Secondly, we utilize a 3D dataset from a contaminated site, combining hydrocarbon concentration as variable of interest and Fluoride concentration as auxiliary variable.

3.1. 2D Dataset: SOC + NDVI

The dataset, sourced from the North American Carbon Program, provides estimates of SOC, pyrogenic carbon, particulate organic carbon, and other soil organic carbon fractions in 469 surface soil samples collected in the Colorado, Kansas, New Mexico, and Wyoming regions of the United States. The data was obtained using Fourier-transform infrared spectroscopy (FTIR) and partial least squares regression (PLSR) to predict SOC fractions in 650 soil samples collected by the United States Geological Survey (USGS) as part of the Geochemical Landscapes project. They validated and reduced the dataset to 469 representative samples. Regarding NDVI, as specified by Ahmed et al. (2017), it represents the NDVI for the months of June and July from 2000 to 2011, derived from bands 1 (red) and 2 (near-infrared) of the Moderate Resolution Imaging Spectroradiometer. An R script was used to download the data, and the "raster" package (Hijmans and Van Etten, 2012) was employed to compute the mean NDVI.

As shown in Figure 3, we randomly divided the soil samples into two sets, one comprising 368 samples and the other 101 samples. The first set was utilized for conducting interpolations, while the second one was reserved for evaluating the models' performance. We have randomly divided the evaluation set into two parts: one for validating the model and the other for testing its performance. We ensured that the test datasets are representative. These datasets comprise exactly the same points for both SOC and NDVI variables.

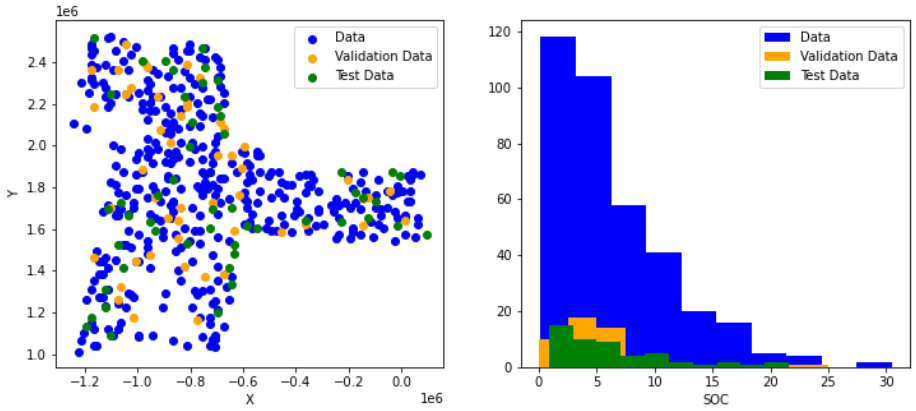

Figure 3: Representation of the spatial and statistical distribution of the data for spatial interpolation, along with validation data to assess the models' performance.

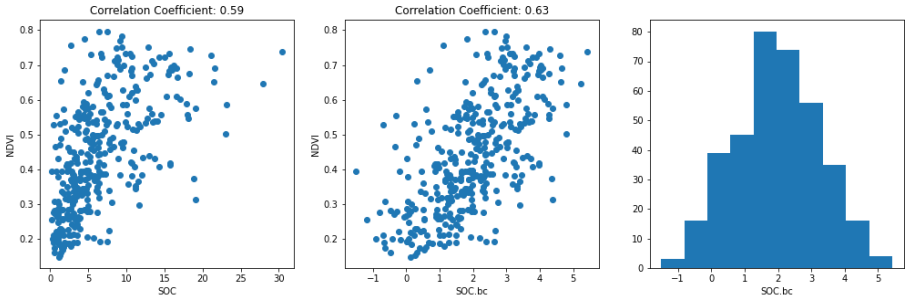

Figure 4: Representation of correlation coefficient ρ between SOC and NDVI, transformed SOC and NDVI, and distribution of transformed SOC.

To measure the links between the two variables, we use the correlation coefficient, which can be defined as follows:

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}, \tag{4}$$

where $\rho(X, Y)$ represents the correlation coefficient between two variables X and Y, $cov(X, Y)$ is the covariance between X and Y,σ_X is the standard deviation of X, and σ_Y is the standard deviation of Y.

We compare the proposed method with ordinary kriging (OK) and co-kriging (COK). However, as depicted in Figure 3, the distribution of SOC does not align with the assumptions of kriging for optimal results. To address this, a Box-Cox transformation (Sakia, 1992) was applied to normalize the variable; The obtained variable is denoted "SOC.bc". Additionally, this improves the correlation coefficient between SOC and NDVI with ρ(SOC.bc, NDVI) = 0.63 compared to ρ(SOC, NDVI) = 0.59 for the original variable, as shown in Fig-

Table 1: Statistical summary of variables SOC, transformed SOC (SOC.bc) and NDVI.

