Aust. Met. Mag. 31 (1983) 1979-184 551.577.32

A new method for estimating the spatial distribution of mean seasonal and annual rainfall applied to the Hunter Valley, New South Wales

M. F. Hutchinson and R. J. Bischof, CSIRO, Division of Water and Land Resources, Canberra

(Manuscript received November 1982; revised March 1983)

An empirical analysis of the spatial distribution of long-term mean seasonal and annual rainfall in the Hunter Valley for the period 1901 to 1975 is presented. Isohyet maps have been derived, using a recently developed method, from Laplacian smoothing spline surfaces fitted to rainfall data at 203 stations. The surface fitting method represents a substantial improvement over previously available subjective, least squares polynomial and weighted interpolation methods. The smoothing spline surfaces are functions of latitude, longitude and elevation with the degree of data smoothing determined by minimising the generalised cross validation (GCV). The method does not require prior standardisation of short-term means to long-term means, affords error estimates and allows testing of the validity of the fitted model. The fitted rainfall surfaces estimate long-term mean annual rainfall at any point in the valley with a percentage root mean square predictive error of 9.2 per cent and estimate long-term mean seasonal amounts at any point with percentage root mean square predictive errors ranging from 11.9 per cent in winter to 18.2 per cent in summer.

Introduction

Isohyet maps so far constructed for the Hunter Valley reflect a progression of methods, from small data sets subjectively analysed, to more explicit methods applied to larger data sets. The maps produced by the New South Wales Premier's Department (1952) and by Tweedie (1963) were obtained simply by plotting rainfall averages on a topographic base and drawing isohyets by freehand interpolation. The annual isohyet map produced by McMahon (1964) used rainfall data from 105 stations standardised to a 50-year period. McMahon developed straight line relationships, where possible, between rainfall and elevation and between rainfall and distance inland from the coast, but relied on freehand interpolation in the southeastern sector where there were too few stations to formulate such relationships.

The rainfall maps presented in this paper have been derived from Laplacian smoothing spline surfaces which express mean annual and mean seasonal rainfall in terms of latitude, longitude and elevation. The surfaces have been fitted to the data using a technique recently developed by Wahba and Wendelberger (1980). Because the number of data points is large, the method has been employed here in its approximate version as indicated in Wahba (1980a,b) and implemented by Hutchinson (1983). The technique takes no account of the physics of rainfall distribution except to assume that rainfall varies smoothly in response to changes in latitude, longitude and elevation. A principal feature of the method is that the degree of data smoothing imposed by each fitted surface has been determined by minimising the predictive error of the surface as measured by generalised cross validation (GCV). Inclusion of elevation as an independent variable gave a significant reduction in predictive error over fitted surfaces which were functions of latitude and longitude only. In addition, the procedure allowed stations to be weighted according to local variance estimates which incorporated length of rainfall record, so that both long and short-term records could be used. This made a much more extensive recording network available for analysis.

The surface fitting technique used in this paper has been described in detail in Wahba and Wendelberger (1980) where it has been applied to simulated irregularly spaced noisy data with impressive results. The technique is sophisticated and requires the use of a computer with a fast memory capacity, in words, exceeding the square of the number of points used to define the surface (Hutchinson 1983). The method is closely related to certain forms of optimum objective analysis methods first presented by Gandin (1965) and cited in Goodin et al. (1979). Unlike the Gandin method however, prior climatologically extracted covariances are not required.

The method is a substantial improvement upon polynomial trend surface analyses of rainfall. Such analyses can adequately estimate areal rainfall from limited networks but persistent patterns in residuals from trend surfaces of different degrees indicate their shortcomings in representing a true rainfall surface over a whole catchment, especially in regions of extrapolation (Edwards 1972; Shulze 1976). Higher degree polynomial surfaces, although capable of fitting larger networks closely, can be erratic in performance in regions away from the data points.

Interpolation methods which involve fitting polynomials or other simple functions to small subregions (Shaw and Lynn 1972; Akima 1978) are complex to implement and tend to fit the data very closely, whether or not this is justified in terms of the amount of noise associated with the data. They can also be unstable in regions where the data network is sparse.

