A geographically weighted deep neural network model for research on the spatial distribution of the down dead wood volume in Liangshui National Nature Reserve (China)

Yuman Sun (1), Ziqi Ao (1), Weiwei Jia (1), Ying Chen (1), Kai Xu (2)

In natural forest ecosystems, there is often abundant down dead wood (DDW) due to wind disasters, which greatly changes the size and structure of forests. Accurately determining the DDW volume (DDWV) is crucial for sustaining forest management, predicting the dynamic changes in forest resources and assessing the risks of natural disasters or disturbances. However, existing models cannot accurately express the significant spatial nonstationarity or complexity in their spatial relationships. To this end, we established a geographically weighted deep neural network (GWDNN) model that constructs a spatially weighted neural network (SWNN) through geographic location data and builds a neural network through stand factors and remote sensing factors to improve the interpretability of the spatial model of DDWV. To verify the effectiveness of this method, using 2019 data from Liangshui National Nature Reserve, we compared model fit, predictive ability and residual spatial autocorrelation among the GWDNN model and four other spatial models: an ordinary least squares (OLS) model, a linear mixed model (LMM), a geographically weighted regression (GWR) model and a deep neural network (DNN) model. The experimental results show that the GWDNN model is far superior to the other four models according to various indicators; the coefficient of determination $R^2$, root mean square error (RMSE), mean absolute error (MAE), Moran's I and Z-statistic values of the GWDNN model were 0.95, 1.05, 0.77, -0.01 and -0.06, respectively. In addition, compared with the other models, the GWDNN model can more accurately depict local spatial variations and details of the DDWV in Liangshui National Nature Reserve.

Keywords: Down Dead Wood Volume (DDWV), Ordinary Least Squares (OLS) Model, Linear Mixed Model (LMM), Geographically Weighted Regression (GWR) Model, Deep Neural Network (DNN) Model, Geographically Weighted Deep Neural Network (GWDNN) Model

Introduction

Coarse woody debris (CWD) is an important material in forest ecosystems (Hagan & Grove 1999) and mainly includes standing dead trees, down dead wood (DDW), snags and large branches (Harmon et al. 1986). DDW is an important component of CWD (Pfeil et al. 2007, Jenkins et al. 2004). In Liangshui National Nature Reserve, DDW often results from windthrows, which are defined as the uprooting of trees by winds (Lavoie et al. 2012). DDW not only has an economic impact (Costa & Ibanez 2005, Kinnucan 2016) but also plays key roles in nutrient cycling, carbon storage, vegetation succession and the maintenance of biodiversity (Chojnacky & Heath 2002, Polo et al. 2015). Therefore, the accurate determination of the spatial distribution of DDW volume (DDWV) is particularly important for disaster prediction and sustainable forest management.

At present, commonly used spatial effect models can be divided into two categories: parametric models and nonparametric models (Fassnacht et al. 2014, Adamiec & Drapela 2017). The former refers to statistical regression methods, such as the ordinary least squares (OLS) model, spatial regression models and the geographically weighted regression (GWR) model. These methods can be used to easily determine the relationships between response and predictor variables. However, spatial effects in data often appear in the form of patches or geographic gradients, which can violate the independence and homogeneity assumptions of OLS and other traditional statistical methods (Green et al. 2005, Subedi et al. 2018). To this end, to include spatial effects in the regression framework, scholars have used spatial regression models, which estimate the covariance matrix to model the spatial autocorrelation of variables in adjacent locations; examples include the spatial lag model.
With the development of computers, nonparametric models, including machine learning methods such as k-nearest neighbor (KNN), artificial neural network (ANN—Deb et al. 2017), support vector machine (SVM), random forest and deep neural network (DNN—Schmidhuber 2015) models, have been found to handle the response variables and predict the complexity of the relationships among variables, and these models are widely used within each domain. However, neural networks rarely consider the spatial weighting problem independently and rarely consider or directly add location coordinates into the network as variables (Chen et al. 2018, Mohammadinia et al. 2019). At present, only geographical general regression neural network (GGRNN) and geographically neural network weighted regression (GNNWR) models combine spatial effects and machine learning. The GGRNN model combines the kernel function of the GWR model, which is fixed as a spatial weight, with the GRNN model (Irfan et al. 2016), and the GNNWR model combines the spatially weighted neural network (SWNN) with the OLS model (Du et al. 2020, Wu et al. 2020, 2021). Both models can improve model fitting and prediction results. However, studies of these models are few. In this paper, the geographically weighted deep neural network (GWDNN) model is introduced, which combines the SWNN with the intermediate output parameters of machine learning (specifically, a DNN). The SWNN matrix can solve the problem of choosing an incorrect weight function in a complex relationship (Du et al. 2020). Furthermore, the form of the weight function does not require a priori assumptions (Du et al. 2020). Theoretically, the GWDNN model achieves stronger fitting accuracy and prediction performance than the general spatial model and can more accurately describe spatial relationships in many domains, especially the complexity of forest ecosystems.

