



# 1 A Spatiotemporal Weighted Regression Model (STWR v1.0) for 2 Analyzing Local Non-stationarity in Space and Time

3 Xiang Que<sup>1,2</sup>, Xiaogang Ma<sup>2,\*</sup>, Chao Ma<sup>2,\*</sup>, Qiyu Chen<sup>3</sup>

4 <sup>1</sup> Computer and Information College, Fujian Agriculture and Forestry University, Fuzhou, Fujian, China

5 <sup>2</sup> Department of Computer Science, University of Idaho, 875 Perimeter Drive MS 1010, Moscow, ID 83844-1010, USA

6 <sup>3</sup> School of Computer Science, China University of Geosciences (Wuhan), 388 Lumo Road, Wuhan 430074, China

7 \*Correspondence to: Xiaogang Ma (max@uidaho.edu); Chao Ma (chao@uidaho.edu)

8

9 **Abstract.** Local spatiotemporal non-stationarity occurs in various natural and socioeconomic processes. Many studies have  
10 attempted to introduce time as a new dimension into the geographically weighted regression model (GWR), but the actual  
11 results are sometimes not satisfied or even worse than the original GWR model. The core issue here is a mechanism for  
12 weighting effects of both temporal variation and spatial variation. In many geographical and temporal weighted regression  
13 models (GTWR), the concept of time distance has been inappropriately treated as time interval. Consequently, the combined  
14 effect of temporal and spatial variation is often inaccurate in the resulting spatiotemporal kernel function. This limitation  
15 restricts the configuration and performance of spatiotemporal weights in many existing GTWR models. To address this  
16 issue, we propose a new spatiotemporal weighted regression (STWR) model and the calibration method for it. A highlight of  
17 STWR is a new temporal kernel function, in which the method for temporal weighting is based on the degree of impact from  
18 each observed point to a regression point. The degree of impact, in turn, is based on the rate of value variation of the nearby  
19 observed point during the time interval. The updated spatiotemporal kernel function is based on a weighted combination of  
20 the temporal kernel with a commonly used spatial kernel (Gaussian or bi-square) by specifying a linear function of spatial  
21 bandwidth versus time. Three simulated datasets of spatiotemporal processes were used to test the performance of GWR,  
22 GTWR and STWR. Results show that STWR significantly improves the quality of fit and accuracy. Similar results were  
23 obtained by using real-world data for the precipitation hydrogen isotopes ( $\delta^2\text{H}$ ) in Northeastern United States. The Leave-  
24 one-out cross-validation (LOOCV) test demonstrates that, comparing with GWR, the total prediction error of STWR is  
25 reduced by using recent observed points. Prediction surfaces of models in this case study show that STWR is more localized



26 than GWR. Our research validates the ability of STWR to take full advantage of all the value variation of past observed  
27 points. We hope STWR can bring fresh ideas and new capabilities for analyzing and interpreting local spatiotemporal non-  
28 stationarity in many disciplines.

29

30 **Key words:** Geographical and temporal weighted regression; Geographically weighted regression; Temporal non-  
31 stationarity; Spatial analysis; Spatiotemporal variations; Spatiotemporal weighted regression.

32

### 33 **1 Introduction**

34 Time, space and attributes are three essential characteristics in geographic entities, and they are recorded to reflect the state  
35 and evolution of various real-world phenomena and processes. Because space and time frame all aspects of the discipline of  
36 geography (Goodchild, 2013), it is important to observe the spatiotemporal variations and explore appropriate analytical  
37 methods to study and reason the internal mechanisms and evolutionary laws. In recent years, new platforms and instruments  
38 have brought increasingly massive spatiotemporal data, such as the time- and geo-tagged sensor monitoring records and  
39 remote sensing images. Those big data create great opportunities for studying human and environmental dynamics from  
40 different perspectives, such as the patterns of human behavior (Chen et al., 2011), environmental risk assessment (Sun et al.,  
41 2015), and disease outbreaks (Takahashi et al., 2008). Nevertheless, although spatiotemporal modeling has been a long-term  
42 research focus in the field of geographical information science (GIScience) (Cressie, 1991; Cressie and Wikle, 2015), the  
43 models are not mature yet and challenges still exist (Fotheringham et al., 2015), which call for further work.

44 In this paper, the technological development and discussion focus on modeling local spatiotemporal variations within  
45 the framework of geographically weighted regression (GWR). GWR is a method for modeling spatially heterogeneous  
46 processes (Brunsdon et al., 1996, 1998; Fotheringham et al., 2003). It has been applied in a variety of areas, such as climate  
47 science (Brown et al., 2012), geology (Atkinson et al., 2003), mineral exploration (Wang et al., 2015), transportation analysis  
48 (Cardozo et al., 2012), crime studies (Cahill and Mulligan, 2007; Wheeler and Waller, 2009), environmental science (Mennis  
49 and Jordan, 2005), and house price modeling (Fotheringham et al., 2015). GWR calibrates a separate regression model at  
50 each location through a data-borrowing scheme, in which distance-weights can be calculated by drawing on data from



51 neighboring observations of each regression point (Fraser et al., 2012). This operation complies with Tobler’s first law of  
52 geography - “Everything is related to everything else, but near things are more related than distant things” (Tobler, 1970).

53 Numerous studies have been devoted to incorporating the temporal dimension into spatial regression (Pace et al., 2000;  
54 Gelfand et al., 2004; Crespo et al., 2007; Cressie and Wikle, 2015). However, most of these studies assume that temporal  
55 effects are constant over space from a global perspective of modeling (Fotheringham et al., 2015). To address that issue,  
56 Crespo et al. (2007) extended GWR by developing spatiotemporal bandwidths that account for varying local spatial effects  
57 across time. Huang and Wu (2010, 2014) proposed a geographical and temporal weighted regression model (GTWR) with a  
58 method of measuring the spatiotemporal ‘closeness’ and a parameter ratio  $\tau$  to deal with different measured units in time  
59 and space. Although the approach can address the issue to some extent, Fotheringham et al. (2015) pointed out that a sole  
60 measurement of integrated spatial and temporal distances can be misleading as location and time are usually measured at  
61 different scales, and he stated that the calculation of distance in three dimensions (time and two-dimensional space) remains  
62 a challenge.

63 A spatiotemporal kernel function, which consists of mixed spatial and time-decay bandwidths, was proposed by  
64 Fotheringham et al. (2015). Nevertheless, the stepwise strategy applied in this function for bandwidth optimization does not  
65 always seem reasonable. In practice, this function needs to first find and fix an optimized spatial bandwidth, then it will find  
66 the optimized temporal bandwidth. After that, the spatiotemporal weight will be calculated. This stepwise search process  
67 means that the function is not able to optimize both temporal and spatial bandwidths at the same time. However, a more  
68 reasonable thought is that the spatiotemporal bandwidth and its weight are simultaneously affected by both spatial and  
69 temporal effects of a process. There should be ways to further improve the spatiotemporal kernel function in Fotheringham  
70 et al. (2015).

71 The aim of this paper to develop a better methodology for the spatiotemporal kernel function. Following Tobler’s first  
72 law, we propose an algorithm, the spatiotemporal weighted regression (STWR). In STWR, the velocity of value change is  
73 higher related if they were in near time and space. Therefore, STWR can borrow data not only from nearby locations, but  
74 also from nearby value variation through time. The latter is what we call as “time distance” in STWR. The time distance is  
75 not the concept of time interval, but the rate of value variation through time. It is a kind of value change that reflects the



76 temporal effect of nearby points to the regression point. Accordingly, our local spatiotemporal regression analysis model can  
77 take advantage of the variation in data to identify temporal non-stationarity, which is an advantage comparing with GWR  
78 and GTWR.

79 Before giving more details about STWR, we can further clarify the meaning of a few concepts. A common issue in the  
80 existing GTWR models is that they use the concept of time interval, instead of the above-mentioned “time distance”, to  
81 calculate temporal and spatiotemporal weights. A time interval is the period between two observed time stages. A time  
82 distance, in the context of STWR, is the rate of value variation between an observed point and a regression point through a  
83 time interval. We can think about the following scenario for a group of points. The values of some points do not change or  
84 change slightly from time A to time B, while a few other points may change greatly in that period. However, many GTWR  
85 models ignore the difference in the value changes of observed points during a period of time, and regard that all these points  
86 have the same temporal effect to their neighbor regression point. It is hard to believe that some unchanged observations  
87 constantly affect their nearby regression points during the observed time interval. Intuitively, different variations of the  
88 observed points have different temporal effects. For example, the faster the house price of a point change, the stronger the  
89 temporal effect is to the house price at its nearby point. Moreover, the rate of value changes at different observed points  
90 (time non-stationary) may also have spatial heterogeneity. The data values observed at different points are results of mixed  
91 spatiotemporal effects and some other unknown factors (including errors). Therefore, using only time interval in the  
92 calculation of temporal and spatiotemporal weights might interpret local spatiotemporal effect imprecisely.

93 There are other issues in the temporal kernel functions and the multiplication form of spatial and temporal kernels used  
94 by the existing GTWR models (Huang et al., 2010; Wu et al., 2014; Fotheringham et al., 2015). When calculating the  
95 spatiotemporal effect, these models generally use time intervals and the common kernel functions to calculate temporal  
96 weights, such as Gaussian kernel or bi-square kernel. However, an appropriate temporal kernel function should not be the  
97 same as the spatial kernel function, because space is in two or three dimensions while time is in one dimension and one  
98 direction. Each regression point can borrow observed points from any directions in space but only use points from the past  
99 rather than from the future. Moreover, these models directly calculate the integrated spatiotemporal weights by using a  
00 multiplication of the spatial and temporal weights. For example, if the temporal effect weight 0.1, and the spatial effect



01 weight is 0.9, then those models will generate a spatiotemporal weight of 0.09. Such a simple multiplication may cause  
02 underestimation of weights. For instance, to calculate the impact (weight) of the historical house price of B on the current  
03 house price of A, there can be many possibilities. One is that if the overall house price changes quickly, then the historical  
04 price of B may have little effect on the price of A, and the weight will be small. Another possibility is that the house prices  
05 of locations around A have not changed much during a long period, then the historical price of B may still have a relatively  
06 big impact on the current price of A. In this case, the weight will be seriously underestimated if the multiplication form of  
07 space and time weights is used.

08 The above-mentioned limitations and issues in GWR and GTWR are the driving force behind our development of  
09 STWR. The remainder of this article is organized as follows. Section 2 introduces the STWR model formulation, including  
10 temporal kernel and spatiotemporal kernel functions. Section 3 describes the methods for bandwidth selection and calibration  
11 when STWR is in operation. Section 4 presents results of applying GWR, GTWR and STWR to three sets of simulated data.  
12 Section 5 presents experiment results with real-world precipitation hydrogen isotope data. In Section 6, we close the article  
13 with a summary of the key findings and a few thoughts for future research.

