Adaptive Spatio-Temporal Interpolation Methods

Jun Gao   **Peter Revesz**

Department of Computer Science and Engineering
University of Nebraska-Lincoln
Lincoln, NE 68588, USA
E-mail: {jgao, revesz}@cse.unl.edu

Abstract

We propose a new adaptive spatio-temporal interpolation method that combines either by a step or a line function existing spatial and temporal interpolation methods. We test the new method using climate data obtained from weather stations in Colorado and Nebraska, for the time period from 1993 to 2003. The experimental results show that in mountainous regions our adaptive spatio-temporal method has a much better performance than both the IDW (Inverse Distance Weighting) and the temporal interpolation methods have in themselves.

Keywords: spatio-temporal interpolation, step function, line function

1. Introduction

Geographic Information System (GIS) applications often require spatio-temporal interpolation of an input dataset, that is, to estimate the unknown values at unsampled location-time pairs. For example, suppose we have some weather stations that have incomplete recording of temperature for some time instances, and our application requires a sequential complete dataset, then we need interpolation to estimate the temperature at the unsampled time instances.

A key issue is the choice of an appropriate interpolation method for a given input data set [1]. In climatology, trend surface analysis [11], IDW [8, 9], splining [3], kriging [2] and shape functions [6] are common methods. Interpolation methods are closely related to visualization techniques and have an increasing presence in advanced scientific databases [7].

Climatology researchers mainly use spatial interpolation methods like IDW without any temporal interpolation method. However, in some situations, spatial interpolation methods are not accurate enough. For example,

(1) In mountainous regions, the assumptions used by the IDW method (see Section 2) do not hold.
(2) Some weather station may not have enough nearby stations for estimation, while the assumption of IDW is based on enough close stations.
(3) Several nearby stations have data for the same time instance, and spatial methods can be used for the estimation, but the estimation accuracy is poor. For example, if we define "nearby" as within 50 miles, but all the nearby stations are between 45 to 50 miles, then the accuracy will be poor.

In this paper we propose a novel spatio-temporal interpolation method. Our main idea is the recognition that temporal methods can be useful in combination with spatial methods in the regions where spatial methods can not work well in themselves.

The rest of the paper is structured as follows. Section 2 gives background of inverse distance weighting. Section 3 describes our adaptive spatio-temporal interpolation method. Section 4 describes our experimental methods. Section 5 discusses the evaluation of the experimental results. Finally, Section 6 presents some ideas for future work.

2. Inverse Distance Weighting

Distance-based weighting methods have been used to interpolate climate data by Legates and Willmont [5], Stallings et al. [10] and others. The main assumption of IDW is that values closer to the unsampled location are more similar to the value to be estimated than values from further away. It is consistent with most spatial data. For example, the maximum and minimum temperatures of one day have their values vary continuously and tend to be more similar to closer locations than farther ones. Hence in order to estimate a value for a particular weather station, the closer the station with known value the more weight it has on the prediction.

The sum of the weights is equal to 1. Weights are assigned proportional to the inverse of the distance between the sampled and unknown weather stations. Hence the larger the distance between sampled and unknown points, the smaller the weight given to the value at the sampled point.

Let λ_i = the weights for the individual locations, and y_i = the variables evaluated in the sampled locations.

IDW interpolations are of the form [4]:

$$y = \sum_{i=1}^{N} \lambda_i y_i \qquad (1)$$

$$\lambda_i = \frac{(\frac{1}{d_i})^p}{\sum_{k=1}^{N} (\frac{1}{d_k})^p}.$$

For simplicity we choose p = 1. Hence,

$$\lambda_i = \frac{\frac{1}{d_i}}{\sum_{k=1}^{N}(\frac{1}{d_k})} \quad (2)$$

3. The New Method

Let E_s be the estimated value using spatial method, E_t the estimated value using temporal method, α the weight of E_s, and β the weight of E_t. We calculate the overall estimation as follows:

$$E = \alpha * E_s + \beta * E_t \quad (3)$$

where $\alpha + \beta = 1$ and $0 \leq \alpha, \beta \leq 1$

For example, in case (1) of Section 1, since spatial method can not work, $\alpha = 0$ and $\beta = 1$. On the other hand, if we do not use temporal method at all, then $\alpha = 1$ and $\beta = 0$. These are the extreme cases.

When we consider this method, the most natural function is a step function as shown in Figure 1. In a step function, we find some threshold θ, and on one side of θ we use IDW with $\alpha = 1$ and $\beta = 0$, and on the other side of θ we use a temporal method with $\alpha = 0$ and $\beta = 1$.