Variable	Count	Mean	Std	Min	25%	Max
SOC	368	6.37	5.06	0.15	2.74	30.47
SOC.bc	368	1.98	1.29	-1.50	1.15	5.42
NDVI	368	0.44	0.16	0.15	0.31	0.80

Table 2: Descriptive statistics for TPH and Fluoride concentrations.

Variable	Count	Mean	Std	Min	25%	50%	75%	Max
TPH (mg/kg)	245	443.81	398.02	0.49	84.82	353.93	721.40	1916.89
Fluoride (mg/kg)	278	37.83	17.51	6.16	22.74	39.44	50.05	82.73

ure 4. Table 1 represents the statistical summary of the SOC, transformed SOC and NDVI variables.

3.2. 3D Dataset: Hydrocarbon + Fluoride

The second dataset is from a site contaminated with Total Petroleum Hydrocarbons and Fluoride. The site is located in France, and six boreholes with a depth of 5m each were conducted. The interpolation grid covers an area of 2612m^2 with a resolution of 2m for the x and y axes and a resolution of 0.1m for the z-axis. This dataset is quite specific as we have very few boreholes, resulting in limited data along the x and y axis. The performance of kriging might be affected by this limitation due to the variographic analysis relying on the variation of values as a function of distances between points.

The datasets come from hyperspectral imagery (HSI) analysis conducted by Tellux. Tellux offers Machine Learning algorithms that correlate indexes from hyperspectral imagery, as demonstrated by Achard and Elin (2019) and Kühn et al. (2004), with Total Petroleum Hydrocarbons (TPH) concentrations obtained from chemical analysis in the laboratory. In this case study, Fluoride also exhibited a correlation between images and concentration analyses conducted in a chemistry laboratory. This provides us with the concentrations of total hydrocarbons and Fluoride at all points along the z-axis of the boreholes.

To evaluate the performance of our model, we proceeded as with the 2D SOC dataset by selecting validation data. However, unlike the 2D dataset, we did not randomly select the validation data. Since we have a significant resolution in the z-axis, randomly selecting validation data would validate the vertical interpolation between the points of each borehole. However, our goal is to demonstrate that the model can capture both the vertical and horizontal spatial and statistical information of the boreholes. Therefore, we selected the validation data in such a way that a purely vertical interpolation of the borehole cannot estimate the removed values. As shown in Figure 5, we removed points from boreholes B5 and B6 based on a change in TPH concentration values that cannot be explained vertically.

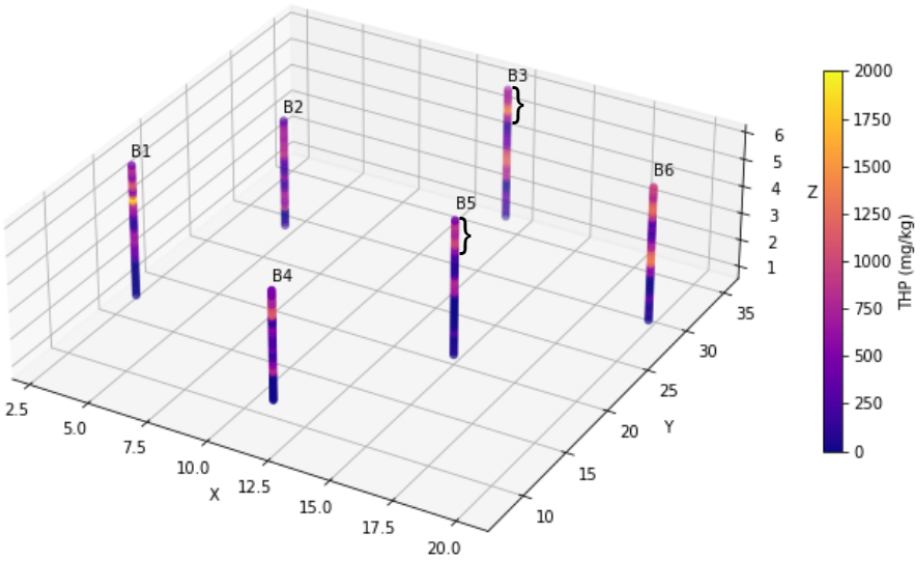

Figure 5: Representation in space of the hydrocarbon concentrations for the 6 boreholes with the selection of test data for total hydrocarbons indicated by the black markers for $B3$ and $B5$.

The correlation coefficient between hydrocarbon and Fluoride for our study site is $\rho = 0.76$. Since both data variables are derived from the same hyperspectral image analysis, we verified that the correlation between the two variables is not solely due to this fact. To do this, we analyzed the correlation coefficient between the laboratory analysis results of soil core samples. The correlation coefficient between these two chemical analysis variables is 0.70, confirming a significant link between the two in our study case. Table 2 summarizes the statistical descriptions of the two variables.