Weighted interpolation methods (Goodin et al. 1979; Lancaster 1979) involve a subjective choice of weighting function which is often defined in terms of a radius of influence beyond which data points are ignored. The degree of data smoothing depends on the choice of weighting function and not directly on the amount of noise associated with the data. In addition, the choice of an optimum radius of influence becomes a problem when the density of the data points varies greatly across the data network. These methods also tend to fit the data very closely.

The Laplacian smoothing spline method works equally well no matter how irregularly the data points are distributed in space and is not subject to erratic behaviour away from the data points. It fits a surface which for a given mean square residual has minimum curvature (Wahba and Wendelberger 1980) so that, away from the data points, the surface tends to be very smooth. Moreover, the method is capable of fitting surfaces closely to extensive data networks. The method provides an estimate of the error associated with surface values at any point in the catchment, as well as providing an estimate of the standard deviation of the data. This standard deviation estimate can be used to test the model in an approximate goodness of fit test if a prior estimate of the standard deviation is available (Wahba 1980a).

Data

The rainfall data used in this study consisted of all available monthly rainfall records between 1901 and 1975 at the 203 Bureau of Meteorology stations for which there were at least five years of record (Fig. 1). Although the surface fitting technique is not constrained by the need for a standard time period of analysis, the rainfall data were analysed for any pronounced time bias in the number of stations records in the period 1901-1975. The distribution of rainfall stations over time (Fig. 2) shows a slight bias towards the period between 1961 and 1975 when 30 per cent of all annual rainfall records were collected.

The spatial distribution of the rainfall stations (Fig. 1) in the valley is influenced by the distribution of population and arable land. Hence the well populated, economically important land of the valley has a higher density of rainfall stations than the remainder of the valley. There are few stations in the rugged high elevation areas on the periphery of the Hunter Valley basin (see Fig. 3).

Fig. 1. Spatial distribution of rainfall recording stations.

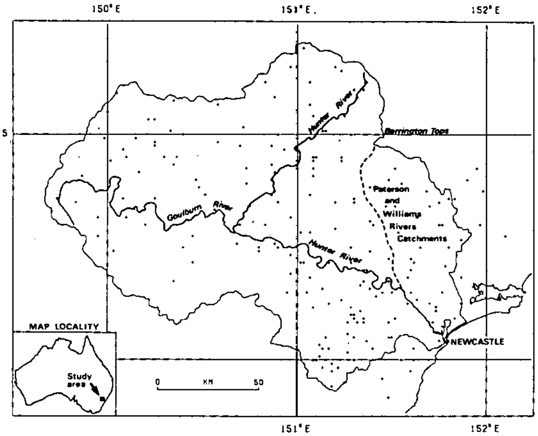
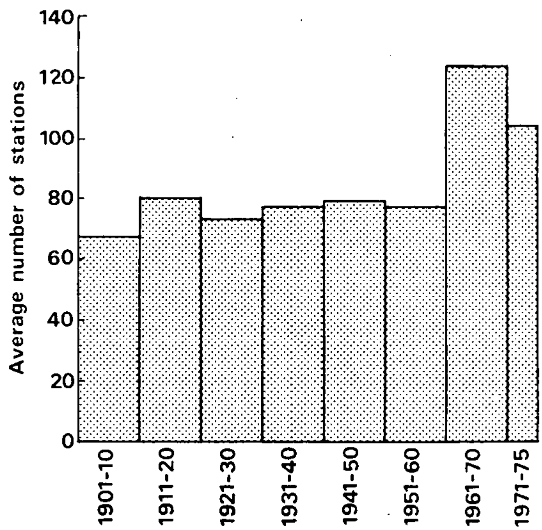


Fig. 2. Distribution of rainfall stations over time.