Let M_i be the absolute difference between IDW estimation value and the original data, M_t be the absolute difference between temporal estimation value and the original data, and σ be the standard deviation of the elevations of the target station and its neighbors. For example, for some θ, if most stations with $\sigma < \theta$ have smaller M_i, while most stations with $\sigma \geq \theta$ have smaller M_t, then a step function works as follows:

$$\begin{cases} \alpha = 1, \ \beta = 0 & \text{if } \sigma < \theta \\ \alpha = 0, \ \beta = 1 & \text{if } \sigma \geq \theta \end{cases} \quad (4)$$

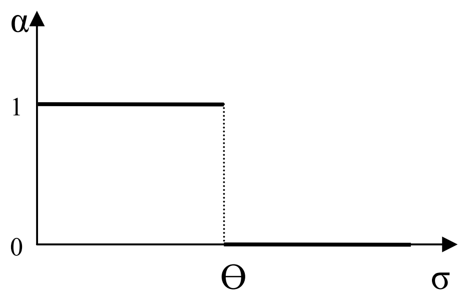

Fig. 1. Step function.

As we state in Section 5, although this natural function has a straightforward intuition, the performance is not as good as for a line function. σ increases, the neighbors are less close to the target station, and we should decrease α. A line function satisfies this intuition by making α vary inversely with σ as follows:

$$\alpha = \begin{cases} 1 - \sigma * \frac{1 - r}{\theta} & \text{if } \sigma < \theta \\ r & \text{if } \sigma \geq \theta \end{cases} \quad (5)$$

where r is a rate constant between 0 and 1. It is illustrated in Figure 2.

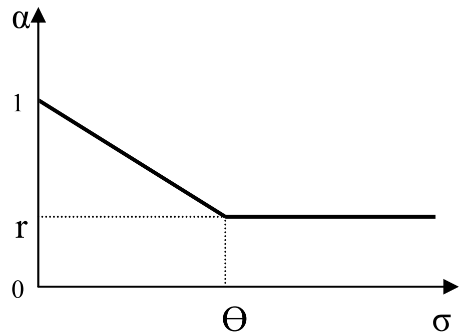

Fig. 2. Line function.

4. Experimental Method

In order to test our idea, we randomly selected weather stations in Colorado and Nebraska and used the minimum daily temperature data of the time period from 1993 to 2003. We estimated the minimum daily temperature using our new method and compared it with the actual data.

The first step is the spatial interpolation. First we choose the closest five stations for each interpolated station. Then we calculate the spatial interpolation value using the IDW method.

The second step is the temporal interpolation. The calculation is similar to the first step except that the distance here is the time distance. For example, if we want to estimate the minimum temperature of one day in 2002, then the distance between that day in 2002 and the same day in 2003 is 365 days.

Once we get the spatial and temporal interpolation values, we apply equation (3) to calculate the final estimation value. In this experiment, we tested both step and line functions to find the best estimation parameters α, β, θ and r.

5. Experimental Results and Evaluation

Several measures are suitable for experimentally comparing the accuracy of interpolation methods. We use mean-absolute-error (MAE) and root-mean-square-error (RMSE).

$$MAE = \frac{\sum_{i=1}^{N} |F(i) - A(i)|}{N}, \qquad RMSE = \sqrt{\frac{\sum_{i=1}^{N} (F(i) - A(i))^2}{N}}$$

Where

$F(i)$: Prediction value,

$A(i)$: Actual measurement,

N : Number of data.

5.1. Evaluation of step functions

Figures 3-5 give the intuition for a step function, which is based on 100 random stations in Colorado. We randomly chose one or two daily minimum temperature for each station. The x-axis is the standard deviation in each figure, while the y-axis is M_i in Figure 3, M_t in Figure 4, and their weighted linear combination in Figure 5. It can be seen that for this data set, 500 seems a reasonable threshold because most stations with $\sigma \leq 500$ have $M_i \leq M_t$.

In order to test the performance of step functions, we compared their MAE and RMSE with those of IDW and temporal estimation methods. Besides 500, we also tried other threshold values (100, 150, …, 1450, 1500). In this experiment, we estimated the minimum daily temperature of 50 stations in Colorado, from May to August 2002. In Table 1, the MAE and RMSE columns summarize the analysis of the various methods.

Compared with line functions in Section 5.2, the performance of step functions is poor. Clearly, IDW and temporal methods have some advantages, and line functions combine those advantages better than step functions do.

