Supplementary Material

S1: Theil-Sen median analysis and Mann-Kendalltrend test (Sen-MK)

Theil-Sen median analysis is a robust non-parametric trend estimation method for estimating the median slope of a set of sample points. Especially when the data contains outliers, it is more accurate than traditional least squares regression methods and is highly suitable for trend analysis of long-term time series data.

$$\beta = Median \ (\frac{y_j - y_i}{x_j - x_i})$$
(5)

Among these, β denotes the slope between points (x_j, y_j) and (x_i, y_i) in the time series, $Median$ as well as the median value.

The Mann-Kendall *(M-K)* test is a non-parametric time series trend test method that does not require measurement values to follow a normal distribution and is unaffected by missing or outliers. It is appropriate for trend significance assessment on lengthy time series data.

$$S = \sum_{i-1}^{n-1} \sum_{j-i+}^{n} sgn(x_j - x_i)$$
(6)

$$sgn(x_j - x_i) = \begin{cases} +1, x_j - x_i > 0 \\ 0, x_j - x_i = 0 \\ -1, x_j - x_i < 0 \end{cases}$$
(7)

$$Z = \begin{cases} \dfrac{S}{\sqrt{\dfrac{n(n-1)(2n+5)}{18}}}, S > 0 \\ 0, S = 0 \\ \dfrac{(S+1)}{\sqrt{\dfrac{n(n-1)(2n+5)}{18}}}, S < 0 \end{cases}$$
(8)

Among them, x_j、 x_i are the corresponding research years, *N* is the number of samples, *S* represents an upward trend in regular representation, while negative values indicate a downward trend. In this test, a bilateral trend test is used to find the critical value "$Z_1 - \frac{a}{2}$" in the normal distribution table at a given level of significance. When $|Z| \leq Z_1 - \frac{a}{2}$, accept the null hypothesis that the trend is not significant; If $/Z/> Z_1 - \frac{a}{2}$, reject the null hypothesis and assume that the trend is significant. This article takes "significance level α=0.05 , The condition of $Z = \pm 1.96$ "is given as the criterion for significant judgment. When the absolute

values of Z are greater than 1.65, 1.96, and 2.58, it indicates that the trend has passed the significance tests with reliability of 90%, 95%, and 99%, respectively.

S2: Random Forest Model (RF) and Result Interpretation Methods

- **Model Introduction:**

In 2001, Leo Breiman developed the Random Forest *(RF)* model, an advanced methodological paradigm grounded in the Classification and Regression Tree *(CART)* framework, capable of performing both classification and regression tasks. This model adeptly integrates the Bagging algorithm within the *CART* decision tree structure. Central to its methodology is bootstrap sampling, entailing repeated random selections with replacement, to construct a robust ensemble of decision trees. The efficacy of the RF model is achieved by either averaging the predictions from these diverse trees or applying a majority voting mechanism. The computational formula for this model is as follows:

$$H(x) = argmax_y \sum_{i=1}^{k} I(h_i(X) = Y) \tag{9}$$

In this context, $H(x)$ represents the final classification outcome of the model. *I* stands for the indicator function. Each individual decision tree is represented by h_i , while *Y* denotes the target variable under consideration.

Relative to other machine learning models, the RF minimizes the impact of multicollinearity in data and demonstrates robust performance with high-dimensional and nonlinear problems. It also allows for the ranking of predictor importance. The generalizability and robustness of RF significantly reduce the risk of overfitting when processing large datasets.

● **Model evaluation**

The RF model is adept for both regression and classification tasks. In classification scenarios, this study employs a combination of the confusion matrix, Receiver Operating Characteristic *(ROC)* curve, and Area Under the Curve *(AUC)* to assess model performance.

①Confusion Matrix:

This matrix comprises accuracy, precision, recall, and the *F1* score.

$$Recall = \frac{TP}{TP + FN} \tag{10}$$

$$Precision = \frac{TP}{TP + FP} \tag{11}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{12}$$

$$Accurarcy = \frac{TP + TN}{TP + TN + FP + FN} \tag{13}$$

In the formula, True Positives *(TP)* and True Negatives *(TN)* represent the correctly identified positive and negative classes, respectively, while False Positives *(FP)* and False Negatives *(FN)* denote the incorrectly identified positive and negative classes. The *F1* score is the harmonic mean of precision and recall, ranging between 0 and 1. A value closer to 1 indicates superior performance of the classifier.