14

## 15 **2 The Core Model of STWR**

### 16 **2.1 The strategy of time distance decay**

17 Since GWR is the background of our work, it is helpful to first give a brief overview of the GWR framework. The basic  
18 formulation of GWR can be described in two equations below (Fotheringham et al., 2003).

$$19 \quad y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i)x_{ik} + \varepsilon_i \quad (1)$$

$$20 \quad \hat{\beta}_k(u_i, v_i) = (X^T W(u_i, v_i) X)^{-1} X^T W(u_i, v_i) y \quad (2)$$

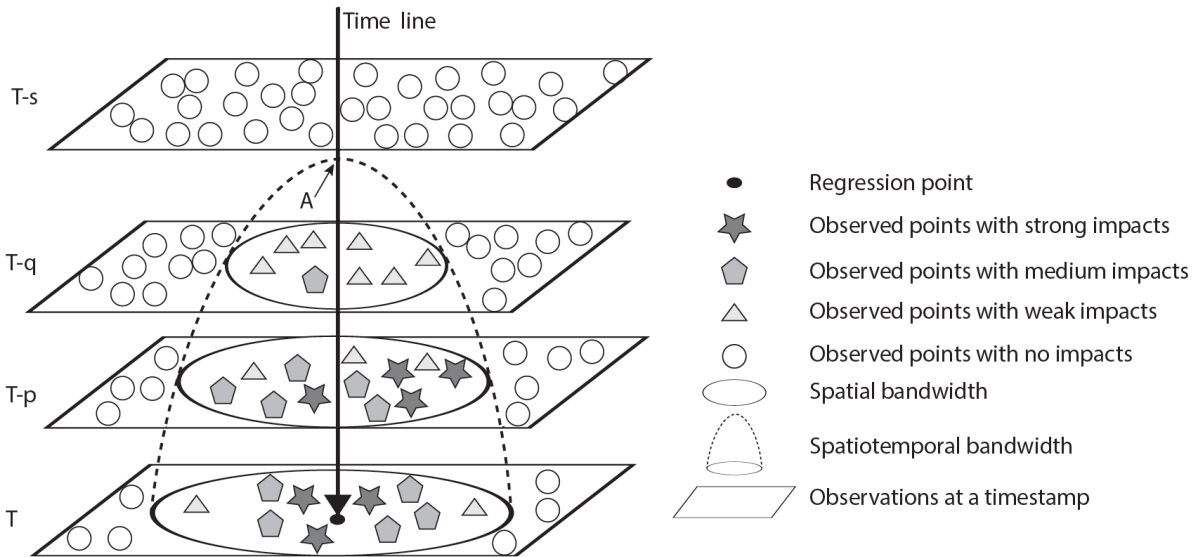
21 In Equation 1,  $y_i$  is a response variable of regression point  $i$  at a location with the coordinates  $(u_i, v_i)$ .  $x_{ik}$  is the  $k^{th}$   
22 dependent variable, and  $\varepsilon_i$  denotes the error term for the  $i^{th}$  observed point. A key difference between GWR and the  
23 traditional global regression method, such as Ordinary Least Squares (OLS), is that GWR allows the coefficient  $\beta_k(u_i, v_i)$   
24 vary spatially to identify spatial heterogeneity. Equation 2 represents the GWR calibration in a matrix form.  $W(u_i, v_i)$  is a  
25 diagonal weighting matrix specific to location  $i$ , which is calibrated by a specified kernel function with a given bandwidth.



26 Every element  $w_i$  in the weighting matrix reflects the impact from another observed point to the regression point. A bigger  
27  $w_i$  value means a higher impact.

28 GWR has a strategy of spatial distance decay impact on a regression point (Brunsdon et al., 1998; Fotheringham et al.,  
29 2003). A similar “time distance decay” strategy was also discussed in several recent GTWR models (Crespo et al., 2007;  
30 Huang et al., 2010; Wu et al., 2014; Fotheringham et al., 2015). Yet, those models did not fully reflect the effect of time  
31 distance decay. Sample points are observed at different time stages, and those data points closer in time distance to a  
32 regression point have more impact on the regression point than those farther away. The time distance refers to the value  
33 variation rate between an observed point and a regression point during a certain time interval. For example, in Fig. 1, there  
34 are four time stages from old to new: T-s, T-q, T-p and T. Through a fitting and calibration process, the spatiotemporal  
35 bandwidth will be fitted, and the spatiotemporal effects (weights) from observed points to a regression points at time stage T  
36 will be calculated by a specific spatiotemporal kernel function. Then, in prediction, the value of a regression point at time  
37 stage T can be estimated. Thus, the observed points at time stage T only have spatial effect on the regression point (Fig. 1).  
38 There is temporal effect from data points at time stages T-p and T-q (shown as stars, pentagons and triangles in the planes of  
39 T-p and T-q in Fig. 1), within a certain spatial bandwidth  $b_{ST}$  at each time stage, to the regression point. The time distance  
40 decay should reflect that different variations of the observed points have different temporal effects. However, as mentioned  
41 in the previous section, many existing GTWR models have applied a strategy of time interval decay instead of time distance  
42 decay. Consequently, they regard that all the observed points have the same temporal effect to their neighbor regression  
43 point.

44



45

46 **Fig. 1.** Spatiotemporal impacts of observed points with different rates of value change on a regression point at time stage T.

47 Temporal bandwidth is the length of time from the intersection point A of the spatiotemporal bandwidth and the time line to  
48 the regression point. Spatial bandwidth and spatiotemporal bandwidth are illustrated in the figure legend.

49

50 Compared to existing GTWR models, the time distance decay strategy of STWR considers the effect of different  
51 variations of observed points through time. For example, some data points may have higher impact on the regression point,  
52 though their spatial distance is farther than other points. Fig. 1 illustrates that the locations of some star-shape points are  
53 farther away from the regression point than some pentagon-shape points at time stage T-p, which denotes that there exist  
54 mixed impacts (spatial impact and temporal impact) on the regression point. The temporal impacts depend on the rate of  
55 value variation, which is the value difference between the observed point and the regression point divided by a time interval  
56 (e.g., [T-p, T] and [T-q, T-p] each is a time interval). If the observed time stage is too long ago or the rate of value variation  
57 is too small, and exceeds the limit of optimized temporal bandwidth for the regression point (as shown by observations at  
58 time stage T-s), the data points at this time stage may have no impact on the regression point. Even though some of those  
59 data points may have huge difference in value and are close to the regression point in space, they are not within the range of



60 the optimized temporal bandwidth. Spatial bandwidths also vary along the time line, and usually the bandwidth gets larger  
 61 when the observation time is closer to the time stage of the regression point (Fig. 1).

## 62 2.2 The spatiotemporal kernel function of STWR

63 We assume that a set of observed points  $O_{\Delta t} = \{O_{N_t}, O_{N_{t-1}}, \dots, O_{N_{t-q}} | \Delta t = [t - q, t]\}$  are collected during a certain time  
 64 interval  $\Delta t$  in a study area, where  $t$  represents the current time stage and  $N_{t-i}, i \in \{0, 1, 2, \dots, q\} (N_t = N_{t-0})$  denotes the  
 65 number of observed points at each recorded time. As the idea described above, we can borrow neighbor points in space and  
 66 their value variation during certain recent time intervals, so we can still use Equation 1 to generate local estimates. The  
 67 weight matrix  $W$  in GWR usually depends on the spatial kernel (Fotheringham et al., 2015). In STWR, we need to consider  
 68 the temporal effect, so the form of  $W$  is different from that in GWR. Correspondingly, we should have a spatiotemporal  
 69 kernel, which can be understood as a temporal extension based on the spatial kernel. However, if we use a multiplication  
 70 form to combine the temporal kernel and the spatial kernel (Huang et al., 2010; Wu et al., 2014; Fotheringham et al., 2015),  
 71 we will face the problem of time and space interaction as mentioned above in the Introduction section. To address that issue,  
 72 we design a weighted average form for the spatiotemporal kernel.

$$73 \quad w_{ijST}^t = (1 - \alpha)k_s(d_{sij}, b_{ST}) + \alpha k_T(d_{tij}, b_T), 0 \leq \alpha \leq 1 \quad (3)$$

74 In Equation 3,  $w_{ijST}^t$  is the weight at time  $t$  and at the observed location  $j$ .  $k_s$  and  $k_T$  are the spatial and temporal kernel,  
 75 respectively, and they both have a value range of 0 to 1.  $\alpha$  is an adjustable parameter to scale the temporal and spatial  
 76 effects, which can be optimized with the bandwidth selections.  $d_{sij}$  and  $d_{tij}$  are the spatial (Euclidean) and temporal  
 77 distance between the regression point  $i$  and an observed data point  $j$ , respectively.  $b_{ST}$  is the spatial bandwidth  $b_s$  at a  
 78 certain time stage  $T$ , and  $b_T$  denotes the temporal bandwidth.

79 The time distance, as mentioned above, is not the time interval but the rate of value variation between an observed point  
 80 and a regression point through a time interval. Following the time distance decay strategy in STWR, we can further derive  
 81 the temporal kernel  $k_T$  as shown below.

$$82 \quad w_{ij\Delta t}^t = \begin{cases} \left[ \frac{2}{1 + \exp\left(-\frac{|(y_i(t) - y_j(t-q))/y_j(t-q)|}{\Delta t/b_T}\right)} - 1 \right], & \text{if } 0 < \Delta t < b_T \\ 0, & \text{otherwise} \end{cases} \quad (4)$$





83 In Equation 4,  $y_{i(t)} - y_{j(t-q)}$  is the subtraction of the regression point  $i$ 's observed value at  $t$  from the point  $j$ 's observed  
 84 value at  $t - q$ , which denotes the value change during the time interval  $\Delta t$ . The internal part of the exponential function is  
 85 negative, in order to make the weight  $w_{ij\Delta t}^t$  range from 0 to 1. The faster the value change rate is, the bigger the weight is,  
 86 which means that the time impact is larger. When the time interval  $\Delta t$  is out of the range  $(0, b_T)$ , the weight will be set to  
 87 zero, which denotes that there is no impact because the observed variation is too far to affect the current moment. For  
 88 example, if the price of a nearby house has changed a long time ago, it may have little or no impact on the present house  
 89 price. But if the house price had a sharp change recently, it will have a big impact on the present house price. Therefore, the  
 90 faster the rate of observed value changes and the shorter the time interval is, the greater the impact on the regression point  
 91 will be. Compared with GTWR models, the advantage of STWR is that the temporal kernel function  $k_T$  can better leverage  
 92 the variation data.

93 To calibrate the weight value  $w_{ijST}^t$ , we need a spatial kernel function. The most widely used kernel functions are bi-  
 94 square and Gaussian (Fotheringham et al., 2003), which are given in Equations 5 and 6, respectively.

95 
$$\text{Bi-square: } w_{ijs} = \begin{cases} \left[1 - \left(\frac{d_{sij}}{b_s}\right)^2\right]^2, & \text{if } d_{sij} < b_s \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

96  
 97 
$$\text{Gaussian: } w_{ijs} = \exp\left[-\frac{1}{2}\left(\frac{d_{sij}}{b_s}\right)^2\right] \quad (6)$$

98 In Equations 5 and 6,  $b_s$  is the spatial bandwidth. Derived from  $b_S$  and  $b_{ST}$ ,  $b_{St}$  is the initial spatial bandwidth at the given  
 99 time stage  $t$  of the regression point (i.e.,  $t$  is the initial time for searching observed points in the past). Many functions can  
 .00 be specified for the change of spatial bandwidth during the time intervals. Because in most cases it will have smooth change  
 .01 during a certain short time interval, we assume that the spatial bandwidth changes linearly along with time, as defined  
 .02 bellow.

