Spatial Simulation of Rainfall Data for Crop Production Modeling in Southern Africa

JA Wright, Institute of Ecology & Resource Management, University of Edinburgh,
J Smith, Scottish Agricultural College, Edinburgh,
SW Gundry, Institute of Ecology & Resource Management, University of Edinburgh,
C Glasbey, Biomathematics and Statistics Scotland.

Abstract: This paper describes a methodology for simulating rainfall across a set of spatial units in areas where long-term meteorological records are available for a small number of sites only. The work forms part of a larger simulation model of the food system in a district of Zimbabwe. The simulation includes a crop production component, which models yields of maize, small grains and groundnuts on the basis of rainfall for 10-day periods or *dekads*. The model simulates production across 30 different spatial units (wards), thereby capturing environmental variability and facilitating modeling of trade in crops between wards. The network of meteorological stations around the district is sparse and few of these have long time series of rainfall records in Zimbabwe. Preliminary analysis of rainfall data for these stations suggested that intra-seasonal temporal correlation was negligible, but that rainfall at any given station was correlated with rainfall at neighbouring stations. This spatial correlation structure can be modeled using a multivariate normal distribution consisting of 30 related variables, representing dekadly rainfall in each of the 30 wards. For each ward, rainfall for each of the 36 dekads in the year was characterised by a mean and standard deviation, which were interpolated from surrounding meteorological stations. A covariance matrix derived from a distance measure was then be used to represent the spatial correlation between wards. Sets of random numbers were then drawn from this distribution to simulate rainfall across the wards in any given dekad. Cross-validation of estimated rainfall parameters against observed parameters for the one meteorological station within the district suggests that the interpolation process works well. The methodology developed is useful in situations where long-term climatic records are scarce and where rainfall shows pronounced spatial correlation, but negligible temporal correlation.

1 INTRODUCTION

This paper develops a methodology for spatial simulation of rainfall for areas where directly measured meteorological data are sparse, by interpolating measurements from surrounding rainfall stations. The methodology characterises the dekadly rainfall distribution across the wards through a multivariate normal distribution. It forms part of a simulation model of the food system in a region of Zimbabwe, which has been developed by an inter-disciplinary research group for use in food aid targeting [Gundry *et al.*, 1998]. This model includes components examining crop production, grain marketing [Vaze *et al.*, 1996], food consumption and onward linkages to human health and nutrition. The study is based in Buhera district, a semi-arid area where dryland production of maize, millet, sorghum, and groundnuts by smallholder farmers predominates. As substantial differences in climate exist within this district, crop production is modelled separately for the 30 different wards that make up the district. In common with several other maize models [e.g. Brisson *et al.*, 1992], the effect of rainfall variability on yields is simulated using a 10-daily time-step or *dekad*. Climate assessment in such areas is problematic, since the network of meteorological stations in Zimbabwe is concentrated in the more productive land further north.

One solution is to use sophisticated interpolation techniques that make use of Cold Cloud Duration data from the NOAA-AVHRR satellite [Herman *et al.*, 1997]. The United States Geological Survey Africa Data Dissemination Service (USGS ADDS) has used this process to provide dekadly rainfall data interpolated to a raster grid since 1995. However, although these data have sufficient spatial resolution for the purposes of the current model, the length of time series available (three years) is too short for use in multiple simulations requiring climatic variability. In addition, the interpolation method used gives rainfall patterns inconsistent with variation in crop production and vegetation within the district. The USGS rainfall data suggest that average rainfall over this period is broadly similar across the district, yet household survey data suggest the widespread cultivation of drought-resistant crops in the south of the district. As the USGS data appear to be inadequate for the wider model, a weather generator was developed to provide input to the crop models. The approach taken here differs from many other published weather generators [e.g. Leenhardt, 1999], in that spatial correlation in rainfall is explicitly modelled.

the values. The data were then standardised by computing means and standard deviations for each dekad at every station. The mean for each dekad was then subtracted from each value and the result was divided by the standard deviation.