②The *ROC* curve:

This is a vital tool for assessing a model's predictive accuracy at varying classification thresholds. It is constructed with the True Positive Rate *(TPR)* on the vertical axis and the False Positive Rate *(FPR)* on the horizontal axis. The *AUC* of the *ROC*, frequently used in evaluating binary classifiers, quantitatively reflects the classifier's effectiveness. A model is considered more proficient if its *ROC* curve is nearer to the upper left corner. The calculation of the *ROC* curve is as follows:

$$TPR = \frac{TP}{TP + FN} \tag{14}$$

$$FPR = \frac{FP}{TN + FP} \tag{15}$$

$$AUC = \frac{\sum_{I \epsilon Positive\ Class} rank_i - M(M + 1)}{2 \times M \times N} \tag{16}$$

M and *N* denote the counts of positive and negative samples, respectively, and $rank_i$ is the ranked probability score for the ith positive sample.

To evaluate the performance of the RF regression model, four commonly used accuracy metrics were introduced: Mean Squared Error *(MSE)*, Root Mean Squared Error *(RMSE)*, Mean Absolute Error *(MAE)*, and the coefficient of determination *(R2)*. The relevant formulas for these metrics are as follows:

$$MSE = \frac{1}{m} \sum_{i=1}^{m} (y_i - \hat{y}_i)^2 \tag{17}$$

$$RMSE = \sqrt{MSE} \tag{18}$$

$$MAE = \frac{1}{m}\sum_{i=1}^{m}|y_i - \hat{y}_i| \tag{19}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{m}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{m}(\bar{y}_i - y_i)^2} \tag{20}$$

In these calculations, m represents the total number of samples (dependent variables), denotes the actual value of each sample, $\hat{y}_i$ is the predicted value, and $\bar{y}_i$ is the average of the actual values of the samples. The *MSE* measures the expected value of the squared differences between actual and predicted values, assessing the variability of the data. The *RMSE* is the square root of *MSE*, indicating the degree of deviation in the predictions. The *MAE* is the average of the absolute errors. Lower values for *MSE*, *RMSE*, and *MAE* indicate higher model accuracy. R^2 represents the goodness of fit of the model; values closer to 1 signify greater predictive accuracy.

- **Model interpretation**

  The importance of features reflects the contribution of each variable to model prediction. It is measured by the percentage increase in the mean square error *(IncMSE%)* of the Out of Bag Data *(OOB)*. Since *IncMSE%* represents the degree to which the accuracy of model predictions decreases after removing variables, the more important the variables are, the higher the *IncMSE%*. Assuming there is a tree in the forest, the *IncMSE%* of *K* tree can be expressed as:

$$IncMSE\ \%\ (i) = \sum_{K=1}^{N}\frac{(OOB_{k2} - OOB_{k1})}{OOB_{k1}} \times 100\ \% \tag{21}$$

IncMSE%(i) represents the importance of feature variable *i*. OOB_{k1} and OOB_{k2} represent the *OOB* errors before and after interference, respectively. For example, if the difference in *IncMSE%* is small after shuffling the *OOB* data, it can be considered that variable *i* is not important. In other words, the lower the *IncMSE%*, the less important the feature variable *i* is in the prediction results.

The Partial Dependency Diagram (PDP) illustrates how specific attributes affect model prediction outcomes at various values. By fixing other non-specified input variables to their marginal distributions, it helps ascertain whether the relationship between features and outcomes is linear, monotonic, or more complex.

The following is the methodology used to calculate PDP:

$$f_{xs}(x_s) = E_{XC}[f(x_s, x_c)] = \int f(x_s, x_c)dp(x_c) \tag{22}$$

In the above formula, f_{xc} is the partial function of the average marginal effect of variable x_s on the predicted variable. x_s is a collection of target features. x_c represents the complementary set of features used in machine learning model. The expected value of $f(x)$ is $E[f(x)]$. While x_s is fixed, x_c changes according to its marginal distribution $dp(x_c)$.