.03 
$$b_{ST} = b_{St} - \tan\theta * \Delta t, \quad -\frac{\pi}{2} < \theta < \frac{\pi}{2} \quad (7)$$

.04 In Equation 7,  $\tan\theta$  denotes the slope of spatial bandwidth change in correspondence to  $\Delta t$ , and  $b_{St}$  denotes the initial  
 .05 spatial bandwidth at  $t$ . Importing Equations 4 to 7, the calibration of Equation 3 can be further derived into Equations 8 and



9, which are our spatiotemporal kernel functions in STWR. Equations 8 and 9 are based on the bi-square and Gaussian kernel, respectively. With the STWR spatiotemporal kernel, we only need to optimize the parameters  $\alpha$  and  $\theta$  instead of the spatial bandwidth  $b_{ST}$ . However, we shall traverse all the observed points at the initial time stage  $t$  to find the optimized spatial bandwidth  $b_{ST}$ . Moreover, we shall also traverse all the time stages to find the optimized temporal bandwidth  $b_T$ .

$$w_{ijST}^t = \begin{cases} \left[ (1-\alpha) * \left[ 1 - \left( \frac{d_{sij}}{b_{St} - \tan \theta * \Delta t} \right)^2 \right] + \alpha * \left( 2 / \left( 1 + \exp\left( -\frac{|(y_{i(t)} - y_{j(t-q)}) / y_{j(t-q)}|}{\Delta t / b_T} \right) \right) - 1 \right) \right], \\ \text{if } \Delta t < b_T, \text{ and } d_{sij} < (b_{St} - \tan \theta * \Delta t) \\ 0, \text{ otherwise} \end{cases} \quad (8)$$

$$w_{ijST}^t = \begin{cases} \left[ (1-\alpha) * \exp\left[ -\frac{1}{2} \left( \frac{d_{sij}}{b_{St} - \tan \theta * \Delta t} \right)^2 \right] + \alpha * \left( 2 / \left( 1 + \exp\left( -\frac{|(y_{i(t)} - y_{j(t-q)}) / y_{j(t-q)}|}{\Delta t / b_T} \right) \right) - 1 \right) \right], \\ \text{if } \Delta t < b_T, \text{ and } d_{sij} < (b_{St} - \tan \theta * \Delta t) \\ 0, \text{ otherwise} \end{cases} \quad (9)$$

12

### 13 3 STWR in Operation

#### 14 3.1 Bandwidth selection and parameter estimation

15 Some goodness-of-fit diagnostics (Loader, 1999) are widely used in general GWR-based models, such as the cross-  
 16 validation (CV) score (Cleveland, 1979; Bowman, 1984) and the Akaike Information Criterion (AIC) (Akaike, 1973;  
 17 Akaike, 1998). For STWR, we use cross-validation (CV) as the default searching criteria and we also calculate the value of a  
 18 corrected version of AIC (Hurvich et al., 1998), the AICc, which is defined bellow.

$$19 \quad AIC_c = 2n \ln(\hat{\sigma}) + n \ln(2\pi) + n \left\{ \frac{n + tr(S)}{n - 2 - tr(S)} \right\} \quad (10)$$

20 In Equation 10,  $n$  is the sample size,  $\hat{\sigma}$  is the estimated standard deviation of the error term, and  $tr(S)$  denotes the trace of  
 21 the hat matrix  $S$  (Hoaglin and Welsh, 1978).



Although there is no need to optimize spatial bandwidth  $b_{ST}$  of the past time stages in STWR, other parameters such as  $\alpha$  and  $\theta$  need to be optimized. Also, we should give the  $b_T$  and initial  $b_{St}$  through trials. For more potential combinations of these parameters for different spatiotemporal processes, a more reasonable limit and optimization procedure is hence needed.

### 3.2 Calibration of STWR

Calibration of the STWR models can be conducted by using weighted least squares. The estimator for the coefficients at location  $(u_i, v_i)$  is shown below.

$$\hat{\beta}_t(u_i, v_i) = [(X_{O_{\Delta t}}^T W_{\Delta t}(u_i, v_i) X_{O_{\Delta t}})^{-1} X_{O_{\Delta t}} W_{\Delta t}(u_i, v_i)] y_{O_{\Delta t}} \quad (11)$$

In Equation 11,  $X_{O_{\Delta t}}$  and  $y_{O_{\Delta t}}$  are observed independent and dependent variables of  $O_{\Delta t}$  respectively.  $X_{O_{\Delta t}}^T$  is the transpose of  $X_{O_{\Delta t}}$ .  $W_{\Delta t}(u_i, v_i)$  denotes the spatiotemporal weight matrix for observed points at different locations to the regression point  $(u_i, v_i)$  at different time stages during  $\Delta t$ . For a better illustration, we show the weight matrix  $W_{\Delta t}$  during the time interval  $\Delta t$  in Fig. 2. The matrix  $W_{\Delta t}$  here is a bit different form the  $W(u_i, v_i)$  in Equation 2. The records in the  $i^{th}$  row of  $W_{\Delta t}$  are the diagonal elements in  $W(u_i, v_i)$ , and only no zero values are used to calibrate the coefficients  $\hat{\beta}_k$  for each regression point. Thus, each row  $r$  of this hat matrix is shown below.

$$r_{it} = X_{it} (X_{\Delta t}^T W_{i\Delta t} X_{\Delta t})^{-1} X_{\Delta t} W_{i\Delta t} \quad (12)$$

In Equation 12,  $X_{it}$  is the  $i^{th}$  row of the matrix of independent variables at  $t$ .  $X_{\Delta t}$  is the matrix of independent variables during a time interval  $\Delta t$ , and  $X_{\Delta t}^T$  is its transpose. Although the  $X_{\Delta t}$  in Equation 12 is equal to the  $X_{O_{\Delta t}}$  in Equation 11 in the fitting and calibration of STWR, we distinguish  $X_{O_{\Delta t}}$  from  $X_{\Delta t}$  here. Because  $X_{O_{\Delta t}}$  is a specific matrix of independent variables of an observed point set  $O_{\Delta t}$  during  $\Delta t$ , while  $X_{\Delta t}$  is a general matrix of independent variables of points during  $\Delta t$ .  $X_{O_{\Delta t}}$  is only used for fitting and calibration of STWR, while  $X_{\Delta t}$  can also be used for prediction in STWR. In other words, we can understand  $X_{O_{\Delta t}}$  as a subclass of  $X_{\Delta t}$ .  $W_{i\Delta t}$  is the  $i^{th}$  row of the weighted matrix  $W_{\Delta t}$ .

### 3.3 Reasonable searching range and procedure of optimization

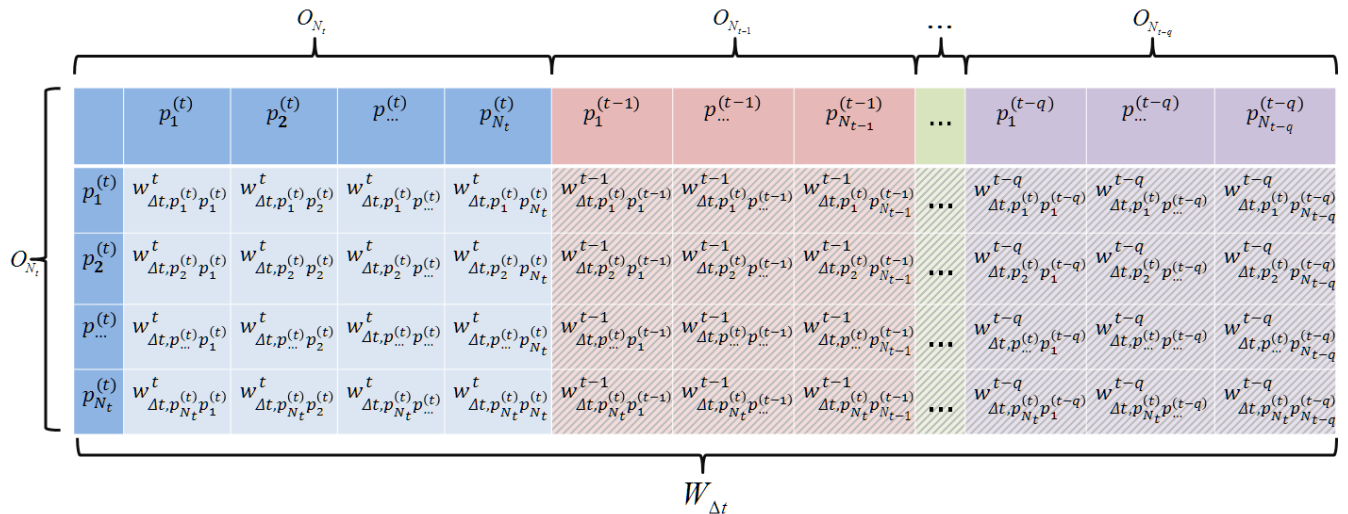
In order to obtain the optimized  $\alpha$  and  $\theta$  for STWR (Equations 8 and 9), the search range should be limited. Here we use the distance from each regression point  $p_i^{(t)}$  to its  $M^{th}$  nearest neighbor as the initial spatial bandwidth  $b_{St}$  at  $t$ . The range of  $b_{St}$  is within a finite set of discrete values, because the maximum number of nearest neighbor is limited to  $N_{t-i}, i \in$



.47  $\{1, 2, \dots, q\}$  for the regression point  $p_i^{(t)}$  ( $N_{t-i}$  is the total number of observed points at  $t - i$ ). We denote that value set for  
 .48  $b_{St}$  as  $BS_{N_t} = \{D_{k+1}, D_{k+2}, \dots, D_{N_t}\}$ , in which the element  $D_U, U \in \{k + 1, k + 2, \dots, N_t\}$  denotes the distance from  $p_i^{(t)}$  to  
 .49 the  $U^{th}$  nearest neighbor, and  $k$  equals to the number of independent variables. Moreover, the searching range of the  
 .50 temporal bandwidth  $b_T$  is also limited to a finite discrete set  $BT_\lambda = \{\Delta t_1, \Delta t_2, \dots, \Delta t_\lambda\}$ , in which the element  $\Delta t_\lambda$  is the time  
 .51 interval from  $t$  to  $t - \lambda$ .

.52 The optimization procedure is to traverse the set  $BT_\lambda$ , and for each step we further traverse the set  $BS_{N_t}$  to get the  
 .53 optimized  $\alpha$  and  $\theta$  through trials. Some trials of  $\theta$  may lead to no solution to Equation 11, because there might be less than  
 .54  $(k + 1)^{th}$  neighbors within the radius of  $b_{St} - \theta \Delta t_\lambda$  from the regression point. Therefore, if it occurs at time stage  $t - \lambda$ ,  
 .55 the spatial bandwidth  $b_{St} - \theta \Delta t_\lambda$  needs to be extended to the distance from its  $(k + 1)^{th}$  nearest neighbor to the regression  
 .56 point, to guarantee the matrix in Equation 11 to be nonsingular.

.57



.58

.59 **Fig. 2.** Weight matrix  $W_{\Delta t}$ . The symbol  $p_k^{(t-i)}, i \in \{0, 1, \dots, q\}, k \in \{1, 2, \dots, N_{t-i}\}$  denotes the  $k^{th}$  observed point at  $t - i$ .