By realizing accurate modeling of the DDWV with stand factors and remote sensing factors, this study demonstrated the potential of GWDNNs to describe spatial relationships. This method establishes a two-layer DNN structure, in which one layer of the neural network is used to establish the spatially weighted matrix, and the other layer is used to estimate a parameter for each variable and realizes precise deconstruction and efficient calculation of spatial nonstationarity. To investigate whether the GWDNN model can improve the prediction of DDWV, the Liangshui National Nature Reserve in the city of Yichun (China) was considered as a study area. The coefficient of determination $R^2$, root mean square error (RMSE), mean absolute error (MAE), Moran’s I and Z-statistic values of five models (OLS model, LMM, GWR, DNN and GWDNN models) were compared to assess model fit, prediction accuracy and residual spatial autocorrelation. The advantages of the GWDNN model in predicting spatial relationships were verified.

### Materials and methods

#### Study area

The study area of Liangshui National Nature Reserve is located in the city of Yichun, Heilongjiang Province, on the east slope of the Dalidailing branch south of the southern Xiaoxing’an Mountains and is surrounded by 6 forest farms of the Daling Forestry Experimental Bureau. The geographical location is $47 \degree 07′ 14′′$-14′′ $′14′′-40′′ $′128° 48′ 08′′-128° 55′ 55′′ 45′′$ E ($\text{Fig. I}$). The total area of the reserve is 12,133 ha, with 98% forest coverage, and the total forest volume is 1.7 million cubic meters. The study area has a temperate continental climate with an average annual temperature of $-0.3 ^\circ C$, and the annual average precipitation is 676.0 mm. The main forest type is mixed forest, and there are large tracts of primitive Pinus koraiensis forest and secondary birch and broad-leaved forest. The main broad-leaved tree species include Betula platyphylla, Quercus mongolica, Pseudolarix amurense, Fraxinus mandshurica and Juglans mandshurica (Zhen et al. 2013).

#### Survey data (variable factors)

The data were collected during an investigation for the forest resource planning and design of the Liangshui reserve in 2019, which is called the second-class survey. For the purpose of forest resource management, 31 compartments and 464 subcompartments were defined. In addition, systematic sampling was conducted every 1 km. There were 150 fixed sample plots. Each plot was round-shaped with an area of 0.06 ha. The 150 sample plots comprised 65 coniferous and broad-leaved mixed forest plots, 31 broad-leaved mixed forest plots, and 34 coniferous relatively pure forests. The main variables recorded in 150 fixed plots were geographic location (latitude and longitude), elevation (from digital elevation model – DEM, m), slope (SLOPE, deg), aspect, average age of the dominant tree species (AGE, years), canopy, average diameter at breast height of the dominant tree species (DBH, cm), total standing forest stock (TSFS, m$^3$) and DDWV (m$^3$). For the determination of DDWV, we recorded tree species, quantity and DBH of the DDW in each sample plot and calculated DDWV through the unitary volume table for Liangshui National Nature Reserve.

The experimental data included the fixed sample plot data and visible light (R, G and B bands) remote sensing images, which were derived from the Chinese Academy of Forestry (CAF)’s LICHy Hyperspectra (AISA Eagle II) airborne observation system (Pang et al. 2016). This system is a push broom imaging system comprising a hyperspectral sensor and a data acquisition unit housed in a rugged control computer. The designed flight altitude is 1050 m, and the flight speed is approximately 160 km h$^{-1}$. The image rate can be up to 160 Hz, the focal length is 18.1 mm, and the field of view (FOV) is 37.7′, with a spectral resolution of 354

---

**Fig. 1** - The study area in Yichun (northeastern China).
9.6 nm and a spatial resolution of 0.52 m. In the imaging process of remote sensing images, due to changes in flight attitude, height and speed, the image pixels are distorted, offset, compressed or stretched relative to the actual position of the ground target. Therefore, we used the ENVI RPC model to perform geometric corrections for these deformations. Furthermore, for atmospheric correction, the ENVI FLAASH module was used to eliminate the absorption and scattering of electromagnetic waves from sunlight and ground objects by the atmosphere (Aquino et al. 2018; Sun et al. 2021). Seven kinds of visible vegetation indexes were extracted (Du & Noguchi 2016, Choi et al. 2020, Boonpook et al. 2021): the visible-band difference vegetation index (VDVI – eqn. 1), the excess green index (EXG – eqn. 2), the red/green ratio index (RGRi – eqn. 3), the excess green minus excess red index (EXGR – eqn. 4), the vegetation index (VEG – eqn. 5), the color index of vegetation (CIVE – eqn. 6), and the normalized green-red difference index (NGRDI – eqn. 7):

\[ VDVI = (2G - R - B)/(2G + R + B) \]  
(1)

\[ EXG = (2 - r - b) \]  
(2)

\[ RGRi = R/G \]  
(3)

\[ EXGR = EXG - 1.4r - g \]  
(4)

\[ VEG = g/(r^b - 2r^{a-1}) \]  
(5)

\[ CIVE = 0.44r - 0.88g + 0.39b + 18.79 \]  
(6)

\[ NGRDI = (G - R)/(G + R) \]  
(7)

where \( R, G \) and \( B \) are the red band, green band and blue band, respectively, and \( r, g \) and \( b \) are the normalized red band, green band and blue band, respectively.

**Variable screening**

The stepwise regression method was used to screen the seven stand factors and seven remote sensing factors. Then, to examine the models for multicollinearity, variance inflation factors (VIFs) were calculated (Miles 2005), and variables with VIFs greater than 10 were eliminated. Descriptive statistics of the dependent variable, stand variables and remote sensing variables are shown in Tab. 1.

