Spatial and temporal rainfall approximation using additive models

C. Zoppou[*]   S. Roberts[†]   M. Hegland[†]

(Received 7 August 2000)

Abstract

We investigate the approximation of Rainfall data using *additive models*. In our model, space and elevation are treated as the predictor variables. The multi-dimensional approximation problem is demonstrated using rainfall data collected by ACTEW Corporation.

[*]CSIRO Land and Water, Canberra, Water Division, ACTEW Corporation, Canberra, AUSTRALIA.

[†]Research School of Information Science and Engineering, Australian National University, Canberra, ACT0200, AUSTRALIA.

Contents

1   Introduction

Rainfall is a fundamental component of any water resource assessment strategy. Many catchment authorities have implemented expensive and elaborate rainfall monitoring networks to capture the spatial and temporal variability in the rainfall. Due to financial constraints, rainfall monitoring networks are generally sparsely distributed within a catchment. To obtain information at a specific location in a catchment, either interpolation or extrapolation of the existing data is required.

Unfortunately, elementary techniques are often used to estimate the spatial and temporal distribution of precipitation from point measurements of

rainfall. Many of these approaches only provide a spatial distribution of rainfall from a set of fixed rainfall gauges. The temporal distribution is ignored by averaging the rainfall over a predetermined period at each rainfall gauge. The crudest method in common use for estimating the precipitation over a region, is to plot contours of equal precipitation manually or with the assistance of a structured grids, called *isohyets* [2]. The average precipitation is computed between successive isohyets. This is simply linear interpolation and does not utilise all the available data. It is also difficult to produce isohyets with sparse data, although with the advent of Geographic Information Systems, the task of plotting isohyets has been automated. Consider the average annual rainfall recorded at the 55 rainfall gauges shown in Figure 1 surrounding and within the Australian Capital Territory.

A common interpolation approach is to use *Theissen polygons* or *Voroni polygons* [10], see Figure 2(a). In this approach, areas closest to a rainfall gauge adopt the rainfall recorded at that gauge. This results in constant rainfall regions with discontinuities between regions. In addition, there is no justification in assuming that point rainfall measurements provide reliable estimates of precipitation in the surrounding region [4], [9]. The *inverse distance weighted* method is an alternative approach. The rainfall at any desired location is interpolated from the given data using weights that are based on the distance from each rainfall gauge and the desired location. This approach produces a smooth rainfall distribution along with the undesirable troughs and peaks located at the rainfall gauges shown in Figure 2(b). *Kriging* [1] and *thin plate splines* [3], which have similar desirable statistical properties, are other examples of interpolation techniques. Kriging requires prior esti-
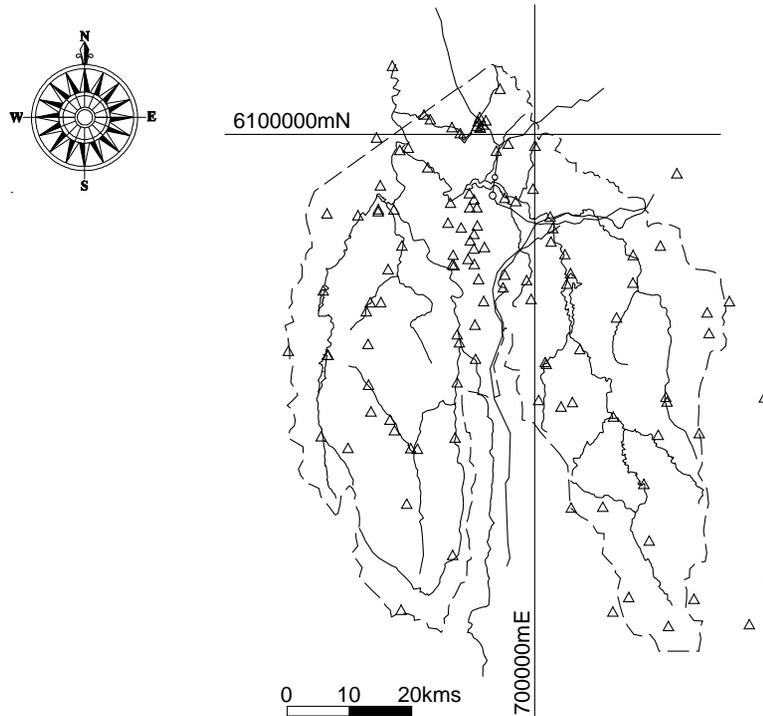
FIGURE 1: The location of the rainfall stations operated by ACTEW Corporation over the period 1870–2000.
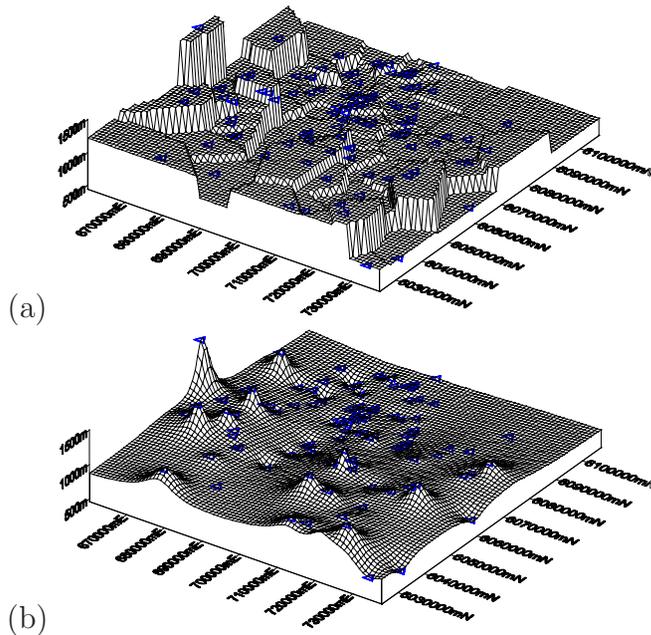
FIGURE 2: Various methods for interpolation rainfall data: (a) Thiessen polygons; (b) inverse distance weighted; (c) Kriging; and (d) finite element thin plate splines (continued next page)