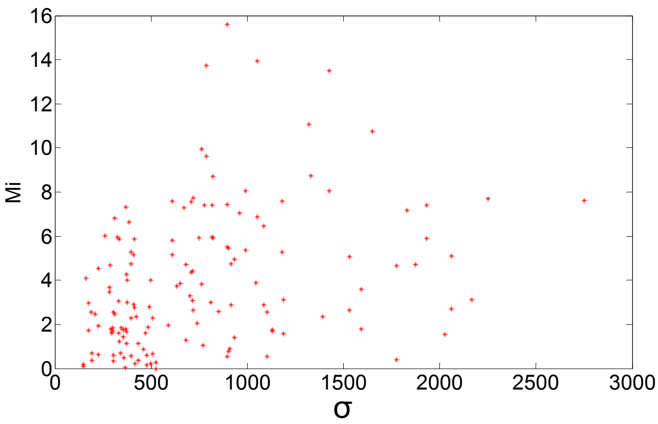

Fig. 3. M_i of step function.

5.2. Evaluation of line functions

In order to test the performance of line functions, we did experiments on 40 threshold values (100, 200, 300, …, 4000) and 10 rates (0.0, 0.1, …, 0.9), and recorded the best combination of those parameters and results in Table 2. We can see from Table 2 that line functions yield much better performance than both the IDW and the temporal methods in themselves. The IDW method has 21% less accurate MAE than the best line function. The temporal method has 39% less accurate MAE than the best line function. Similarly, the IDW method has 15% and the temporal method has 42% less accurate RMSE than the best line function has.

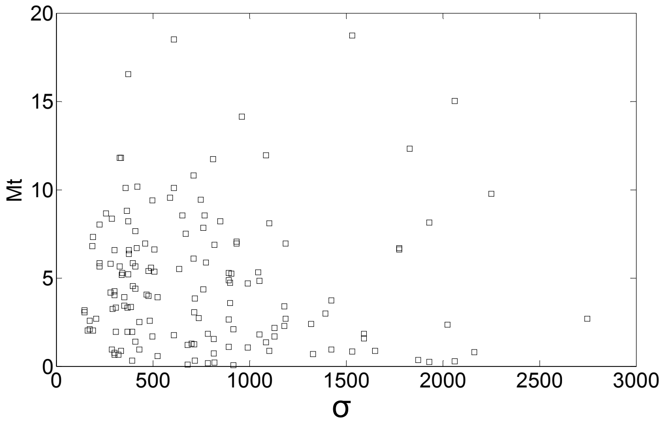

Fig. 4. M_t of step function.

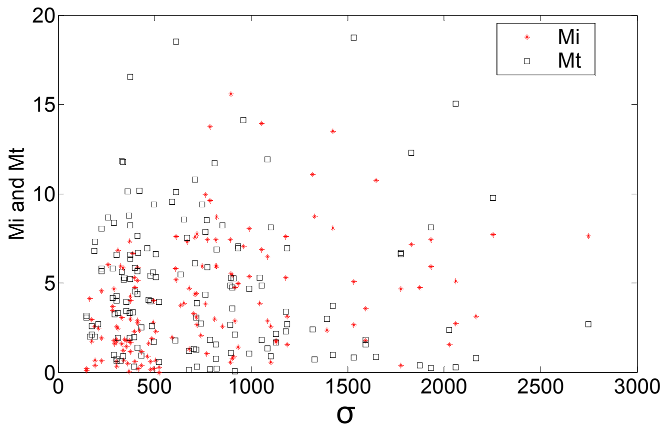

Fig. 5. M_i and M_t of step function.

Table 1. Comparison of step function, IDW and temporal methods

	Best parameters	MAE	Method's MAE / Best method's MAE	RMSE	Method's RMSE / Best method's RMSE
Step Function	Θ = 950	3.8452	1.00	4.6988	1.00
IDW	N = 5, p = 1	4.1958	1.09	4.8912	1.04
Temporal		4.8114	1.25	6.0262	1.28

Table 2. Comparison of line function, IDW and temporal methods

	Best parameters	MAE	Method's MAE / Best method's MAE	RMSE	Method's RMSE / Best method's RMSE
Line Function	r = 0.3, Θ = 1400	3.4598	1.00	4.2586	1.00
IDW	N = 5, p = 1	4.1958	1.21	4.8912	1.15
Temporal		4.8114	1.39	6.0262	1.42

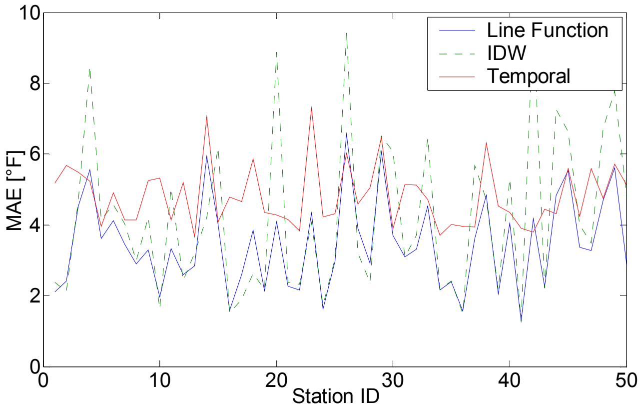

Fig. 6. MAE of 50 Colorado stations.