Surface fitting

A Laplacian smoothing spline function of degrees latitude, degrees longitude and elevation in kilometres was fitted to the mean annual rainfall data set and to each mean seasonal rainfall data set using programs developed by Hutchinson (1983). Stations were weighted according to the assumption that each mean data value is distributed independently with variance σ^2/n where n is the number of years of record and σ is the local standard deviation estimate obtained from the data. Thus mean data values with small standard deviation estimates or long periods of record tended to be fitted more closely than mean data values with large standard deviation estimates or short periods of record. The assumption of independence is strictly true only for stations with non-overlapping periods of record since rainfall data at different stations can be highly correlated. The assumption is not seriously violated in this case because the data contain many shorter-term records obtained at different times throughout the period 1901 to 1975.

The degree of data smoothing imposed by each fitted surface and the scaling of elevation to be in units of km were determined by minimising the generalised cross validation (GCV). The degree of smoothing represents a trade-off between data infidelity, as measured by the mean square residual from the data points weighted according to the variance estimates described above, and surface roughness, as measured by the total curvature of the fitted surface (Wahba and Wendelberger 1980). For a given degree of smoothing the GCV is an estimate of the mean square predictive error of the fitted surface obtained by (implicitly) removing each data point in turn and determining how well a surface fitted to the remaining data points with the given degree of smoothing estimates the missing value.

An error analysis for the fitted surfaces is presented in Table 1. The root mean square predictive errors in this table are given simply by $(GCV)^{1/2}$. Bartlett's homogeneity of variance test applied to the rainfall data transformed by taking logarithms (see §§10.21, 11.17 of Snedecor and Cochran 1967) indicates that the standard deviation of both seasonal and annual rainfall in the Hunter Valley is approximately proportional to the mean. The standard error associated with individual point estimates using the fitted surfaces therefore should not be taken to be the average value for the region as represented by the root mean square predictive error in the first column of Table 1, but rather it should be expressed as a percentage as given by the percentage root mean square error in the second column of Table 1. The root mean square predictive errors are naturally larger than the root mean square residual errors. As might be expected, the annual rainfall values were fitted with smaller percentage root mean square predictive error and with smaller average percentage residual error than the seasonal rainfall values. The standard deviation estimates obtained from the fitted surfaces compare favourably in Table 1 with estimates obtained from the data by taking the square root of the mean of all local data point variance estimates. This gives ground for confidence in the validity of the fitted model (Wahba 1980a).

Mapping

The summer (December-February), winter (June-August) and annual isohyet maps presented in Figs 4, 5, 6 were produced from the fitted surfaces using mean elevation data on a regular 0.02° x 0.02° latitude-longitude grid. The elevation data were obtained by first collecting mean elevation data for the region on 2 km x 2 km grids from 1:25 000 topographic maps. Because these maps belong to different Universal Transverse Mercator mapping zones, an iterative procedure developed by Hutchinson (1983) was used to transform the elevation data to a regular 0.02° x 0.02° grid. The values calculated by the procedure are those of a Laplacian smoothing spline function of latitude and longitude fitted to the height data with a prescribed root mean square residual of 20 m (Fig. 3). Rainfall values were then calculated on the same regular 0.02° x 0.02° grid using the calculated elevation values and the fitted rainfall surfaces. Finally, digitised isohyets were calculated from these rainfall values (Hutchinson 1983) and mapped by the computer program MAPROJ (Hutchinson 1981) on a rectangular projection with standard parallel 32°30'S.

It should be possible to obtain reasonable seasonal and annual rainfall estimates at any point in the Hunter Valley, a region which is the subject of ongoing research (Erskine and Bell 1982; Waterman et al. 1982), by linear interpolation from these maps.

Table 1. Error analysis for the fitted rainfall surfaces.

	Root mean square predictive error (mm)	Percentage of network mean	Root mean square residual error (mm)	Average % residual error	Surface standard deviation estimate (mm)	Data standard deviation estimate (mm)
Summer (Dec-Feb)	50	18.2	28	7.5	37	34
Autumn (Mar-May)	27	13.3	13	4.6	19	33
Winter (Jun-Aug)	20	11.9	10	4.2	14	22
Spring (Sep-Nov)	21	11.7	13	5.6	16	22
Annual	75	9.2	33	3.0	50	53

Fig. 3. Elevation map for the Hunter Valley (m).