**Parametric methods**

**OLS model**

The traditional linear regression model is based on the OLS method to establish a linear relationship, as shown in eqn. 8:

\[ Y_{OLS} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + e \]  
(8)

where \( Y \) is the dependent variable vector, \( X \) is the independent variable vector, and \( \beta \) is the OLS model coefficient vector.

**LMM**

The LMM can be used to incorporate fixed effects and a single compartment random effect, with 31 levels (Bates 2005), as expressed in eqn. 9:

\[ Y_{LMM} = \beta_0 + \beta_1 Y + e \]  
(9)

where \( Y \) is an \( n \times 1 \) dependent variable vector, \( X \) is an \( n \times p \) matrix of the \( p \) - 1 predictors (with the first column as 1 to estimate the intercept), \( n \) is the number of sample plots, \( p \) is the number of fixed-effect parameters, \( \beta \) is a \( p \times 1 \) vector of unknown fixed-effect parameters, \( Z \) is a known \( n \times q \) design matrix for the \( q \) random effects, and \( y \) is a \( q \times 1 \) vector that includes the empirical best linear unbiased predictors (EBLUPS) for the random effects. A diagonal structure was used for the covariance matrix of the compartment random-effects, while an exponential correlation structure was proposed for the residual (between-plots) covariance matrix (Zhang & Gove 2005).

**GWR model**

The GWR model is an extension of the traditional linear regression model and can establish local parameter estimates (Brunsdon et al. 1996). The GWR model is defined as follows (eqn. 10):

\[ Y_{GWR} = \beta_0(u_v) + \sum_{i=1}^{k} \beta_i(u_v) x_i + e \]  
(10)

where \( Y \) is the independent variable vector, \( (u_v) \) denotes the spatial coordinates of location \( i \), \( \beta_i(u_v) \) is the intercept, and \( \beta_i(u_v) \) is the coefficient of \( k \) explanatory variables. The estimation of parameters is obtained by eqn. 11:

\[ \hat{\beta}(u_v) = \frac{X W(u_v) Y}{X W(u_v) X} \]  
(11)

where \( W(u_v) \) is an \( n \times n \) weight matrix, the off-diagonal elements of which are zero, and the diagonal elements of which are the spatial weights. GWR spatial weights can be estimated by spatial kernel functions, with Gaussian and bisquare kernel functions currently being the most common methods (Fotheringham et al. 1998). The specific expression for the Gaussian function is (eqn. 12):

\[ W_{ij} = \exp \left( -\frac{d_{ij}^2}{2b^2} \right) \]  
(12)

for the bisquare function (eqn. 13):

\[ W_{ij} = \left( 1 - \frac{d_{ij}^2}{b^2} \right)^2 \]  
(13)

where \( b \) is a non-negative decreasing function describing the functional relationship between the weight and the distance, called the bandwidth, and \( d_i \) is the Euclidean distance. The model parameter estimation and prediction accuracies largely depend on the bandwidth choice. We used the corrected Akaike information criterion (AICc) to select the GWR model with the optimal bandwidth (Subedi et al. 2018).

**Nonparametric models**

**DNN model**

With the development of computer technology, DNNs have a wide range of applica-

### Tab. 1 - Descriptive statistics of the dependent variable, stand variables and remote sensing variables

| Variables                        | Name                  | Mean  | SD   | Min  | Q1   | Q3   | Max     |
|----------------------------------|-----------------------|-------|------|------|------|------|---------|
| Dependent variable               | Down dead wood volume | 3.44  | 4.72 | 0.01 | 0.73 | 3.9  | 25.63   |
| Stand variables                  | Elevation (DEM, m)    | 401.06| 76.22| 270  | 344  | 430  | 638     |
|                                  | Slope (SLOPE, deg)    | 8.58  | 5.21 | 1    | 6    | 10   | 36      |
|                                  | Aspect                | 5.45  | 2.31 | 1    | 4    | 7    | 9       |
|                                  | Average age of dominant tree species (AGE, years) | 98.08 | 47.57 | 30 | 55 | 150 | 170 |
|                                  | Canopy                | 0.53  | 0.14 | 0.3  | 0.4  | 0.6  | 0.9     |
|                                  | Average diameter at breast height of dominant tree species (DBH, cm) | 40.36 | 18.36 | 11.6 | 26 | 54 | 91     |
|                                  | Total standing forest stock (TSFS, m²) | 14.94 | 8.17 | 0.89 | 9.4 | 18.45 | 45.97  |
| Remote sensing variables         | Visible-band difference vegetation index (VDVI) | 0.33  | 0.03 | 0.27 | 0.31 | 0.36 | 0.41    |
|                                  | Excess green index (EXG) | 0.5   | 0.05 | 0.39 | 0.47 | 0.54 | 0.64    |
|                                  | Red-green ratio index (RGRi) | 0.58  | 0.05 | 0.44 | 0.54 | 0.61 | 0.7     |
|                                  | Excess green minus excess red index (EXGR) | 0.6   | 0.09 | 0.41 | 0.55 | 0.67 | 0.86    |
|                                  | Vegetation index (VEG) | 1.95  | 0.01 | 1.67 | 1.88 | 2.02 | 2.29    |
|                                  | Color index of vegetation (CIVE) | 18.39 | 0.02 | 18.33 | 18.38 | 18.4 | 18.44   |
|                                  | Normalized green-red difference index (NGRDI) | 0.27  | 0.04 | 0.18 | 0.25 | 0.3  | 0.39    |
DNNs are designed to fit the complex function, where the expression of the transfer function is in the form of full connections. For a hidden layer, the expression of the transfer function is the rectified linear unit (ReLU) function, which is computationally efficient, has good sparsity and does not require unsupervised pretraining. The parameter initialization strategy enables the neural network model to learn useful information during the training process. However, the neural network model is prone to fall into local optima and is easily overfitted because the initial model is complex and the amount of noise in the training set is too large. It is necessary to use the dropout algorithm to discard neurons in the hidden layer with a certain probability in the iterative training process, but their weights will be retained. The network participates in the next training iteration, and the weights are updated, which is equivalent to using multiple models to train the same dataset. Finally, the weights are assigned to the neurons to optimize the network structure.