4. Results and Discussion

In this section, we delve into the results of our method on the two presented datasets. Similar to the previous section, this part is divided into two sections. First, we compare our proposed two-variable method with the single-variable DIP-SI method, as well as with OK and COK. We use three metrics to compare the results of our models: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R2) score. For each pair of observations y_i (the actual value) and $\hat{y}_i$ (the predicted value), the formulas are defined as follows:

- Mean Absolute Error (MAE):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

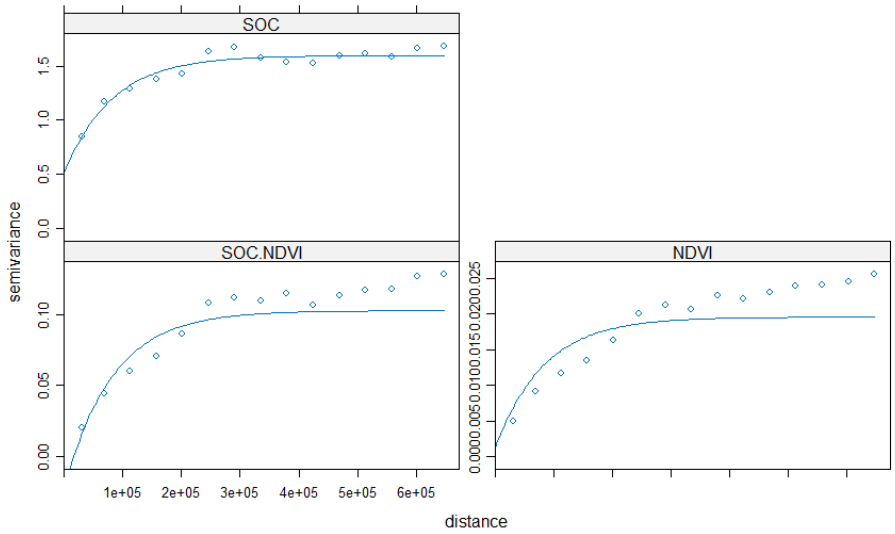

Figure 6: SOC and NDVI variograms and the SOC.NDVI co-variogram for co-kriging.

- Root Mean Squared Error (RMSE):

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

- R-squared (R2) Score:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

In these formulas, n represents the number of observations, and $\bar{y}$ is the mean of the actual values. These metrics are commonly used to assess the performance of predictive models.

4.1. 2D Dataset: SOC + NDVI

OK and COK for the SOC and NDVI variables were performed based on the methods outlined in this source[1]. To ensure the optimal results of OK and COK, we use the transformed data and then perform a back transformation afterward.

Figure 6 shows the variograms of SOC and NDVI and co-variogram for both.

[1] https://zia207.github.io/geospatial-r-github.io/cokriging.html

Table 3: erformance metrics of DIP-SI and proposed method, evaluated on the validation data.

Method	RMSE	MAE	R2
DIP-SI	3.67	2.68	0.42
This Paper	3.58	2.63	0.45

Table 4: Performance metrics of several methods, SOC for OK and SOC + NDVI for others, evaluated on test data

Method	RMSE	MAE	R2
OK	4.27	2.78	0.33
COK	4.07	**2.55**	0.39
DIP-SI	4.13	2.89	0.38
This Paper	**3.79**	2.56	**0.47**

For the proposed method, we do not need to perform a variographic analysis before interpolation. In this case study, we use the same parameters as in our previous work, namely the bicubic function as the upsampling function and an encoding size of 10. This demonstrates the flexibility of the DIP-SI method. However, in this paper, we introduce a new parameter that can be adjusted in our model. This parameter is denoted as θ, and it manages the interaction between the two variables during training. We conducted tuning parameters to select the optimal value for θ. It is important to note that there are several methods to determine the value of θ. It is even possible to use a value that varies during training. The choice of whether θ should be a fixed or trainable parameter remains an open problem. In this paper, we chose to fix the value of θ. To compare the proposed method with DIP-SI, we performed an interpolation with the value of θ set to 0. Indeed, if the value is 0, it means that the auxiliary variable does not contribute to the learning process, which is equivalent to spatial interpolation with a single variable. The value of θ obtained using the method described in Section 2.2 in this case study for interpolation with two variables is 0.03. Table 3 presents the performances of DIP-SI and proposed method CO-DIP-SI on validation data.

