Journal of Hydrology 508 (2014) 410–417

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

On using smoothing spline and residual correction to fuse rain gauge observations and remote sensing data

Chengcheng Huang [a], Xiaogu Zheng [a,*], Andrew Tait [b], Yongjiu Dai [a], Chi Yang [a], Zhuoqi Chen [a], Tao Li [a], Zhonglei Wang [a]

[a] *College of Global Change and Earth System Science, Beijing Normal University, Beijing 100875, China*
[b] *The National Institute of Water and Atmospheric Research, Wellington, New Zealand*

ARTICLE INFO

Article history:
Received 1 April 2013
Received in revised form 13 November 2013
Accepted 14 November 2013
Available online 21 November 2013
This manuscript was handled by Corrado Corradini, Editor-in-Chief, with the assistance of Okke Batelaan, Associate Editor

Keywords:
Thin-plate smoothing spline
Residual correction
Gauge observation
Cressman weight
Remote sensing data

SUMMARY

A new approach is presented to construct daily gridded precipitation fields with high spatial resolution by fusing gauge precipitation observations and existing gridded precipitation, including remote sensing precipitation products and reanalysis data. The approach comprises of the following two steps: first, gauge observations are used as the response variable, and a bivariate thin-plate smoothing spline and an existing gridded precipitation field are used as explanatory variables, to estimate the precipitation trend surface which is better than using the gridded precipitation field only; then the Cressman weight is modified and applied to correct the correlated residual field to ensure the interpolated precipitation is close to observations. An approach for estimating the error covariance matrix of the interpolated precipitation field is also provided.

An observed daily precipitation dataset from New Zealand is then applied to validate the proposed approach. The results suggest that the proposed interpolation approach can produce precipitation surfaces with high spatial resolution and smaller interpolation errors in both data sparse and data dense areas.

1. Introduction

Daily gridded precipitation data with high spatial resolution are required to drive distributed hydrological models (Xie et al., 2007). An estimation of precipitation based solely on gauge networks is hardly optimal (Haberlandt, 2007). Recently, satellite data have been frequently used in precipitation modeling (Arkin and Xie, 1994; Chen and Dudhia, 2001; Robock et al., 2003; Xie and Arkin, 1995; Yu et al., 1999). Despite the high spatial resolution of satellite data, there is often a large space–time variable bias in satellite precipitation estimates. To improve the quality of the estimation of daily precipitation, a better strategy is to combine gauge observations and satellite precipitation by applying reliable interpolation methods (Adler et al., 2003; Janowiak and Xie, 1999; Shen et al., 2010; Xie and Arkin, 1997; Xie and Xiong, 2011; Jones et al., 2009).

Basically, there are two approaches for merging the gauge observations and satellite precipitation data. The first approach is to fit a partial thin-plate smoothing spline model to the gauge data with the satellite data as a covariate, and applying the least cross validation principle to estimate the precipitation field (e.g. Basher and Zheng, 1998). An advantage of this approach is that the error in the data sparse area is minimized, because the least cross validation principle is applied (Wahba, 1990). However, fitting a smoothing spline could cause the residuals to be significantly larger than the range of the observation error. In other words, the estimation error in data dense areas is too large. In this situation, the interpolated surface often looks physically unreasonable in data dense areas.

The second approach is to treat the satellite precipitation as a trend surface (also called the 'background' or 'first guess') and to apply the Simple Kriging method to correct the correlated residual field (e.g. Chen et al., 2002; Kyriakidis et al., 2001). An advantage of this approach is that the fitted precipitation values are close to observations as a result of residual correction. That is, the estimation error in data dense areas can be satisfactory. However, the error in the data sparse area is usually larger than that derived by the first approach, because the satellite precipitation is a poorer trend surface than that produced by the first approach. Another disadvantage is that, in Simple Kriging theory, the residuals are assumed to be Gaussian (Martin and Simpson, 2005), but this is generally not the case for daily precipitation data.

In this paper, a new approach is proposed by modifying the second approach described above. First, the trend surface is changed from the satellite precipitation to the surface constructed by partial

* Corresponding author. Tel.: +86 10 5880 7541.
E-mail address: x.zheng@bnu.edu.cn (X. Zheng).

0022-1694/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jhydrol.2013.11.022

smoothing spline. Then the traditional Cressman weight is modified and applied to correct the correlated residual field. Since the prediction error of the trend surface generated by partial thin-plate smoothing spline with the satellite precipitation as a covariate is generally less than that of the satellite precipitation only, the error of the trend surface is reduced. Besides, the residual correction using Cressman weight does not require Gaussian residuals, so it could be a better method to correct correlated residuals of precipitation data than the Simple Kriging correction. Finally, since the error variance of the spline can be estimated, it can be used to estimate the error covariance matrix of the corrected precipitation field.

In this study, interpolation approaches are validated using observations only, because the true precipitation is only available at the observation network. The selected interpolation approach is then applied to construct a precipitation field with very high spatial resolution (for example $0.01^\circ \times 0.01^\circ$), such that the observed precipitation intensity at a gauge site is close enough to the true precipitation intensity in the grid box which contains the gauge site. Finally, the interpolated precipitation fields can be scaled to lower resolutions, for example $0.05^\circ \times 0.05^\circ$.

