Thank you for the detailed comments and suggestions. Please check below the point-by-point response to the main concerns listed in the referee report.

Comment 1. The main innovation of STWR is using the rate of value variation of the nearby observed point during the time interval to represent the time distance. However, the value variation between the estimated point and the observed points is not only influenced by the time variation but also the difference of geographical locations. How to distinguish whether this effect is caused by time or space? Further, the value variation not only occurs during the time but also occurs across space. Why not also consider the value variation across space?

Reply 1.

①We can use $y_{i(t)} - y_{j(t-q)}$ to represent the value variation between the regression point and the observation point that have time difference of Δt (q). Suppose that the variation contains two parts caused by time and space, and they are $f_t(\Delta y_{j(t-q)})$ and $f_s(y_{i(t)} - y_{j(t)})$ respectively. $f_t(\Delta y_{j(t-q)})$ is not affected by spatial effects, because the location of point j does not change during Δt. $f_s(y_{i(t)} - y_{j(t)})$ is not affected by temporal effects, because $y_{i(t)}$ and $y_{j(t)}$ are observed at the same time. In theory, if we get the value $y_{j(t)}$, we may determine if the variation caused by time or space, because both f_t and f_s need the value $y_{j(t)}$. The y value of the location j at t (i.e. $y_{j(t)}$) is often unavailable or may not exist, we use the $y_{i(t)} - y_{j(t-q)}$ to approximate $\Delta y_{j(t-q)}$ within the local spatiotemporal bandwidth when employing the k_T to calculate the temporal weights. (Please see relevant explanations in the reply 1 of the first reviewer). This may introduce some errors because of the different locations of i and j, but the errors are limited. Consequently, the value variation between the estimated point and the observed point in different times is mainly temporal effect, the spatial effect is limited and ignored here.

② The STWR algorithm is based on the assumptions and framework of the GWR model. When calculating the spatial weights, we use the same k_s employed in GWR, whose spatial impacts is calculated by the spatial distance d_{sij} between i and j. We introduce the value variation to better identify or capture the heterogeneities caused by the same time interval but different temporal effects, that is, the temporal heterogeneity of the rate of value change. The heterogeneities of this part were not considered in the previous GTWR. As for the calculation of spatial weights, the main reason that we did not consider the value variation across space is to be consistent with the GWR model, i.e. following the assumption that as long as the spatial distances between observation points to the regression point are equal, their spatial weights are the same. There may be other factors, such as anisotropy or value variation across space, that may have some additional spatial impacts on the regression point. The reasons we follow GWR's assumptions are: (a) In the optimization procedure, the model will adaptively adjust its spatial bandwidth according to the density of sampling points, and to the value variations in the space. If the value variations across the space are small, the adaptive spatial bandwidth will be large. It means that the optimization procedure already uses the information about value variation across the space. (b) If the variation $y_{i(t)} - y_{j(t)}$ was used to build a new spatial distance, which will violate the aforementioned assumption of GWR, the prediction and calibration process should be changed. Because $y_{i(t)}$ value that is required in the calculation of the new distance does not exist in prediction, spatial weights from surrounding observed points should be estimated by interpolation or other methods (just like the interpolation of temporal weights) that may bring other uncertainties or errors. Evaluating and comparing these uncertainties is not the scope of this paper in our plan. (c) If the $|y_{i(t)} - y_{j(t)}|/d_{sij}$ was used as a new spatial distance for calculating the spatial weights, we have to deal with the special case when $y_{i(t)}$ equal to $y_{j(t)}$, because the spatial kernels (such as bi-square and Gaussian) are different form the temporal kernel of STWR. In other words, if $y_{j(t-q)}$

is close or equal to $y_{i(t)}$ when employing our temporal kernel k_T, the output temporal weight is close or equal to 0. The underly meaning is explainable, because when the value variation gets close or equal to 0, the influence from observed point to the regression point gets weak or disappear. If $y_{i(t)}$ is close or equal to $y_{j(t)}$ when employing the bi-square or Gaussian kernel, the meaning may be difficult to understand, because when the new spatial distance $|y_{i(t)} - y_{j(t)}|/d_{sij}$ is close to 0, the output spatial weights will be large, which is inconsistent with the fact that the weaker influences it should have when the smaller value variation across space. Besides, the bi-square or Gaussian kernel have no solutions when $y_{i(t)}$ is equal to $y_{j(t)}$. If the numerator and denominator are swapped (i.e. $d_{sij}/|y_{i(t)} - y_{j(t)}|$), the $y_{i(t)}$ can not be equal to $y_{j(t)}$, while it is normal that $y_{j(t)}$ may be equal to $y_{i(t)}$. Therefore, if we consider combing the $y_{i(t)} - y_{j(t)}$ with d_{sij} to build a new spatial distance, we may probably need to design a new appropriate spatial kernel, which requires more difficult theoretical knowledge on describing the local spatial effects.