Table 4 presents the performance metrics for different methods, including Ordinary Kriging (OK), Co-Kriging (COK), DIP-SI, and the proposed method. The results suggest that the proposed method outperforms the other methods, achieving the lowest RMSE and the highest R2 Score. Co-Kriging and DIP-SI exhibit comparable performance on R2 score, while Ordinary Kriging shows slightly lower performance across the metrics. However, there seems to be a disparity in the MAE values for different methods compared to other metrics. Since MAE is less sensitive to extreme values than other measures, it provides an average estimation of errors that does not take into account very high values.

Figure 3 illustrates the spatial distribution of SOC soil samples used for interpolation, as well as the distribution of validation data and test data. Figure 7

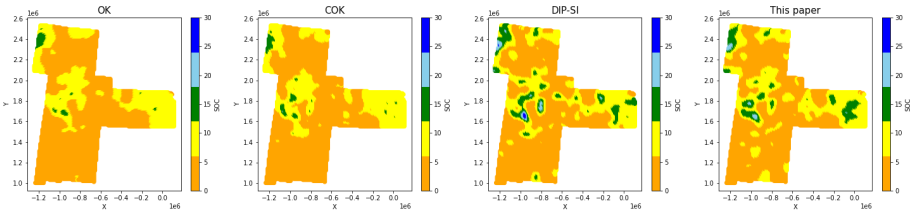

Figure 7: Spatial representation of the SOC interpolation result for the different methods.

Table 5: Performance metrics of several methods, TPH + Fluoride for OK, COK, DIP-SI and the proposed method

Method	RMSE	MAE	R2
OK	321.44	261.92	0.35
COK	320.10	258.79	0.34
DIP-SI	252.51	175.04	0.50
This Paper	**206.12**	**134.12**	**0.67**

demonstrates that kriging-based techniques, namely OK and COK, exhibit less variability in the estimates than Deep Learning-based techniques. This could be attributed to the smoothing effect of kriging on interpolation results. It is also noteworthy that the DIP-SI method shows more variability than the proposed method. However, referring to the metrics, we can suggest that this might be associated with overestimation. In the next dataset, we demonstrate that the two-variable method helps control overfitting of the neural network in our case. Even though the used dataset aligns with the expectations of kriging, the proposed method shows interesting performances. Notably, no prior spatial analysis or pre-training is required for interpolation.

4.2. 3D Dataset: Hydrocarbons + Fluoride

In this section, we delve into a 3D case study by interpolating the hydrocarbon concentration data from boreholes combined with Fluoride concentration data. We do not use the bicubic upsample function, as in the 2D case, due to the need to employ the tricubic function and the potentially significant associated computation time. Therefore, we employ the Nearest function, which is less computationally demanding and easier to implement. And similar to the first case study, we compare our approach with $\theta = 0$, which is equivalent to single-variable interpolation. We also compare the proposed method with kriging methods, namely OK and COK. The target variable TPH has a highly skewed distribution, as is common in cases of hydrocarbon contamination. Consequently, we performed a log-normal transformation of the variable before conduction kriging. The optimal value of θ obtained using the method described in Section 2.2 for this case study is 0.3.

According to the results in Table 5, the Deep Learning methods outperform

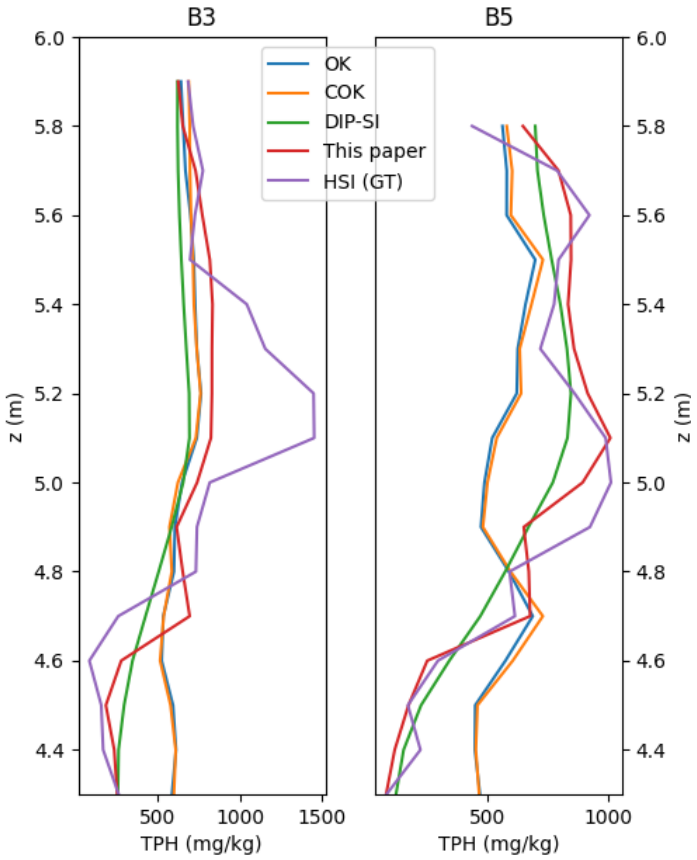

Figure 8: TPH log concentration on the validation data for boreholes B3 and B5, comparing various methods against the Hyperspectral Imaging (HSI) data, which is our ground truth.