In the DNN model, 4 layers of the neural network are designed to fit the complex relationships between the independent and dependent variables. Among them, the five selected variables conform to the hidden layer, the 2 layers in the middle are the hidden layers, and the prediction results are the output layer. The hidden layer uses a fully connected layer. The activation function is the rectified linear unit (ReLU) function, which is computationally efficient, has good sparsity and does not require unsupervised pretraining. The parameter initialization strategy enables the neural network model to learn useful information during the training process. However, the neural network model is prone to fall into local optima and is easily overfitted because the initial model is complex and the amount of noise in the training set is too large. It is necessary to use the dropout algorithm to discard neurons in the hidden layer with a certain probability in the iterative training process, but their weights will be retained. The network participates in the next training iteration, and the weights are updated, which is equivalent to using multiple models to train the same dataset. Finally, the weights are assigned to the neurons to optimize the network structure.

In the GWR model, spatial weighting is achieved with a function such as the Gaussian function or double square function, but its structure is simple, and it is difficult to accurately estimate nonstationarity due to the difficulty in choosing the correct kernel function for complex relationships. To address this problem, using the superior fitting ability of neural networks, Wu et al. (2020) proposed an SWNN to construct the nonstationary weight matrix and calculated the kernel weights as the weights of a complex problem. According to this approach, the kernel weights of point i are computed as follows (eqn. 17):

\[ w_i = SWNN([d_1, d_2, ..., d_n]) \]  

where \([d_1, d_2, ..., d_n]\) are the distances from point i to all the samples. Consequently, the definition of the GWDNN model can be illustrated as shown in Fig. 2. The entire GWDNN contains three steps: (1) predicting the spatial weights through the coordinates (SWNN); (2) predicting the parameter \(\alpha\) by using the DNN in OLS form; and (3) obtaining the predicted values by multiplying the independent variable, the estimated spatial weights and the parameter \(\alpha\).

\[ Y_{\text{GWDNN}} = w_1(u, v) + w_2(u, v)X_1 + w_3(u, v)X_2 + w_4(u, v)X_3 + w_5(u, v)X_4 \]  

where \(w(u, v)\) is a 6×6 diagonal matrix representing the geographical weights for observation i (eqn. 16), and \(\alpha\) is the intermediate output layer parameter predicted by the DNN (eqn. 16):

\[ w(u, v) = \begin{bmatrix} w_1(u, v) & 0 & \cdots & 0 \\ 0 & w_2(u, v) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_5(u, v) \end{bmatrix} \]  

In the GWR model, spatial weighting is achieved with a function such as the Gaussian function or double square function, but its structure is simple, and it is difficult to accurately estimate nonstationarity due to the difficulty in choosing the correct kernel function for complex relationships. To address this problem, using the superior fitting ability of neural networks, Wu et al. (2020) proposed an SWNN to construct the nonstationary weight matrix and calculated the kernel weights as the weights of a complex problem. According to this approach, the kernel weights of point i are computed as follows (eqn. 17):

\[ w_i = SWNN([d_1, d_2, ..., d_n]) \]  

where \([d_1, d_2, ..., d_n]\) are the distances from point i to all the samples. Consequently, the definition of the GWDNN model can be illustrated as shown in Fig. 2. The entire GWDNN contains three steps: (1) predicting the spatial weights through the coordinates (SWNN); (2) predicting the parameter \(\alpha\) by using the DNN in OLS form; and (3) obtaining the predicted values by multiplying the independent variable, the estimated spatial weights and the parameter \(\alpha\).

The entire training processes of the DNN and GWDNN are shown in Fig. 3. This experiment uses simple cross-validation, and the dataset is randomly divided into training sets and validation sets. During the training step, when the training process meets the early stopping condition or when the number of epochs reaches the set maximum value, the training step is completed, and the best network weights are saved. When the training process is complete, the validation set is used to estimate the predictive power of the model.
The validation set does not participate in the model training process (Du et al. 2020). In this paper, the MAE is used as the training loss function of the DNN and GWDNN models, and the MAE of the validation set is used as the index of overfitting. By setting the maximum number of epochs to view the trend in the MAE of the training set and the verification set, we can find the optimal model parameters under the optimal number of iterations.

The OLS model and LMM were implemented with the “nlme” library in R, and the GWR model was implemented using the “mgcv” library in Python. The DNN and GWDNN models were implemented using the “TensorFlow” and “Keras” libraries in Python.