Comment 2. The authors indicate that the current GTWR model directly calculates the integrated spatiotemporal weights by using a multiplication of the spatial and temporal weights, which may cause underestimation of weights. This is easily misunderstood. The GTWR model also uses a scale parameter to handle the difference between time and space, which is the same as the proposed STWR model. Please correct or give more explanation.

Reply 2. The composite spatiotemporal weights might be underestimated in the current GTWR models by using the multiplication kernel. Because both outputs of the spatial kernel and the temporal kernel range from 0 to 1, and the multiplied value is never bigger than the smaller one of the spatial and temporal kernels, which means that the composite spatiotemporal impacts are never greater than the single spatial impacts and the single temporal impacts. However, the real combined spatiotemporal impacts, may be higher than the single spatial impacts or the temporal impacts, or at least may be higher than the smaller ones. Moreover, multiplication makes the weight decay faster. The role of the adjustable parameter α used in STWR is different from the scale parameter τ ($\tau = \frac{u}{\lambda}$) in GTWR. The parameter α is used for adjusting the outputs of the spatial kernel k_s and the temporal kernel k_T, which means measuring the relative strength of the spatial and temporal impacts on the regression point. However, the scale parameter τ is used for linearly adjusting the inconsistency of the distance between time and space, because of the differences of their units, scales, or metrics, etc. Specifically, GTWR uses parameters u and v to generate the spatiotemporal distance d_{ij}^{ST} (given in the following **Equation 1**). And then substituting the d_{ij}^{ST} into the spatial kernel (Gaussian), its composited weights were obtained (**Equation 2**, we use w to replace the α in the original formulation, which is easier to understand in symbol). This equation, after transformation, is equal to the multiplication form of two Gaussian kernels (i.e. the spatial kernel and temporal kernel). Therefore, the scale parameter τ in GTWR only adjusts the differences between time distances and space distances, which does not change the multiplication form of the spatiotemporal kernel. In contrast, the parameter α in STWR (**Equation 3**) is used to adjust the effects of the two kernels k_s and k_T, and the adjusted composite spatiotemporal weight w_{ijST}^{t} may be larger than the smaller one of the output values of $k_s(d_{sij}, b_{ST})$ and $k_T(d_{tij}, b_T)$.

$$\boldsymbol{d_{ij}^{ST} = \lambda[(u_i - u_j)^2 + (v_i - v_j)^2] + \mu(t_i - t_j)^2} \qquad \textbf{(1)}$$

$$
\begin{aligned}
\boldsymbol{w_{ij}} = \ & \boldsymbol{exp}\left\{-\left(\frac{\boldsymbol{\lambda}\left[(\boldsymbol{u_i}-\boldsymbol{u_j})^2+(\boldsymbol{v_i}-\boldsymbol{v_j})^2\right]+\boldsymbol{\mu}(\boldsymbol{t_i}-\boldsymbol{t_j})^2}{\boldsymbol{h_{ST}^2}}\right)\right\} \\
& = \boldsymbol{exp}\left\{-\left(\frac{\left[(\boldsymbol{u_i}-\boldsymbol{u_j})^2+(\boldsymbol{v_i}-\boldsymbol{v_j})^2\right]}{\boldsymbol{h_S^2}}\right)+\frac{(\boldsymbol{t_i}-\boldsymbol{t_j})^2}{\boldsymbol{h_T^2}}\right\} \\
& = \boldsymbol{exp}\left\{-\left(\frac{(\boldsymbol{d_{ij}^S})^2}{\boldsymbol{h_S^2}}+\frac{(\boldsymbol{d_{ij}^T})^2}{\boldsymbol{h_T^2}}\right)\right\} \\
& = \boldsymbol{exp}\left\{-\frac{(\boldsymbol{d_{ij}^S})^2}{\boldsymbol{h_S^2}}\right\}\times \boldsymbol{exp}\left\{-\frac{(\boldsymbol{d_{ij}^T})^2}{\boldsymbol{h_T^2}}\right\} \\
& = \boldsymbol{w_{ij}^S} \times \boldsymbol{w_{ij}^T} \qquad \textbf{(2)}
\end{aligned}
$$

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

We give more explanation in the revised manuscript, please see lines 99-105 (Page 4,5) and lines 174-179 (Page 8).

Comment 3. As new platforms and instruments have brought increasingly massive spatiotemporal data, deep learning and neural networks have also been integrated with geostatistical models to handle spatial and temporal non-stationary relationships, such as geographically neural network regression (GNNWR), geographically and temporally neural network regression (GTNNWR). These neural network-based models can even capture the complex non-linearity in the non-stationary relationship. Some discussion or comparison between STWR with these models should be added.