the kriging methods, likely attributed to the fact that Deep Learning methods do not require transforming the variable to achieve a normal distribution. The limited number of points on the x and y axis of this dataset also has an impact on the results of kriging methods, due to the difficulty of variographic modelling. We also observe that the addition of the auxiliary variable does not provide a significant improvement to the co-kriging method. The proposed method outperforms the single variable method DIP-SI on all metrics, with a lower RMSE and MAE and a higher R2. This suggests that the model is capable of capturing spatial and statistical information from the auxiliary variable to enhance the interpolation of the variable of interest. However, it is important to note that we only excluded validation data for the TPH, and therefore, the presence of Fluoride values may have contributed to the improvement in interpolation.

It is also important to note that despite the fact that the DIP-SI method yields less favorable results than the proposed method, it is still capable of

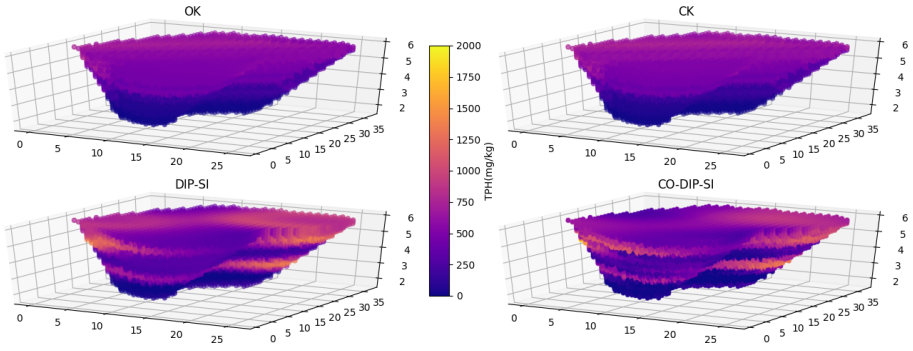

Figure 9: Spatial interpolated maps of univariate methods OK, DIP-SI and multivariate methods COK and proposed method CO-DIP-SI for TPH concentrations.

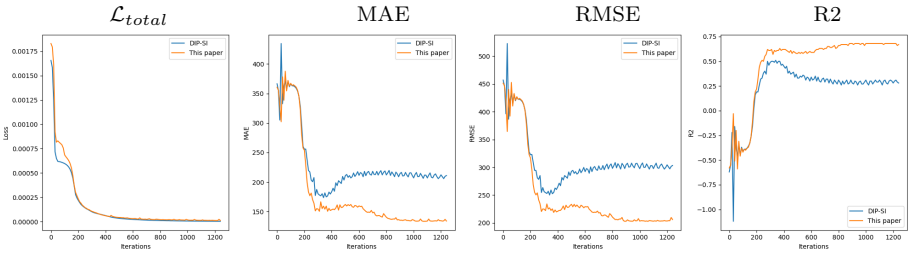

Figure 10: Evolution of the metrics on the validation data based on the model iterations

capturing the spatial and statistical information from the different boreholes. Considering that we excluded validation data so that a vertical interpolation cannot determine the values, we still observe an increase in concentration values, whether for borehole B3 or B5. This is evident in Figure 8. This figure also illustrates that the interpolation from DIP-SI is smoother compared to the proposed method, as seen in the concentration peak at 5.1m, where the proposed method adheres more closely to the actual data.

Figure 9 shows the spatial distribution maps of univariate methods OK and DIP-SI, and multivariate methods COK and the proposed CO-DIP-SI. These maps suggest that the Deep Learning-based methods can capture the vertical variations of boreholes, as shown in Figure 5. On the other hand, geostatistical methods tend to smooth the results, which is also due to the variographic modeling that faces challenges with the limited number of boreholes. Comparing the results of DIP-SI and the proposed method, it is evident that the proposed method presents even more detail than DIP-SI. This seems to indicate that the model effectively retrieves auxiliary information from the Fluoride variable.