(c)

(d)
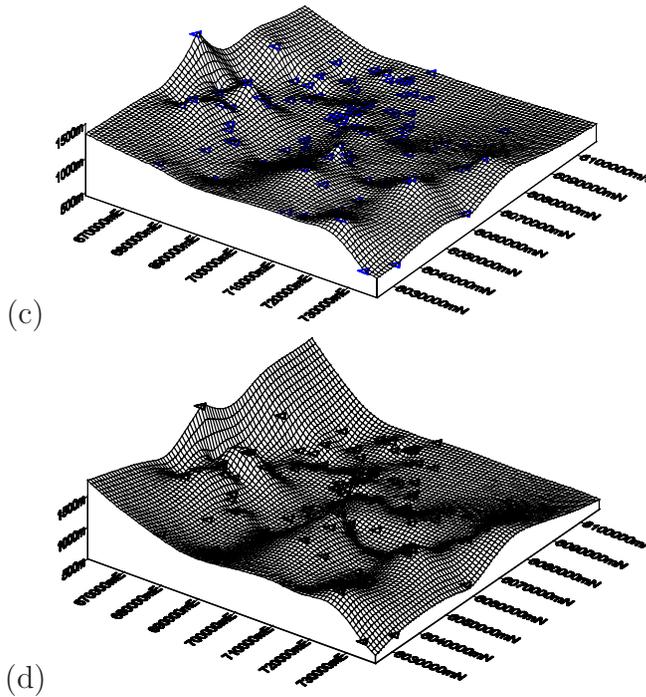
FIGURE 2: (continued) Various methods for interpolation rainfall data: (c) Kriging; and (d) finite element thin plate splines.

mation of the auto-covariance structure, which can be difficult to estimate. Naive implementation of thin plate splines and Kriging do not scale linearly with the number of data points. This has been recently addressed for thin plate splines for one, two and three-dimensional data sets with the development of the *finite element thin plate spline* [7, 6]. However, for moderately higher dimensional data sets (4–5 dimensions) interpolation is difficult. *Additive models* [5] provide a class of models that can be used to approximate multi-dimensional data sets.