**Evaluation of DDWV models**

Based on ground-truthed DDWV samples, a total of 104 samples were randomly selected from among the 150 samples from Liangshui National Nature Reserve for model training, and 26 samples were used for model verification. The R², RMSE and MAE values were used to assess model fit and prediction ability.

Spatial effects include spatial autocorrelation and nonstationarity. Ignoring the spatial effects of a model will misleadingly reduce the significance of a test and the model’s predictive ability (Li et al. 2007). To investigate the spatial autocorrelation in the model residuals, this study calculated the spatial autocorrelation index (i.e., Moran’s I) for the residuals of the five models (eqn. 18):

\[
I = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (x_i - \bar{x}) (x_j - \bar{x})
\]

where \(x_i\) and \(x_j\) are the values of the sample locations \(i\) and \(j\) (where \(i \neq j\)), respectively, \(n\) is the number of sample plots, \(\bar{x}\) is the average of the observed values, and \(w_{ij}\) is the weight estimated according to the distance between sample locations \(i\) and \(j\).

The global Moran’s I can reveal the overall spatial autocorrelation of the study area, but if the significance of the spatial autocorrelation on each plot must be checked, the global Moran’s I must be localized (Tiefel...d model is the worst. The evaluation index results for the LMM and GWR model are similar. However, the fitting accuracy of the LMM is slightly higher than that of the GWR model, whereas prediction accuracy

### Model evaluation

Tab. 4 presents the training sets and validation sets of the 5 models for fitting and predicting DDWV. For both the training set and the validation set, the GWDNN model is the best performing model, and the OLS model is the worst. The evaluation index results for the LMM and GWR model are similar. However, the fitting accuracy of the LMM is slightly higher than that of the GWR model, whereas prediction accuracy

### Hyperparameter analysis of the nonparametric models

The setting of the parameters in the DNN and GWDNN models is particularly important in the training process. The DNN is fully connected with a total of 4 layers. The network layer includes 1 input layer, 2 hidden layers, and 1 output layer. The numbers of neurons are set to \(n \times 5\), \(n \times 2\), \(n \times 2\), \(n \times 1\); the initial learning rate (LR) is set to 0.001; the maximum number of epoch iterations is 2000; the batch size is 100; and the dropout rate of the dropout layer is 0.5. Two branches are set in the middle of the GWDNN structure: one branch is used to find the spatial weight, and the number of features in each layer is \([n \times 130, n \times 128, n \times 128, n \times 6]\); the other branch is used to find the parameters, and the number of features in each layer is \([n \times 6, n \times 64, n \times 64, n \times 6]\), among which the independent variables and the set 1 are used to obtain the output layer \(n+1\), the initial LR is set to 0.001, the maximum number of epoch iterations is 2000, the batch size is 5, and the dropout rate of the dropout layer is 0.4. The specific DNN and GWDNN hyperparameters are shown in Tab. 3.

### Results

**Model coefficients**

The dependent variable and the five standardized independent variables SLOPE, AGE, TSFS, VDVI, and VEG were used to establish the five models. The regression coefficients are significant at the \(\alpha\) level of 0.05, which shows the strong statistical significance of the five selected variables. The VIF values indicate no multicollinearity among the selected variables. Because the design of the GWDNN model is similar to that of the OLS model, we also calculated the coefficients for the GWDNN model (Tab. 2).

### Tab. 2 - OLS model, LMM, GWR model and GWDNN model results.

| Model     | Estimate | Intercept | SLOPE | AGE | TSFS | VDVI | VEG |
|-----------|----------|-----------|-------|-----|------|------|-----|
| OLS       | Coefficient | 3.58      | 1.39  | 2.65 | -2.04 | 1.34 | -1.79 |
|           | SE       | 0.31      | 0.33  | 0.42 | 0.41  | 0.34 | 0.37 |
|           | P-value  | <0.001    | <0.001| <0.001| <0.001| <0.001| <0.001|
|           | VIF      | -         | 1.19  | 1.81 | 1.67  | 1.23 | 1.33 |
| LMM       | Coefficient | 3.54      | 1.29  | 2.55 | -2.08 | 1.36 | -1.68 |
|           | SE       | 0.37      | 0.39  | 0.44 | 0.41  | 0.36 | 0.39 |
|           | P-value  | <0.001    | 0.002 | <0.001| <0.001| <0.001| <0.001|
|           | Mean_Coefficient | 3.28      | 1.16  | 2.46 | -2.08 | 1.29 | -1.58 |
| GWR       | Min_Coefficient | 2.6      | 0.14  | 1.95 | 3.66  | 0.36 | 2.37 |
|           | Median_Coefficient | 3.15      | 1.32  | 2.24 | -1.82 | 1.13 | 1.57 |
|           | Max_Coefficient | 4.43      | 1.81  | 3.74 | -1.19 | 2.43 | -0.78 |
| GWDNN     | Min_Coefficient | -2.2     | -0.14 | -0.25 | -9.26 | -0.16 | -2.85 |
|           | Median_Coefficient | 2.18     | 0.04  | 1    | -0.42 | 0.32 | -0.9 |
|           | Max_Coefficient | 6.98     | 0.35  | 10.98 | 10.36 | 0.94 | -0.29 |

### Model evaluation

Tab. 4 presents the training sets and validation sets of the 5 models for fitting and predicting DDWV. For both the training set and the validation set, the GWDNN model is the best performing model, and the OLS model is the worst. The evaluation index results for the LMM and GWR model are similar. However, the fitting accuracy of the LMM is slightly higher than that of the GWR model, whereas prediction accuracy