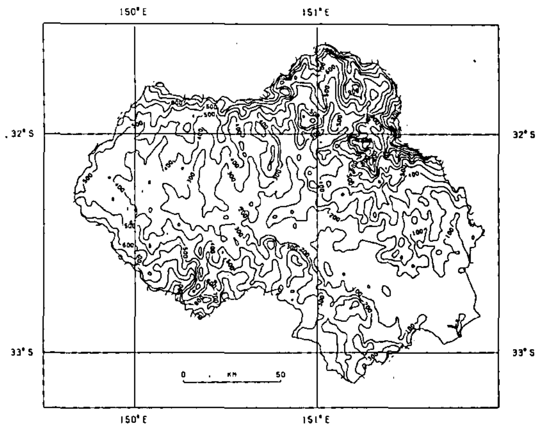


Fig. 4. Summer (December-February) rainfall for the Hunter Valley (mm).

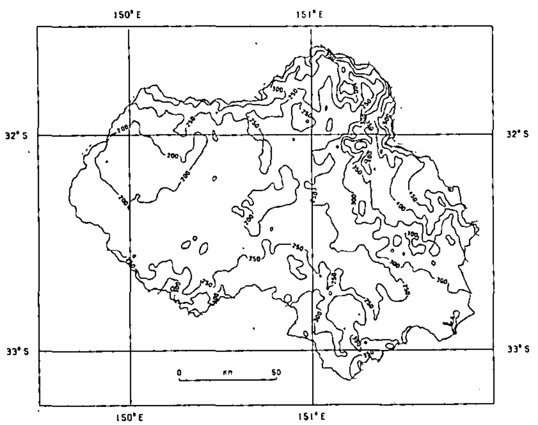
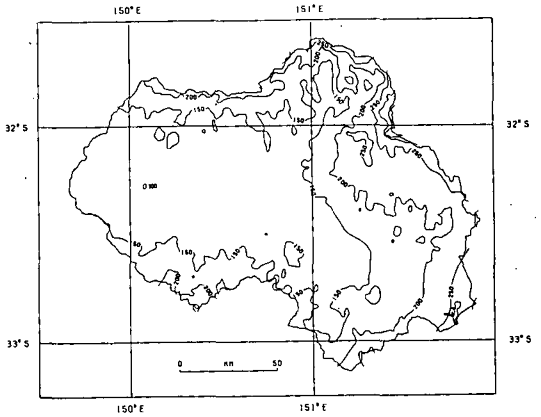


Fig. 5. Winter (June-August) rainfall for the Hunter Valley (mm).



Copies of all seasonal and annual rainfall maps at a scale of 1:250 000 with isohyet intervals of 25 mm and 50 mm respectively are available from the authors. These isohyet intervals approximate the root mean square predictive errors in Table 1. Coefficients of the fitted surfaces, from which it is possible to calculate more accurate rainfall estimates for any point in the valley, are also available from the authors. It would be impractical to present the coefficients in this paper.

Discussion of spatial patterns

The annual rainfall pattern confirms the analysis by Tweedie (1963). It consists of a dry central and western valley core, becoming wetter at differing height gradients until the valley escarpment is reached in the north and the south and becoming wetter towards the coast in the southeast. The wettest area is the Barrington Tops with an average annual rainfall exceeding 1700 mm, while the driest region is the Goulburn River valley with only 550 mm. The complexity of the isohyets in the eastern, northeastern and southern sectors reflects the influence of a rugged complex topography. The relative simplicity of the isohyets in the western sector reflects a less complex topography.