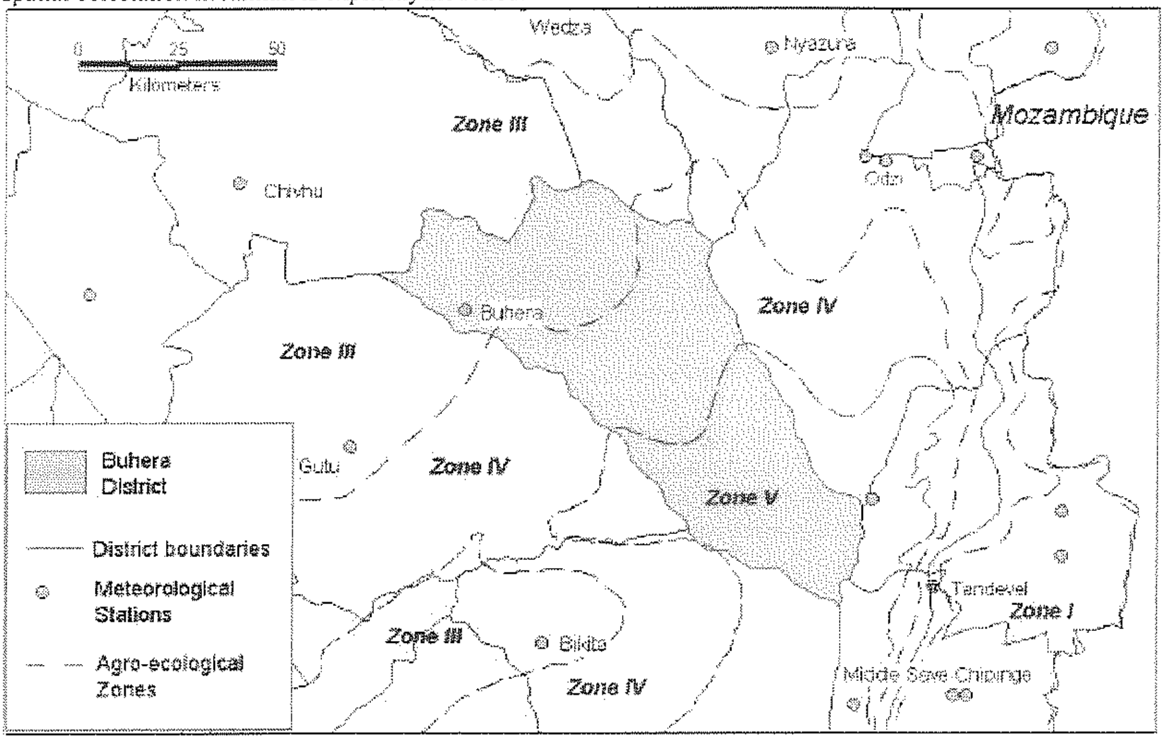

Figure 1: Map of Buhera district agro-ecological zones and meteorological stations used in interpolation.

2 METHODS

2.1 Data preparation

Dekadly rainfall data for 64 Zimbabwean meteorological stations were obtained from the USGS ADDS, covering the period 1952-1992. Since rainfall is known to contain a greater orographic component in the Eastern Highlands area to the east of Buhera, the analysis was restricted to 8 stations immediately surrounding Buhera but outside this upland region. This included one rainfall station lying in the northern part of the district itself. The location of the meteorological stations used is shown in Figure 1, together with agro-ecological zone boundaries for the area. The agro-ecological zones vary from Zone I (the wettest, with typical annual rainfall over 1,000 mm) to Zone V (the driest, with typical annual rainfall less than 600 mm). Buhera district straddles three of these zones: Zone III in the north through to Zone V in the south.

As the rainfall data had a skewed distribution, a logarithmic transformation was used to normalise

2.2 Assessment of temporal correlation

The transformed, standardised data were examined for evidence of spatial and temporal correlation. Correlation co-efficients were calculated using lags of up to 8 dekads for these stations, both including and excluding dry season dekads. Dekadly rainfall data were also aggregated by month and the same procedure was repeated using lags of up to 8 months. This analysis suggested little evidence of temporal correlation: for example, the correlation coefficient between rainfall in the current month and rainfall in the previous month was only 0.2. However, a preliminary exploration of the change in rainfall covariance with distance suggested that spatial correlation did exist (see later analysis). Consequently, the simulation made use of this spatial correlation structure, but not temporal correlation.