Not only do the results show that the two-variable method is much more effective than the DIP-SI method, but we have also observed that the use of the second variable helps control model overfitting. As noted by Wang et al. (2023), DIP-based methods tend to overfit, and this is evident in the case of the

Table 6: Number of trainable parameters for different configurations of the proposed neural network f

Dataset	**Output Shape**	**# Parameters**
2D (SOC+NDVI)	136×160	14×10^6
3D (TPH+Fluoride)	$28 \times 40 \times 60$	25×10^6
3D	$1000 \times 1000 \times 100$	8×10^9

DIP-SI method where the model's performance decreases after a certain number of iterations. Figure 10 illustrates this, as from around the 400th iteration, the model's performance decreases for all metrics (RMSE, MAE, or R2), while the proposed method continues to learn and improves its performance. However, it is important to note that, due to the very limited number of data points we have, we were unable to allocate separate validation and test data for our model. It is therefore crucial to consider this in cases like ours, and it is beneficial to seek the opinion of a domain expert regarding the interpolated maps for additional validation.

Regarding the complexity of the model and computational capacity, the application of this method in a real-time setting depends on the size of the grid to be interpolated. Table 6 shows us that in the two case studies we have, the number of trainable parameters is manageable. In a scenario where we have a grid of $1000 \times 1000 \times 100$ to interpolate, we estimate that it would involve 8 billion parameters. This may be limiting for real-time applications but is still feasible and at the same scale as the most-used deep neural networks, such as ResNet-50, U-net, VGG-16, and their 3D variants. For reference, ResNet-50 has approximately 25×10^6 parameters (He et al., 2016), while U-net has 31×10^6 parameters (Ronneberger et al., 2015), and VGG-16 has 138×10^6 parameters (Simonyan and Zisserman, 2014). Moreover, implementations of convolution layers using memory-efficient cuDNN have been also investigated with success for 3D neural networks, such as in (Çiçek et al., 2016).

5. Summary and conclusions

In this paper, we proposed a spatial interpolation method that can serve as an alternative to co-kriging. As a spatial interpolation method with a generative neural network, the proposed method allowed for the inclusion of an auxiliary variable. To achieve this, we introduced a CNN with an encoder-decoder architecture but with the distinctive feature of having two decoders to produce two output maps. Additionally, we introduced a loss function with a parameter to control the relationships between the two variables, and this parameter can be adjusted according to specific cases.

We tested our method on two datasets. The first was a 2D dataset of SOC from the North American Carbon Program, coupled with NDVI data. Despite SOC data aligning with the expectations of kriging, the proposed method demonstrated superior performance in terms of RMSE and R2 compared to

other methods. The second dataset involved a contaminated site in France, featuring Hydrocarbons and Fluoride concentrations. Results indicated better performance with our proposed method despite the limited number of points on the x and y axis with only 6 boreholes, and importantly, the addition of the second variable helped mitigate overfitting issues often associated with DIP-based approaches. We have also shown that real-time application for substantial interpolation grids requires significant computational capacity. And in cases where we have a very limited number of observed values, validation with a domain expert can be useful to avoid over-interpretation.

There are several avenues for improving our method. One potential enhancement could involve the parameter θ. The value of θ can be considered dynamic, allowing it to evolve during training rather than remaining fixed. Additionally, the proposed method can serve as a foundation for conditional generation. Given that neural networks are initialized randomly, model outputs can vary between executions. This variability can be harnessed for conditional generation tasks.

Acknowledgements

The authors would like to thank the agency of ecological transition ADEME in France and TELLUX company for the funding of this research work. We would like to express our sincere gratitude to the anonymous reviewers for their valuable feedback and constructive comments, which significantly contributed to the enhancement of the quality and clarity of this manuscript. The first author would like to thank la Maison Yvonne for the inspiration.

References

Achard, V., Elin, C., 2019. Automatic mapping of hydrocarbon pollution based on hyperspectral imaging, in: IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium, IEEE. pp. 5768–5771.

Ahmed, Z., Woodbury, P., Sanderman, J., Hawke, B., Jauss, V., Solomon, D., Lehmann, J., 2018. Stocks of surface soil organic carbon fractions, great plains region, USA, 2007-2010. ORNL DAAC, Oak Ridge, Tennessee, USA.

Ahmed, Z.U., Woodbury, P.B., Sanderman, J., Hawke, B., Jauss, V., Solomon, D., Lehmann, J., 2017. Assessing soil carbon vulnerability in the Western USA by geospatial modeling of pyrogenic and particulate carbon stocks. Journal of Geophysical Research: Biogeosciences 122, 354–369.

Bangroo, S., Najar, G., Achin, E., Truong, P.N., 2020. Application of predictor variables in spatial quantification of soil organic carbon and total nitrogen using regression kriging in the North Kashmir forest Himalayas. Catena 193, 104632.

Bougourzi, F., Distante, C., Dornaika, F., Taleb-Ahmed, A., 2023. PDAtt-Unet: Pyramid dual-decoder attention unet for covid-19 infection segmentation from CT-scans. Medical Image Analysis 86, 102797.

Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O., 2016. 3d u-net: learning dense volumetric segmentation from sparse annotation, in: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19, Springer. pp. 424–432.