In practice, part of function $f_{xs}(x_s)$ is the average of all training data.

$$f_{xs}(x_s) = \frac{1}{n}\sum_{i=1}^{n} f(x_s, x_c^i) \tag{22}$$

x_c^i is the feature in the ith sample, excluding x_s; N is the number of observation samples in the dataset.

S3: *Multivariate geographic regression model (MGWR)*

Scale is a critical dimension in understanding and describing urban phenomena. The evolution of each phenomenon is often determined by spatial processes operating at multiple scales. As Goodchild aptly noted, "Scale is among the most significant issues in the study of geographic matters. Studies have shown that implementing policies based solely on a singular, average, globally applicable perspective may lead to misleading outcomes. In the pursuit of high-quality development, it is imperative to explore the future of cities from a nuanced, multi-scaled perspective and to formulate corresponding policies.

In spatially related analyses, the feature values of observational entities are often derived from their geographic locations, which serve as sampling units. The inherent non-stationarity of spatial data imparts unique characteristics to it. Prior to 1996, studies of geographic spatial relationships predominantly employed the Ordinary Least Squares (OLS) method. However, as a global regression model, OLS is based on the assumption of spatial homogeneity and tends to overlook the nuanced characteristics of actual spatial contexts. The advent of Geographically Weighted Regression (GWR), introduced by Fotheringham, marked a significant shift. GWR transcends the limitations of OLS by utilizing locally weighted least squares for individual parameter estimation. This method adeptly captures the diversity and variability of variables in geographic space by incorporating locally varying parameters. The formulation of GWR is as follows:

$$y_i = \sum_r \beta_r(u_i, v_i)x_{ir} + \varepsilon_i \tag{23}$$

In this model, (u_i, v_i) represents the centroid coordinates of the ith grid in the study area. y_i denotes the observed values (dependent variables) of the objects of interest within the ith grid. x_{ir} refers to the rth influencing factor (independent variable) within the ith grid that affects the characteristics of the observed objects. β_r signifies the regression coefficient of the rth independent variable in the ith grid, and ε_i accounts for the random error term.

Although GWR considers the influence of spatial scales on variables, the spatial non-stationarity of inter-element relationships often manifests across different scales. A

uniform, global bandwidth in GWR can lead to increased regression bias. To address these scale effects, Fotheringham et al. innovatively developed the Multiscale Geographically Weighted Regression (MGWR) model in 2017, building upon the foundations of GWR and generalized additive models. This model allows for variable-specific spatial smoothing levels, enabling each variable to undergo regression with its distinct bandwidth. This approach significantly reduces noise in parameter estimation and enhances the accuracy of spatial process simulations. The smaller the bandwidth assigned to a variable, the greater its spatial heterogeneity. Following enhancements by Yu et al. in 2020, MGWR has been extensively utilized in a variety of empirical research. The precise computational formula is delineated as follows:

$$y_i = \beta_0(u_i, v_i) + \sum_{j=1}^{k} \beta_{bwj}(u_i, v_i)x_{ij} + \varepsilon_i \tag{24}$$

In the formula, bwj represents the bandwidth used for the regression coefficient of the jth variable; β_{bwj} denotes the regression coefficient of this independent variable within the ith grid at this bandwidth. y_i signifies the feature value of the observational entity in the ith grid; (u_i, v_i) represents the centroid coordinates of the ith grid; x_{ij} signifies the factor within the ith grid that influences the characteristic values of observation objects associated with the jth variable; k denotes the number of observation samples; ε_i represents the observational error value and $\beta_0(u_i, v_i)$ represents the intercept term; The Akaike Information Criterion (AIC) or the Akaike Information Criterion Corrected (AICc), along with the R^2 value, are commonly employed standards for assessing the goodness of fit of regression models. Lower AIC and AICc values, coupled with a higher R^2, indicate a better fit of the model.