2   Additive Models

The additive model assumes that the data has the form $(y^i, x_1^i, x_2^i, \ldots, x_d^i)$ where $i = 1, 2, \ldots, N$ are the N data points, y is the response variable and $\mathbf{x} = (x_1, x_2, \ldots, x_d)$ is the d-dimensional vector space of predictor variables. The form of the additive model with second-order interaction terms is given by

$$f(\mathbf{x}) = \overline{y} + \sum_{j=1}^{d} f_j(x_j) + \sum_{1 \leq j \leq k \leq d} f_{j,k}(x_j, x_k) \tag{1}$$

in which $\overline{y}$ is the mean response. Interpreting $f_j(x_j) = f_{j,j}(x_j, x_j)$ then

$$f(\mathbf{x}) = \overline{y} + \sum_{1 \leq j < k \leq d} f_{j,k}(x_j, x_k).$$

This model assumes that the high dimensional data is well approximated by a sum of low dimensional contributions. The additive model requires the estimation of d one-dimensional functions f_j and $d(d-1)/2$ two-dimensional functions $f_{j,k}$. The original high-dimensional problem has been reduced to a set of coupled one and two-dimensional problems.

The estimation of f_j and $f_{j,k}$ becomes a smoothing problem because the projection of the multi-dimensional data onto lower dimensional space can produce data with large gradients and perhaps discontinuities. The fitting of an interpolant through this data would be difficult. Therefore, a smoothing problem is solved. Since the additive model does not pass through the data points it is not an interpolant, it does however, approximate the multi-dimensional data set by a smoothing function.

Suppose all the functions $f_{l,m}$ except one are known, then the unknown $f_{l,m}$ is a two-dimensional smoothing function of x_j, x_k which fits the data

$$y - \overline{y} - \sum_{l,m \neq j,k} f_{j,k}.$$

Any suitable smoother can be used, such as thin plate or finite element thin plate splines.

3   Application to ACT Rainfall Data

The additive model is used to approximate rainfall data collected by ACTEW Corporation. The location of the 55 rainfall gauges are shown in Figure 1. The rainfall gauges are distributed throughout the Cotter River catchment, to the south west of Canberra and Queanbeyan River catchment, which lies to the south east of Canberra. Reservoirs in these two catchments provide the potable water for the residents of Canberra, with a population of 300,000 and Queanbeyan, to the west of Canberra, which has a population of 26,000.

For the rainfall data, the two spatial dimensions and time are obvious predictor variables. Hutchinson [8] found that when interpolating annual mean rainfall, an appropriately scaled elevation makes a dominant contribution to the accuracy of the fitted surface. Wind direction, which is controlled by topography, can significantly influence the variability in rainfall [4]. All these parameters can be treated as predictor variables when approximating the response variable, rainfall. To avoid the generation of negative rainfall, the logarithm of the response variable could also be used. This would result in a *generalised additive model* [5]. However, taking logarithms of zero rainfall is problematic.

Rainfall recorded during a storm that occurred on the 10th of February 1996 was approximated. All fifty-five rainfall gauges were operating during the storm. For each rainfall gauge, a one-dimensional interpolating function $y^i = f^i(t)$ in which y^i is the rainfall recorded at the $i-$th rainfall gauge and

t is the time, is fitted to the rainfall data. Interpolation will ensure that rainfall is conserved. This data is interpolated at a particular time $t = T$ and an additive model is fitted through the values $y^i(T)$ using the location and elevation of the rainfall stations as the predictor variables.

The additive model is used to interpolate the rainfall data on a 32×32 rectangular mesh covering the catchment. Contour plots, illustrated in Figure 3 were produced from the interpolated rainfall for the February storm. The plots represent the interpolated rainfall over an eight minute interval commencing at 2:00pm on the 10th of February, 1996. In each plot the contour interval is 5 mm/hr with the minimum contour level arbitrarily set to 5 mm/hr. For this event, with the exception of an intense rainfall burst in the southern ACT catchment, most of the rain fell within Canberra. These results demonstrate that the additive model is capable of providing detailed information on the spatial and temporal variability in the rainfall.