.60 The symbol  $w_{\Delta t, p_m^{(t)} p_n^{(t-i)}}, i \in \{0, 1, \dots, q\}, m \in \{1, 2, \dots, N_t\}, n \in \{1, 2, \dots, N_{t-i}\}$  denotes the weight of the  $n^{th}$  point  $p_n^{(t-i)}$  at

.61  $t - i$  to the  $m^{th}$  point  $p_m^{(t)}$  at  $t$ . The symbol  $O_{N_{t-i}}, i \in \{0, 1, \dots, q\}$  denotes a set of points observed at  $t - i$ .  $\Delta t$  denotes all

.62 the time intervals of the weight matrix. In the central and right parts of the figure, the records with background shading

.63 indicate weight values affected by temporal effects.



:64

### :65 **3.4 Steps of using STWR for prediction**

:66 In this paper, STWR is used to predicate the current values of regression points with known coordinates. The prediction  
:67 formulas of STWR are more complicated than GWR because the spatial distance is calculated directly from the regression  
:68 point to each observed data point, while the time distance between the regression point and the data points observed in the  
:69 past cannot be calculated directly. Therefore, we specify a few steps for the prediction in STWR. First, we need to have the  
:70 optimized initial spatial bandwidth  $b_{st}$ , the optimized  $\alpha$  and  $\theta$ , the optimized number of time stages model used and the  
:71 fitted weight matrix. Second, all data points within the limited distance of spatial bandwidth at the latest time stage should be  
:72 found for the regression point. Third, all the temporal weights of these data points need to be retrieved from the established  
:73 weight matrix (Fig. 2). Fourth, we use these retrieved weights to calculate (e.g., use mean value or inverse distance  
:74 weighting value) the temporal weight on the regression point. Fifth, by combining with the calculated spatial weight and the  
:75 optimized  $\alpha$  and  $\theta$ , we can calculate the spatiotemporal weight on the regression point. Then the value of the regression  
:76 point can be calculated.

:77

## :78 **4 Experiments with Simulated Data**

### :79 **4.1 Simulation design**

:80 To verify the performance of STWR and compare with the results of GWR and GTWR, several groups of simulated data  
:81 were used in this study to represent different types of heterogeneity in space and time. All the data and code used in the  
:82 experiments are shared on GitHub. Web links are provided at the end of this manuscript.

:83 For GTWR, we only compared with the results generated by algorithms in Huang et al. (2010) and Wu et al. (2014),  
:84 because we did not find the software package of Fotheringham et al. (2015). The data generating process (DGP) and the  
:85 spatial heterogeneity are introduced here. The basic DGP is a linear model shown in Equation 1 and the study area is a  
:86 regular  $25 \times 25$  lattice. We defined three initial surfaces to represent the spatial heterogeneity of parameters (Fig. 3), which  
:87 were generated by Equations 13, 14 and 15, respectively (Fotheringham et al., 2017). Through Equation 1, the two  
:88 independent variables  $x_1$  and  $x_2$  were initially generated randomly from the normal distribution  $x_1^{initial} \sim N(100, 8)$  and



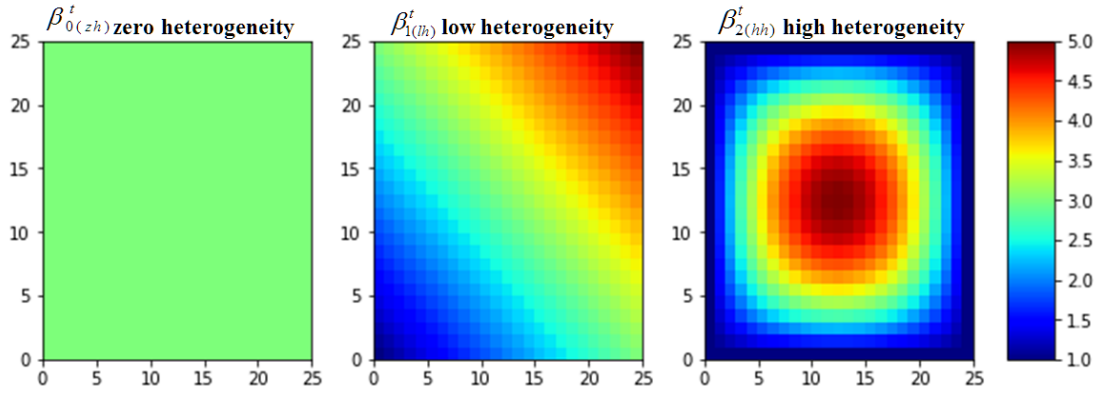
.89  $x_2^{initial} \sim N(50,6)$ , respectively. They can be set as any other values, and the mean values of both distributions may change  
 .90 over time. The error term was generated from a normal distribution  $\varepsilon \sim N(0,0.5)$ .

.91 
$$\beta_{0(zh)}^t = 3 \quad (13)$$

.92 
$$\beta_{1(th)}^t = 1 + \frac{1}{12}(u, v) \quad (14)$$

.93 
$$\beta_{2(hh)}^t = 1 + \frac{1}{324}[36 - (6 - u/2)^2][36 - (6 - v/2)^2] \quad (15)$$

.94



.95

.96 **Fig. 3.** Three simulated initial surfaces for representing spatial heterogeneity of parameters.

.97

.98 Several trends were designed to simulate the value change. For a better simulation, we assumed that value variation can  
 .99 also be spatial heterogeneity. To distinguish from the heterogeneity of the coefficient surface, three other heterogeneity trend  
 .00 functions were defined by Equations 16, 17 and 18.

.01 
$$T_1 V^{t+\Delta t} = V^t + \varphi * \sin(v/4) \Delta t^{npower} \quad (16)$$

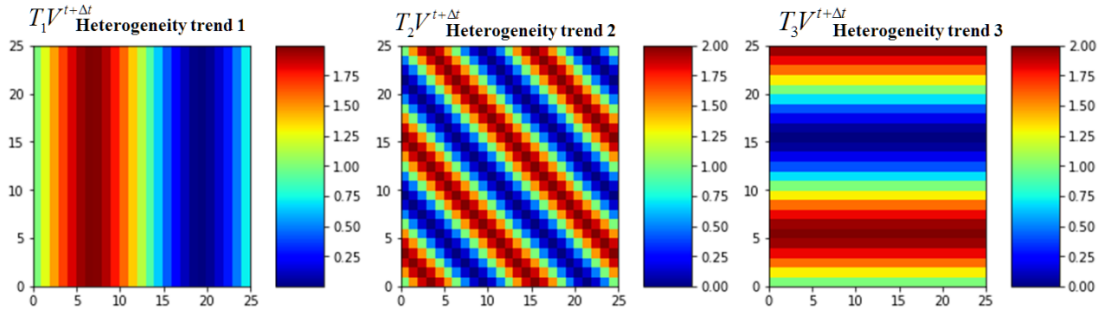
.02 
$$T_2 V^{t+\Delta t} = V^t + \varphi * \sin[1/10\pi u] \Delta t^{npower} \quad (17)$$

.03 
$$T_3 V^{t+\Delta t} = V^t + \varphi * \sin[1/6\pi(u + v)] \Delta t^{npower} \quad (18)$$

.04 In the above equations,  $V^t$  denotes the value at time stage  $t$ ,  $\varphi$  is used for adjusting the magnitude of change,  $\Delta t^{npower}$   
 .05 denotes value change with the  $n^{th}$  power of time interval, and  $T_i V^{t+\Delta t}, i \in \{1,2,3\}$  denotes the  $V$  value at time stage  $t +$



06  $\Delta t$ , which is the result of the  $i^{th}$  trend function from the  $V^t$ . Fig. 4 shows these trends when  $\varphi$ ,  $V^t$ , and  $\Delta t^{power}$  are set to  
 07 one.



08  
 09 **Fig. 4.** Three heterogeneity trend surfaces.

10  
 11 Our goal of this experiment was to test model performance by using sample data from the simulation process at  
 12 different time. Three case studies were designed for different situations. Besides the spatial heterogeneity trends, in our  
 13 simulation design we assumed that the mean values of two independent variables  $x_1$  and  $x_2$  also changed over time, which  
 14 were generated by Equations 19 and 20, respectively.

15 
$$T_1 x_m^{t+\Delta t} = x_m^t \pm \eta_1 * \Delta t \quad (19)$$

16 
$$T_2 x_m^{t+\Delta t} = x_m^t \pm \eta_2 * \Delta t \quad (20)$$

17 In the above two equations,  $x_m^t$  denotes the mean of an independent variable  $x$  at time stage  $t$ ,  $T_i x_m^{t+\Delta t}$ ,  $i \in \{1,2\}$  denotes  
 18 the mean of  $x$  at time stage  $t + \Delta t$ , and  $\eta_1$  and  $\eta_2$  are two parameters for adjusting the rate of change. At each time stage  
 19 during the simulations, the independent variables  $x_1$  and  $x_2$  are generated by a normal distribution with new means of  
 20  $T_1 x_m^{t+\Delta t}$  and  $T_2 x_m^{t+\Delta t}$ , respectively.

21 **4.2 Results with simulated data**

22 We compared the results of OLS, GWR, GTWR, and STWR. A total of 333 random sample points for five time stages  
 23 ( $t_0, t_1, t_2, t_3, t_4$  from old to new) were collected from the 25×25 lattice generated in the above-mentioned DGP. To simplify  
 24 the calculation process, we set  $\theta$  of Equation 7 to zero. Due to the limitation of paper length, in the comparison below the



25 STWR results only include those generated by the spatiotemporal kernel in Equation 8. The objective is to compare the  
26 predicted results with the true value at the latest time stage.

#### 27 4.2.1 Case study 1

28 The time interval of observations in case study 1 was one unit, such as one second or one day. The value change of  $x_1$  and  
29  $x_2$  were generated by  $\eta_1 = 0.5$  and  $\eta_2 = 0.1$ , and were affected by  $T_1V$  with  $\varphi = 0.5$  and  $npower = 1$ . This means that  
30  $x_1$  and  $x_2$  only changed slightly over time. Table 1 presents the results of the global OLS, GWR, GTWR and STWR at the  
31 latest time stage, i.e., stage 5. It shows that the sum of squared errors (SSE) of prediction in STWR is much lower than the  
32 other models in at least one magnitude. In addition, the AICc scores (Equation 10) also shows that STWR outperforms  
33 GTWR and GWR. As shown in Table 1, the R2 (R-squared) value increases from 13.8% in OLS to 94.2% in GWR, 94.9%  
34 in GTWR, and 99.3% in STWR. The estimated standard error, Sigma, reduces to 4.292 in STWR from 23.331 in GTWR.  
35 Also, Fig. 5 shows that both the prediction surface ( $Y\_pred$ ) and the prediction error surface ( $Pred\_Error$ ) of STWR are more  
36 accurate than those in GWR. Due to the limitation of the software package in Huang et al. (2010) and Wu et al. (2014), we  
37 did not generate images for GTWR in Fig. 5, but the result can be seen from the Sigma value in Table 1.