### Tab. 3 - DNN model and GWDNN model hyperparameter settings.

| Hyper-   | DNN model | GWDNN model |
|----------|-----------|-------------|
| parameter|           |             |
| Input    | n=5       | n=130 (A), n=6 (BC) |
| Hidden1  | n=32      | n=128, n=64 |
| Hidden2  | n=32      | n=128, n=64 |
| Hidden3  | -         | n=128       |
| Output   | n=1       | n=1         |
| LR       | 0.001     | 0.001       |
| Epoch maximum | 2000 | 2000 |
| Dropout  | 0.5       | 0.4         |
| Batchsize| 10        | 5           |
| Epoch stop| 1175   | 792         |
exhibits the opposite pattern. The non-parametric models perform much better than the parametric models in fitting and prediction accuracy. Compared with that of the GWR model, the RMSE of the DNN model was decreased by 28% and 41%, and that of the GWDNN model was decreased by 61% and 45%.

Tab. 5 - Global Moran’s I values of the residuals of the five models.

| Model   | Moran’s I | Z-statistic |
|---------|-----------|-------------|
| OLS     | 0.24      | 2.29        |
| LMM     | 0.06      | 0.8         |
| GWR     | 0.05      | 0.64        |
| DNN     | 0.17      | 1.95        |
| GWDNN   | -0.01     | -0.06       |

Tab. 4 - Simple cross-validation model evaluation of the training set and validation set.

| Models | Training set | Validation set |
|--------|--------------|----------------|
|        | $R^2$ | RMSE  | MAE  | $R^2$ | RMSE  | MAE  |
| OLS    | 0.57 | 3.08   | 2.26 | 0.43 | 3.60   | 2.57 |
| LMM    | 0.71 | 2.52   | 1.81 | 0.46 | 3.51   | 2.69 |
| GWR    | 0.69 | 2.59   | 1.91 | 0.52 | 3.31   | 2.34 |
| DNN    | 0.84 | 1.86   | 1.33 | 0.83 | 1.96   | 1.40 |
| GWDNN  | 0.95 | 1.05   | 0.77 | 0.85 | 1.82   | 1.39 |

Fig. 4 - The spatial correlation between residuals for the five models.

Fig. 5 - Spatial distribution of (a) the ground-truth DDWV (down dead wood volume, m$^3$), and (b-f) the results of the five predicting models.
OLS. The spatial correlation of the residuals of the DNN, which does not consider spatial effects, is high. When considering the spatial weights in the GWDNN model, the spatial autocorrelation is eliminated very well. To compare the spatial relationship of the residuals among the models, residual spatial correlation diagrams of the five models were constructed at 500 m intervals (Fig. 4). Moran’s I is relatively stable across step lengths for the GWDNN model and close to zero at all step lengths in the other models. This finding shows that the GWDNN model has a good ability to maintain spatial stability.

**Mapping of the spatial distribution of DDWV**

With reference to the ground-truthed DDWV data, the results of the five different models were compared. According to the general kriging interpolation in ArcGIS® v. 10.4 (ESRI, West Redlands, CA, USA) geo-statistical analysis, the spatial interpolation process is similar to the weighted moving average (Klobučar 2010). The cartography can be visualized by using the symbolic representation principle (Chen et al. 2018) to more directly depict the spatial variation and details of the DDWV in Liangshui National Nature Reserve (Fig. 5).

The overall trend in the measured DDWV and the spatial distributions from the five models are consistent. The DDWV values ranged from 2.00 to 7.00 m. To facilitate comparison, the values were divided into seven levels by equal intervals of 1 m (Fig. 5). The maps show the various spatial distributions of DDWV. All six maps show that the western part of the Liangshui reserve is a high DDWV region, with values greater than 7 m, while low DDWV regions are located near the river and road in the middle and western parts of Liangshui National Nature Reserve, with values ranging from 0.00 to 2.00 m. The DNN model predicted poor levels in this region. The map resulting from the GWDNN model exhibited more explicit spatial variations than the other maps.

On the whole, the GWDNN model results are closer to the measured data and can describe the spatial distribution of DDWV in more detail than the other model results. Such maps can guide resource allocation for forest management. The forest DDWV mapping results of these models were insufficient for evaluation in this study, as we were limited by sample size. Future verification work should be conducted; conventionally, such verification work is performed by using independent sample sets or acknowledged high-accuracy results such as airborne data, especially unmanned aerial vehicle LiDAR data (Chen et al. 2018, Li et al. 2019).

**Discussion**

**Stand and remote sensing predictors**

Currently, in the already existing models for mortality and CWD distribution, the predictors are usually stand factors, which have been confirmed to have great potential in previous studies (Spletich et al. 1999; Zhen et al. 2013, Bassett et al. 2015). In other forest models, multispectral characteristics, texture characteristics and vegetation index are used as predictive variables, and the near infrared, which is more sensitive than other wavelengths, is used as the band of vegetation to compute the normalized difference vegetation index (NDVI) and the enhanced vegetation index (DEVI; Li et al. 2017). These remote sensing factors have pioneering significance in machine learning and local regression estimation. However, because there is only one visible light band and no near-infrared band in many data, scholars have studied visible light vegetation indexes, such as the VSDEI, in place of the NDVI. The present study represents the first time in which the visible vegetation index and terrain and stand factors have been used to establish a DDWV model. The results show that visible remote sensing factors can be used to extract important information for DDWV estimation.