Dhaini, M., Roudaut, F.J., Garret, A., Arzur, R., Chereau, A., Varenne, F., Honeine, P., Mignot, M., Exem, A.V., 2021. Hyperspectral imaging for the evaluation of lithology and the monitoring of hydrocarbons in environmental samples, in: RemTech (International event on Remediation, Coasts, Floods, Climate, Seismic, Regeneration Industry), Ferrara, Italy.

Ding, L., Qin, Z., Jiang, L., Zhou, J., Zhu, Z., 2022. A validation approach to over-parameterized matrix and image recovery. arXiv preprint arXiv:2209.10675 .

Dowd, P.A., Pardo-Igúzquiza, E., 2023. The many forms of co-kriging: A diversity of multivariate spatial estimators. Mathematical Geosciences , 1–27.

Elsayed, K., 2015. Optimization of the cyclone separator geometry for minimum pressure drop using co-kriging. Powder Technology 269, 409–424.

Exem, A.V., Kassem, P., Honeine, P., Mignot, M., 2023. High-resolution characterization of total hydrocarbons by infrared hyperspectral imaging in an alluvial soil, in: NICOLE Fall Workshop 2023 (Innovative solutions for sustainable redevelopment and land stewardship of contaminated sites and sediments), Malmö, Sweden.

Feray, C., Jacquemoud, S., Honeine, P., Exem, A.V., 2023. Hyperspectral characterization of soil matrix effects by coupling physical models and machine learning methods. Poster at the 13th IEEE Workshop on Hyperspectral Image and Signal Processing : Evolution in Remote Sensing (WHISPERS), Athens, Greece.

Feurer, M., Hutter, F., 2019. Hyperparameter optimization. Automated machine learning: Methods, systems, challenges , 3–33.

Gia Pham, T., Kappas, M., Van Huynh, C., Hoang Khanh Nguyen, L., 2019. Application of ordinary kriging and regression kriging method for soil properties mapping in hilly region of Central Vietnam. ISPRS International Journal of Geo-Information 8, 147.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2020. Generative adversarial networks. Communications of the ACM 63, 139–144.

Haq, M.A., 2022. Cnn based automated weed detection system using uav imagery. Computer Systems Science & Engineering 42.

Haq, M.A., Rahaman, G., Baral, P., Ghosh, A., 2021. Deep learning based supervised image classification using uav images for forest areas classification. Journal of the Indian Society of Remote Sensing 49, 601–606.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778.

Heldmann, F., Berkhahn, S., Ehrhardt, M., Klamroth, K., 2023. Pinn training using biobjective optimization: The trade-off between data loss and residual loss. Journal of Computational Physics 488, 112211.

Hijmans, R., Van Etten, J., 2012. Geographic analysis and modeling with raster data. R package version 2.0-12.

Inaguma, H., Higuchi, Y., Duh, K., Kawahara, T., Watanabe, S., 2021. Orthros: Non-autoregressive end-to-end speech translation with dual-decoder, in: ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE. pp. 7503–7507.

Jo, H., Pyrcz, M.J., 2022. Automatic semivariogram modeling by convolutional neural network. Mathematical Geosciences 54, 177–205.

Journel, A.G., Huijbregts, C.J., 1976. Mining geostatistics .

Kim, H.R., Kim, K.H., Yu, S., Moniruzzaman, M., Hwang, S.I., Lee, G.T., Yun, S.T., 2019. Better assessment of the distribution of As and Pb in soils in a former smelting area, using ordinary co-kriging and sequential gaussian cosimulation of portable X-ray fluorescence (PXRF) and ICP-AES data. Geoderma 341, 26–38.

Kirkwood, C., Economou, T., Pugeault, N., Odbert, H., 2022. Bayesian deep learning for spatial interpolation in the presence of auxiliary information. Mathematical Geosciences 54, 507–531.

Koziel, S., Bekasiewicz, A., Couckuyt, I., Dhaene, T., 2014. Efficient multiobjective simulation-driven antenna design using co-kriging. IEEE Transactions on Antennas and Propagation 62, 5900–5905.

Kühn, F., Oppermann, K., Hörig, B., 2004. Hydrocarbon index–an algorithm for hyperspectral detection of hydrocarbons. International Journal of Remote Sensing 25, 2467–2473.

Le, H., Pino, J., Wang, C., Gu, J., Schwab, D., Besacier, L., 2020. Dual-decoder transformer for joint automatic speech recognition and multilingual speech translation. arXiv preprint arXiv:2011.00747 .

Leenaers, H., Burrough, P.A., Okx, J.P., 2020. Efficient mapping of heavy metal pollution on floodplains by co-kriging from elevation data, in: Three Dimensional Applications in GIS. CRC Press, pp. 37–50.

