Estimating the Values of Missing Data Related to Infrastructure Condition States Using Their Spatial Correlation

Journal Article

Author(s):
Moghtadernejad, Saviz; Jin, Yuchuan; Adey, Bryan T. (iD)

ASCE

Estimating the Values of Missing Data Related to Infrastructure Condition States Using Their Spatial Correlation

Saviz Moghtadernejad, Ph.D.[1]; Yuchuan Jin[2]; and Bryan Tyrone Adey, Ph.D.[3]

Abstract: Infrastructure managers consistently monitor the condition of their assets to predict their deterioration speed and determine the optimal time to execute preventive interventions. However, despite the recent progress in more frequent and accurate monitoring of assets and storage of the related results, in practice, real-world data often contains errors and discrepancies such as missing data or faulty entries. This problem can happen owing to collection errors during routine inspections or inconsistency of data storage formats in different years. Because the quality of data plays a significant role in the accuracy of deterioration prediction and the resulting intervention programs, it is important to improve condition state predictions by imputing the values of missing information. This paper examines the efficiency of different models that use the spatial correlation of infrastructure assets in predicting the value of missing data. The models include univariate and multivariate Kriging, a hybrid artificial neural network (ANN)-Kriging model, and the bidirectional long short term memory (bi-LSTM) neural network, which can model the data with spatial correlation or a sequential relationship. The results confirm that the condition indicator values can be estimated with reasonably low levels of error. **DOI: [10.1061/(ASCE)IS.1943-555X.0000726.](https://doi.org/10.1061/(ASCE)IS.1943-555X.0000726)** *This work is made available under the terms of the Creative Commons Attribution 4.0 International license, [https://creativecommons.org/licenses/by/4.0/.](https://creativecommons.org/licenses/by/4.0/)*

Author keywords: Kriging model; bidirectional long short term memory neural network; artificial neural network-Kriging model; Insufficient or missing data.

Introduction

Modern societies depend on well-functioning transportation infrastructure. As infrastructure assets continually deteriorate, decision-makers have to be able to accurately predict their deterioration speed to determine the optimal maintenance programs. For this purpose, they continually inspect the condition of the assets in a network using the values of condition indicators. Consequently, the accuracy of deterioration prediction and the resulting intervention decisions depends on the quality of the collected data, which are usually stored in the form of geographical information system (GIS) indexed data. However, although technology advances have facilitated the frequent and accurate monitoring of assets and storage of the related results, in practice, real-world data is not always suitable for future predictions, since it often does not exist in sufficient quantity, is missing information, or contains faulty entries.

[1]Postdoctoral Fellow, Institute of Construction and Infrastructure Management, Eidgenössische Technische Hochschule (ETH) Zurich, Zurich 8093, Switzerland (corresponding author). ORCID: https://orcid.org/0000-0001-9860-5523. Email: moghtadernejad@ibi.baug.ethz.ch

[2]Graduate Student, Division of Transport Planning, Kungliga Tekniska högskolan (KTH) Royal Institute of Technology, Stockholm 114 28, Sweden. Email: yuchuanj@kth.se

[3]Professor, Institute of Construction and Infrastructure Management, Eidgenössische Technische Hochschule (ETH) Zurich, Zurich 8093, Switzerland. ORCID: https://orcid.org/0000-0002-4932-5901. Email: adey@ibi.baug.ethz.ch

This problem can happen for three major reasons:
1. Errors during the inspection campaigns: for instance, the value of a condition indicator for one segment is not measured owing to malfunction in equipment, sensors, etc., or is incorrectly entered and saved in the repository (Lethanh et al. 2016).
2. Poor planning or coordination in data storage: for example, using several independent sets of GIS indexed data or dividing roads/rail tracks into different segments in consecutive years.
3. When maintenance is being done on a road/rail segment during an inspection campaign, condition data are generally not collected on that segment.

Regardless of the reason behind missing the condition indicator data, it is desirable to be able to estimate these values and create a complete dataset. For example, when digital tools are used to predict the evolution of the condition of all assets over time, it is imperative to have a complete dataset of the condition of the road/rail segments at $t = 0$. If the data are missing because of maintenance activities being carried out on that road/rail segment, the condition state of that segment would be restored to the target state according to the maintenance strategy. However, if the information is missing for one of the other two reasons, appropriate models should be used to impute this missing data.

Consequently, when facing missing data, first one should check the available intervention reports during that year to determine if the road/rail segments with missing condition indicator data have undergone some maintenance activities. This may also be verified if sufficient data are available regarding the historical condition state of the assets. If it is determined that an intervention has been executed on the segments with missing data, then the condition state of the segments would be adjusted either based on the maintenance strategy or to the best state to account for the new state of the segments after the interventions.

In case it is confirmed that no intervention has been executed on the segments, or there is no information available to make a

determination, to be conservative, it may be assumed that the segments have not undergone maintenance activities, and appropriate models should be used to estimate the missing condition indicator values.

There are two general approaches to deal with missing data. The first approach is possible in the presence of information on the value of missing data from previous inspection campaigns, where behavioral models are used to extrapolate the past data (Anastasopoulos and Mannering 2015; Chu and Durango-Cohen 2007; Nakat and Madanat 2008). This approach has been the main focus of researchers in the past in managing transportation infrastructure. For example, Ben-Akiva et al. (1993) and Ben-Akiva and Ramaswamy (1993) predicted the deterioration of infrastructure assets using a statistical approach when panel data sets were incomplete. Chu and Durango-Cohen (2008) suggested using the auto-regressive integrated moving average (ARIMA) model and a Kalman filter to deal with missing data related to pavement condition indicators. Other researchers (Hong and Prozzi 2006; Kobayashi et al. 2012; Lethanh et al. 2015), estimated the parameters of predictive models using Markov models in the absence of complete data sets.

The second approach is to use techniques to interpolate the data using the values that have been recorded for the nearby points. This technique is useful when dealing with rail and road data and is based on the idea that because road/rail segments are longitudinally connected if the segments are sufficiently close, the related values of condition indicators will be spatially correlated. For example, to find the value of a condition indicator for a road segment, a weighted mean of the values of the two nearest road segments can be used, where the weights are negatively correlated with the distance of the segments to the target segment (Al-Zou'bi et al. 2015). Research on these approaches is not abundant, however. Notable work in this area includes that of Farhan and Fwa (2013, 2014, 2015), who used a linear regression model with a stochastic error term to estimate the missing information using temporally correlated data. Sabillon-Orellana (2021) studied the process of data imputation for several popular models including linear interpolation, moving average models (simple, exponential, and ARIMA), regression models (mean, deterministic, and stochastic), and spline interpolation to determine the best model for pavement texture data imputation in terms of accuracy and time efficiency. According to the results, linear interpolation provided the best results, followed by moving average models. However, the results by spline interpolation and regression models were not as satisfactory as the other models.

Al-Zou'bi et al. (2015) categorized the techniques to estimate the missing pavement condition indicators into two groups, namely model-based and model-free techniques. The model-based techniques, as the name suggests, require a mathematical model to estimate the missing data. This means that a time- or parameter-dependent function should be initially defined to use these techniques, such as the work done by Paterson (1986) and Stampley et al. (1995). Model-free techniques rely solely on interpolation techniques such as the work done by Kestler et al. (1994), who investigated the spatial dependency of falling weight deflectometer test values at various distances on a road segment using geostatistical analysis. In another study by Lethanh et al. (2016), a univariate Kriging model was used to predict the missing values of road segment condition indicators based on their spatial correlation.

Despite the merits of the available literature in estimating the values of missing condition indicator data for transportation infrastructure, some improvements can be made in the exploitation of the spatial correlation of data points. For example, although using a univariate Kriging model has shown relatively high levels of accuracy in missing data estimation, research on using the value of other condition state indicators as assisting information is nonexistent in dealing with data related to the transportation networks. This may also be helpful in situations when data on the target condition indicator are insufficient. For this reason, it is interesting to compare the results of using a multivariate Kriging model (co-Kriging) with a univariate model. Moreover, the current literature has not compared the efficiency of different linear and nonlinear techniques in the estimation of spatially correlated data.

In recent years, the application of neural networks in value prediction and pattern recognition is receiving increasing attention. Consequently, it is interesting to investigate the efficiency of combining the Kriging model and an artificial neural network (ANN) model in the estimation of missing condition indicator values related to transportation infrastructure. Another possibility that needs further investigation is using a bidirectional long short-term memory (bi-LSTM) neural network model, which has shown great potential in extracting information using the latent correlation among the target point and its nearby points (Bengio et al. 1994).

