A Comparison Study of Design Rainfall Mapping Using Ordinary Kriging and Kriging with External Drift

Khuluse S.[1,2], Stein A.[3] and Debba P.[1]

[1]CSIR, Pretoria - South Africa
[2]Harvard University, Boston - USA
[3]University of Twente, ITC - Enschede, The Netherlands

1. Introduction

Early development of Extreme Value Theory (EVT) was motivated by a need to quantify the probability of unusually large (or small) events in hydrology and climatology. The current challenge in modelling extremes is how to adequately account for variation of the processes in both space and time. The statistic of interest in EVT is the 1-in-N year return level or design value as it is known in hydrology. This is an important quantity as it helps to identify areas where flood protection infrastructure needs to be erected or improved as well as identifying communities located in high flood-risk areas. This necessitates a regional, rather than an in-situ estimate of the N-year design rainfall value. In this study we focus on estimating the N-year design rainfall surface in a case where the number of sampled sites is small using inverse distance weighting, ordinary kriging and kriging with external drift methods. Does incorporating additional explanatory information, given a spatially sparse sample, lead to any improvement to an estimate of the design rainfall surface?

2. Materials and Methods

The data consists of observed daily rainfall data from fifteen weather stations in the Western Cape province of South Africa. Seven stations had observation periods of fifty years, whilst for other stations the periods are shorter. The sampled area is bounded between latitudes $-34.058\,^{\circ}\mathrm{S}$ and $-32.463\,^{\circ}\mathrm{S}$ and longitudes $18.157\,^{\circ}\mathrm{E}$ and $26.493\,^{\circ}\mathrm{E}$, covering an area of approximately $132000\,\mathrm{km}^2$. The Western Cape is classified as a winter rainfall region. Rainfall in this area is associated with of orography and a cold frontal system that comes in from the Southern Atlantic ocean [4]. As our interest is in large rainfall events, we restricted our study to South Africa's winter months, June to August. The altitude for the whole Western Cape province was obtained from a $100\,\mathrm{m} \times 100\,\mathrm{m}$ DEM, by re-sampling it to $1\,\mathrm{km} \times 1\,\mathrm{km}$ pixel resolution. These were transformed to point data. Distances to the coast were calculated from the point data, where points which had altitude values smaller than $5\,\mathrm{m}$ were defined as the coast.

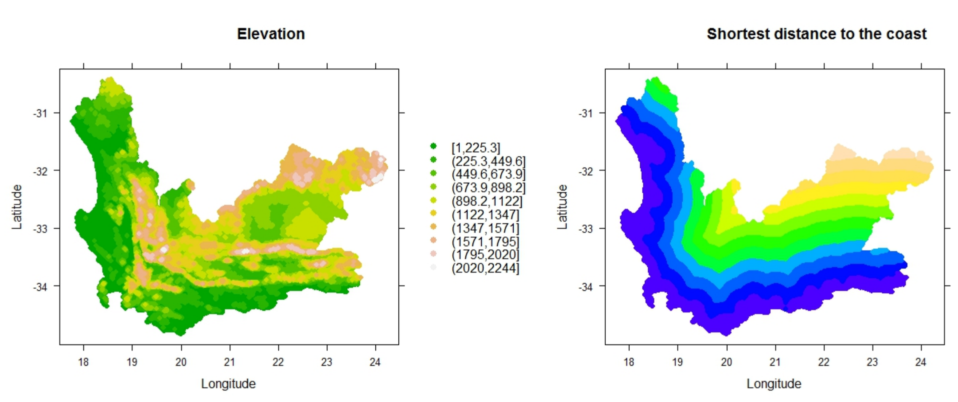

Figure 1: *Densely sampled covariate information - Altitude and distances to the coast*

Given a random sample of observations, the extreme value limit theorem states that the distribution of the largest (or smallest) member of that sample can be approximated by the Generalized Extreme Value distribution [1]. To derive an estimate of N-year design values at each site, the point process approach to EVT was followed. Temporal dependence of extremes was treated using a variant of the runs-declustering technique where extremal indices are obtained through applying the method of [3]. This was especially implemented for sites where the extremal index was below 0.9, which was evidence of temporal correlation of threshold exceedances at those sites. Parameter estimates at each site were then used to obtain in-situ estimates of the 25 to 50 year design rainfall (at 5 yearly intervals). Using these in-situ design rainfall values, the next task was to compare different methods for obtaining a 50-year 24-hour design rainfall surface over the Western Cape, given the small spatial sample constrain.