2.3 Assessment of spatial correlation

To examine spatial correlation, correlation coefficients were also calculated for each pair of rainfall stations using transformed, standardised data from all dekads.

This spatial correlation structure was simulated in two stages. Firstly, dekadly means and standard deviations for each ward were estimated from the station observations using a distance decay function. The reciprocal of distance squared was used to derive the weight for each station:

$$P_w = (\Sigma_{i=1..8}\, p_i D_{iw}^{-2}) / (\Sigma_{i=1..8}\, D_{iw}^{-2}) \qquad [1]$$

{where P_w = dekadly mean or standard deviation of precipitation for ward w; p_i = dekadly mean or standard deviation for each station; and D_{iw} = 3-dimensional distance between station i and ward w}

The 3-dimensional distance term in this equation (D_{iw}) was derived as follows:

$$D = \sqrt{(d^2 + ke^2)} \qquad [2]$$

{where D = 3-dimensional distance, d = horizontal distance between two locations, k = a weighting factor for elevation, and e = difference in elevation between two points}

The weight factor for elevation (k) in this equation was estimated using a cross-validation process. This involved omitting the data for each station in turn and estimating dekadly means and standard deviations for this omitted station from the other stations' data. The value of the weight was chosen so as to minimise the difference between estimated and observed dekadly rainfall parameters across all stations through an optimisation process.

Once a weight had been derived for elevation, this was used to compute three-dimensional distances between all pairs of stations. The rainfall covariance for each pair of stations was then plotted against distance. A function was fitted through simple regression that described the effect of distance on the rainfall covariance between points.

2.4 Generation of rainfall scenarios

Distances based on equation [2] above were then calculated between each of the wards in the district using a Geographical Information System. For each pair of wards, the covariance for rainfall was estimated based on the relationship derived with distance. A FORTRAN program was then used to sample random numbers from a truncated multivariate normal distribution for the 30 wards. The same covariance matrix was used to generate the values for each dekad, but means and standard