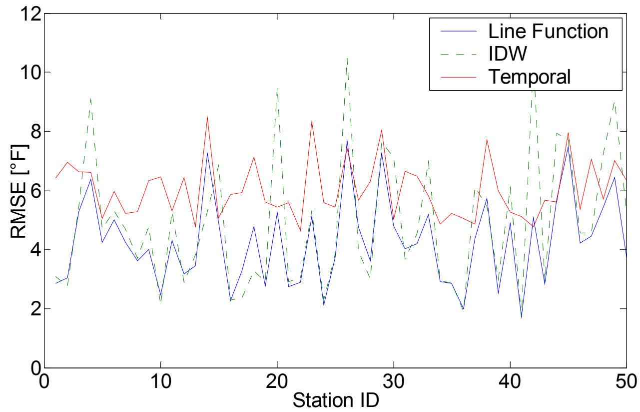

Fig. 7. RMSE of 50 Colorado stations.

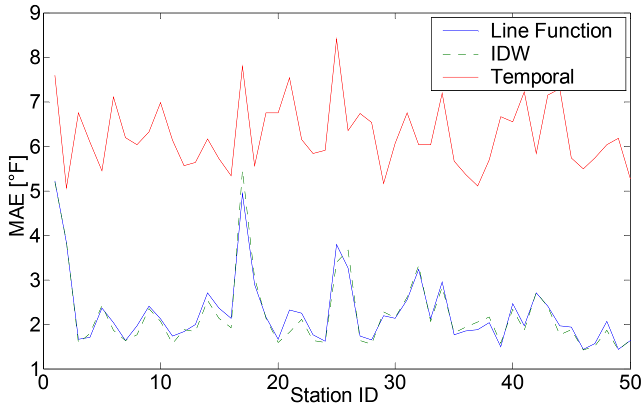

Fig. 8. MAE of 50 Nebraska stations.

The MAE and RMSE of each weather station in Colorado are shown in Figure 6 and Figure 7, respectively. Virtually all economic sectors and many public and private activities are affected in some measure by changes in weather and climate. Hence applying our adaptive spatial-temporal interpolation method to achieve more accurate weather and climate predication is a matter of considerable economic and social significance.

We did experiments on Nebraska weather stations too and show the result in Figure 8. In this case the IDW method yields the best performance among the three tested methods. This result is not surprising in a plain area like Nebraska, because the weather stations in plain have a better chance of having a close neighbor with the similar height than weather stations in mountains have.

6. Conclusion and Future Work

Given the experimental results above, we conclude that in mountainous regions, our adaptive spatio-temporal interpolation method has a much better performance than traditional spatial interpolation and temporal interpolation methods. In the future, we plan to apply our method to other climatic variables like precipitation, mean temperatures, and many others. We also plan to look into other spatial methods like polynomial regression interpolation, and developing spatio-temporal methods based on regression.

7. References

[1] P. A. Burrough and R. A. McDonnell. *Principles of Geographical Information Systems*. Oxford University Press, 1998.

[2] C. V. Deutsch and A. G. Journel. *GSLIB: Geostatistical Software Library and User's Guide*, 2^{nd} ed., Oxford University Press, 1998.

[3] J. E. Goodman and J. O'Rourke, editors. *Handbook of Discrete and Computational Geometry*. CRC Press, New York, 1997.

[4] K. Johnston, J. M. V. Hoef, K. Krivoruchko, and N. Lucas. *Using ArcGIS Geostatistical Analyst*. ESRI Press, 2001.

[5] D. R. Legates and C. J. Willmont. Mean seasonal and spatial variability in global surface air temperature. *Theoretical Application in Climatology*, 41:11-21, 1990.

[6] L. Li and P. Revesz, Interpolation Methods for Spatiotemporal Geographic Data. *Journal of Computers, Environment, and Urban Systems*, 28(3): 201-227, 2004.

[7] P. Revesz. *Introduction to Constraint Databases*, Springer, New York, 2002

[8] P. Revesz and L. Li. Constraint-Based Visualization of Spatiotemporal Databases. In M. Sarfraz, editor, *Advances in Geometric Modeling*, pages 263-276. John Wiley Inc., 2003.

[9] D. A. Shepard. A two-dimensional interpolation function for irregularly spaced data. In *Proc. 23nd National Conference ACM*, pages 517-524, 1968.

[10] C. Stallings., R. L. Huffman, S. Khorram, and Z. Guo. Linking Gleams and GIS. In *Proc. American Society of Agricultural Engineers*, 1992.

[11] E. G. Zurflueh. Applications of two-dimensional linear wavelength filtering. *Geophysics*, 32:1015-1035, 1967.