Error analysis for the interpolation of monthly rainfall used in the generation of SILO rainfall datasets

Stephen Jeffrey

Climate Impacts and Natural Resource Systems
Queensland Centre for Climate Applications
Department of Natural Resources and Water
80 Meiers Road, Indooroopilly Qld 4068 Australia

27 October 2006

The SILO interpolation system was subject to an error analysis in 1999. The results were published in:

> "Using spatial interpolation to construct a comprehensive archive of Australian climate data", Jeffrey, S.J., Carter, J.O., Moodie, K.B. and Beswick, A.R., Environmental Modelling and Software, 16(4), 309-330 2001.

The preceding paper outlined the algorithm used to generate the interpolated monthly rainfall datasets which are available via the SILO database. The rainfall interpolation system described in that paper has since been upgraded and consequently the error statistics reported for rainfall are no longer representative of the datasets currently provided by SILO.

SILO monthly rainfall datasets are interpolated by Ordinary Kriging observational datasets. The procedure is summarised as follows:

1. Observational data is normalised to minimise seasonal and orographic effects. It has been previously reported[1] that rainfall can be reliably interpolated across a range of timescales when raised to an appropriate power and subsequently normalised:
$$value = \frac{obs.value^{power} - mean}{std.deviation}$$
The procedure used to calculate the normalisation parameters is described below.
2. The normalised data is interpolated using Ordinary Kriging. All stations are independently cross validated.
3. Stations with large residuals (observed - cross validated) are removed from the data set.
4. The revised dataset is re-interpolated and the normalisation reversed.
5. The monthly rainfall surface is generated and interpolation statistics logged.

---

[1] Hutchinson, M., Richardson, C., and Dyke, P. (1993). Normalisation of rainfall across different time steps. In *Management of Irrigation and Drainage Systems*, Volume 9, pp 432-439. Irrigation and Drainage Division, ASCE, US Department of Agriculture.

Interpolating rainfall is problematic due to the fact that rainfall is highly variable. Interpolation accuracy can be improved by reducing the inherent variability in observational datasets by normalising the data. Normalisation effectively removes the variability due to persistent effects such as orography. Accumulated rainfall can be normalised as the frequency histogram of rainfall (raised to an appropriate power) is approximately normal. The power required to achieve an approximate normal distribution may vary both spatially and with respect to the accumulation period. Therefore the parameters required to fit a normal distribution to observational data must be fitted locally and subsequently interpolated. The interpolation is required as the fitted power parameter is typically unstable and requires a significant amount of smoothing.

A Maximum Likelihood procedure has been used to estimate the power required to fit a truncated normal distribution, and also the mean and standard deviation of the fitted distribution. A truncated distribution is fitted as small rainfall amounts are unreliably recorded, so a truncation threshold is used to eliminate these values.

Normalisation parameters have been computed for monthly rainfall. The steps required to compute these values are as follows:

1. The normalisation power is computed for each station using Maximum Likelihood. An individual value is computed for each month. (January, February etc).

   A station must have at least 40 years of data for a distribution to be fitted. The fitted power parameter is constrained to lie in the range [0.4 - 0.6].
2. The power parameters are spatially interpolated using a two-dimensional smoothing spline. Stations with a Kolmogorov-Smirnov goodness-of-fit statistic less than 0.8 are excluded from the interpolation. This reduces the number of stations with fitted parameters from about 6000 to about 1300. The remaining data points are then interpolated with a severe degree of smoothing imposed. Aggressive smoothing is necessary as the resulting spatial variation in the power parameter must be relatively small in relation to the station density. The degree of smoothing of individual data points is controlled through the associated variance estimate: each point's variance is determined by it's deviation from the mean of the 10 nearest neighbours. Large deviations result in high variances and hence, the impact of an individual data point on the surface is minimised.
3. Using the interpolated power parameter, the Maximum Likelihood estimators for the mean and standard deviation are computed for each station. Individual values are computed for each month. (January, February etc).
4. The fitted means and standard deviations are spatially interpolated using a three-dimensional smoothing spline. Smoothing is enforced using the variance procedure described above. In this case however, 20 neighbouring stations are used.

An error analysis has been performed to assess the accuracy of the upgraded SILO interpolation system (as described above). The results (averaged over all stations and dates) are as follows:

Error statistic	1880-2006	2001-2005
Root Mean Square (mm)	20.5	19.1
Mean Absolute (mm)	11.0	10.1
Mean (mm)	-0.80	-0.67
r^2	0.87	0.90