Consequently, this paper advances the current literature by
- Considering the auxiliary variables (values of different condition indicators) in models to explore their influence on the estimation of the missing values;
- Proposing a novel hybrid ANN-Kriging model to enhance the capability of the Kriging-based models in using spatial correlation of condition indicators; and
- Investigating the performance of different models (simple and complex) in estimating missing data related to infrastructure assets.

In this investigation, the spatially correlated data from two real-world case studies are taken to estimate the values of the condition indicators using four models: univariate Kriging model, co-Kriging model, ANN-Kriging hybrid model, and bi-LSTM neural network model. The models are compared, and their merits and limitations are discussed.

The structure of the rest of the paper is as follows. The "Methodology" section provides the details of different models to estimate the value of missing data. In the next section, the first case study related to a road network is presented. It includes the data preparation and processing procedure and the results of estimating the condition indicator values using the models presented in the "Methodology" section. The procedure is then repeated for a case study related to a rail network. The discussion section provides the overall results and the merits and limitations of the models. Finally, the conclusions and future research directions are summarized.

Methodology

This section provides a detailed explanation of the models that are used in this study to estimate the missing condition indicator values.

Univariate Kriging Model

Kriging is a well-known model that estimates the value of an indicator at a target point using the spatial correlation of the surrounding data points. It determines the optimal weights to adjust the influence of nearby points in estimating the value of the target point, i.e., it uses a linear combination of the observed values. This model is widely used in geoscience and was named by the French mathematician Georges Matheron on the basis of work by Danie G. Krige (Krige 1951; Lichtenstern 2013; Matheron 1963).

Bennett et al. ([1984](#)) reviewed several methods for estimating the missing information in spatial statistics and concluded that Kriging models were preferable. Lethanh et al. ([2016](#)) implemented univariate Kriging to estimate the missing data related to pavement condition indicators using the spatial correlation between the data points. Shtiliyanova et al. ([2017](#)) used a Kriging-based interpolation method to fill data gaps in the temporal dimension of air temperatures. The study suggested that the Kriging model could successfully extract the information in time-series data, in addition to geographical spatial data. Yang et al. ([2018](#)) proposed a spatiotemporal Kriging-based data estimation approach that estimated the missing data from multiple sensors including Bluetooth and remote traffic microwave sensors. These studies all suggested that they could estimate the missing data with a relatively high level of accuracy.

Univariate Kriging models use n observation attributes, $z_1, z_2, \ldots, z_n$, and their spatial coordinates, $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, to estimate the attributes z_0 at point (x_0, y_0). In this model, it is assumed that the spatial attribute z of an arbitrary point (x, y) in the space can be represented with the expected attribute of this point $c(x, y)$ and a random error $R(x, y)$

$$z(x, y) = c(x, y) + R(x, y) \tag{1}$$

where the expected attribute is represented by a function $f_k(x, y)$ and a coefficient a_k, such that $c(x, y) = \sum_{k=1}^{m} a_k f_k(x, y)$; and $R(x, y)$ = random error with a mean equal to 0 and satisfies

$$Cov[R(x_i, y_i), R(x_j, y_j)] = C(d_{ij})$$

where C = covariance function related to the distance between point i and j. A commonly used covariance function is indicated in Eq. ([2](#)):

$$C(d_{ij}) = \exp(-\rho d_{ij}^2) \tag{2}$$

where ρ = scaling coefficient; and d_{ij} = Euclidean distance between (x_i, y_i) and (x_j, y_j). The Kriging model requires the estimated value $\hat{z}_0$, which should be the best linear unbiased estimation of the true value z_0. Consequently, the difference between $\hat{z}_0$ and z_0 should be minimized, and the error is constrained to have a mean equal to 0. The estimated attribute $\hat{z}_0$ at point (x_0, y_0) is shown as follows:

$$\hat{z}(x_0, y_0) = \sum_{i=1}^{n} w_i z(x_i, y_i) = \sum_{i=1}^{n} w_i \left[\sum_{k=1}^{m} a_k f_k(x_i, y_i) + R(x_i, y_i) \right] \tag{3}$$

From Eqs. ([1](#)) and ([3](#)), it is inferred that

$$\sum_{i=1}^{n} w_i \sum_{k=1}^{m} f_k(x_i, y_i) = \sum_{k=1}^{m} f_k(x_0, y_0) \tag{4}$$

The minimum estimation of variance is constrained such that

$$\min_{\lambda} Var(\hat{z}_0 - z_0) = \min_{\lambda} Var\left(\sum_{i=1}^{n} w_i z_i - z_0 \right) \tag{5}$$

To calculate the optimal λ_i value, one must convert Eq. ([4](#)) into matrix format. A linear function is used in Eq. ([6](#)) for showing the matrix equation of the univariate Kriging

$$
\begin{bmatrix}
0 & \gamma(d_{12}) & \gamma(d_{13}) & & \gamma(d_{1n}) & 1 & x_1 & y_1 \\
\gamma(d_{21}) & 0 & \gamma(d_{23}) & \cdots & \gamma(d_{2n}) & 1 & x_2 & y_2 \\
\gamma(d_{31}) & \gamma(d_{32}) & 0 & & \gamma(d_{3n}) & 1 & x_3 & y_3 \\
& \vdots & & \ddots & & \vdots & & \\
\gamma(d_{n1}) & \gamma(d_{n2}) & \gamma(d_{n3}) & & 0 & 1 & x_n & y_n \\
1 & 1 & 1 & \cdots & 1 & 0 & 0 & 0 \\
x_1 & x_2 & x_3 & & & 0 & 0 & 0 \\
y_1 & y_2 & y_3 & & y_n & 0 & 0 & 0
\end{bmatrix}
\cdot
\begin{bmatrix}
w_1 \\ w_2 \\ w_3 \\ \vdots \\ w_n \\ \lambda \\ a_1 \\ a_2
\end{bmatrix}
=
\begin{bmatrix}
\gamma(d_{01}) \\ \gamma(d_{02}) \\ \gamma(d_{03}) \\ \vdots \\ \gamma(d_{0n}) \\ 1 \\ x_0 \\ y_0
\end{bmatrix}
\tag{6}
$$

where $\gamma(d_{ij})$ = semi-variogram model value for distance d_{ij} (see [Cressie 1985](#); [Olea 2006](#); [Jian et al. 1996](#)) for more information]; d_{ij} = distance between point i and j; x_i, y_i = coordinates of point i; w_i = weight for point i; λ = Lagrange multiplier, which is used to minimize the estimation error; a_1, a_2 = coefficients of the linear expected attributes function; and x_0, y_0 = coordinates of the estimated point. When the weights $w_1, w_2, \ldots, w_n$ are calculated, the values are taken into Eq. ([3](#)) to calculate the estimated attribute value $\hat{z}_0$.

It is worth mentioning that most Kriging models use the semi-variogram function instead of the covariance function since $\gamma(d) = sill - C(d)$, where the $sill$ = value on the y axis that the semi-variogram model attains at the range (range is the distance where the model first flattens out).

Consequently, the steps of the univariate Kriging model are as follows:
1. Determine the neighbor points (input points) of the target point (output point);
2. Calculate the distance between all input point pairs and fit the parameters of the semi-variogram model;
3. Calculate the semi-variogram values among input points and the semi-variogram values between the target point (output) and all input points using the fitted model;
4. Calculate the weights [Eq. ([6](#))]; and
5. Calculate the estimated attributed value of the output point [Eq. ([3](#))].

Co-Kriging Model

Univariate Kriging uses the spatial information of the target data point and its neighborhood, with respect to one variable. In the real world, however, there are normally multiple variables (indicators) that may have correlations. Moreover, in some situations the target variable has insufficient data and cannot support the estimation process; hence, variables that have a spatial correlation with the target variable could be considered to assist the estimation process. Zarei et al. ([2011](#)) claimed that it is necessary to use multivariate geostatistical models when there are insufficient data related to the target variable.

Co-Kriging is a multivariate Kriging model that can consider the influence of auxiliary variables (values of different condition indicators) to estimate the value of the target point. The application of co-Kriging has been investigated by several researchers; for example, Knotters et al. ([1995](#)) compared the efficiency of univariate Kriging with co-Kriging and a hybrid Kriging-regression model, which could consider auxiliary variables, to interpolate horizon depth with censored observations. Both multivariate models showed a better performance than univariate Kriging. Yates and Warrick ([1987](#)) used co-Kriging to estimate gravimetric moisture content with two additional features: namely, the bare soil surface

© ASCE

04022041-3

J. Infrastruct. Syst.

temperature and the percentage sand content. The results suggest that using highly correlated auxiliary variables in the co-Kriging model will significantly improve estimation accuracy, whereas using poorly correlated auxiliary variables will decrease the performance even in comparison with the univariate Kriging model.