The model coefficients of traditional parametric models include a substantial amount of information (Tab. 2). The regression coefficients of the five selected variables are significant at an α level of 0.05. Among them, the model coefficients of SLOPE, AGE and VDVI are positive, indicating positive correlations with the dependent variable (DDWV). Thus, for example, when holding the other predictor variable constant, the steeper the slope, the higher the DDWV. The coefficients for TSFS and VEG are negative, indicating negative correlations with the dependent variable (DDWV). This result indicates that DDWV in the study area is more strongly associated with larger trees, commonly older trees, with low TSFS (Zhen et al. 2015). Because the GWDNN model was constructed in the form of an OLS model, statistics were calculated for the coefficients of the GWDNN model. In the GWDNN model, the coefficients of VEG were negative values, while the coefficients of the other variables exhibited positive and negative values. In other words, the variation in the GWDNN coefficients was greater than that of the GWR coefficients, which might explain the significant increases in fitting performance and prediction performance achieved with the GWDNN model.

**Model comparison**

The fitting accuracy and prediction accuracy were compared among the five models based on the R², RMSE, MAE, Moran’s I, Z-statistic, combined accuracy, and residual spatial autocorrelation results, and revealed that the GWDNN model is the best. In many studies, machine learning has been proven to be significantly superior to the GWR model and the LMM (Behrens et al. 2018, Chen et al. 2018). However, there are many differences between the LMM and GWR model. For example, Quirós-Segovia et al. (2016) discovered that the quality of the height-diameter model of GWR and LMM regressions were comparable. Wei et al. (2019) studied the spatial distribution of PM2.5 and found that the LMM was better than the GWR model; however, the GWR model has been shown to perform better than the LMM in many other studies (Liu et al. 2014, Zhang & Gove 2005). For the GWDNN model, a multi-layer DNN was constructed to consider the influence of the spatial weights, the GWR model residual space. However, by changing the spatial weights, the GWDNN model achieves accurate deconstruction and efficient calculation of spatial nonstationarity. The GWDNN model achieves excellent performance. In terms of predicting the spatial distribution of DDWV, the GWDNN model results are the closest to the ground-truthed distribution. In addition to offering statistical graphics and GIS mapping capabilities, the GWDNN model can intuitively provide key information on the distribution of DDWV, which is of great significance for forest decision-making and management planning.