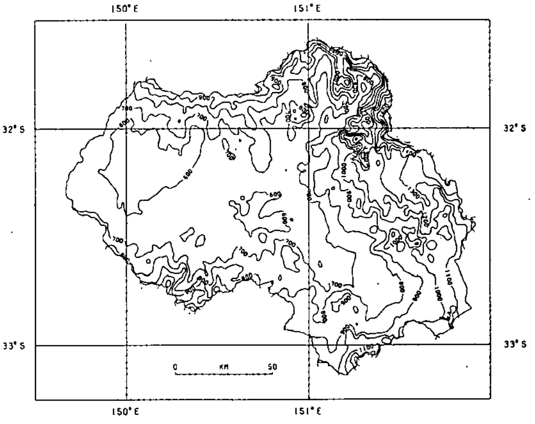
The seasonal maps, as indicated in Tweedie (1963), are broadly similar to the annual map. The stronger influence of elevation evident in summer reflects the inflow of warm saturated air of equatorial origin from the northeast. The flatter distribution in winter reflects more stable conditions. Most of the very severe flood events are due to the southern penetration of tropical cyclones and storms giving rise to heavy rainfall particularly in the upper reaches of the valley (see Gentilli 1971). The coastal influence is stronger in winter in response to the inflow of maritime air masses of moderate temperature, high moisture content and high conditional instability. The autumn and spring maps (not shown here) show intermediate patterns, with autumn conforming more closely to the summer pattern and spring conforming more closely to the winter pattern.

The annual map represented in Fig. 6 is also broadly similar to the map produced by McMahon (1964) but significant differences between the two maps do exist. These range from 250 mm in the Barrington Tops region to 50 mm in the Goulburn River valley. The Barrington Tops is a complex pattern of hills and valleys which affect the rainfall regime. This is reflected in the isohyets of the surface fitted map, but not in McMahon's map. Similar discrepancies can be seen in the isohyet patterns of the Paterson and Williams river catchments.

Validation

Because the surface fitting procedure allows the data points to be weighted according to variance estimates which incorporate length of record, short-period means did not have to be first standardised to long-period means using regressions with rainfall data at nearby long-period stations. This has allowed an extensive network of stations to be included in the analysis (Fig. 1). Moreover, the temporal station

Fig. 6. Annual rainfall for the Hunter Valley (mm).



coverage is reasonably uniform for the period 1901 to 1975 (Fig. 2), so the resulting surface fits to the data should be estimates of the mean rainfall for this period.

This claim has been verified by comparing long-term (1901-1975) estimates with surface estimates of mean annual rainfall at stations with short periods of records in four different areas in the valley (Table 2). Each long-term estimate was obtained from the long-term mean at a nearby station via a linear regression equation developed for the common years of record. Table 2 shows good agreement between the surface estimates and the long-term estimates. Percentage differences are well within the 9.2 per cent annual root mean square predictive error given in Table 1. In most cases the fitted surface value tends toward the long-term estimate obtained from a nearby long-term station.

No attempt has been made to incorporate into the surface fitting analysis long-term trends in precipitation for the Hunter Valley, despite the evidence for a significant increase in summer rainfall after 1945 (Pittock 1975). In view of the good agreement between surface estimates and long-term regression estimates in Table 2, this has not significantly influenced the validity of the fitted surface mean rainfall estimates for the period 1901 to 1975.

Conclusion

The rainfall analysis presented here for the Hunter Valley offers several improvements over rainfall analyses based on subjective, least squares polynomial and weighted interpolation methods. These include:

(a) the analysis is completely objective and explicit;
(b) prior standardisation of records to a given period is not required;
(c) each fitted surface is valid for the entire catchment;
(d) the surfaces are not subject to erratic behaviour away from the data points;
(e) the surfaces afford variance estimates which may be used to check the validity of the model; and
(f) the surfaces afford percentage predictive error estimates which allow the surface values to be used in subsequent applications with a precisely defined degree of confidence.

References

Akima, H. 1978. A method of bivariate interpolation and smooth surface fitting for irregularly distributed data points. *A.C.M. Trans. Math. Software 4*, 148-59.

Edwards, K. A. 1972. Estimating areal rainfall by fitting surfaces to irregularly spaced data. *Proc. International Symposium on the Distribution of Precipitation in Mountainous Area 2.* World Meteorological Organization. No. 326. 565-87.