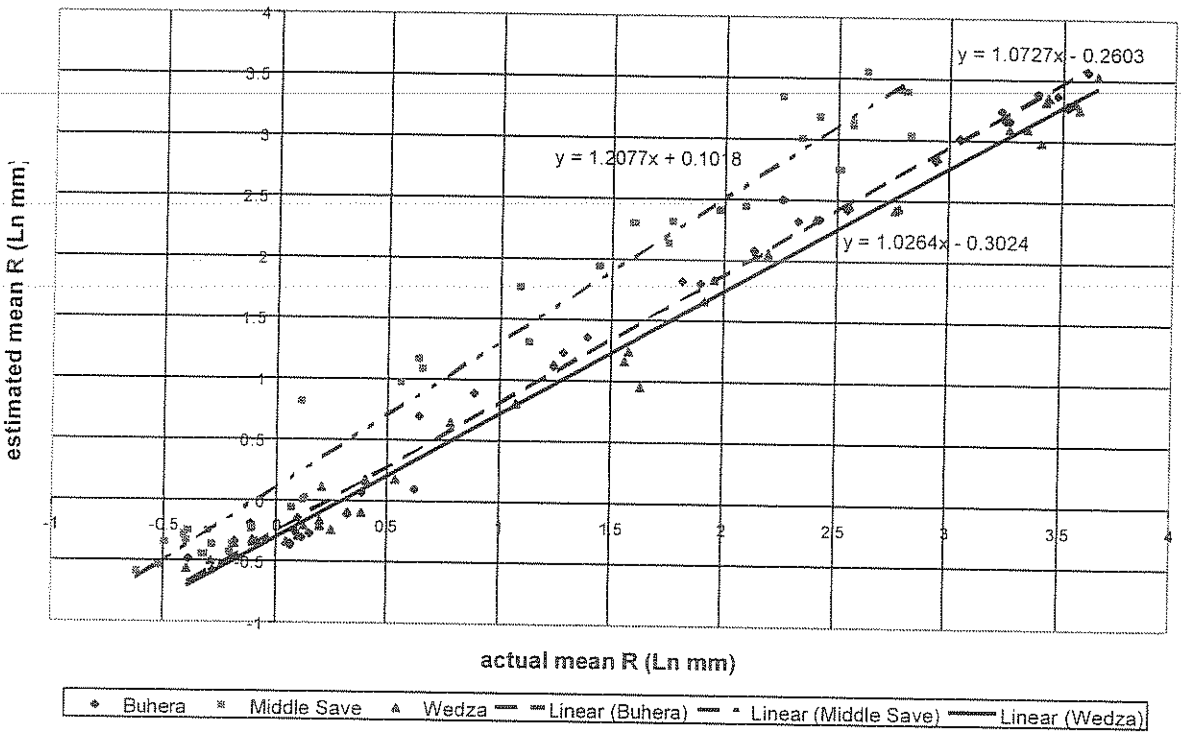

Figure 2: Results of cross-validation - estimated versus actual mean transformed rainfall by dekad for 3 sample meteorological stations

deviations varied from one dekad to the next.

3 RESULTS

The cross-validation exercise suggested a weight for difference in elevation of 140,000. Actual versus estimated mean rainfall for each dekad is shown in Figure 2. Cross-validation results for three of the eight original rainfall stations used are shown. When the Buhera station (which lies within the district as shown in Figure 1) was omitted and the remaining data points were used to estimate the mean and standard deviation of transformed rainfall by dekad, the model 'goodness of fit' remained high. When estimates of the mean and standard deviation were compared with observed data for this station, these gave a Root Mean Square Error of 0.30, suggesting that the method of estimating dekadly rainfall parameters worked well at least in part of the district. Similarly, the omission of the data point at Wedza to the north of the district did not adversely affect the model 'goodness of fit'. However, the omission of the Middle Save meteorological station led to an over-estimation of mean rainfall by dekad. This is because Middle Save is the only data point lying in the drier agro-ecological zone (Zone V in Figure 1) to the south of Buhera and the other data points inadequately represent this zone. This data point is therefore one of the most critical to overall model performance.

A linear regression model was used to relate the rainfall covariance between two points to 3-dimensional distance, which had an R^2 of 0.41. This relationship with distance is illustrated for the correlation co-efficient for rainfall at any two points in Figure 3. Despite the relatively high 'goodness of fit' of this model, the rainfall stations are generally further apart than the wards and so the correlation between wards that are close together is somewhat difficult to gauge. In the absence of more detailed information, the linear relationship between covariance and distance was extrapolated to represent such shorter distances.

Figure 4 shows sample rainfall scenario for one year. In this illustration, total annual rainfall derived from the dekadly model output has been linked to ward boundaries and the lower rainfall to the south of the district is clearly shown.

4 DISCUSSION & CONCLUSIONS

This paper illustrates a method of simulating rainfall across an arbitrary number of sites in situations where long-term meteorological records are scarce. The methodology is based on the assumption that the distribution of rainfall has remained largely unchanged since the 1950s. Although it has been suggested that rainfall patterns have changed since this date, in terms of national rainfall patterns, Gommes and Petrassi [1994: p. 16] note that there is 'no marked negative