38

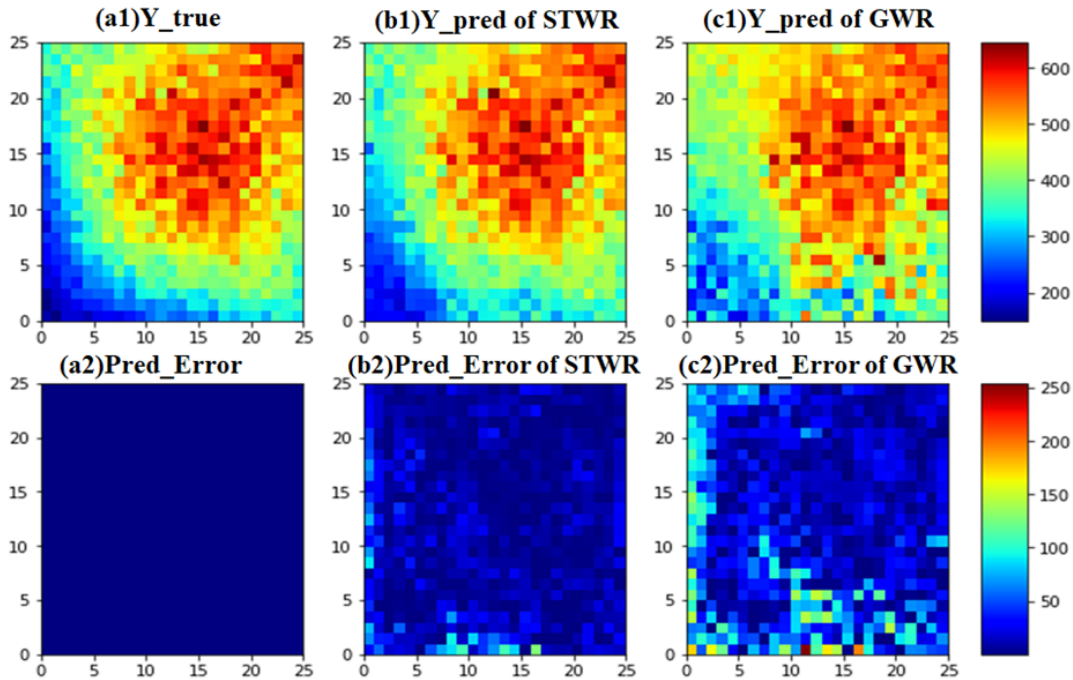
39

**Table 1.** Results of case study 1 at time stage  $t_4$ .

Time stage $t_4$	SSE	AICc	R2	Sigma
OLS	676366.268	805.455	0.138	
GWR	45674.420	705.529	0.942	33.277
GTWR	40056.823	616.641	0.949	23.331
STWR	5761.109	528.860	0.993	4.293

40





41  
42 **Fig. 5.** Comparing prediction results of STWR and GWR in case study 1. Images a1, b1, and c1 are the simulation surfaces  
43 of true Y, the predicted surface of Y by STWR, and the predicted surface of Y by GWR, respectively. Images a2, b2, and c2  
44 are the surface of simulation error, the surface of prediction error of STWR, and the surface of prediction error of GWR,  
45 respectively.

46  
47 **4.2.2 Case study 2**

48 The time interval of observations in case study 2 was 10 units. The value change of  $x_1$  was generated by  $\eta_1 = 0.5$  and  
49 affected by  $T_3V$  with  $\varphi = 0.5$  and  $npower=2$ .  $x_2$  was generated by  $\eta_2=2$  and affected by  $T_2V$  with  $\varphi = 1$  and  $npower$   
50  $= 1$ , which denotes that  $x_1$  and  $x_2$  changed fast over time. Table 2 shows the results of the global OLS, GWR, GTWR and  
51 STWR at the time stage 5. The SSE value in STWR is much lower than other models, and STWR has the highest R2 value  
52 0.995. The Sigma value of STWR is 13.299, which is the lowest and less than one-fifth of the Sigma in GWR and less than  
53 one-sixth of the Sigma in GTWR. Besides, the AICc scores show that STWR significantly outperforms GTWR and GWR.



54 STWR utilized data from the latest three time stages to calibrate the model. The initial spatial bandwidth  $b_{St}$  of STWR  
55 was three nearest neighbors, which was smaller than the one in GWR with 15 nearest neighbors. The optimized  $\alpha$  of STWR  
56 was 0.08, which shows that the effect of used observed points to their local regression points was mainly determined by their  
57 spatial distance. In this case, the GWR outperforms GTWR, which may due to the higher ratio of value change. Compared  
58 with the  $y_{true}$  surface, the predict surface of STWR is much better than GWR (Fig. 6). For the same reason as mentioned in  
59 case study 1, we did not generate images for GTWR in Fig. 6.

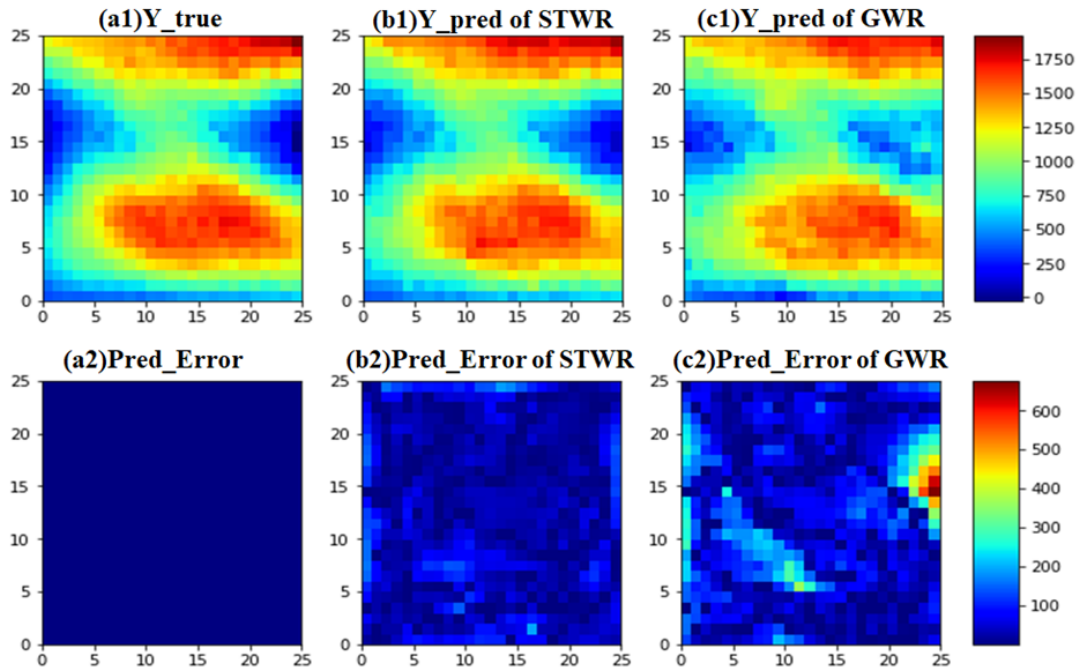
60

61

**Table 2.** Results of case study 2 at time stage  $t_4$ .

Time stage $t_4$	SSE	AICc	R2	Sigma
OLS	5085961.816	938.610	0.494	
GWR	300088.969	840.178	0.970	87.201
GTWR	627011.021	895.6621222	0.938	127.821
STWR	52688.545	709.573	0.995	13.299

62



63

64

65

66

67

68

69

**Fig. 6.** Comparing prediction results of STWR and GWR in case study 2. Images a1, b1, and c1 are the simulation surfaces of true Y, predicted surface of Y by STWR, and predicted surface of Y by GWR, respectively. Images a2, b2, and c2 are the surface of simulation error, the surface of prediction error of STWR, and the surface of prediction error of GWR, respectively.

68

69

### 4.2.3 Case study 3

70

71

72

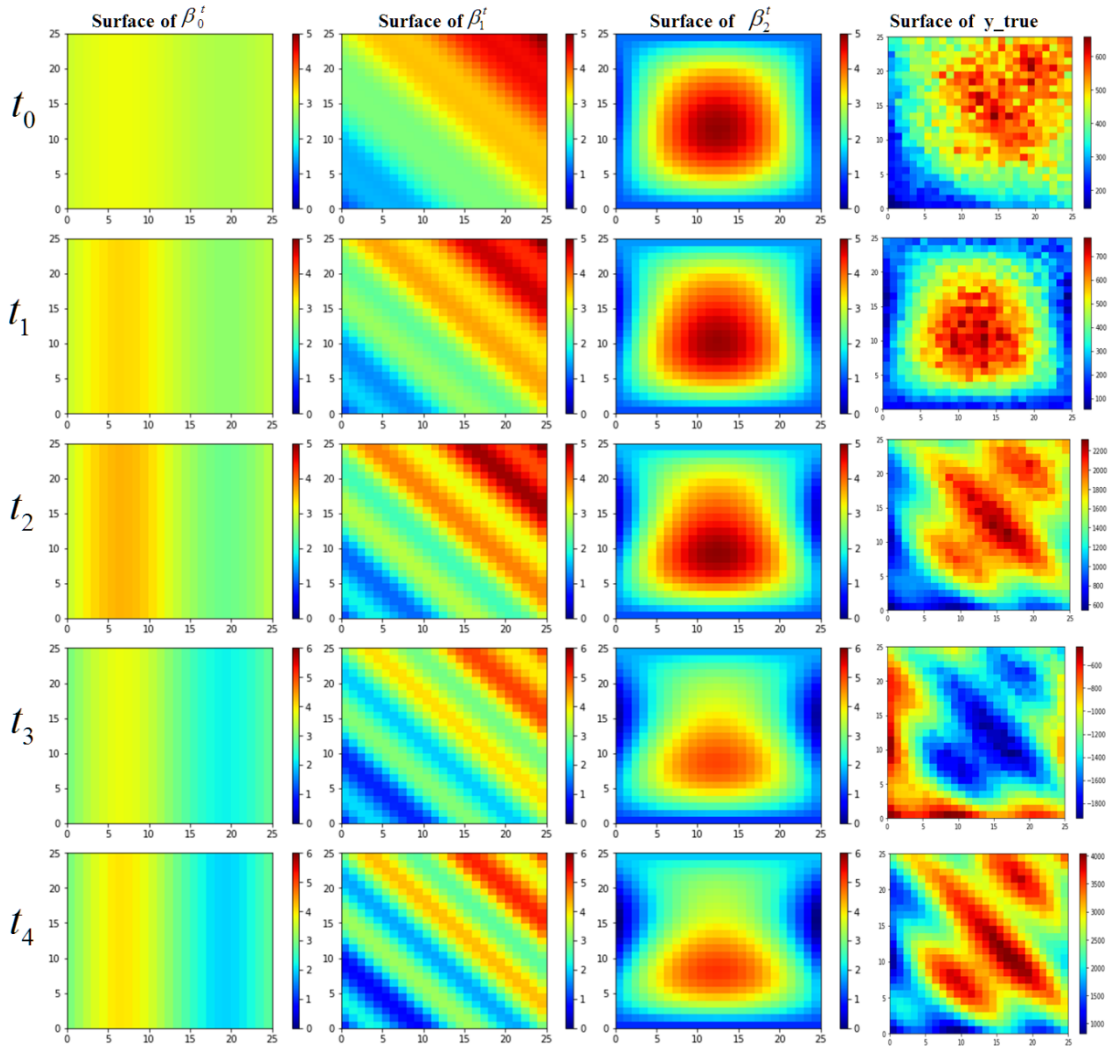
73

74

75

76

The time interval of observations in case study 3 was 200 units. In both case studies 1 and 2, the coefficients in Equation 1 were unchanged. In contrast, in case study 3, three surfaces of coefficients changed over time, which were generated by the trends  $T_1V$ ,  $T_2V$ , and  $T_3V$ , respectively. The variations of coefficients were assumed to be slow. The  $\varphi$  and  $npower$  in each trend were set to be 0.2 and 1, respectively. Both  $\eta_1$  and  $\eta_2$  were set to be 0.5. The dynamic process of the three surfaces of coefficients and the  $y\_true$  surface at each time stage are shown in Fig. 7. The process in case study 3 is more complicated than a general process, but it may be closer to reality.