The co-Kriging model uses the covariance between two or more spatially correlated variables within a region. In general, co-Kriging is used when the investigated indicator variable is sparse, but the related indicator variables are abundant. The variables used in co-Kriging should be autocorrelated, and the main variable and assisting variables should be cross-related (Cressie 2015). The matrix format of a two-variable co-Kriging model can be written as

$$\begin{bmatrix} G_{AA} & G_{AB} & 1_n & 0 \\ G_{BA} & G_{BB} & 0 & 1_m \\ 1'_n & 0' & 0 & 0 \\ 0' & 1'_m & 0 & 0 \end{bmatrix} - \begin{bmatrix} w \\ \eta \\ \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} \gamma_{0A} \\ \gamma_{0AB} \\ 1 \\ 0 \end{bmatrix} \quad (7)$$

where G_{AA} = semi-variogram matrix with shape $n \times$ that contains the values of $\gamma_A(d)$ calculated from variable A; $G_{AB} = G'_{BA}$ is a cross-variogram matrix (see Cressie 1985; Myers 1982 for more information) with shape $n \times m$ that contains the values of $\gamma_{AB}(d)$ calculated from variables A and B; G_{BB} = semi-variogram matrix with shape $m \times m$ that contains the values of $\gamma_B(d)$ calculated from variable B; 1_n and 1_m = matrix with all elements equal to 1 with the shapes $n \times n$ and $m \times m$, respectively; w = column vector with n weights assigned to variable A; η = column vector with m weights assigned to variable B; γ_{0A} = column vector with n semi-variogram values calculated by $\gamma_A(d)$, where the distance is from the output point to all the input points of variable A; γ_{0AB} = column vector with m cross-variogram values calculated by $\gamma_{AB}(d)$, where the distance is from the output point to all the input points of variable B; and λ_1 and λ_2 = column vectors with Lagrange parameter for variables A and B, respectively.

The estimated attribute $\hat{z}_0$ at point (x_0, y_0) for a two-variable co-Kriging model is as follows:

$$\hat{z}_0 = \sum_i^n w_i A_i + \sum_j^m \eta_j B_j \quad (8)$$

Similar to the univariate Kriging model, the steps of co-Kriging are as follows:
1. Determine the neighbor points (input points) of the target point (output point);
2. Calculate the distance between all input point pairs and fit the parameters of semi-variogram models $\gamma_A(d)$ and $\gamma_B(d)$ and cross-variogram model $\gamma_{AB}(d)$;

3. Calculate the semi-variogram values and cross-variogram values among input points for variables A and B, and calculate the semi-variogram values and cross-variogram values between output point and all input points for variables A and B using the fitted model;
4. Calculate the weights based on the co-Kriging formula [Eq. (7)]; and
5. Calculate the estimated attributed value of the output point [Eq. (8)].

ANN-Kriging

ANNs are analytical models that handle problems whose solutions are not simply articulated using a mathematical model of how the output varies with the input. ANN can derive meaning from a complicated data set, which can be very helpful in value prediction and pattern recognition. Studies suggest that the combination of an ANN with the Kriging model can improve the capacity for estimation of missing data points. For example, Demyanov et al. (1998) successfully implemented a two-step algorithm, namely direct neural network residual Kriging (DNNRK), on climate data. Tapoglou et al. (2014) combined ANNs, Kriging, and fuzzy logic algorithms to simulate the distribution of hydraulic head in groundwater using spatial and temporal attributes. The results suggest that their model could be successfully implemented in aquifers where geological characteristics are uncertain. In a recent study, Yasrebi et al. (2020) compared the accuracy of the ordinary Kriging (OK) model with a hybrid OK-ANN to estimate elemental distribution based on lithogeochemical data in porphyry deposits. According to the study, the correlation coefficient between the estimated results and the raw data was above 84%, which shows the capability of the OK-ANN technique in the estimation of elemental distribution related to porphyry deposits using geochemical data.

However, most researchers who have combined ANN and Kriging did not deeply hybridize the two models; rather they only stacked them in series, similar to connecting two singular layers. In this study, however, a deep hybrid ANN-Kriging model is proposed (Fig. 1) and its application is investigated in the estimation of missing condition indicators related to road and railway networks.

The proposed hybrid ANN-Kriging model replaces the semi-variogram fitting in the Kriging with a neural network model that has great potential for fitting any type of function regardless of whether it is linear or nonlinear. This model needs two sets of input, neighborhood points that help in estimating the target point value, and the basic coordinate and the auxiliary attribute of the target point. The information from the neighborhood points are converted to the semi-variogram matrix, and target point information is converted to a semi-variogram value vector. The weight vector related
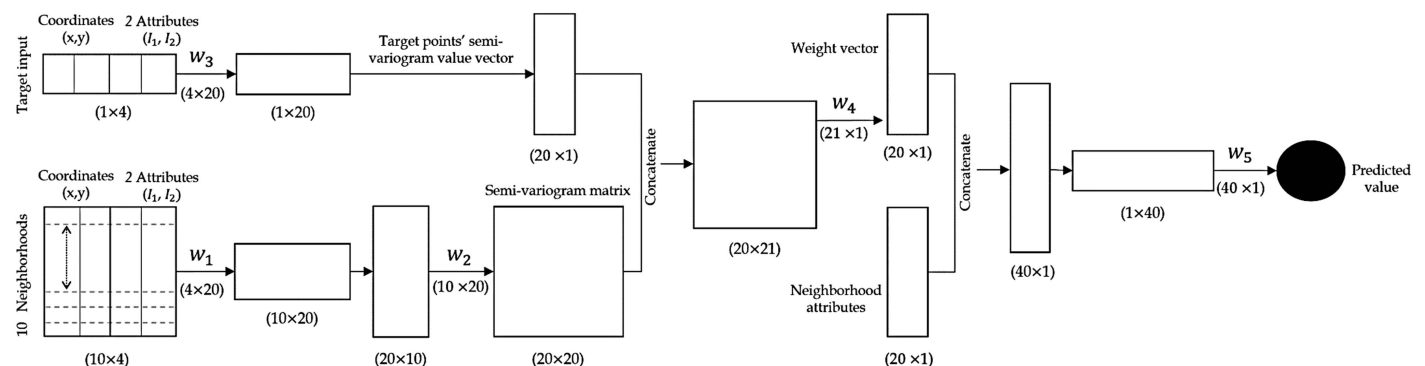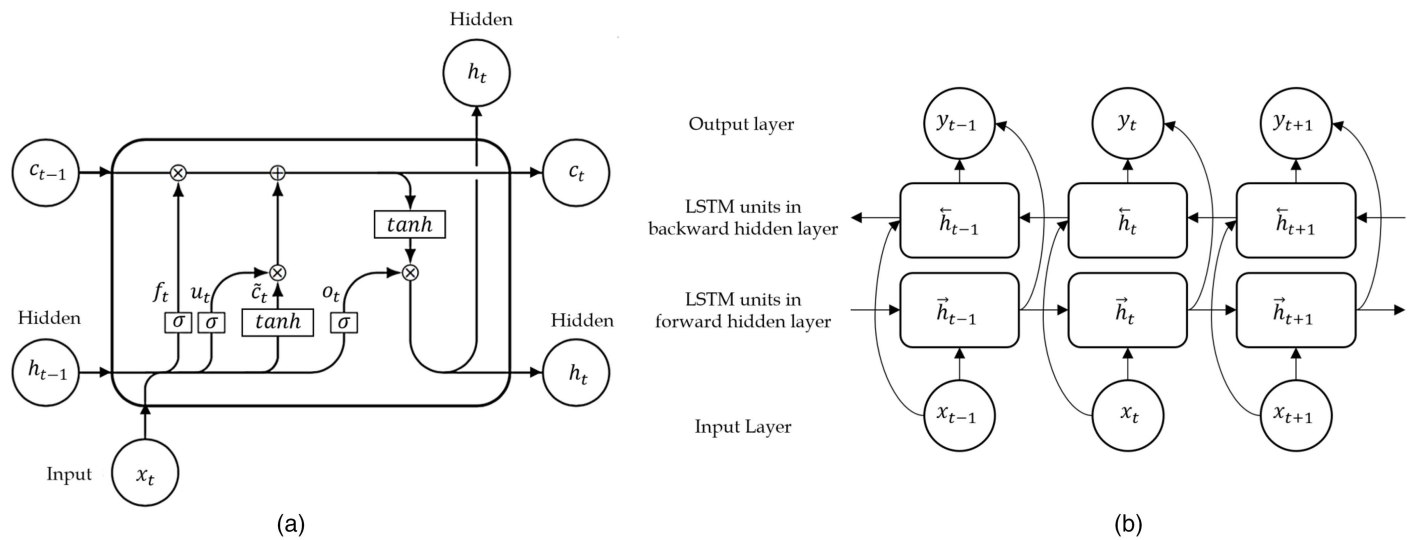

Fig. 1. ANN-Kriging hybrid model architecture with two auxiliary variables and 10 neighborhoods.

