Thin Plate Smoothing Spline Interpolation of Long Term Monthly Mean Rainfall for the Wet Tropics Region of North-Eastern Australia

P.A. Hancock[a], **M.F. Hutchinson**[b], **S.M. Turton**[c] **and A.L. Lewis**[d]

[a]*School of Tropical Environmental Studies and Geography, James Cook University, Townsville, Qld, 4811, Australia (penlepen@cres.anu.edu.au)*
[b]*Centre for Resource and Environment Studies, Australian National University, Canberra, ACT,0200, Australia*
[c]*School of Tropical Environmental Studies and Geography, James Cook University, Townsville, Qld, 4811, Australia*
[d]*School of Tropical Environmental Studies and Geography, James Cook University, Townsville, Qld, 4811, Australia*

Abstract: This study aimed to interpolate 78 year monthly mean rainfall surfaces with accurate standard error surfaces for the Wet Tropics region using thin plate smoothing spline functions of position and elevation. Two error covariance models were applied to the smoothing spline model to obtain accurate pointwise standard error estimates. These models, a non-diagonal error covariance (NDEC) model and a pre-standardised diagonal error covariance (PDEC) model, were specifically designed to provide accurate estimates of the error covariance structure for rainfall data consisting of rainfall means estimated from records that have missing years within the specified time period. The accuracy of the rainfall surfaces was assessed by comparing surface values with data withheld from the interpolation procedure. The accuracy of the estimated point standard errors was assessed by comparing estimated standard errors with the actual errors of the withheld data. It was found that the accuracy of the interpolated rainfall surfaces for the two models was similar, with the NDEC model performing slightly better in the wet season and the PDEC model performing better in the dry season. The NDEC model performed better in all months in estimating the standard errors for the withheld data. It was therefore the preferred model, although further examination of sound statistical procedures to interpolate rainfall with minimal error is indicated.

Keywords: Rainfall; Australia; Tropics; Thin plate spline interpolation; Standard error

1. INTRODUCTION

There is strong emphasis in recent literature on the need for improved techniques to spatially predict monthly mean rainfall. Numerous past studies have identified monthly mean rainfall as a primary determinant of biotic and abiotic responses [Mackey et al., 1989]. This knowledge has been employed to develop methodologies involving the use of predicted rainfall means to in turn predict a number of environmental characteristics and processes controlled by rainfall, such as vegetation types [Mackey et al., 1989]. These techniques are particularly useful for areas where vegetation and geomorphological characteristics have not been adequately sampled and mapped, as is often the case in protected areas [Mackey et al., 1989]. Furthermore, Hutchinson [1995] states that estimation of the spatial distribution of monthly, seasonal and annual mean rainfall provides a basis for spatially interpolating and simulating actual weather conditions at various time steps.

Mathematical interpolation of rainfall data using thin plate spline functions of position and elevation have been shown in recent studies by Hutchinson [1995] and Price et al. [2000] to be a promising tool for obtaining accurate spatial predictions of rainfall. The study by Hutchinson [1995] developed pointwise estimates of the prediction errors of an elevation dependent thin plate spline fitted to a set of long term annual mean rainfall data obtained from south-eastern Australia. The study reported that the estimated mean rainfall over the total study area had a standard error of just 4% [Hutchinson, 1995].

There is no record in recent literature of the application of thin plate splines with pointwise error estimates to rainfall in tropical regions. Considering that rainfall in tropical regions is known to be more spatially variable than rainfall in higher latitude regions [Jackson, 1978] there is every reason to expect a reduction in the accuracy with which thin plate splines can predict long term mean rainfall in the tropics compared to temperate areas, given a comparable raingauge network. The prediction accuracy of the surfaces produced in this study may be further limited in comparison to the Hutchinson [1995] study, due to the greater spatial and temporal variability of monthly rainfall in comparison to annual rainfall. It was therefore anticipated that, by obtaining pointwise estimates of the prediction error using the techniques developed by Hutchinson [1995], this study would allow a quantitative assessment of the accuracy of thin plate splines as a tool for predicting long term mean rainfall in the Wet Tropics region. This would facilitate the incorporation of the climatic component into management and research of the Wet Tropics World Heritage Area.

2. METHODS

This study used the methods developed by Hutchinson [1995] to interpolate thin plate spline surfaces with pointwise standard errors for Wet Tropics monthly mean rainfall. These surfaces were constructed by fitting thin plate spline functions of position and elevation to 78 year monthly rainfall means estimated from serially incomplete rainfall records obtained from 265 rainfall stations in the Wet Tropics region of north Queensland. The accuracy of the thin plate spline surfaces and their standard error estimates was tested using the validation procedure described in Hutchinson [1995]. Standard error surfaces corresponding to the long term mean rainfall surfaces were then developed. A summary of these procedures is given in the following paragraphs.