.77

.78

**Fig. 7.** Dynamic process of three surfaces of coefficients and the  $y\_true$  surface at five different time stages.

.79

.80

.81

.82

.83

.84

Results of these comparisons in case study 3 show that STWR outperforms both GWR and GTWR in accuracy of model and effectiveness of simulation process (Fig. 8a). Along with the change of the coefficients and the increase of  $x_1$  and  $x_2$ , the R2 values of both GWR and GTWR are consistent in the five time stages, showing an overall downward trend. But the R2 of STWR is stable and is at a high level among the five time stages. At the beginning stage  $t_0$ , the R2 values of the three models are similar because there are no previous observations that can be used by STWR and GTWR.

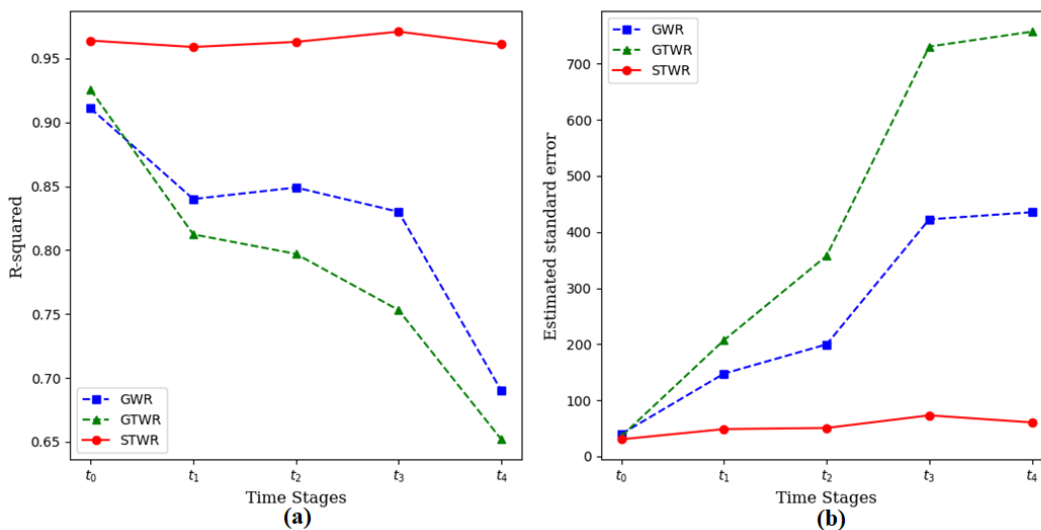


85 The small difference among these models at  $t_0$  may be caused by their different searching range of spatial bandwidth.  
86 Starting from time stage  $t_1$ , STWR and GTWR can borrow points from previous observations. At time stage  $t_1$ , STWR  
87 outperforms both GWR and GTWR, and the advantage of STWR becomes more obvious in the later stages.

88 It may seem strange that GWR can outperform GTWR (Fig. 8), but that is reasonable for the process in case study 3.  
89 The change of this process is faster; and the time interval of observations is bigger than the previous case studies. STWR is  
90 not only able to deal with time intervals, but also to make full use of the value variation of observed points for calibration. In  
91 contrast, GTWR only uses the time interval information and all the observed points to calibrate, which may cause problems  
92 when the observed values are significantly different in spatial distribution or the time intervals are long. GTWR makes use of  
93 points from previous time stages without considering their variation, but if the actual values are quite different from previous  
94 observations at the current time stage, all the point values for the calibration of GTWR will become smooth. Thus, GWR  
95 outperforms GTWR in this situation because GWR only uses the current data points for model calibration.

96 STWR is better for estimation than GWR and GTWR because its Sigma value is much smaller. As shown in Fig. 8b,  
97 the Sigma of STWR was half of GWR at time stage  $t_1$ , and even less than a third of GWR at time stage  $t_4$ . The results show  
98 that the advantage of STWR is obvious comparing with GWR and GTWR.

99



00



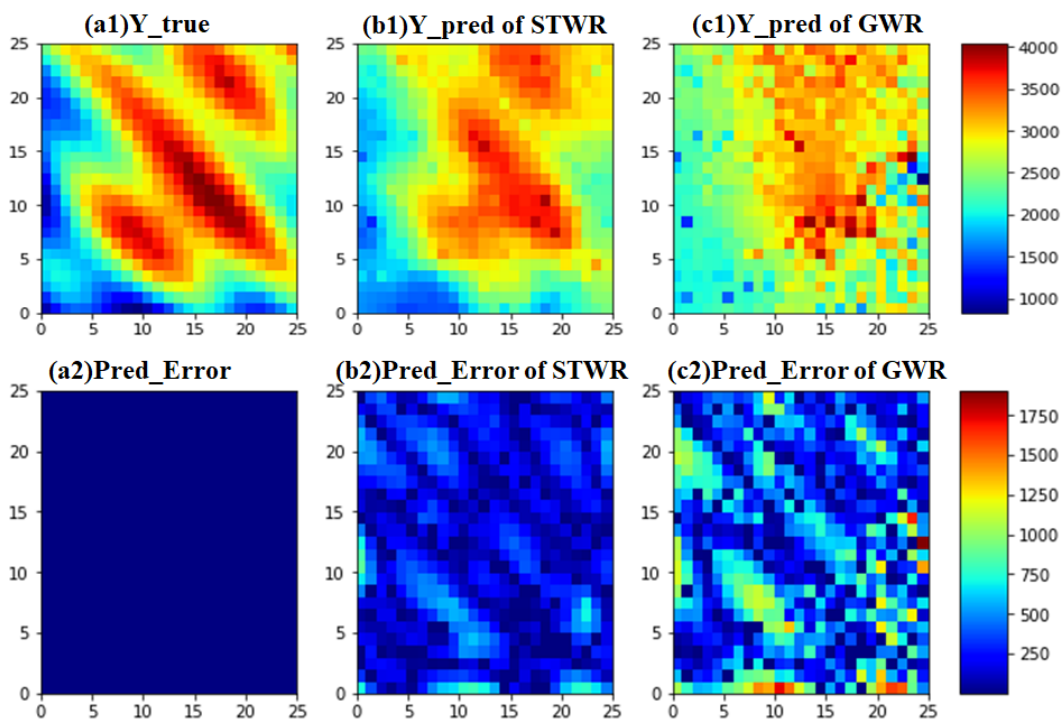
.01 **Fig. 8.** Comparing and evaluating the performance of GWR, GTWR and STWR at five time stages. (a) Comparing the R2  
.02 value of different models; (b) Comparing the Sigma value of different models.

.03

.04 At  $t_4$ , STWR used data from all the past time stages to calibrate the model, and its optimized (initial) spatial bandwidth  
.05  $b_{St}$  was derived from four nearest neighbors, which was smaller than the one in GWR with 25 nearest neighbors. The  
.06 optimized  $\alpha$  of STWR was 0, which means that STWR only borrowed points from past time stages without considering  
.07 their temporal weights to each regression point at  $t_4$ . The predict surfaces at time stage  $t_4$  is shown in Fig. 9. The  $Y_{pred}$   
.08 surface of STWR is much better than GWR, especially in the middle and bottom left parts of the surface. The  $Pred\_Error$  of  
.09 STWR is also much lower than GWR at almost every location. In this case, the  $\alpha$  of STWR at each time stage was 0, 0.96,  
.10 0, 0.07, and 0, respectively. These values indicate that the temporal effects are different at each stage. They also show that  
.11 the value of  $\alpha$  can be adaptive to scale the temporal and spatial effects (see Equation 3).

.12 As Fig. 10 shows, the optimized bandwidths are quite different among these models, and the bandwidths of GWR and  
.13 GTWR are larger than the initial bandwidth of STWR at each time stage. The optimized bandwidth for each time stage refers  
.14 to an optimized number of the nearest neighbors (see Section 3.3). As GTWR considers all the nearest neighbors from  
.15 different time stages, the optimized numbers of the nearest neighbors (bandwidth) grow fast, and exceed the GWR model at  
.16 time stage  $t_2$ . However, the actual distance from the observed points to the regression points is not necessarily farther. The  
.17 initial optimized numbers of the nearest neighbor of STWR are smaller than those in GWR and GTWR, which means that  
.18 the initial spatial bandwidth is narrower than the bandwidth of GWR and GTWR. Nevertheless, due to the strategy of  
.19 borrowing points from nearby neighbors of past observations, the total points for model calibration in STWR may still be  
.20 more than GWR and GTWR. Therefore, the initial optimized numbers of the nearest neighbors in STWR are kept at a lower  
.21 level, which means it is more localized than GWR in this sense.

.22



.23

.24

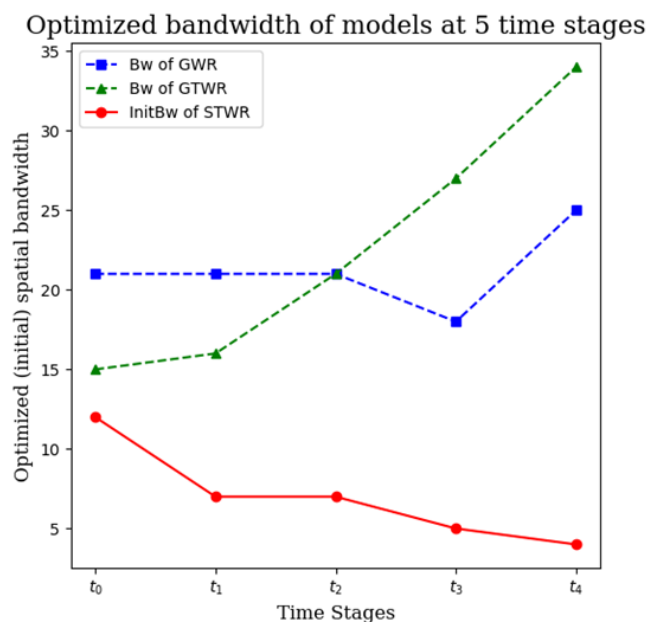
.25

.26

.27

.28

**Fig. 9.** Comparing prediction results of STWR and GWR in case study 3. Images a1, b1, and c1 are the simulation surfaces of true  $Y$ , the predicted surface of  $Y$  by STWR, and the predicted surface of  $Y$  by GWR, respectively. Images a2, b2, c2 are the surface of simulation error, the surface of prediction error of STWR, and the surface of prediction error of GWR, respectively.



.29

.30 **Fig. 10.** Optimized bandwidths (or initial bandwidths) of GWR, GTWR and STWR for the five time stages in case study 3.

.31

### .32 **5 Experiments with Real-world Data**

.33 To further test the performance of STWR, we used data of precipitation  $\delta^2\text{H}$  isotopes in Northeastern United States in  
.34 another case study. We chose  $\delta^2\text{H}$  data in three days from October 29 to 31, 2012, which have enough spatiotemporal data  
.35 for the test. Here in the comparison the STWR results only include those generated by the spatiotemporal kernel in Equation  
.36 8. Data and code used here are shared on Zenodo (See DOI and web links in the ‘Code and data availability’ section at the  
.37 end of the main text of this article).