Fig. 2. Structure of LSTM units: (a) LSTM unit; and (b) bi-LSTM.

to each neighborhood is calculated based on the results from ANN, and the final estimated value is calculated by multiplying the weight vector and the attributes of the neighborhood. When using multiple auxiliary variables in the ANN-Kriging model, the dimension of the semi-variogram matrix is the same as that of the co-Kriging model [Eq. (7)]. However, the 0 and 1 elements in the semi-variogram matrix in Eq. (7) are ignored in this model, because they are used in the Kriging and co-Kriging models only to minimize the errors.

Bi-LSTM Neural Network

Bi-LSTM is a variant of recurrent neural network (RNN), which can estimate an unknown value in a sequence by making full use of its neighborhood. LSTM captures long-time dependencies. Vanishing or exploding gradient problems may happen in RNNs when modeling long-time dependencies in sequential data, and LSTM solves these problems by adding the forget gate, update gate, and output gate in the recurrent unit. Bi-LSTM is a combination of a forward LSTM and a backward LSTM and can extract the information before and after the estimation point in the sequential data, hence making full use of the neighborhood to estimate the target point value (Graves et al. 2013).

The model is widely used in areas dealing with sequential data structures, including speech recognition, text translation, and natural language processing. For example, Graves et al. (2005) carried out experiments on the speech corpus with a bi-LSTM network and compared the performance with unidirectional LSTM and RNN. The results proved that bidirectional LSTM outperforms on tasks with sequential structure compared with the other two. Huang et al. (2015) proposed a bi-LSTM model with a conditional random field (CRF) layer for sequence tagging and produced state-of-the-art accuracy on this type of task. In a study by Fan et al. (2014), a deep neural network (DNN) was combined with bi-LSTM to capture the correlation or co-occurrence information in the speech utterance sequence. The model outperformed individual DNN and bi-LSTM and also conventional RNN and LSTM.

The application of bi-LSTM in the estimation of the road/rail missing data is unprecedented, so it is interesting to test its potential in capturing sequence characteristics of condition indicators and finding patterns in the sequences where long roads/rails are involved.

Fig. 2(a) shows the structure of an LSTM unit and how the cells are updated in the model. Eqs. (9)–(14) explain the meaning of notations in the figure

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \tag{9}$$

$$u_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \tag{10}$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \tag{11}$$

$$\tilde{c}_t = \tanh(W_c[h_{t-1}, x_t] + b_c) \tag{12}$$

$$c_t = f_t \odot c_{t-1} + u_t \odot \tilde{c}_t \tag{13}$$

$$h_t = o_t \odot \tanh(c_t) \tag{14}$$

where f_t, u_t, o_t = forget gate, update gate, and output gate, respectively; $\tilde{c}_t$ = candidate cell state, which is used to update the cell state; c_t = current cell state; c_{t-1} = previous cell state; h_t = current hidden state; h_{t-1} = previous hidden cell state; x_t = current time step input; W_f, W_i, W_o, W_c = weighted matrices; b_f, b_i, b_o, b_c = bias terms (all these parameters are trainable); $\odot$ = Hadamard product; σ = sigmoid function; and tanh = hyperbolic tangent function.

The forget gate f_t decides the information that is to be dropped from the previous hidden cell state h_{t-1} and the current time step input x_t. The update gate u_t decides the information that is to be stored. The candidate cell state $\tilde{c}_t$ contains new information of the previous hidden state h_{t-1} and the current time step input x_t. The updated cell state c_t is the sum of the Hadamard product of the previous cell state c_{t-1} and the forget gate f_t and the Hadamard product of the candidate cell state $\tilde{c}_t$ and the update gate u_t. The outputs of the sigmoid function and tanh function are between 0 and 1. They are used to decide the percentage of the input information that is to be retained and the percentage of the previous information that is to be dropped. The updated hidden cell state h_t is the Hadamard product of the output gate o_t and the updated cell state c_t. The last hidden cell state is passed to an activation function to get the final prediction (Bengio et al. 1994; Hochreiter and Schmidhuber 1997).

Bi-LSTM extracts the previous and later information by processing the forward and backward directions using two separate hidden layers and combines the two layers at the corresponding

© ASCE

04022041-5

J. Infrastruct. Syst.

time step to the output layer. Fig. 2(b) shows the structure of bi-LSTM. In this figure, x_t = input sequence; $\overrightarrow{h}_t$ = forward hidden units; $\overleftarrow{h}_t$ = backward hidden units; and y_t = output sequence. When calculating the forward hidden layer, bi-LSTM will iterate from $t = 1$ to T, and for the backward hidden layer, it will iterate from $t = T$ to 1. Then, it updates the output layer related to the corresponding time step. Eqs. (15)–(17) illustrate the process

$$\overrightarrow{h}_t = \mathcal{H}(W_{x\overrightarrow{h}}x_t + W_{\overrightarrow{h}\overrightarrow{h}}\overrightarrow{h}_{t-1} + b_{\overrightarrow{h}}) \tag{15}$$

$$\overleftarrow{h}_t = \mathcal{H}(W_{x\overleftarrow{h}}x_t + W_{\overleftarrow{h}\overleftarrow{h}}\overleftarrow{h}_{t+1} + b_{\overleftarrow{h}}) \tag{16}$$

$$y_t = W_{\overrightarrow{h}y}\overrightarrow{h}_t + W_{\overleftarrow{h}y}\overleftarrow{h}_t + b_y \tag{17}$$

where $\mathcal{H}$ = hidden layer function, derived from Eqs. (9)–(14); $W_{x\overrightarrow{h}}, W_{\overrightarrow{h}\overrightarrow{h}}, W_{x\overleftarrow{h}}, W_{\overleftarrow{h}\overleftarrow{h}}, W_{\overrightarrow{h}y}, W_{\overleftarrow{h}y}$ = weighted matrix related to each input and hidden unit; and $b_{\overrightarrow{h}}, b_{\overleftarrow{h}}, b_y$ = bias terms. All these parameters are trained during the model training process.

Stacking multiple bi-LSTM layers can give a better performance on the prediction. Assuming the hidden layer functions are the same for each bi-LSTM layer, using the output from the previous bi-LSTM layer as the input for the next bi-LSTM layer, one can build a deep bi-LSTM neural network with n layers as indicated in Eqs. (18) and (19)

$$h_t^n = \mathcal{H}(W_{h^{n-1}h^n}h_t^{n-1} + W_{h^n h^n}h_{t-1}^n + b_h^n) \tag{18}$$

$$y_t = W_{h^N y}h_t^N + b_y \tag{19}$$

where h_t^n = hidden vector sequences concatenated by $\overrightarrow{h}_t^n$ and $\overleftarrow{h}_t^n$ at each bi-LSTM layer, computed from $n = 1$ to N and $t = 1$ to T iteratively; $W_{h^{n-1}h^n}, W_{h^n h^n}, W_{h^N y}h_t^N$ = weighted matrix; and b_h^n, b_y = bias terms.

In the following sections, the above models are used in two separate case studies related to infrastructure networks to estimate the value of missing condition indicators, and the results are compared. When lacking information on the maintenance activities and the historical condition state of the assets, in both studies, it is assumed that the missing data related to rail/road segments are not missing because of maintenance activities being executed on those segments.