2.1. Thin Plate Spline Modeling

Thin plate splines are discussed in detail in Wahba [1990]. The elevation dependent model used here supposes that there are n observed rainfall means given by:

$$r_i = f(x_i, y_i, h_i) + \varepsilon_i \qquad i = 1.....n \tag{1}$$

where r_i is the observed rainfall value, f is an unknown, smooth, continuous function of the longitude x_i, the latitude y_i and the elevation h_i at station i [Hutchinson, 1995]. The error term ε_i is assumed to have zero mean with covariance structure given by:

$$E(\mathbf{e}\mathbf{e}^{\mathrm{T}}) = \mathbf{V}\sigma^2 \tag{2}$$

where $\mathbf{e}^{\mathrm{T}} = (\varepsilon_{1,...,}\varepsilon_n)$ and $\mathbf{V}$ is a known, positive definite $n \times n$ matrix. The matrix $\mathbf{V}$ can be estimated from the data, but σ^2 is usually unknown [Hutchinson, 1995].

The unknown function f in equation (1) is estimated by a thin plate smoothing spline function using the program ANUSPLIN [Hutchinson, 1995].

2.2. Estimating the Error Covariance Structure

It was important in this study was to obtain an accurate estimate of the error covariance structure $\mathbf{V}$ for the Wet Tropics data set, in order to obtain accurate pointwise standard error estimates. For many of the rainfall stations in the Wet Tropics data set, the rainfall records have missing years of record during the specified 78 year period. The means calculated for these stations are therefore only estimates of the true 78 year 'standard period' mean [Hutchinson, 1995]. Given that over 50% of the rainfall stations in the Wet Tropics data set have fewer than 40 years of record, and some have less than 10 recorded years, there is likely to be significant error in estimating the standard period mean at some stations. It has been shown by Hutchinson [1995] that, when dealing with long term means calculated from rainfall records with missing years, the errors in estimating the standard period mean are significantly correlated. It was therefore necessary to account for this correlation when estimating the error covariance structure for the Wet Tropics data set.

Hutchinson [1995] describes two methods for incorporating correlation into the estimate of the error covariance structure. These two models are described here as the non-diagonal error covariance (NDEC) model and the pre-standardised diagonal error covariance (PDEC) model. The NDEC model contains a more complete description of the error covariance structure than the PDEC model, but it is more time consuming to develop. This study applied both of these models to the Wet Tropics data set and compared the accuracy of each by looking at errors of withheld data. A brief summary of the two models is given below.

Non-diagonal error covariance (NDEC) model.
The NDEC model focuses on estimating the non-diagonal error covariance structure **V** in equation 2. According to Hutchinson [1995], the matrix **V** can be decomposed into two components

$$\mathbf{V}=\mathbf{R}+\mathbf{S} \tag{3}$$

R accounts for the correlated errors in estimating the standard period mean induced by missing records. Its elements are given by

$$r_{ij} = \left(\frac{n_{ij}}{n_i n_j} - \frac{1}{78} \right) \rho_{ij} \sigma_i \sigma_j \tag{4}$$

where n_i and n_j are the number of years of record at stations i and j for a given month, n_{ij} is the number of years in common, ρ_{ij} is the Pearson correlation coefficient between monthly rainfall at each location, and σ_i^2 and σ_j^2 are the observed variances of monthly rainfall at each location [Hutchinson, 1995]. To guarantee that **R** is positive definite, the ρ_{ij} were estimated by fitting an exponential function of interstation separation d_{ij} to the observed pairwise correlations [Hutchinson, 1995). The general form of this model is

$$\rho_{ij} = \exp(-d_{ij} / a) \tag{5}$$

where a is a constant factor estimated by linear least squares regression [Hutchinson, 1995].

S accounts for the independent errors in the spatial model given by the smooth function f, assumed to be due to local effects, or microscale effects below the spatial resolution of the data network. Hutchinson [1995] assumes that these errors could be modelled by setting the diagonal elements of **S** to

$$s_{ii} = \sigma_i^2 / m \tag{6}$$

where σ_i^2 is the observed rainfall variance at station i and m is a constant scale factor. The value of m was then estimated by adjusting its value until the minimum GCV model estimate of the unknown σ^2 produced by ANUSPLIN, was found to be 1.0, making the relative error structure **V** absolute [Hutchinson, 1995].