Erskine, W. and Bell, F. C. 1982. Rainfall, floods and river channel changes in the upper Hunter. *Aust. geogr. Stud. 20,* 183-96.

Gandin, L. S. 1965. *Objective analysis of meteorological fields.* Israel Program for Scientific Translations. Jerusalem. 242 pp.

Table 2. Comparison of surface fit estimates and long-term estimates.

Long-term station no.	Period of record	Short-term station no.	Record mean	Long-term estimate (LTE)	Surface fit estimate (SFE)	Correln coeff. of regressn	Percentage difference SFE and LTE
061002	1901-1975		885	885	880		−0.6
	1961-1968	061157	901	858	869	0.97	1.3
061068	1939-1974		1218	1218	1233		1.2
	1961-1972	061136	1715	1544	1583	0.98	2.5
061070	1901-1968		681	681	676		−0.7
	1961-1975	061143	685	709	675	0.84	−4.8
	1960-1975	061171	641	655	638	0.93	−2.6
	1960-1967	061180	683	653	677	0.80	3.7
062009	1901-1975		607	607	591		−2.6
	1935-1954	061084	566	593	603	0.96	1.7
	1965-1975	062069	594	571	578	0.87	1.2

Gentilli, J., ed. 1971. Climates of Australia and New Zealand. *World Survey of Climatology, Vol. 13, Ch. 5.* Elsevier Publishing Co., Amsterdam-London-New York.

Goodin, W. R., McRae, G. J. and Seinfeld, J. H. 1979. A comparison of interpolation methods for sparse data: application to wind and concentration fields. *Jnl appl. Met., 18,* 761-77.

Hutchinson, M. F. 1981. MAPROJ — A computer map projection system. *CSIRO Div. Land Use Research Tech. Paper 39,* 38 pp.

Hutchinson, M. F. 1983. Surface fitting and contouring programs for small to large data sets. *CSIRO Div. Water and Land Resources Tech. Paper.*

Lancaster, P. 1979. Moving weighted least-squares methods. *Proceedings of the NATO Advanced Study Institute on Polynomial and Spline Approximation, Calgury,* B. Sahney, ed., D. Reidell Publising Co. 103-20.

McMahon, T. 1964. Hydrological features of the Hunter Valley, NSW. *The Hunter Valley Research Foundation Monographs No. 20.* Newcastle, NSW.

NSW Premier's Dept. 1952. *The Upper Hunter region — a preliminary survey of resources.* N.S.W. Govt Printer, Sydney.

Pittock, A. B. 1975. Climatic change and patterns of variation in Australian rainfall. *Search 6,* 498-504.

Schulze, R. E. 1976. On the application of trend surfaces of precipitation to mountainous areas. *Water S.A., 2,* 110-18.

Shaw, E. M. and Lynn, P. P. 1972. Areal rainfall evaluation using two surface fitting techniques. *Bull. IAHS, 17,* 419-33.

Snedecor, G. W. and Cochran, W. G. 1967. *Statistical Methods, 6th ed.* Iowa State University Press. 593 pp.

Tweedie, A. D. 1963. Climate of the Hunter Valley. In General Report on the Lands of the Hunter Valley, pp. 62-80. *Land Research Series No. 8.* CSIRO, Melbourne.

Wahba, G. 1980a. Ill-posed problems: numerical and statistical methods for mildly, moderately and severely ill-posed problems with noisy data. *Proc. International Symposium on Ill-Posed Problems,* Nashed (ed.). Academic Press.

Wahba, G. 1980b. Spline bases, regularization and generalized cross validation for solving approximation problems with large quantities of noisy data. University of Wisconsin, Madison *Statistical Department Technical Report No. 597,* 1-8.

Wahba, G. and Wendelberger, J. 1980. Some new mathematical methods for variational objective analysis using splines and cross validation. *Mon. Weath. Rev., 108,* 1122-43.

Waterman, P., Tebbutt, J., Dawson, P. and Gorrie, I. 1982. *Directory of Resource and Environmental Research for the Hunter Region.* Centre for Resource and Environmental Studies, Australian National University.