Case Study I: Estimating the Value of Condition Indicators for Road Segments

This investigation is related to predicting the "surface defect" values in a road network using the models introduced in "Methodology" section. The data used in this study were taken from an inspection campaign conducted on a road network in 2009, in which three road indicators—surface defects and longitudinal and transversal unevenness—were measured. There are 1,290 roads in the network, and each road is divided into 100-m-long segments. The surface defects indicator refers to the surface damage without rut depth. The observations for all indicators were recorded by high-speed inspection cars, with one value recorded to represent the average state for each 100-m road segment (Lethanh et al. 2016). The raw collected values were then converted to a continuous value in the range (0, 5), with 0 being the best state. This was done following the available standards and guidelines, and the results were stored in an inventory that was used for this study. The inventory includes information regarding the position (coordinates and the axis distance), the date, and the indicator values

Table 1. Descriptive statistics of the variables

Variables	Data points	Mean	Standard deviation	Min	Median	Max
Surface defects[a]	35,022	1.01	0.46	0	0.8	5
Longitudinal unevenness	35,330	1.52	0.84	0	1.27	5
Transversal unevenness	35,339	0.98	0.92	0	0.75	5

[a]Main variable.

related to inspections for each road segment. Separate columns provide information on the start and end points of the road segments, where the difference between the points shows the length of the segment (100 m). Because the coordinates are one-dimensional, the estimation results will not be influenced by the choice of the reference point (start, middle, or end point) as long as the choice stays consistent. In this study, the start points are used as reference points for each road segment.

For the univariate Kriging, only the values of the surface defect in the neighborhood of the target point were considered for the target value estimation. For the co-Kriging, ANN-Kriging, and bi-LSTM, two auxiliary variables (longitudinal and transversal unevenness) were also used. In imputation of the missing values for each 100-m segment using the neighboring points, it was assumed that the traffic loads and surface type were the same on each road in the network.

The correlation coefficient between surface defects and longitudinal unevenness is 0.159, and the correlation coefficient between surface defects and transversal unevenness is 0.141. The descriptive statistics of the variables are shown in Table 1.

Data Processing

In the first step, each road segment was given a unique road ID by combining two columns that provide information on the road name and the axis distance. An illustration of the division of a road into 100-m segments and the lanes in both directions of traffic, A and B, is presented in Fig. 3.

Subsequently, the roads were clustered separately by their unique ID. This means that only the segment records within a specific road were used to estimate the value of the condition indicator for the target segment in that road. The roads that were short and included a only few segments were removed from the analysis because they did not provide sufficient information from the neighboring points. In this study, the minimum number of segments was different for each neighborhood size. Consequently, for neighborhood sizes of 6, 7, 8, 9, and 10, the minimum segments required were 7, 8, 9, 10, and 11, which resulted in the availability of 736, 619, 551, 491, 426, and 369 roads for the analysis, respectively.
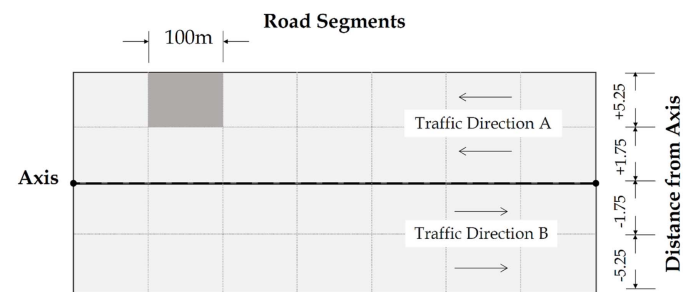
Fig. 3. Representation of the distance from the road start point (AxisDist_begin) and the axis distance (SideDist).

In the next step, to convert the one-directional coordinates into (X, Y) format, the starting point of the road segments was set as X and the Y value was set as 0, assuming that the road segments within the same lane (same axis distance) follow a straight line. Finally, the data points were split into training and test sets with 80% and 20% share of the total data points, respectively. The target point value was estimated, and the estimation error was calculated in a rolling process. For example, with a neighborhood size of 4, the first 5 consecutive road segments were taken, the middle point (segment) was removed, the remaining points were used to estimate the value of the middle point, and the estimation error for the middle point was calculated. Then the first road segment was put aside, the next five consecutive road segments were taken, and the process was repeated until the last five consecutive road segments were reached. Then the estimation errors were averaged to represent the estimation error for the specific road.

All models were coded in Python from scratch. The implementation of the univariate Kriging and co-Kriging models followed guidelines from Spatialanalyst.net (2020a, b). Tensorflow 2.5.0 and Pytorch were used to implement bi-LSTM and ANN-Kriging, respectively.

In constructing the bi-LSTM model, the number of input neurons was based on the number of target point neighbors. In this study, two hidden layers were used, with eight and 16 hidden neurons, respectively. The single output neuron provided the final prediction of the target value for the road segment.

Results

Figs. 4–6 illustrate the mean squared error (MSE) of the test set in estimating the value of surface defects for road segments, using four models: namely, univariate Kriging, co-Kriging, ANN-Kriging, and bi-LSTM. In univariate Kriging, only the surface defect values for the neighboring points were considered; thus the MSE values for the univariate Kriging are the same in all three figures.

In Fig. 4, the values related to the longitudinal unevenness of the target segment, and the neighboring points are also considered in the prediction process for co-Kriging, bi-LSTM, and ANN-Kriging models. As can be observed, the MSE for the co-Kriging model increases when the values of this condition indicator are considered. The reason is that surface defects and longitudinal unevenness of the road segments do not have a strong linear correlation. Bi-LSTM shows a better performance than the univariate and co-Kriging models, although as the number of the neighborhood points increases, the performance of the bi-LSTM regresses. The reason behind this might be related to the insufficiency of the data related to roads with a large number of segments to train the bi-LSTM model. Moreover, the efficiency of this model is more
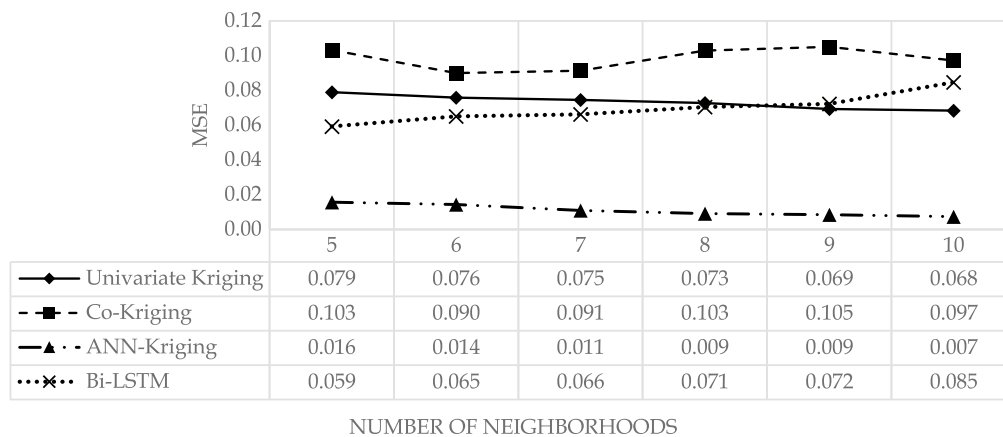

	5	6	7	8	9	10
Univariate Kriging	0.079	0.076	0.075	0.073	0.069	0.068
Co-Kriging	0.103	0.090	0.091	0.103	0.105	0.097
ANN-Kriging	0.016	0.014	0.011	0.009	0.009	0.007
Bi-LSTM	0.059	0.065	0.066	0.071	0.072	0.085

NUMBER OF NEIGHBORHOODS

Fig. 4. Mean squared error for estimating the surface defects, using longitudinal unevenness as an auxiliary variable.