.38 In the experiments, we collected a total of 782 measurements from 116 sites located in Northeastern United States  
.39 during the three-day period, and prepared the data on a daily average. The daily precipitation, mean temperature, and  
.40 elevation were used as explanatory variables. The model derived from Equation 1 is represented below.

$$.41 \quad y_i = \beta_0 + \beta_1 ppt + \beta_2 tmean + \beta_3 height + \varepsilon_i \quad (22)$$

.42 In Equation 22,  $ppt$  denotes the daily total precipitation (rain + melted snow),  $tmean$  denotes daily mean temperature, and  
.43  $height$  is the elevation value. After data preprocessing, there were 272 points for model calibration and 73 points values on





October 31, 2012. For the first day, both GTWR and STWR took no information from the past. Therefore, we only show the results of SSE, R2 and the optimized initial neighbor (bandwidth) in the model comparisons for the second and third day (D2 and D3) in Tables 3. The SSE of STWR is the lowest at both days. GWR shows a slightly higher SSE than GTWR at D2 and D3. The R2 of STWR is the highest at both days among these models. GWR has lower R2 than GTWR at D2, and almost the same R2 as GTWR at D3.

Similar to the experiments on three simulation datasets, the result here shows that STWR outperforms GTWR and GWR. In the experiment, the number of optimized initial neighbors of STWR was smaller than that of GWR and GTWR. The optimized  $\alpha$  of STWR was 0 at both D2 and D3. The optimized temporal bandwidths of STWR (number of time stages model used) in both D2 and D3 were 2, which means that the STWR in this case only borrowed data points from the latest 2 time stages for D2 and D3. In the result (Table 3), an interesting part to see is that the numbers of optimized initial neighbors of STWR are smaller than the spatial bandwidths of GWR for D2 and D3. The reason is that STWR borrowed points from past time stages in the calculation, which led to narrower bandwidths to some extent.

**Table 3.** Results of model performance with real-world data.

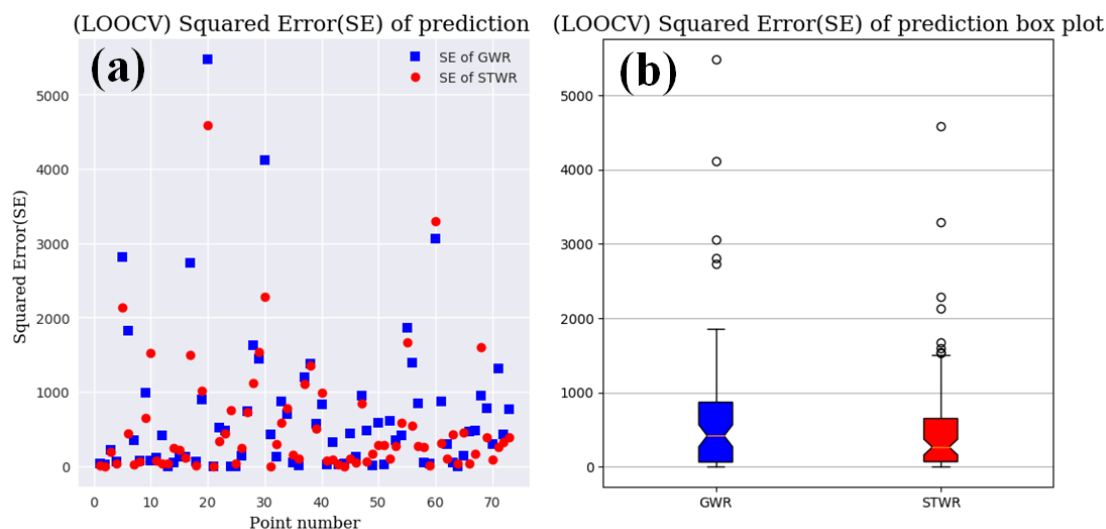
Model	SSE-D2	SSE-D3	R2-D2	R2-D3	Neighbor -D2	Neighbor -D3
<b>OLS</b>	58711.528	52669.399	0.595	0.502		
<b>GWR</b>	33576.400	33043.921	0.769	0.688	52	43
<b>GTWR</b>	32659.808	31967.850	0.775	0.698	37	31
<b>STWR</b>	24022.226	25118.096	0.834	0.763	16	16

We adopted Leave-one-out cross-validation (LOOCV) at D3 for the comparison between STWR and GWR. The squared errors (SE) of prediction are shown in Fig. 11. The prediction results of STWR are better than GWR for most points. The mean SE of STWR is smaller than GWR. Moreover, the SE of STWR shows a narrower regional trend, which indicates



.62 that STWR is more robust than GWR. In addition, the total SSE of GWR and STWR are 50216.510 and 39724.995,  
.63 respectively. Therefore, the result further validates that the quality of predication in STWR is better than GWR.

.64



.65

.66 **Fig. 11.** LOOCV results of STWR and GWR. (a) Squared error of prediction at each point (leave out); (b) Box plot of the  
.67 LOOCV results of GWR and STWR.

.68

.69 In Fig. 12, the predicted  $\delta^2\text{H}$  surface at D3 is broadly similar between the GWR and STWR calibrations. The  
.70 percentages of explanation of variance in GWR and STWR are similar, which are 68.8% and 76.3%, respectively. However,  
.71 like the experiment results with simulated data (Fig. 10), STWR has narrower initial bandwidth, which generates more  
.72 localization in the predicted  $\delta^2\text{H}$  surface than GWR. For instance, the lower (light yellow and blue parts) or higher (orange  
.73 parts) predicted values of  $\delta^2\text{H}$  are more concentrated in the  $\delta^2\text{H}$  surface of STWR than that of GWR (Fig. 12).

.74

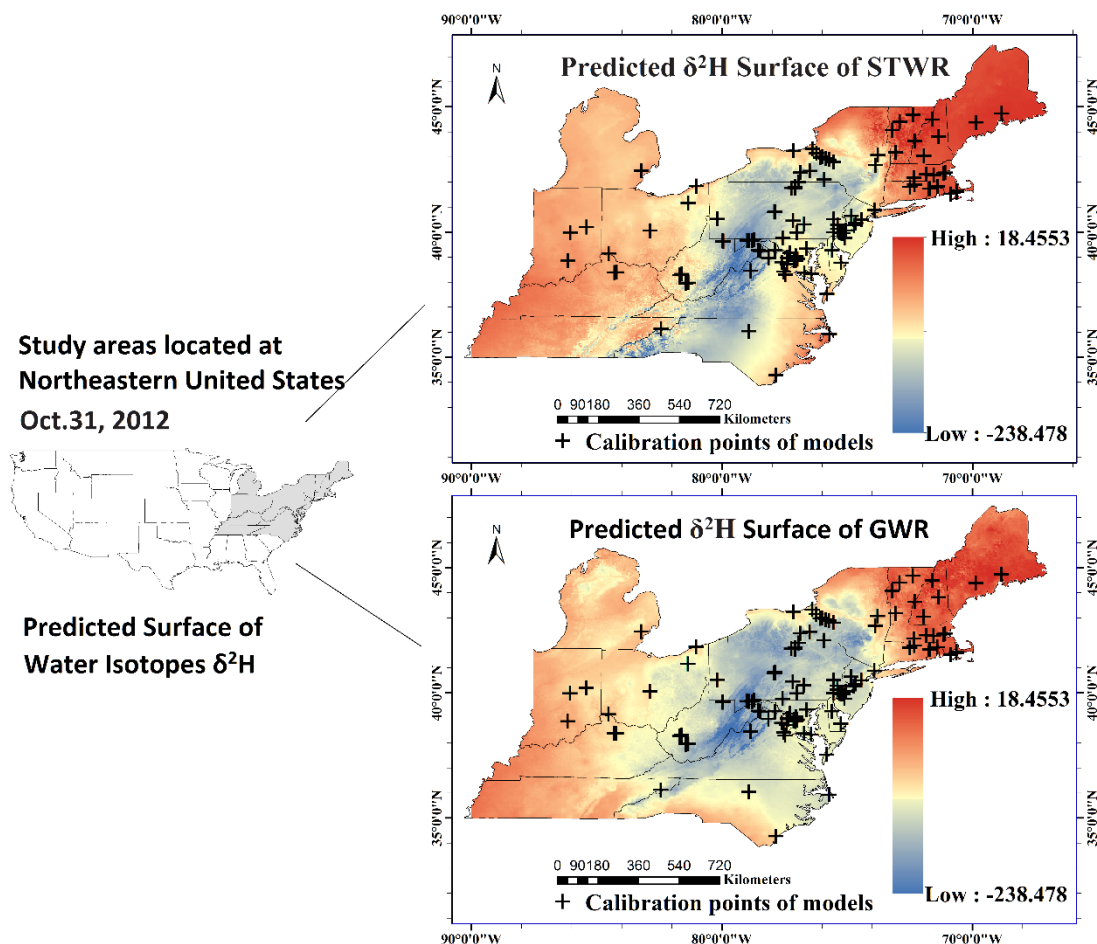


Fig. 12. Predicted  $\delta^2\text{H}$  surfaces of STWR and GWR at D3.

.75

.76

.77

## .78 6 Discussion and Conclusions

.79 Spatiotemporal data analysis is important in many scientific studies. Due to the complexity of spatiotemporal models,  
.80 spatiotemporal effect may not be fully taken into account when the temporal and spatial information is manipulated  
.81 simultaneously. In particular, the models for the effect of spatial dynamics should not be simply adapted for modeling the  
.82 effect of temporal dynamics. Although the GTWR model can borrow points from the near recent, without careful  
.83 consideration of temporal effect, the performance of GTWR may even be worse than GWR. Increasingly, many scientific  
.84 issues are not just about spatial non-stationary but involve many spatiotemporal processes. It is necessary to review the



.85 limitation of the current spatiotemporal models and make new extensions. The aim of the STWR model developed in this  
.86 study is to advance the work and discussion in that direction.

.87 The temporal kernel and the spatiotemporal kernel functions are two important contributions of STWR. The temporal  
.88 kernel in STWR applies an improved sigmoid form (see Equation 4), which is different from the methods for temporal effect  
.89 analysis in previous GTWR models. The temporal weight generated by the STWR temporal kernel is limited as a value  
.90 between 0 and 1. The spatial weight in STWR is also limited as a value between 0 and 1. The STWR spatiotemporal kernel  
.91 function has a weight adjustment parameter  $\alpha$  to scale the temporal and spatial weights (Equation 3). In practice,  $\alpha$  can be  
.92 obtained through optimization. This form of weighted average between temporal and spatial effects in the STWR  
.93 spatiotemporal kernel is a big improvement comparing with the multiplication form in previous GTWR models. The  
.94 advantage of the STWR spatiotemporal kernel has been proven in four case studies with both simulated and real-world  
.95 datasets.

.96 Though the performance of STWR is outstanding, the models can still be further extended. A big topic is about the time  
.97 distance. In the current STWR, the time distance represents the rate of value variation between an observed point and a  
.98 regression point through a time interval. Nevertheless, we can also use time distance to represent the rate of value variation  
.99 at each observed point object through time. Note that, from an object-oriented perspective, here we differentiate the point  
.00 objects from locations, although the point objects have geospatial coordinates as part of their attributes. Following that new  
.01 definition of time distance, the  $y_{i(t)} - y_{j(t-q)}$  in the STWR temporal kernel (Equation 4) can be replaced by  $\Delta y_{j(t-q)}$   
.02 (value variation of an observed point object during  $\Delta t$ ). A scenario of interest is that, the observed point objects in the past  
.03 time stages (such as those shown in Fig. 1) may move to new locations, have no value for a few time stages, or even  
.04 disappear, so the  $\Delta y_{j(t-q)}$  may not exist. We can use object-based methods to address issues caused by that scenario. For  
.05 example, each point object can be assigned with a unique ID, and then the observed value of the point object at each time  
.06 stage can be retrieved by using its ID. With this new definition of time distance, the temporal weight on a regression point  
.07 object is determined by the rate of value variation of its nearby point objects. Several different scenarios for a regression  
.08 point object at current time stage  $t$  are discussed here.