**Conclusion**

DDWV was predicted by using parametric and nonparametric models of stand and remote sensing factors. The nonparametric models were found to be far superior to the parametric models in terms of fitting and verification accuracy, and the Moran I and Z-statistic values of the DNN model were lower than those of the LMM and the GWR model in terms of the stability of the model residual space. However, by changing the network structure to consider the spatial weights, the GWDNN model showed greater advantages in fitting and verification, and produced ideal residuals by verifying the spatial autocorrelation. Through the application of GIS technology in the study area, clear spatial distribution information was obtained; thus, this method enables very good assessment of the damage caused by natural disasters and can provide key information on forest resources for decision-making and management plans and be used to prevent and reduce the disturbance due to natural disas-
Chen L, Ren C, Zhang B, Wang Z, Xi Y (2018). Esti-
mation of forest above-ground biomass by ge-
ographically weighted regression and machine
learning with Sentinel Imagery. Forests 9 (10):
582. - doi: 10.3390/f9100582
Choi S, Lee S, Kang Y, Choi DY, Choi J (2020). Use
of unmanned aerial vehicle imagery and deep
learning to classify upland crop in small scale
agricultural land. Journal of the Korean
Society of Surveying, Geodesy, Photo-
grammetry and Cartography 38: 671-679. [on-
line] URL: http://www.koreascience.or.kr/artic-
el/JAK02020020165721397.page?lang=en
Chopincky DC, Heath LS (2002). Estimating down
deadwood from FIA forest inventory variables in
Maine. Environmental Pollution 116 (7): 525-
530. - doi: 10.1016/S0269-7491(02)00243-3
Costa S, Ibáñez L (2005). Can woodland storage be
profitable? French experience after the wind-
storms in 1999. Journal of Forest Economics 11:
161-176. - doi: 10.1016/j.jfe.2005.08.001
Deb D, Singh JP, Deb S, Datta D, Ghosh A, 
Chaurasia RS (2017). An alternative approach
for estimating above ground biomass using Re-
sourcesat-2 satellite data and artificial neural
network in Bundelkhand region of India. Envi-
ronmental Monitoring and Assessment 189 (11):
631. - doi: 10.1007/s10661-017-6307-6
Du M, Noguchi N (2016). Multi-temporal moni-
toring of wheat growth through correlation
analysis of satellite images, unmanned aerial
vehicle images with ground variable. IFAC-Pa-
PapersOnline 49: 5. - doi: 10.1016/j.ifacol.2016.
10.002
Du Z, Wang Z, Wu S, Zhang F, Liu R (2020). Geo-
graphically weighted network regression for the
accurate estimation of spatial non-
stationarity. International Journal of Geograph-
ical Information Science 34: 1533-1577. -
doi: 10.1080/09629419.2019.1707834
Fassnacht FE, Hartig F, Latifi H, Berger C, Her-
nandez J, Corvalan P, Koch B (2014). Impor-
tance of sample size, data type and prediction
method for remote-sensing-based estimations of
aboveground forest biomass. Remote Sens-
ing of Environment 154: 102-114. - doi: 10.1016/
J.rse.2014.07.028
Fotheringham AS, Charlton ME, Brunsdon C
(1998). Geographically weighted regression: a
natural evolution of the expansion method for spatial
data analysis. Environment and Planning
A 30: 1905-1927. - doi: 10.1068/a301905
Green JL, Hastings A, Arzberger P, Ayala FJ, Cot-
tingham KL, Cuddington K, Davis F, Dunne JA,
Fortin MJ, Gerber L, Neubert M (2005). Com-
plexity in ecology and conservation: mathe-
matical, statistical, and computational challenges.
BioScience 55: 501-510. - doi: 10.1641/0006-3568
200505035 [doi:C1ECAMC2].c020
Hagan JM, Grove SL (1999). Coarse woody de-
bris: humans and nature competing for trees.
Journal of Forestry 97: 6-11. [online] URL:
http://academic.oup.com/jof/article/97/6/4613
946
Harmon ME, Franklin JF, Swanson FJ, Sollins P,
Gregory SV, Lattin JD, Anderson NH, Cline SP,
Aumen NG, Sedell JR, Lienkaemper GW, Cro-
mack J Jr, Cummins KW (1986). Ecology of
coarse woody debris in temperate ecosystems.
Advances in Ecological Research 15: 133-302.
-doi: 10.1016/0065-2504(86)90011-X
Irfan M, KoI A, Thomas H, Sedighi M (2016).
Geographical General Regression Neural Network
(GGRRNN) tool for geographically weighted re-
gression analysis. In: Proceedings of the 8th In-
nernational Conference on “Advanced Geo-
graphic Information Systems, Applications, and
Services”. Venice (Italy), 24-28 April 2016. GEO-
Processing 2016, pp. 154-159. [online] URL:
http://orca.cardiff.ac.uk/91693
Jenkins MA, Webster CR, Parker GR, Spetch MA
(2004). Coarse woody debris in managed cen-
tral hardwood forests of Indiana, USA. Forest
Science 50: 781-789. [online] URL: http://acadm
eda.forestscience.oregonstate.edu/jfs/article/50/4/7
17267
Ketkar N (2017). Introduction to Keras. In: “Deep
Learning with Python: A Hands-on Introduc-
tion” (N Ketkar ed). Apress, Berkeley, CA, USA,
pp. 15-30. - doi: 10.978-1-4842-2766-4_7
Kinnucan HW (2016). Timber price dynamics af-
after a natural disaster: hurricane Hugo revisited.
Journal of Forest Economics 25: 115-129. -
doi: 10.1016/j.jfe.2016.09.002
Klobučar D (2010). Using geostatistics in forest
management. Sumarski Listi 134: 249-259
Lavoie S, Ruel-JC, Bergeron Y, Harvey BD (2012).
Windthrow after group and dispersed tree re-
tention in eastern Canada. Forest Ecology and
Management 269: 158-167. - doi: 10.1016/j.
eco.2011.12.018
Lecun Y, Bengio Y, Hinton G (2015). Deep learn-
ing. Nature 521: 436-444. - doi: 10.1038/nature
14534
Li H, Calder CA, Cressie N (2007). Beyond Mor-
an's I: testing for spatial dependence based on
the spatial autoregressive model. Geographical
Analysis 39: 357-375. - doi: 10.1111/0016-4367.
2007.00708.x
Li J, Jin M, Li H (2019). Exploring spatial influence
of remotely sensed PM2.5 concentration using a
developed deep convolutional neural net-
work model. International Journal of Environ-
mental Research and Public Health 16: 454. -
doi: 10.3390/ijerph16120454
Liu C, Zhang L, Li J, Lin X (2014). Spatial modeling
of the carbon stock of forest trees in Hei-
 longjiang Province, China. Journal of Forestry
Research 25: 269-280. - doi: 10.1007/s11676-014-
0458-x
Li J, Zhang L (2011). Modeling and prediction of
tree height-diameter relationships using spatial
autoregressive models. Forest Science 57: 252-
264. - doi: 10.1093/forests/57.3.252
Mennis J (2006). Mapping the results of geo-
graphically weighted regression: A developed
deep convolutional neural net-
work model. International Journal of Environ-
mental Research and Public Health 16: 454. -
doi: 10.3390/ijerph16120454
Miles J (2005). Tolerance and variance inflation
factor. Encyclopedia of Statistics in Behavioral
Science, North Yorkshire, UK, pp. 12. - doi:
10.1002/0470010703.sbea110025
Mohammadian A, Saedian B, Pradhan B, Gha-
eni Z (2019). Prediction mapping of human lep-
hemia. OUP emic.oup.com/forestscience/article/50/6/781/464
87-Miles J (2005). Tolerance and variance inflation
factor. Encyclopedia of Statistics in Behavioral
Science, North Yorkshire, UK, pp. 12. - doi:
10.1002/0470010703.sbea110025
Mohammadian A, Saedian B, Pradhan B, Gha-
eni Z (2019). Prediction mapping of human lep-
hemia. OUP emic.oup.com/forestscience/article/50/6/781/464