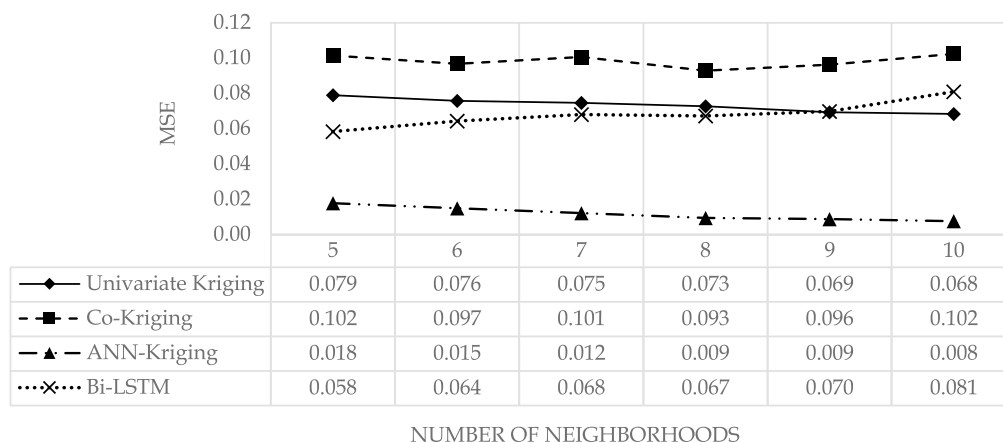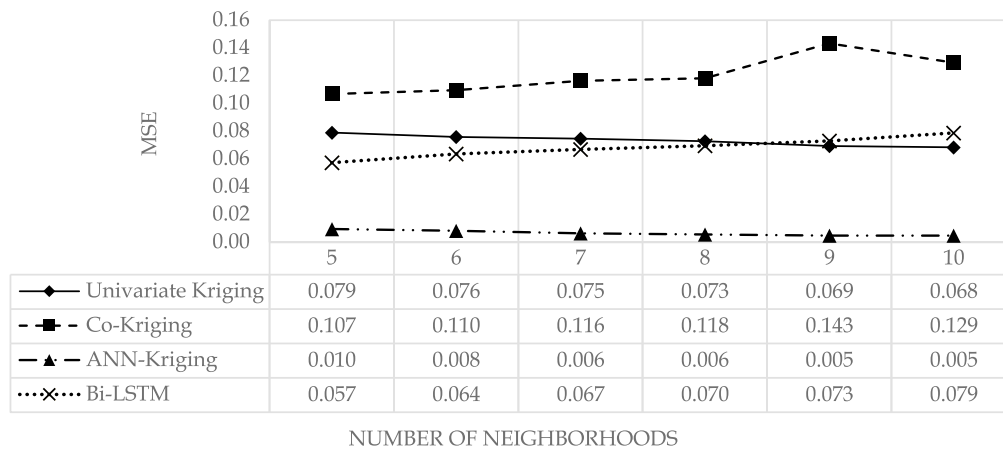|	5	6	7	8	9	10
Univariate Kriging	0.079	0.076	0.075	0.073	0.069	0.068
Co-Kriging	0.102	0.097	0.101	0.093	0.096	0.102
ANN-Kriging	0.018	0.015	0.012	0.009	0.009	0.008
Bi-LSTM	0.058	0.064	0.068	0.067	0.070	0.081

NUMBER OF NEIGHBORHOODS

Fig. 5. Mean squared error for estimating the surface defects, using transversal unevenness as an auxiliary variable.

Fig. 6. Mean squared error for estimating the surface defects, using longitudinal and transversal unevenness as auxiliary variables.

	5	6	7	8	9	10
◆ Univariate Kriging	0.079	0.076	0.075	0.073	0.069	0.068
■ Co-Kriging	0.107	0.110	0.116	0.118	0.143	0.129
▲ ANN-Kriging	0.010	0.008	0.006	0.006	0.005	0.005
✕ Bi-LSTM	0.057	0.064	0.067	0.070	0.073	0.079

NUMBER OF NEIGHBORHOODS

significant in high-dimensional datasets, but the road data used in this example contain only three features. The ANN-Kriging model shows significant improvement over univariate Kriging (around 80%), and its performance improves as the number of neighborhood points increases.

The situation is the same when transversal unevenness is considered as an auxiliary variable (Fig. 5). The MSE of the co-Kriging is higher than the MSE of the univariate Kriging, meaning that the correlation between the surface defects and transversal unevenness is not strong enough to improve the estimation accuracy in comparison to univariate Kriging.

Fig. 6 illustrates the average MSE for the scenario in which both the longitudinal and transversal unevenness values are used as auxiliary variables to estimate the value of the surface defects in road segments. The results suggest that considering both of these variables improves the performance ANN-Kriging model. For the bi-LSTM, by comparing the results in Fig. 6 with the second scenario (Fig. 5), whose MSE values are generally lower than the first scenario (Fig. 4), we see an improvement in the results with 5, 6, 7, and 10 neighboring points and a decrease in performance with 8 and 9 neighborhoods. The performance of the co-Kriging model decreases even more than when single auxiliary variables are used (regardless of the neighborhood size), which confirms that using the values of poorly correlated variables decreases the prediction performance of co-Kriging models.

Case Study II: Estimating the Value of Condition Indicators for Railway Track Segments

This case study is related to predicting the condition state of the track segments in a rail network. An inventory of the network assets including 23,140 tracks was created in the year 2020, where each track has several segments. The data contain information related to track condition states, track material, type, and life cycle information (auxiliary variables). The auxiliary variables include renewal, tamping, grinding, rail replacement, and "other" maintenance requirements, in addition to delay and safety risks. The condition state of each segment is attributed based on a weighted average of these auxiliary variables that include the lifecycle expenditures on different interventions and the related consequences. The range value of the condition indicators ranges is (1, 5), where 1 represents the best state and 5 represents the worst. These converted values were then recorded in an inventory that was used in this study.

In the inventory, the tracks have a unique ID along with the information related to each particular segment of the track. The ID and coordinate information of the segments are stored in the column "Nr./ID" with the format "ID, start point—end point." The track segments do not necessarily have the same length. The coordinate of the first segments in tracks start from 0 to the length of the segment, and the next segment follows that value. For example, the segment with Nr./ID "110359, 0–13.068" belongs to the track with the ID 110359; it is the first segment of the track, as it starts from 0, and it has a length of 13.063 meters, as the end point is 13.068. In this study, the start points are used as reference points for each track segment, although as the coordinates are one-dimensional, the estimation results are not influenced by the choice of the reference point as long as it is consistent.

The goal of this study is to compare the efficiency of the previously introduced models in estimating the value of the condition states for the track segments. For univariate Kriging, only the values of condition state indicators in the neighborhood of the target point are considered; for co-Kriging, ANN-Kriging, and BiLSTM, the auxiliary variables are also used. The descriptive statistics of the variables are shown in Table 2.

Data Processing

The tracks were initially clustered separately by their unique track IDs. This means that only the segment records within a specific track were used to estimate the value of the condition indicator for the target segment in that track. The tracks that were short and included only a few segments were removed from the analysis

Table 2. Descriptive statistics of the variables

Variables	Data points	Mean	Standard deviation	Min	Median	Max
Condition state[a]	736,249	3.18	1.26	1.2	3.19	5
Renewal requirements	736,249	3.02	1.15	1.06	3.00	5
Tamping requirements	736,249	1.98	1.17	1	1.06	5
Grinding requirements	736,249	2.10	1.32	1	1.58	5
Rail replacement	248,368	3.06	1.90	1	3.37	5
Other maintenance requirements	736,249	3.88	1.57	1	5.00	5
Delay risks	593,242	2.34	1.75	1	1.00	5
Safety risks	736,249	2.83	1.87	1	1.99	5

[a]Main variable.

because they did not provide sufficient information from the neighboring points. Excluding those tracks also eliminated the bias for comparing the models. The maximum number of neighborhoods to consider was set as 10; therefore each track needed to have at least 11 records. In total, of 23,140 tracks, 9,031 had more than 11 records and were used in the analysis.

In the next step, the Nr./ID column was split into the track ID and the track coordinate. The coordinate of the track was normalized to reduce the measurement error when calculating the distance between the points. Finally, the data points were split into training and test sets, with 80% and 20% share of the total data points, respectively; the target point value was estimated; and the estimation error was calculated in a rolling process (as in Case Study I).

In constructing the bi-LSTM model, the number of input neurons was based on the number of target point neighbors. In this study, two hidden layers were used, with eight and 16 hidden neurons respectively.

In this study, the number of auxiliary variables used in the co-Kriging, bi-LSTM, and ANN-Kriging models was decided for each track separately, such that for each track, the variable combination with the best performance was selected to calculate the average MSE. Before testing different variable combinations, the spatial autocorrelation (Moran's I) and the Pearson correlation coefficient were calculated to check the sufficiency of the variables. The variables that did not satisfy the requirements (Moran's I values within the 95% confidence interval and Pearson coefficients larger than 0.6) were ignored.

Results

The results of the MSE for the test set for each model using different neighborhood numbers are shown in Fig. 7. The results suggest that ANN-Kriging, bi-LSTM, and co-Kriging all outperform univariate Kriging, since the main and auxiliary variables have strong enough correlations and the models can use the additional information from the auxiliary variables. As in the previous case study, bi-LSTM performs better than co-Kriging because the structure of the neural network has a stronger ability to extract the potential correlation between the estimation point and its neighborhood, and ANN-Kriging shows the highest accuracy among the four models.

Regarding the number of neighbors, a trend appears for co-Kriging and bi-LSTM models where the estimation accuracy improves with increasing the neighborhood size, while this is reversed for univariate and ANN-Kriging models.