Liu, D., Wang, J., Shan, Q., Smyl, D., Deng, J., Du, J., 2023. Deepeit: Deep image prior enabled electrical impedance tomography. IEEE Transactions on Pattern Analysis and Machine Intelligence 45, 9627–9638.

Liu, X., Zhao, W., Wan, D., 2022. Multi-fidelity co-kriging surrogate model for ship hull form optimization. Ocean Engineering 243, 110239.

Ma, H., Zhang, Q., Zhang, C., Wu, B., Fu, H., Zhou, J.T., Hu, Q., 2023. Calibrating multimodal learning, in: International Conference on Machine Learning, PMLR. pp. 23429–23450.

Matheron, G., 1963. Traité de géostatistique appliquée. 2. le krigeage. Editions Technip.

Milillo, T.M., Sinha, G., Gardella Jr, J.A., 2012. Use of geostatistics for remediation planning to transcend urban political boundaries. Environmental pollution 170, 52–62.

Ochie, K.I., Rotimi, O.J., 2018. Geostatistics–kriging and co-kriging methods in reservoir characterization of hydrocarbon rock deposits, in: SPE Nigeria Annual International Conference and Exhibition, SPE. pp. SPE–193483.

Ouabo, R.E., Sangodoyin, A.Y., Ogundiran, M.B., 2020. Assessment of ordinary kriging and inverse distance weighting methods for modeling chromium and cadmium soil pollution in e-waste sites in Douala, Cameroon. Journal of Health and Pollution 10, 200605.

Rakotonirina, H., Honeine, P., Atteia, O., Exem, A.V., 2024. Spatial interpolation and conditional map generation using deep image prior for environmental applications. Mathematical Geoscience .

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation, in: Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18, Springer. pp. 234–241.

Sakia, R.M., 1992. The Box-Cox transformation technique: a review. Journal of the Royal Statistical Society Series D: The Statistician 41, 169–178.

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 .

Tao, H., Liao, X., Zhao, D., Gong, X., Cassidy, D.P., 2019. Delineation of soil contaminant plumes at a co-contaminated site using BP neural networks and geostatistics. Geoderma 354, 113878.

Tomar, N.K., Jha, D., Ali, S., Johansen, H.D., Johansen, D., Riegler, M.A., Halvorsen, P., 2021. DDANet: Dual decoder attention network for automatic polyp segmentation, in: Pattern Recognition. ICPR International Workshops and Challenges: Virtual Event, January 10-15, 2021, Proceedings, Part VIII, Springer. pp. 307–314.

Tran, Q.N., Yang, S.H., 2022. Video frame interpolation via down–up scale generative adversarial networks. Computer Vision and Image Understanding 220, 103434.

Tziachris, P., Metaxa, E., Papadopoulos, F., Papadopoulou, M., 2017. Spatial modelling and prediction assessment of soil iron using kriging interpolation with pH as auxiliary information. ISPRS International Journal of Geo-Information 6, 283.

Ulyanov, D., Vedaldi, A., Lempitsky, V., 2018. Deep image prior, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 9446–9454.

Wadoux, A.M.C., Padarian, J., Minasny, B., 2019. Multi-source data integration for soil mapping using deep learning. Soil 5, 107–119.

Wang, H., Li, T., Zhuang, Z., Chen, T., Liang, H., Sun, J., 2023. Early stopping for deep image prior. Transactions on Machine Learning Research URL: `https://openreview.net/forum?id=231ZzrLC8X`.

Wang, Y., Peng, Y., Li, W., Alexandropoulos, G.C., Yu, J., Ge, D., Xiang, W., 2022. DDU-Net: Dual-decoder-U-Net for road extraction using high-resolution remote sensing images. IEEE Transactions on Geoscience and Remote Sensing 60, 1–12.

Yaman, B., Hosseini, S.A.H., Akcakaya, M., 2021. Zero-shot physics-guided deep learning for subject-specific MRI reconstruction, in: NeurIPS 2021 Workshop on Deep Learning and Inverse Problems.

Yang, L., Shami, A., 2020. On hyperparameter optimization of machine learning algorithms: Theory and practice. Neurocomputing 415, 295–316.

Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang, Q., Wang, J., et al., 2020. Deep learning in environmental remote sensing: Achievements and challenges. Remote sensing of Environment 241, 111716.

Zhu, F., Honeine, P., 2016. Bi-objective nonnegative matrix factorization: Linear versus kernel-based models. IEEE Transactions on Geoscience and Remote Sensing 54, 4012 – 4022. doi:10.1109/TGRS.2016.2535298.

Zhu, F., Honeine, P., Chen, J., 2020. Pixel-wise linear/nonlinear nonnegative matrix factorization for unmixing of hyperspectral data, in: Proc. 45th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain. pp. 4737–4741. doi:10.1109/ICASSP40776.2020.9053239.