4   Conclusions

In this work, the application of an additive model is presented. Its main advantage is that it is relatively simple to include any number of predictor variables that are thought to influence the response variable in the additive model. With an efficient additive model, it is possible to obtain real time simulations of rainfall distributions. These could be used in a rainfall-runoff model to provide more timely and accurate predictions of major flood events.
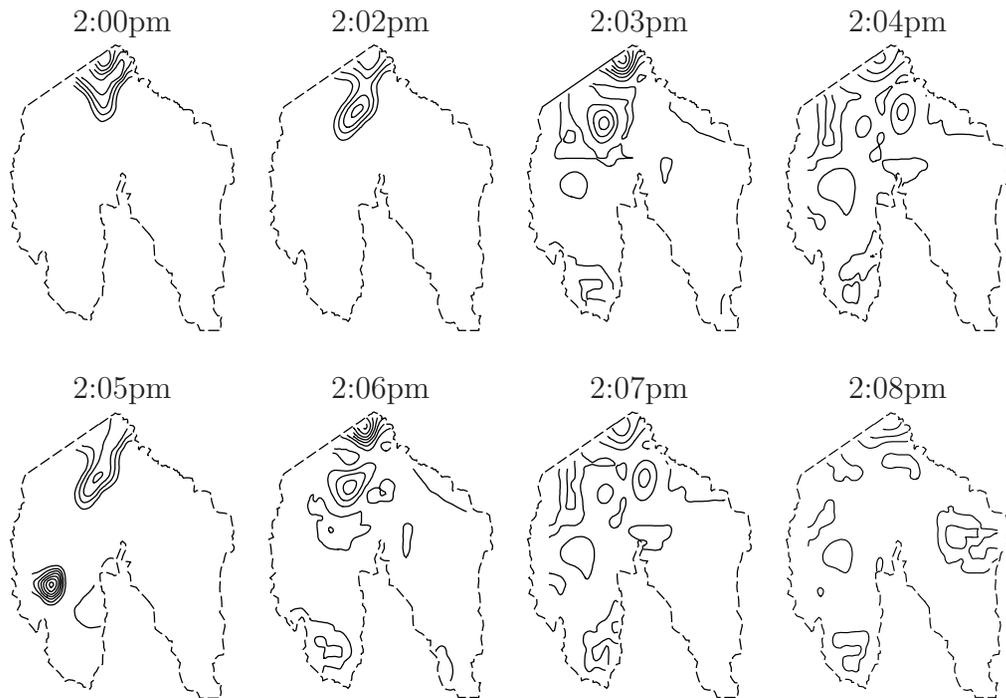
FIGURE 3: Contours of interpolated rainfall using an additive model for the 10th February 1996 storm.

The ability to produce real time simulations have obvious applications in flood forecasting.

References

[1] J. P. Delhomme. Kriging in the hydrosciences. *Advances in Water Resources*, 1(5):251–266, 1978. C1601

[2] S. L. Dingman. *Physical Hydrology*. Prentice-Hall, New York, 1994. C1601

[3] J. Duchon. Splines minimizing rotation-invariance semi-morms in Soblev spaces. *Lecture Notes in Mathematics*, 571:85–100, 1977. C1601

[4] D. C. Goodrich, J-M. Faures, D. A. Woolhiser, L. J. Lane, and S. Sorooshian. Measurement and analysis of small-scale convective storm rainfall variability. *Journal of Hydrology*, 173:283–308, 1995. C1601, C1607

[5] T. J. Hastie and R. J. Tibshirani. *Generalized Additive Models*. Monographs Chapman and Hall, 1990. C1605, C1607

[6] M. Hegland, S. Roberts, and J. Atlas. Finite element thin plate splines for surface fitting. In B.J. Noye, M.D. Teubner, and A.W. Gill, editors, *Computational Techniques and Applications: CTAC97*, pages 289–296, Singapore, 1998. World Scientific Publishing. C1605

[7] M. Hegland, S. Roberts, and I. Altas. Finite element thin plate splines for data mining applications. *submitted* to *Mathematical Methods for Curves and Surfaces II,* Eds. M. Daehlen, T. Lyche, and L.S. Schumaker, 1998. C1605

[8] M.F. Hutchinson. Interpolating mean rainfall using thin plate smoothing splines. *International Journal on Geographical Information Systems*, 9(4):385–403, 1995. C1607

[9] Z. Sen. Average areal precipitation by percentage weighted polygon method. *Journal of Hydrologic Engineering*, American Society of Civil Engineers, 3(1):69–72, 1998. C1601

[10] A.H. Thiessen. Precipitation averages for large areas. *Monthly Weather Review*, 39(7):1082–1084, 1911.

C1601