Reply 3. With many successful applications of deep learning and neural network in many fields, its combinations with the traditional geospatial tools is becoming a promising research topic. Geographic neural network weighted regression (GNNWR) (Du et al., 2020) is a new attempt to combine the OLS and GWR with Artificial neural networks (ANNs). Geographic and temporal neural network regression (GTNNWR) (Wu et al., 2020) is based on the GNNWR with combing a new ANNs based method to calculate the spatiotemporal distance. Our STWR algorithm is based on the GWR with a new temporal distance and spatiotemporal kernel. There are four main differences between the GTNNWR/GNNWR and STWR: ① The basic formulation of GNNWR is defined as **Equation (4)**. The $w_0(u_i, v_i)$ and $w_k(u_i, v_i)$ denote the geographical weight of the constant coefficient β_0 and coefficient β_k, respectively. It assumed that the multiplication of $w_p(u_i, v_i)$ and β_p is equal to $\beta_p(u_i, v_i)$ $(0 \le p \le k)$. The combined $\beta_p(u_i, v_i)$ is thought as the same as the coefficients of GWR. But in STWR and GWR, the weights and the estimated coefficients are separated. The weights mainly reflect the degree of the influences from the observed points to the regression point, while the coefficient values reflect the relationships between the independent variable and dependent variable. ② GTNNWR and GNNWR use the proposed ANNs based method (**Equation 5**) to calculate the weighted matrix, which is quite different from the kernel functions used in GWR and STWR models. Although GTNNWR and GNNWR use the idea of pointwise regression, they do not consider how to "borrow points" from nearby neighbors and do not have the concept of bandwidth. Without spatial bandwidth, all observation points in the study area may have impacts on the regression point, which might violate the Tobler's first law of geography (Tobler, 1970). It may be difficult to understand the relationships between the influence weight and the spatial distances, especially when the study area and the data amounts are large. STWR has spatial bandwidths and follows the Tobler's first law of geography, which can help analyze the affected range of local regression points. ③ The data points will be divided into training set (including validation set) and test set for the GTNNWR and GNNWR, which might require more data points. Thus, it may not be appropriate for analyzing fewer amounts of data points (data acquisitions of many geoscience processes are difficult and costly). STWR and GWR do not need to divide data points into the training set (including validation set) and test set, which requires less data points than GNNWR and GTNNWR. ④ Although GTNNWR utilizing a method named spatiotemporal proximity neural network (STPNN) to calculate the spatiotemporal distance, the obtained integrated spatiotemporal distance is lack of explanation, and it is also impossible to tell apart which parts of the calculated weight is affected by time or space. Besides, there is no concept of temporal bandwidth in GTNNWR. Thus, it cannot tell us how old the historical observation points that will have impacts on the regression point. But STWR has temporal bandwidth, and it can distinguish the strength of temporal weight and spatial weight. Therefore, we can analyze the characteristics of the local interaction of time and space according to the temporal bandwidth, spatial bandwidth, and the adjustment parameter α, etc.

$$\boldsymbol{y_i = w_0(u_i, v_i)\beta_0 + \sum_{k=1}^{p} w_k(u_i, v_i)\beta_k x_{ik} + \varepsilon_i \, , i = 1, 2, \ldots, n} \qquad \textbf{(4)}$$

$$\boldsymbol{W_i = W(u_i, v_i) \; = \; SWNN([d_{i1}^s, d_{i2}^s, \ldots, d_{in}^s \,]^T)} \qquad \textbf{(5)}$$

Our STWR algorithm, especially the new concept of the time distance, may also be integrated with the machine learning methods, which is our future work.

We add the discussions on the differences between STWR and GTNNWR/GNNWR to the Section 6, please see lines 489-523 (Page 28,29)

References:

Du, Z., Wang, Z., Wu, S., Zhang, F. and Liu, R. Geographically neural network weighted regression for the accurate estimation of spatial non-stationarity. *International Journal of Geographical Information Science*, 34:7, 1353-1377, 2020. DOI: 10.1080/13658816.2019.1707834

Wu, S., Wang, Z., Du, Z., Huang, B., Zhang, F. and Liu, R. Geographically and temporally neural network weighted regression for modeling spatiotemporal non-stationary relationships. *International Journal of Geographical Information Science*, 1-27. 2020. DOI: 10.1080/13658816.2020.1775836

Tobler, W. R.: A computer movie simulating urban growth in the Detroit region, *Economic geography*, 46, 234-240, 1970.

Thanks again for your comments.