09 (1) The location of an observed point object  $j$  is fixed through time (e.g., a fixed sensor). If the value of  $j$  is observed  
10 at both time stages  $t$  and  $t - q$ , then  $\Delta y_{j(t-q)}$  can be calculated directly. If the value of  $j$  is observed at  $t$  but not  
11 observed at  $t - q$ , we can use interpolation to generate a value for  $j$  at  $t - q$ . If the value of  $j$  is not observed at  $t$ , but  
12 the variation in the past is observed, we can use prediction methods to generate a value for  $j$  at  $t$ .

13 (2) The location of  $j$  is not fixed through time (i.e.,  $j$  moves). The moving point objects can still have temporal  
14 effects to the regression point, then the  $\Delta y_{j(t-q)}$  can be calculated. The spatial effect, however, depends on whether  $j$   
15 moves out of the spatial bandwidth from the regression point or not.

16 (3)  $j$  disappears or appears at a certain time stage. If  $j$  does not appear until the current time stage  $t$ , the  $\Delta y_{j(t-q)}$  can  
17 be set to be 0. If  $j$  appears in a past time stage (e.g.,  $t - q$ ) but it disappears before or at  $t$ , we can ignore the impact of  $j$   
18 for the regression point object.

19 There are other possibilities for the further improvement of STWR. The first is about the optimization of  $\theta$  in the  
20 spatiotemporal kernel (Equations 8 and 9). The slope  $\theta$  indicates that the variation of the spatial bandwidth is in a linear  
21 form, but it may not be a perfect solution. In many situations, the change of the spatial bandwidth over time may not be  
22 linear. The second is about making predications for future time stages. In this paper, we only predict values for points at the  
23 current time stage  $t$ . Extensions can be made in STWR to predict values for points in future time stages beyond  $t$ . The third  
24 future work is about exploring multiple spatial and temporal bandwidths of models. Different variables may have different  
25 spatial and temporal bandwidths due to their unique characteristics. Correspondingly, we may need more bandwidths to  
26 capture the different non-stationarities of those independent variables, to better represent the spatiotemporal heterogeneity.

27 In short, the core contribution of STWR is the clarification of the ‘time distance’ concept and the new temporal kernel  
28 and spatiotemporal kernel functions based on this concept. Our experiments show that STWR outperforms GWR and GTWR  
29 in analyzing and interpreting local spatiotemporal non-stationarity. We hope STWR can bring fresh ideas and new  
30 capabilities for spatiotemporal data analysis in many disciplines.

31

32 **Code and data availability**



33 The Python source code of STWR v1.0, the data used in the experiments and all the case studies (written in Jupyter  
34 Notebook) were archived on Zenodo and made freely accessible via <http://doi.org/10.5281/zenodo.3637689>. Data source of  
35 water isotopes  $\delta^2\text{H}$  is on the website: [http://wateriso.utah.edu/waterisotopes/pages/spatial\\_db/SPATIAL\\_DB.html](http://wateriso.utah.edu/waterisotopes/pages/spatial_db/SPATIAL_DB.html). The data  
36 of daily precipitation and mean temperature were collected from the PRISM Climate Group  
37 (<http://www.prism.oregonstate.edu>), and the elevation data were collected from the GMTED2010  
38 ([https://topotools.cr.usgs.gov/gmted\\_viewer/viewer.htm](https://topotools.cr.usgs.gov/gmted_viewer/viewer.htm)) at U.S. Geological Survey (USGS).

39

#### 40 **Author Contribution.**

41 X.Q., X.M. and C.M. developed the algorithm, X.Q. implemented and coded the algorithm. X.Q. prepared the manuscript  
42 with contributions from all co-authors.

43

#### 44 **Competing interests.**

45 The authors declare that they have no conflict of interest.

46

#### 47 **Acknowledgement.**

48 The research presented in this paper was partially supported by the National Science Foundation under Grant No. 1835717,  
49 the China Scholarship Council under Grant No. 201807870006, and the Fujian Provincial Department of Education under  
50 Grant No. KLA18025A. The authors thank Prof. Stewart Fotheringham and other colleagues at the Spatial  
51 Analysis Research Center (SPARC) of Arizona State University for their insightful comments and suggestions during a  
52 seminar about the STWR model.

53

54

#### 55 **References**

56 Akaike, H.: Information theory and an extension of the maximum likelihood principle. In: Selected papers of hirotugu  
57 akaike, Springer, 1998.



- 58 Akaike, H.: Maximum likelihood identification of Gaussian autoregressive moving average models, *Biometrika*, 60, 255-  
59 265, 1973.
- 60 Atkinson, P. M., German, S. E., Sear, D. A., and Clark, M. J.: Exploring the relations between riverbank erosion and  
61 geomorphological controls using geographically weighted logistic regression, *Geographical Analysis*, 35, 58-82, 2003.
- 62 Bowman, A. W.: An alternative method of cross-validation for the smoothing of density estimates, *Biometrika*, 71, 353-360,  
63 1984.
- 64 Brown, S., Versace, V. L., Laurenson, L., Ierodiaconou, D., Fawcett, J., and Salzman, S.: Assessment of spatiotemporal  
65 varying relationships between rainfall, land cover and surface water area using geographically weighted regression,  
66 *Environmental Modeling & Assessment*, 17, 241-254, 2012.
- 67 Brunsdon, C., Fotheringham, A. S., and Charlton, M. E.: Geographically weighted regression: a method for exploring spatial  
68 nonstationarity, *Geographical analysis*, 28, 281-298, 1996.
- 69 Brunsdon, C., Fotheringham, S., and Charlton, M.: Geographically weighted regression, *Journal of the Royal Statistical*  
70 *Society: Series D (The Statistician)*, 47, 431-443, 1998.
- 71 Cahill, M. and Mulligan, G.: Using geographically weighted regression to explore local crime patterns, *Social Science*  
72 *Computer Review*, 25, 174-193, 2007.
- 73 Cardozo, O. D., García-Palomares, J. C., and Gutiérrez, J.: Application of geographically weighted regression to the direct  
74 forecasting of transit ridership at station-level, *Applied Geography*, 34, 548-558, 2012.
- 75 Chen, J., Shaw, S.-L., Yu, H., Lu, F., Chai, Y., and Jia, Q.: Exploratory data analysis of activity diary data: a space-time GIS  
76 approach, *Journal of Transport Geography*, 19, 394-404, 2011.
- 77 Cleveland, W. S.: Robust locally weighted regression and smoothing scatterplots, *Journal of the American statistical*  
78 *association*, 74, 829-836, 1979.
- 79 Crespo, R., Fotheringham, S., and Charlton, M.: Application of geographically weighted regression to a 19-year set of house  
80 price data in London to calibrate local hedonic price models, 2007.
- 81 Cressie, N. and Wikle, C. K.: *Statistics for spatio-temporal data*, John Wiley & Sons, 2015.
- 82 Cressie, N. A.: *Statistics for Spatial Data*. New York: John Willey & Sons, 1993.



- 83 Fotheringham, A. S., Brunson, C., and Charlton, M.: Geographically weighted regression: the analysis of spatially varying  
84 relationships, John Wiley & Sons, 2003.
- 85 Fotheringham, A. S., Crespo, R., and Yao, J.: Geographical and temporal weighted regression (GTWR), *Geographical*  
86 *Analysis*, 47, 431-452, 2015.
- 87 Fotheringham, A. S., Yang, W., and Kang, W.: Multiscale geographically weighted regression (mgwr), *Annals of the*  
88 *American Association of Geographers*, 107, 1247-1265, 2017.
- 89 Fraser, L. K., Clarke, G. P., Cade, J. E., and Edwards, K. L.: Fast food and obesity: a spatial analysis in a large United  
90 Kingdom population of children aged 13–15, *American journal of preventive medicine*, 42, e77-e85, 2012.
- 91 Gelfand, A. E., Ecker, M. D., Knight, J. R., and Sirmans, C.: The dynamics of location in home price, *The journal of real*  
92 *estate finance and economics*, 29, 149-166, 2004.
- 93 Goodchild, M. F.: Prospects for a space–time GIS: Space–time integration in geography and GIScience, *Annals of the*  
94 *Association of American Geographers*, 103, 1072-1077, 2013.
- 95 Hoaglin, D. C. and Welsch, R. E.: The hat matrix in regression and ANOVA, *The American Statistician*, 32, 17-22, 1978.
- 96 Huang, B., Wu, B., and Barry, M.: Geographically and temporally weighted regression for modeling spatio-temporal  
97 variation in house prices, *International Journal of Geographical Information Science*, 24, 383-401, 2010.
- 98 Hurvich, C. M., Simonoff, J. S., and Tsai, C. L.: Smoothing parameter selection in nonparametric regression using an  
99 improved Akaike information criterion, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60,  
100 271-293, 1998.
- 101 Loader, C. R.: Bandwidth selection: classical or plug-in?, *The Annals of Statistics*, 27, 415-438, 1999.
- 102 Mennis, J. L. and Jordan, L.: The distribution of environmental equity: Exploring spatial nonstationarity in multivariate  
103 models of air toxic releases, *Annals of the Association of American Geographers*, 95, 249-268, 2005.
- 104 Pace, R. K., Barry, R., Gilley, O. W., and Sirmans, C.: A method for spatial–temporal forecasting with an application to real  
105 estate prices, *International Journal of Forecasting*, 16, 229-246, 2000.





- i06 Sun, T. Y., Conroy, G., Donner, E., Hungerbühler, K., Lombi, E., and Nowack, B.: Probabilistic modelling of engineered  
i07 nanomaterial emissions to the environment: a spatio-temporal approach, *Environmental Science: Nano*, 2, 340-351,  
i08 2015.
- i09 Takahashi, K., Kulldorff, M., Tango, T., and Yih, K.: A flexibly shaped space-time scan statistic for disease outbreak  
i10 detection and monitoring, *International Journal of Health Geographics*, 7, 14, 2008.
- i11 Tobler, W. R.: A computer movie simulating urban growth in the Detroit region, *Economic geography*, 46, 234-240, 1970.
- i12 Wang, W., Zhao, J., Cheng, Q., Carranza, E.J.M.: GIS-based mineral potential modeling by advanced spatial analytical  
i13 methods in the southeastern Yunnan mineral district, China. *Ore Geology Reviews*, 71, 735-748.  
i14 <https://doi.org/10.1016/j.oregeorev.2013.08.005>, 2015.
- i15 Wheeler, D. C. and Waller, L. A.: Comparing spatially varying coefficient models: a case study examining violent crime  
i16 rates and their relationships to alcohol outlets and illegal drug arrests, *Journal of Geographical Systems*, 11, 1-22, 2009.
- i17 Wu, B., Li, R., and Huang, B.: A geographically and temporally weighted autoregressive model with application to housing  
i18 prices, *International Journal of Geographical Information Science*, 28, 1186-1204, 2014.