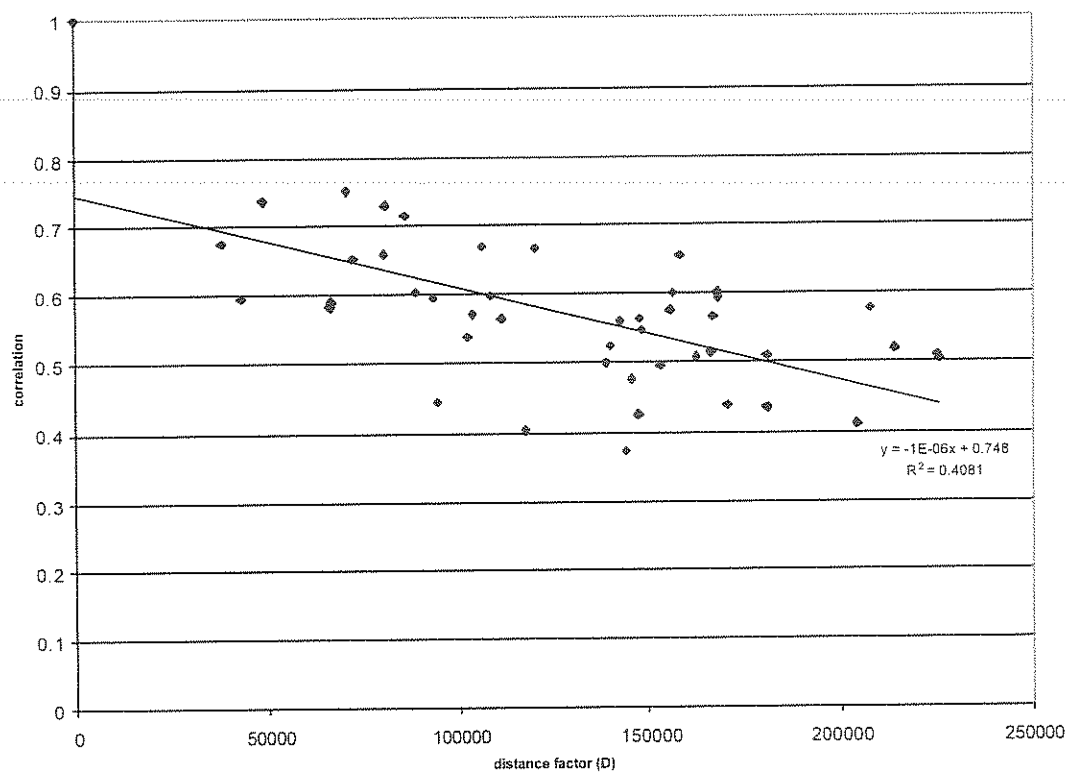

Figure 3: Relationship between 3-dimensional distance and rainfall correlation between pairs of meteorological stations

trend in rainfall' for a group of African countries including Zimbabwe. The Zimbabwean rainfall time series shows a slight increase up to 1979 and a slight decline thereafter [ibid., p. 13], suggesting that nationally at least this assumption is justified. The method adopted here is useful where significant spatial correlation exists, but negligible short-term temporal correlation. However, the methodology could be generalised to include temporal correlation. The absence of temporal correlation in this data set is surprising, as drought years in southern Africa are known to be related to El Nino-Southern Oscillation (ENSO) events. For example, earlier work [Cane *et al*, 1994] identified a close relationship between sea surface temperatures measured using remote sensing and maize yields in Zimbabwe, whilst the Southern Oscillation Index has been used for crop production forecasting in South Africa [de Jager *et al.*, 1998]. On this basis, below average rainfall totals would be expected throughout the dekads of ENSO event years, yet no evidence of temporal correlation was found here.

The simulated rainfall distributions produced for the 30 wards are now being used to model maize, small grain, and groundnut yields within the district.

The methodology can be transferred to other situations where point data have a spatial correlation structure and interpolation to contiguous polygons is required.