Discussion

By observing the results from both case studies, the following can be concluded:

- When using the co-Kriging model, one has to pay attention to the correlation between the target indicator and the auxiliary variables, since a low correlation can decrease performance and make the predictions worse than in a univariate model. The changes related to the performance of co-Kriging with respect to neighborhood size are case dependent and do not follow a specific trend.

- ANN-Kriging significantly improves the performance of univariate Kriging, even in situations where the condition indicators are not highly correlated. It seems that the ability of the ANN model to fit the spatial correlation among neighborhood points and the target point is more efficient than that of the semivariogram model in Kriging.

- In comparison to Kriging models, bi-LSTM can better extract the hidden relationship between the target point and its neighborhood. The characteristic of bi-LSTM that can deal with sequential data makes it a suitable model for estimating the missing indicator values related to the road and track segments. However, it must be noted that sufficient training data must be available to exploit the model's true potential. Moreover, the efficiency of deep learning models such as bi-LSTM is more significant in high-dimensional datasets; hence the second example could better showcase the potential of this model in capturing sequence characteristics of condition indicators and finding patterns in the data where long tracks were involved.

- There is an interesting trend in the changes related to the performance of ANN-Kriging and univariate Kriging with respect to neighborhood size. If one model's performance is improved by increasing the neighborhood size, the other's performance will increase and vice versa. This makes sense, of course, since the hybrid model is a combination of the traditional geostatistical model and the NN model. Hence, they share the same core, although the hybrid model performs remarkably better.
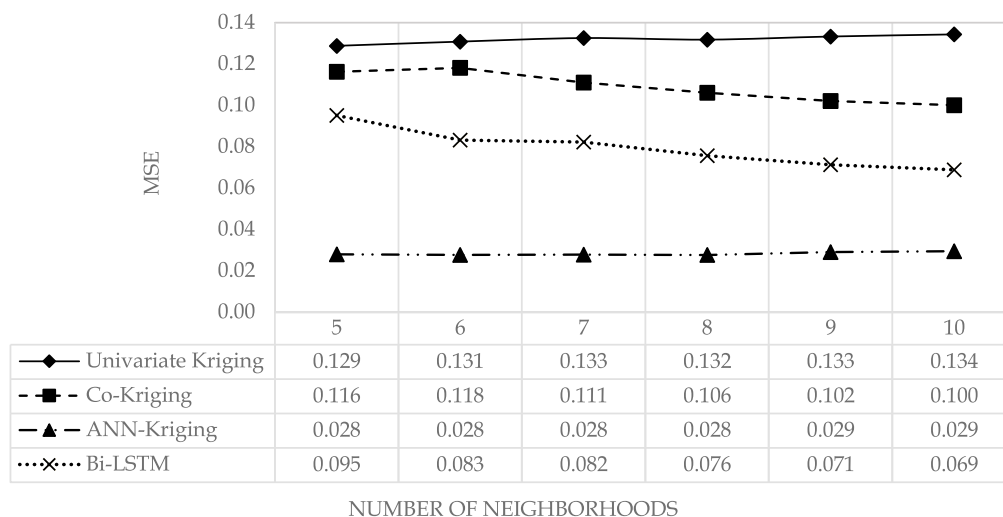

	5	6	7	8	9	10
—◆— Univariate Kriging	0.129	0.131	0.133	0.132	0.133	0.134
– ■ – Co-Kriging	0.116	0.118	0.111	0.106	0.102	0.100
– ▲ – ANN-Kriging	0.028	0.028	0.028	0.028	0.029	0.029
⋯×⋯ Bi-LSTM	0.095	0.083	0.082	0.076	0.071	0.069

NUMBER OF NEIGHBORHOODS

Fig. 7. Mean squared error for estimating the condition state of the track segments for different neighborhood numbers.

- In both case studies, for simplicity, it was assumed that one segment within each neighborhood contained missing information, and a rolling process method, as explained earlier, was used to estimate the average imputation errors for each road/rail segment. However, it must be noted that missing data scenarios can be more complex, and there can be situations where two or more consecutive segments are missing. Although all four models can deal with scenarios with multiple missing data points, the accuracy of the results might be affected depending on the availability of neighborhood information to impute the missing data.

Conclusions

In this paper, four different models, univariate Kriging, co-Kriging, ANN-Kriging, and bi-LSTM, were introduced to investigate the value of missing condition indicator data in transportation infrastructure, where the data loss is not due to the execution of maintenance activities. The univariate Kriging model exploits the spatial correlation between data points, while co-Kriging explores both the spatial correlation and the influence of the correlated auxiliary variables. ANN-Kriging is used to enhance the capability of extracting the spatial correlation among data points while investigating the influence of correlated auxiliary variables. Bi-LSTM uses a sequential data storage structure to extract the potential correlation among the data points and the correlated auxiliary variables.

The efficiency of these models was investigated in two case studies. The first case study was related to a road network where the goal was to estimate the value of surface defects at a target point, using the values of its neighborhood and two auxiliary variables, longitudinal and transversal unevenness. The goal of the second case study was to estimate the condition state of the rail segments that were attributed based on their lifecycle expenditure, using the value of the neighboring points and seven other auxiliary variables. The results of both studies suggested that all models could estimate the value of target points with reasonably high levels of accuracy. However, the performance of the ANN-Kriging model was significantly better than the other three models. Bi-LSTM outperformed the univariate and co-Kriging models, while the performance of the co-Kriging model relied on the level of correlation among the auxiliary variables and the target indicator. As expected, when this correlation was weak, using the auxiliary variables did not help the estimation accuracy and even decreased the performance in comparison to a univariate model.

The introduced models can be used in practice and help infrastructure managers in charge of monitoring the condition states of the road and rail networks to overcome issues related to real-world data, such as missing data or faulty entries. This will reduce the prospective costs incurred by having to remeasure the indicator values at that position. The models are fairly easy to implement and require only the asset coordinates and the related condition indicator values as input. These models will help in having a complete and relatively accurate data set and improve the decisions related to the optimal time to execute corrective interventions on the network assets.

These models can be used to estimate the missing condition indicator values when it can be confirmed that the missing data is not related to the execution of maintenance activities on those segments, or as a conservative decision when there is no information available to determine that.

A future study should consider maintenance history and combine the temporal and spatial information to improve the estimation accuracy. Of course, this is possible only when the temporal data and the maintenance records are available.

Another potential study would be implementing the bi-LSTM model in a case study with more auxiliary variables, where the geospatial correlations of the assets are available (rather than just one-dimensional coordinates) and comparing the results with the proposed ANN-Kriging model, to see if a well-trained bi-LSTM can outperform the ANN-Kriging model.

Data Availability Statement

Some or all data, models, or code generated or used during the study are proprietary or confidential in nature and may be provided only with restrictions. The restriction applies to the original railway and road data, and if requested they will be anonymized.

References

Al-Zou'bi, M. M., C. M. Chang, S. Nazarian, and V. Kreinovich. 2015. "Systematic statistical approach to populate missing performance data in pavement management systems." *J. Infrastruct. Syst.* 21 (4): 04015002. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000247.

Anastasopoulos, P. C., and F. L. Mannering. 2015. "Analysis of pavement overlay and replacement performance using random parameters hazard-based duration models." *J. Infrastruct. Syst.* 21 (1): 04014024. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000208.

Ben-Akiva, M., F. Humplick, S. Madanat, and R. Ramaswamy. 1993. "Infrastructure management under uncertainty: Latent performance approach." *J. Transp. Eng.* 119 (1): 43–58. https://doi.org/10.1061/(ASCE)0733-947X(1993)119:1(43).

Ben-Akiva, M., and R. Ramaswamy. 1993. "An approach for predicting latent infrastructure facility deterioration." *Transp. Sci.* 27 (2): 174–193. https://doi.org/10.1287/trsc.27.2.174.

Bengio, Y., P. Simard, and P. Frasconi. 1994. "Learning long-term dependencies with gradient descent is difficult." *IEEE Trans. Neural Networks* 5 (2): 157–166. https://doi.org/10.1109/72.279181.

Bennett, R. J., R. P. Haining, and D. A. Griffith. 1984. "The problem of missing data on spatial surfaces." *Ann. Assoc. Am. Geogr.* 74 (1): 138–156. https://doi.org/10.1111/j.1467-8306.1984.tb01440.x.

Chu, C.-Y., and P. L. Durango-Cohen. 2007. "Estimation of infrastructure performance models using state-space specifications of time series models." *Transp. Res. Part C Emerging Technol.* 15 (1): 17–32. https://doi.org/10.1016/j.trc.2006.11.004.