Pre-standardised diagonal error covariance (PDEC) model.
Rather than estimating the non-diagonal error covariance structure, the PDEC model makes use of the strong relationship between the rainfall patterns of station pairs in the data network to estimate the missing records using linear regression [Hutchinson, 1995]. Using this model, every station is regressed with every other station in the network. For each station, the minimum error regression is chosen to estimate the missing records. The estimated missing records are then used in the calculation of the 78 year mean in addition to the original records to give 'pre-standardised means' that incorporate more information from the data network than the observed means. Following the linear regression procedure, it is argued by Hutchinson [1995] that the errors are no longer correlated, because the correlation was originally induced by the missing years in the rainfall records. Given that it is unlikely that all the missing records can be estimated by the regression procedure, the PDEC model is regarded as an approximate method of estimating the error covariance structure.

2.3. Validation

In order to compare the error estimation accuracy of the NDEC and PDEC models for the Wet Tropics data set, this study used the validation procedure described in Hutchinson [1995]. Accordingly, the ANUSPLIN program SELNOT was used to select a spatially representative sample of 38 data points to be withheld from the analysis. The performance of each model was then assessed by comparing the actual root mean square (RMS) residual across the 38 data points with the RMS residual estimated by both the NDEC and PDEC models.

3. RESULTS

3.1. Estimating the Error Covariance Structure

It was found that the calibration of the parametric models of interstation correlation, given by Equation 5, greatly improved if observed pairwise correlations were calculated on a seasonal basis, rather than on a monthly basis. Further significant improvements resulted from the elimination of station pairs with less than 70 years of overlapping record. The resulting exponential curves were found to be $\rho_{ij} = \exp(-d_{ij}/400)$ for the wet season (November to April) and $\rho_{ij} = \exp(-d_{ij}/256)$ for the dry season (October to May). Plots of these models are shown in Figures 1 and 2.

The monthly mean surface characteristics for surfaces produced using the NDEC model were found to be highly sensitive to changes in the expression for the s_{ii}. The reliability of the fitted surfaces, as indicated by the signal values [Hutchinson and Gessler, 1994], was found to improve when the s_{ii} were set to $(\sigma_i^2)^{1/3}/m$ during emperical testing. For the June surface, no value of m could be found for which the ANUSPLIN estimate of σ^2 converged to 1.0 using the original expression, $s_{ii} = \sigma_i^2/m$. Convergence could only be achieved using the cube root transformation on the observed variances. Given the general improvement in surface reliability produced from the cube root transformation, further analysis of the NDEC model was conducted using this transformation for all months.

3.2. Validation

Comparing the actual RMS of the withheld data for both models, shown in Table 1, it can be seen that the NDEC model estimates the data values slightly more accurately than the PDEC model in the wet season, whilst the PDEC model is more accurate in the dry season. With regard to error estimation accuracy, the NDEC model outperformed the PDEC model for almost all months. The NDEC model was therefore chosen to produce rainfall surfaces and corresponding standard error surfaces for the Wet Tropics region. Further analysis of the standard error surfaces produced by the NDEC model revealed that, according to the pointwise standard error estimates, the thin plate spline rainfall surfaces produced using the NDEC model are capable of predicting rainfall throughout the Wet Tropics region with a relatively high level of accuracy. This was particularly the case for the wet season surface, for which 70% of the area was found to have estimated standard errors of les than 17% of the predicted long term mean.

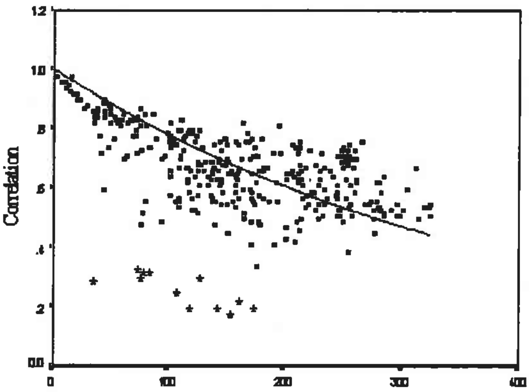

Figure 1. Observed pairwise interstation correlation versus interstation separation for the wet season, using station pairs with 70 or more years of overlapping record.

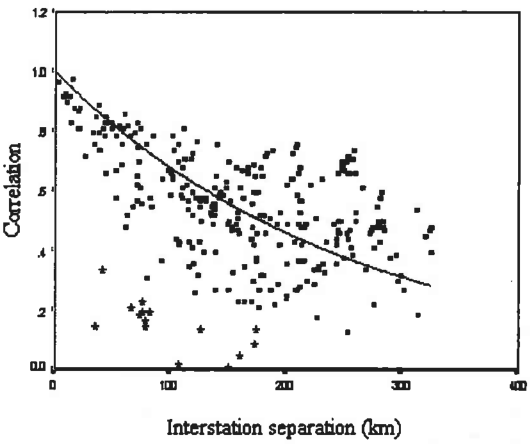

Figure 2. Observed pairwise interstation correlation versus interstation separation for the dry season, using station pairs with 70 or more years of overlapping record.