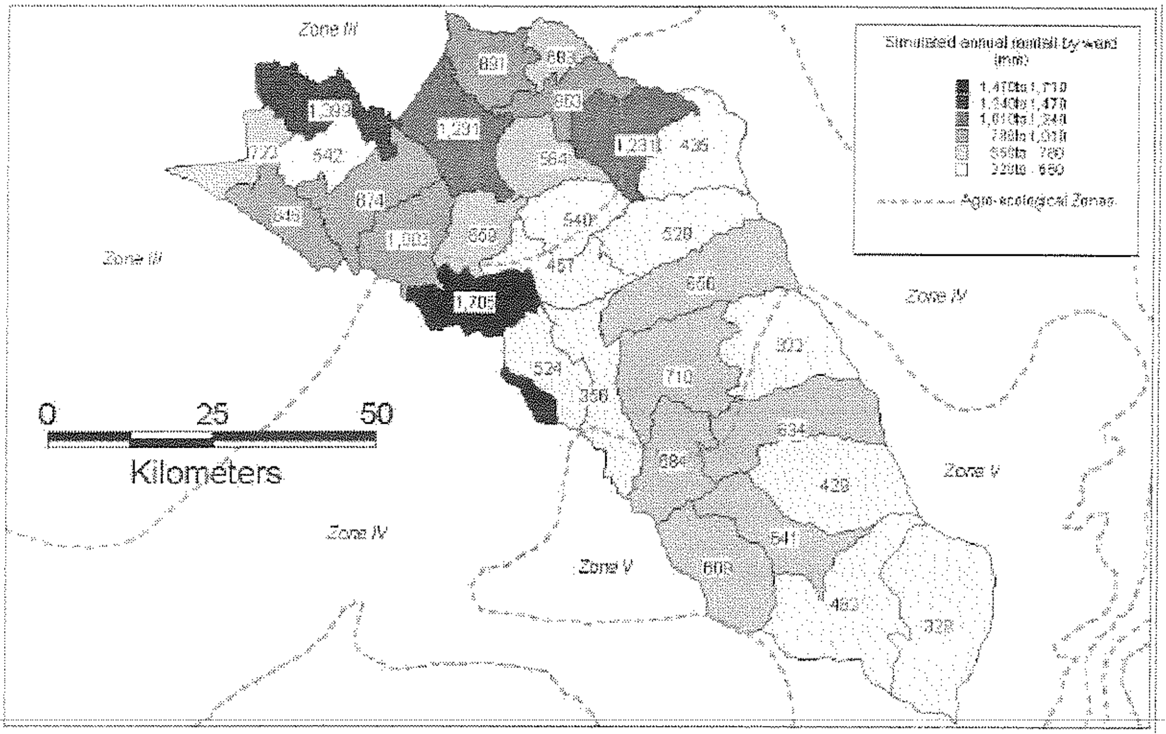

Figure 4: Sample model output, showing total annual rainfall scenarios by ward

5 ACKNOWLEDGEMENTS

This research was funded by the Commission of the European Communities DGXII, Science and Technology for Developing Countries Programme (ref: TS3*-CT92-0048). The authors wish to acknowledge the assistance of E. Austin of Biomathematics and Statistics Scotland in developing the FORTRAN program for this paper.

6 REFERENCES

Brisson, N., B. Seguin, and P. Bertuzzi, Agro-meteorological soil water balance for crop simulation models. *Agricultural and Forest Meteorology* 59 (3-4): 267-287.

Cane, M.A., G. Eshel, and R.W. Buckland, Forecasting Zimbabwean maize yield using Eastern Equatorial Pacific sea-surface temperature. *Nature* 370 (6486): 204-205, 1994.

De Jager, J., A.B. Potgieter, and W.J. van den Berg, Framework for forecasting the extent and severity of drought in maize in the Free State province of South Africa. *Agricultural Systems* 57 (3): 351-365.

Gommes, R., and F. Petrassi, Rainfall variability and drought in sub-Saharan African since

1960. *FAO Agrometeorology Series Working Paper* 9, 1994

Gundry S.W., J. Wright, and A. Ferro-Luzzi, Simulating the food and nutrition system in rural Zimbabwe to support targetting of emergency aid'. Proceedings, *ModSim97, the International Congress on Modelling and Simulation*, Hobart, Tasmania, 3: 1018-1022, 8-11 December, 1997.

Herman, A., V.B. Kumar, P.A. Arkin, and J.V. Kousky, Objectively determined 10 day African rainfall estimates created for Famine Early Warning Systems. *International Journal of Remote Sensing* 18 (10): 2147-2159, 1997.

Leenhardt, D., Weather generation and climate analysis for crop modelling at a Mediterranean site. Proceedings, *International Symposium Modelling Cropping Systems,* Lleida, Spain: 99-100, 21-23 June, 1999.

Vaze, P.B., S. Kudhlande, J. Wright, and S. Gundry, A spatial analysis of household grain purchases in Zimbabwe's liberalized marketing system'. *Outlook on Agriculture* 25 (1): 37-42, 1996.