Chu, C.-Y., and P. L. Durango-Cohen. 2008. "Estimation of dynamic performance models for transportation infrastructure using panel data." *Transp. Res. Part B Methodol.* 42 (1): 57–81. https://doi.org/10.1016/j.trb.2007.06.004.

Cressie, N. 1985. "Fitting variogram models by weighted least squares." *J. Int. Assoc. Math. Geol.* 17 (5): 563–586. https://doi.org/10.1007/BF01032109.

Cressie, N. 2015. *Statistics for spatial data.* Hoboken, NJ: Wiley.

Demyanov, V., M. Kanevsky, S. Chernov, E. Savelieva, and V. Timonin. 1998. "Neural network residual kriging application for climatic data." *J. Geog. Inf. Decis. Anal.* 2 (2): 215–232.

Fan, Y., Y. Qian, F.-L. Xie, and F. K. Soong. 2014. "TTS synthesis with bidirectional LSTM based recurrent neural networks." In *Proc., 15th*

Annual Conf. of the International Speech Communication Association. Washington, DC: International Speech Communication Association.

Farhan, J., and T. F. Fwa. 2013. "Airport pavement missing data management and imputation with stochastic multiple imputation model." *Transp. Res. Rec.* 2336 (1): 43–54. https://doi.org/10.3141/2336-06.

Farhan, J., and T. F. Fwa. 2014. "Augmented stochastic multiple imputation model for airport pavement missing data imputation." *Transp. Res. Rec.* 2449 (1): 96–104. https://doi.org/10.3141/2449-11.

Farhan, J., and T. F. Fwa. 2015. "Improved imputation of missing pavement performance data using auxiliary variables." *J. Transp. Eng.* 141 (1): 04014065. https://doi.org/10.1061/%28ASCE%29TE.1943-5436.0000725.

Graves, A., S. Fernández, and J. Schmidhuber. 2005. "Bidirectional LSTM networks for improved phoneme classification and recognition." In Vol. 3697 of *Proc., 15th Int. Conf. Artificial Neural Networks: Formal Models and Their Applications—ICANN 2005*, edited by W. Duch, J. Kacprzyk, E. Oja, and S. Zadrożny. Berlin: Springer. https://doi.org/10.1007/11550907_126.

Graves, A., A. Mohamed, and G. Hinton. 2013. "Speech recognition with deep recurrent neural networks." In *Proc., 2013 IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, 6645–6649. New York: IEEE.

Hochreiter, S., and J. Schmidhuber. 1997. "Long short-term memory." *Neural Comput.* 9 (8): 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

Hong, F., and J. A. Prozzi. 2006. "Estimation of pavement performance deterioration using Bayesian approach." *J. Infrastruct. Syst.* 12 (2): 77–86. https://doi.org/10.1061/(ASCE)1076-0342(2006)12:2(77).

Huang, Z., W. Xu, and K. Yu. 2015. "Bidirectional LSTM-CRF models for sequence tagging." Preprint, submitted August 9, 2015. http://arxiv.org/abs/1508.01991.

Jian, X., R. A. Olea, and Y.-S. Yu. 1996. "Semivariogram modeling by weighted least squares." *Comput. Geosci.* 22 (4): 387–397. https://doi.org/10.1016/0098-3004(95)00095-X.

Kestler, M. A., M. E. Harr, R. L. Berg, and D. M. Johnson. 1994. "Spatial variability of falling weight deflectometer data: A geostatistical analysis." In *Proc., 4th Int. Conf., Bearing Capacity of Roads and AirfieldsFHWA*. Washington, DC: Transportation Research Board.

Knotters, M., D. J. Brus, and J. H. O. Voshaar. 1995. "A comparison of kriging, co-kriging and kriging combined with regression for spatial interpolation of horizon depth with censored observations." *Geoderma* 67 (3–4): 227–246. https://doi.org/10.1016/0016-7061(95)00011-C.

Kobayashi, K., K. Kaito, and N. Lethanh. 2012. "A statistical deterioration forecasting method using hidden Markov model for infrastructure management." *Transp. Res. Part B Methodol.* 46 (4): 544–561. https://doi.org/10.1016/j.trb.2011.11.008.

Krige, D. G. 1951. "A statistical approach to some basic mine valuation problems on the Witwatersrand." *J. Chem. Metal. Mining Soc. South Africa* 52 (6): 119–139. https://doi.org/10.10520/AJA0038223X_4792.

Lethanh, N., K. Kaito, and K. Kobayashi. 2015. "Infrastructure deterioration prediction with a Poisson hidden Markov model on time series data." *J. Infrastruct. Syst.* 21 (3): 04014051. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000242.

Lethanh, N., C. Richmond, and B. T. Adey. 2016. "Investigation of the ability to estimate values of road section condition indicators based on their spatial correlation." *J. Infrastruct. Syst.* 22 (3): 04016006. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000290.

Lichtenstern, A. 2013. *Kriging methods in spatial statistics*. Munich, Germany: Technische Universität München.

Matheron, G. 1963. "Principles of geostatistics." *Econ. Geol.* 58 (8): 1246–1266. https://doi.org/10.2113/gsecongeo.58.8.1246.

Myers, D. E. 1982. "Matrix formulation of co-kriging." *J. Int. Assoc. Math. Geol.* 14 (3): 249–257. https://doi.org/10.1007/BF01032887.

Nakat, Z. S., and S. M. Madanat. 2008. "Stochastic duration modeling of pavement overlay crack initiation." *J. Infrastruct. Syst.* 14 (3): 185–192. https://doi.org/10.1061/(ASCE)1076-0342(2008)14:3(185).

Olea, R. A. 2006. "A six-step practical approach to semivariogram modeling." *Stochastic Environ. Res. Risk Assess.* 20 (5): 307–318. https://doi.org/10.1007/s00477-005-0026-1.

Paterson, W. 1986. "International roughness index: Relationship to other measures of roughness and riding quality." *Transp. Res. Rec.* (1084): 49–59.

Sabillon-Orellana, C. A. 2021. *Evaluation of data imputation techniques in pavement texture processing*. Austin, TX: Univ. of Texas.

Shtiliyanova, A., G. Bellocchi, D. Borras, U. Eza, R. Martin, and P. Carrère. 2017. "Kriging-based approach to predict missing air temperature data." *Comput. Electron. Agric.* 142 (Part A): 440–449. https://doi.org/10.1016/j.compag.2017.09.033.

Spatialanalyst.net. 2020a. "Co-Kriging." Accessed March, 10, 2021. http://spatial-analyst.net/%0AILWIS/htm/ilwisapp/cokriging_algorithm.htm.

Spatialanalyst.net. 2020b. "Universal Kriging." Accessed March, 10, 2021. http://spatial-analyst.%0Anet/ILWIS/htm/ilwisapp/universal_kriging_algorithm.%0Ahtm.

Stampley, B. E., B. Miller, R. E. Smith, and T. Scullion. 1995. *Pavement management information system concepts, equations, and analysis models*. Washington, DC: Transportation Research Board.

Tapoglou, E., G. P. Karatzas, I. C. Trichakis, and E. A. Varouchakis. 2014. "A spatio-temporal hybrid neural network-Kriging model for groundwater level simulation." *J. Hydrol.* 519 (Part D): 3193–3203. https://doi.org/10.1016/j.jhydrol.2014.10.040.

Yang, H., J. Yang, L. D. Han, X. Liu, L. Pu, S. Chin, and H. Hwang. 2018. "A Kriging based spatiotemporal approach for traffic volume data imputation." *PLoS One* 13 (4): e0195957. https://doi.org/10.1371/journal.pone.0195957.

Yasrebi, A. B., A. Hezarkhani, P. Afzal, R. Karami, M. E. Tehrani, and A. Borumandnia. 2020. "Application of an ordinary kriging–artificial neural network for elemental distribution in Kahang porphyry deposit, Central Iran." *Arabian J. Geosci.* 13 (15): 1–14. https://doi.org/10.1007/s12517-020-05607-0.

Yates, S. R., and A. W. Warrick. 1987. "Estimating soil water content using cokriging." *Soil Sci. Soc. Am. J.* 51 (1): 23–30. https://doi.org/10.2136/sssaj1987.03615995005100010005x.

Zarei, A., M. Masihi, and K. Salahshoor. 2011. "Comparison of different univariate and multivariate geostatistical methods by porosity modeling of an Iranian oil field." *Pet. Sci. Technol.* 29 (19): 2061–2076. https://doi.org/10.1080/10916461003681703.