

## Supplementary material

The Supplementary material is the recommended parameters for the models used in this paper:

### OLS

Default setting of the *LinearRegression* function of the *python sklearn* package.

### GWR

Function for calculating the geographical weights: the *Gaussian* function.

The criterion for choosing the bandwidth: the *Akaike* information criterion (AICc).

The range for searching the appropriate bandwidth: 0.1 m - 50,000 m.

The average bandwidth selected in this study: 721.8 m.

### FCNN

The number of hidden layers: **4**.

Hidden layer 1:

The number of neurons: **200**.

The activation function: **ReLU**.

Hidden layer 2:

The number of neurons: **120**.

The activation function: **ReLU**.

Hidden layer 3:

The number of neurons: **100**.

The activation function: **ReLU**.

Hidden layer 4:

The number of neurons: **20**.

The activation function: **none**.

Regularization type: **the L2 regularization**.

Regularization weight: **0.00005**.

Initial learning rate: **0.5**.

Learning rate decay: **0.99996**.

Moving average decay: **0.99996**.

Batch size: **32**.

Principal loss function:  $\sum(Y - Y^*)^2$

## CNN

The number of convolutional layers: **2**.

Convolutional layer 1:

Convolution kernel size: **3**.

Convolution stride: **1**.

Depth: **8**.

Initializer of the weights: **the truncated normal initializer** with the **mean** value of **0.0** and the **standard deviation** value of **0.05**.

Initializer of the bias: **the constant initializer** with the value of **0.01**.

The activation function: **ReLU**.

Padding: **SAME**.

Convolutional layer 2:

Convolution kernel size: **3**.

Convolution stride: **1**.

Depth: **16**.

Initializer of the weights: **the truncated normal initializer** with the **mean** value of **0.0** and the **standard deviation** value of **0.05**.

Initializer of the bias: **the constant initializer** with the value of **0.01**.

The activation function: **ReLU**.

Padding: **SAME**.

The number of pooling layers: **0**.

The algorithm of the attention block: **the Softmax attention**.

The number of fully-connected layers: **2**.

Fully-connected layer 1:

The number of neurons: **128**.

The activation function: **ReLU**.

Dropout probability: **0.5**.

Fully-connected layer 2:

The number of neurons: **64**.

The activation function: **none**.

Dropout probability: **0**.

Regularization type of the fully-connected layers: **the L2 regularization**.

Regularization weight: **0.0001**.

Initial learning rate: **0.3**.

Learning rate decay: **0.9999**.

Moving average decay: **0.9999**.

Batch size: **32**.

Principal loss function:  $\sum(Y - Y^*)^2$