The proposed approach is demonstrated using a New Zealand daily precipitation dataset. The result shows that the proposed approach is indeed better than the other two, in the sense of smaller errors in both data dense and sparse areas.

This paper is arranged in 6 sections. The data and methodology used in this study are documented in Sections 2 and 3 respectively. The main results of the application of this methodology are stated in Section 4. The discussion and the conclusions are included in Sections 5 and 6 respectively.

2. Data

Daily precipitation data at about 650 rainfall gauges (see Fig. 1 for their positions) over New Zealand from 2007 to 2009 were supplied by the National Institute of Water and Atmospheric Research of New Zealand. For computational efficiency, the dataset was divided into two regions: the North Island and the South Island (see Fig. 1 for more details).

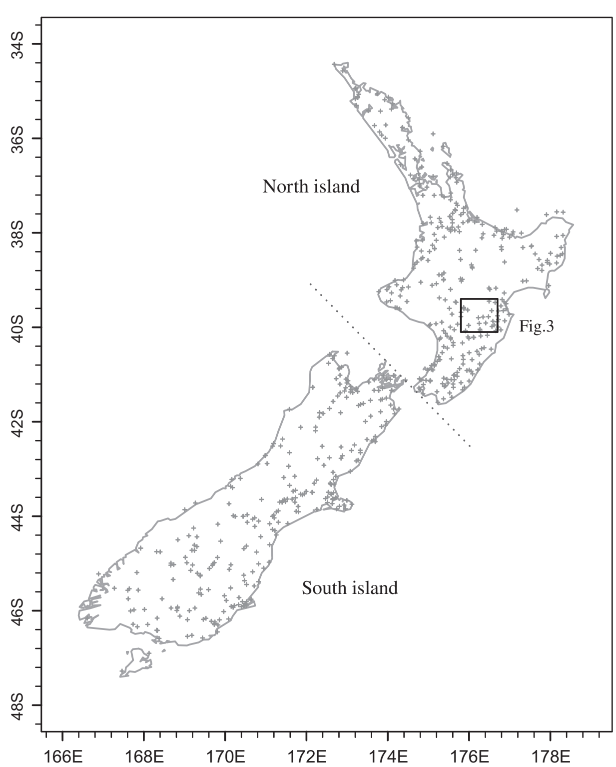

Fig. 1. Location of rainfall gauges and two regions in New Zealand.

The NOAA CPC Morphing Technique data (CMORPH, with a resolution of 0.25° and 3-hourly, Joyce et al., 2004) were also used in this study. CMORPH is a method that produces global precipitation estimates from passive microwave and infrared data. CMORPH exhibits excellent skill in depicting the spatial patterns of precipitation, especially those associated with orographic effects (Xie et al., 2007).

3. Methodology

3.1. Trend surface

In this study, the thin-plate smoothing spline (e.g. Hancock and Hutchinson, 2006; Wahba, 1990; Zheng and Basher, 1995) is used for estimating a precipitation trend surface, with CMORPH precipitation as a covariate:

$$p(x) = f(x) + \alpha \cdot C(x) + \varepsilon(x) \tag{1}$$

where x is a location in a region, p is the observed precipitation at a time step, $C(x)$ is the CMORPH data for the grid which contains x, f is a bivariate thin-plate smoothing spline (see Eq. (2.4.9) in Wahba, 1990), α is a regression coefficient, and ε is the residual with spatially invariant variance σ^2 .

For every day, the coefficients of the spline f, α and the residual variance σ^2 are estimated by fitting Eq. (1) to the daily precipitation data. The routine "gam" in the R package "mgcv" (Wood, 2001; Wood and Augustin, 2002) is used to achieve this goal. Once these parameters are estimated, the estimated value of precipitation at any location x is also derived by using the routine "predict" in the R package "mgcv".

3.2. Corrected fields

Residuals estimated by the partial thin-plate smoothing spline often fall outside the range of the observation error, which is considered as the major disadvantage of smoothing spline approaches. In order to improve the precipitation interpolation accuracy near gauge sites, residual correction may be necessary.

In this paper, the Cressman weight (Stephens and Stitt, 1970) is modified for the correction of the trend surface. For more details, the fitted precipitation at any site x is corrected to:

$$p_a(x) = p_b(x) + \sum_j (p_o(x_j) - p_b(x_j)) \cdot (w_j^2(x) \Big/ \sum_k w_k(x)) \tag{2}$$

where p_o is the observed precipitation and p_b is the estimated trend surface by Eq. (1); the subscripts j and k run over the gauge sites around the site x within the influence radius R; w_j is defined as:

$$w_j(x) \equiv \begin{cases} \frac{R^2 - d_j^2(x)}{R^2 + d_j^2(x)} & 0 \leqslant d_j(x) < R \\ 0 & d_j(x) \geqslant R \end{cases} \tag{3}$$

where $d_j(x)$ is the Euclidian distance between sites x and x_j. It is worth mentioning that, in order to improve the smoothness and accuracy of the corrected precipitation field, the weight in Eq. (2) is chosen as $(w_j^2(x)/\sum_k w_k(x))$ (e.g. Zhang et al., 2009) rather than $(w_j(x)/\sum_k w_k(x))$ (e.g. Stephens and Stitt, 1970). More detailed discussions are documented in Appendix A. The error statistics for corrected precipitation are documented in Appendix B (refer to Fig. 2 for the flow chart for constructing the daily precipitation field).