The dry season surface was found to have higher standard error estimates, with 70% of the area having standard errors less than 46% of the predicted long term mean.

Table 1. Estimated and actual root mean square residuals of 38 data points, withheld from monthly surfaces fitted to the remaining 227 data points using the NDEC model and the PDEC model.

Month	PDEC model		NDEC model	
	Estimated RMS residual	Actual RMS residual	Estimated RMS residual	Actual RMS residual
Jan	154.2	113.3	121.1	108.4
Feb	178.4	96.5	127.3	88.0
Mar	143.2	134.5	122.3	129.3
Apr	73.5	76.3	71.5	84.5
May	70.2	57.3	53.6	66.5
Jun	49.8	23.9	38.6	30.0
Jul	30.1	16.1	29.9	23.4
Aug	29.7	21.1	29.0	35.3
Sep	25.3	19.9	25.9	30.4
Oct	26.0	18.7	24.3	21.3
Nov	61.4	36.2	40.1	35.5
Dec	90.6	57.1	67.7	59.1

4. DISCUSSION

4.1. Estimating the Error Covariance Structure

Figures 1 and 2 demonstrate that temporal correlations can be effectively calibrated when an appropriate time scale is chosen. While past studies of interstation rainfall correlation in the tropics over shorter time periods found that models of interstation correlation did not reveal substantial structure [Jackson, 1974], the plots produced for the Wet Tropics data set show that the noise can be resolved to yield stable structures when long time periods of 70 years or more are considered. Figures 1 and 2 also reveal seasonal differences in the strength of the correlation structure. It is evident that dry season rainfall has a weaker correlation structure than wet season rainfall, demonstrating the more extreme spatial variability of rainfall during the dry season in the Wet Tropics. This is a possible reason for the poorer performance of the NDEC model in estimating withheld data for the dry season.

The improvement in signal values resulting from the cube root variance transformation illustrates the possible need to transform this expression when applying the NDEC model to different data sets. Given the high local variability of rainfall in the Wet Tropics, it might be expected that a variance damping transformation on the diagonal weightings would be appropriate to prevent underweighting of high rainfall values. Further examination of an optimal error structure is warranted.

4.2. Validation

The superior accuracy of the NDEC model in estimating the RMS of the withheld data is consistent with the results obtained by the Hutchinson [1995] study. In the case of the Hutchinson [1995] study, the difference between the two models was not as marked as in this study. For this study, the accuracy of the PDEC model was likely to have been inhibited by its susceptibility to outliers resulting from the linear regression procedure. These outliers arose from short records and poor regression fits. Outliers were found to have a significant influence on the ANUSPLIN estimate of σ^2 used in the calculation of the pointwise standard error estimates [Hutchinson, 1995].

5. CONCLUSION

The results of this study show that thin plate splines are capable of predicting long term mean rainfall in the Wet Tropics with a useful degree of accuracy. This indicates that the 78 year time period considered was long enough to resolve the inherent spatial and temporal variability in the Wet Tropics rainfall patterns, allowing the development of stable spatial models of long term rainfall. The NDEC model is the preferred model for this data set, given that it provided more accurate estimates of the standard errors for the withheld data. However, further examination of the error covariance structure is warranted.

6. ACKNOWLEDGEMENTS

The author gratefully acknowledges the assistance of Dr Jennifer Kesteven in detecting and correcting the many errors in location and elevation of the source precipitation data, Adella Edwards for her help with preparation of the images, and the financial assistance given by the Rainforest CRC Project 2.5 Climate Measuring and Monitoring, and the Department of Tropical Environmental Studies and Geography James Cook University.

7. REFERENCES

Hutchinson, M.F. Interpolating mean rainfall using thin plate smoothing splines, *International Journal of Geographic Information Systems*, 9 (4), 385-403, 1995.

Hutchinson M.F. and P.E. Gessler., Splines – more than just a smooth interpolator, *Geoderma*, 62, 45-67, 1994.

Jackson, I.J., 'Coastal differences in the patterns of variability of tropical rainfall: some characteristics and implications', *Journal of Hydrology*, 38, 273-287, 1978.

Mackey, B.G., H.A. Nix, J.A. Stein and S.E. Cork, S.E., Assessing the Representativeness of the Wet Tropics Queensland World Heritage Property, *Biological Conservation*, 50, 279-303, 1989.

Price, D.T., D.W. Mckenney, I.A. Nalder, M.F. Hutchinson, and J.L. Kesteven, A comparison of two statistical methods for spatial interpolation of Canadian monthly mean climate data, *Agricultural and Forest Meteorology,* 101, 81-94.

Wahba, G., Spline models for observational data, In *CBMS-NSF Regional Conference Series in Applied Mathematics*, Philadelphia: Society for industrial and applied mathematics, 169 pp, 1990.