Direct estimation of the variogram model with only 15 sites leads to unreliable parameter estimates. The technique used to increase the sample size for estimating the variogram is implemented by considering return level estimates for return periods 25 to 50 years (in intervals of 5 years) as temporal replicates of the return level surface. Design values at each site are assumed to be a sample from a random variable which forms continuous surface over the study area. This continuous surface or spatial random field is denoted as $\{Y(\mathbf{s},\mathbf{t}) : \mathbf{s} \in \mathbf{D} \subset \mathbb{R}^2, \mathbf{t} \in \mathbf{T}\}$, where D is a fixed, continuous subset of a two-dimensional plane and $\mathbf{t} \in \mathbf{T}$ is an index for the temporal component. In this regionalisation method as described in [5], are space-time observations, with space relating to their geographical position and time represented by a sequence of equally-spaced return periods, $t_i \bigcap t_j = \emptyset$ for $i \neq j$. To satisfy the intrinsic stationarity requirement, the data is standardized by a ratio of the overall average design rainfall (m_0) and the average design rainfall for that specific period $(m_{t_j}), j = 1, 2, \ldots, p$. The standardized design value is determined as

$$\tilde{Y}(\mathbf{s},\mathbf{t}) = Y(\mathbf{s},\mathbf{t}) \times \frac{m_0}{m_{t_j}} \tag{1}$$

Standardization by the ratio of the overall average return level estimate to the one for that particular return period is done to remove the effect of different expectations for each return period. The p standardized values at each location are extended in space, by displacing the set of locations by a fixed distance c repetitively for $p-1$ instances, in this case $100\,\mathrm{km}$. The variogram is modelled using this extended sample. The penta-spherical model variogram model was chosen [2]. Estimation in this variogram model parameters is by weighted least squares (WLS) [2], with weights given as the ratio of the number of point pairs for a particular distance lag to the square of that lag distance. For predicting specifically the 50-year 24-hour design rainfall surface, estimates of the partial sill and nugget for the 50-year return period are obtained by multiplying parameter estimates of the variogram model for the pooled standardized data with the square of the reciprocal of the standardization factor.

In ordinary kriging (OK), design values at unsampled locations will be predicted by the predictor

$$Y(s_0) = \sum_{i=1}^{p} \lambda_i y_i, \quad \text{with} \quad \sum_{i=1}^{p} \lambda_i = 1$$

If there is evidence of a global spatial trend, then kriging weights λ_i are a linear function of explanatory variables, which in the case of KED include both geographical coordinates and other densely sampled external variables, such as Altitude and distances to the coast in our case. We compare the results of IDW, ordinary and KED kriging in the next section.

3. Results and Discussion

In the first step of our two-tiered modelling approach an extreme value distribution is fitted to threshold exceedances at each site to obtain a return level estimates corresponding to that site. For 9 of the 15 sites, the dispersion index is below 0.9 indicating weak temporal dependence of high rainfall values. This clustering effect violates distributional assumptions. To circumvent this problem the declustering technique [3] is employed at each of the 9 sites. Goodness-of-model-fit at each site is evaluated through quantile-quantile and return level plots. Model fits at each site seem adequate, however the level of uncertainty is large for sites which had few large excesses which deviate substantially from the rest of the observations. Removal of these large observations can lead to improved precision, however this strategy is avoided here to avoid obtaining overly conservative return level estimates. 25-year 24-hour design rainfall estimates are shown in Fig 2. These show variation in space, with higher values observed as one proceeds eastward of the bounding box. This warrants our next step to model this observed spatial dependence.

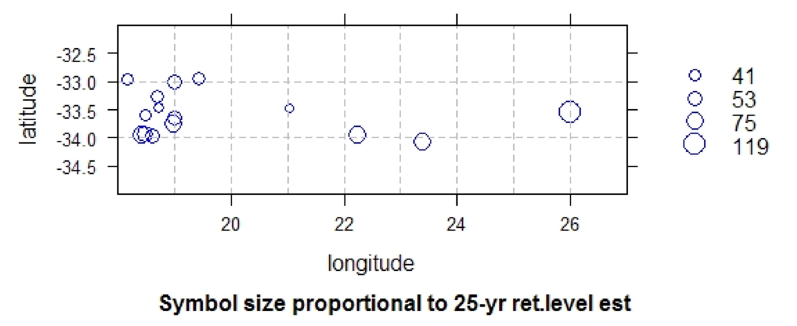

Figure 2: *Estimates of 25-yr 24-hr design rainfall at sampled sites*

We model spatial autocorrelation in pooled values through the penta-spherical model. Prior to this, global trends were explored. No statistically significant relationship was found between pooled design rainfall values and the individual feature space variable - altitude and distances to the coast. However, trend models where either one of these covariates was included with second order terms of the geographical coordinates yielded relatively stronger correlation coefficients as shown in Table 1. The two covariates cannot be included in the same surface trend model in the same global trend model as they are highly correlated ($\rho = 0.76$). In looking at regional trend five cases are explored: no trend, 1^{st} and 2^{nd} order linear trend in geographic space only, as well as 1^{st} order linear trend in geographic space including altitude (KED Alt) or distance to the coast (KED Coast). Reduction in sills for empirical variograms of residuals compared to that of pooled design values indicate that 20% of variation in design values can be accounted for by a regional trend. The lowest empirical variogram sill corresponds to KED Alt. Penta-spherical variogram parameter estimates for each of the five cases explored are obtained through a WLS procedure as discussed in the methodology section.

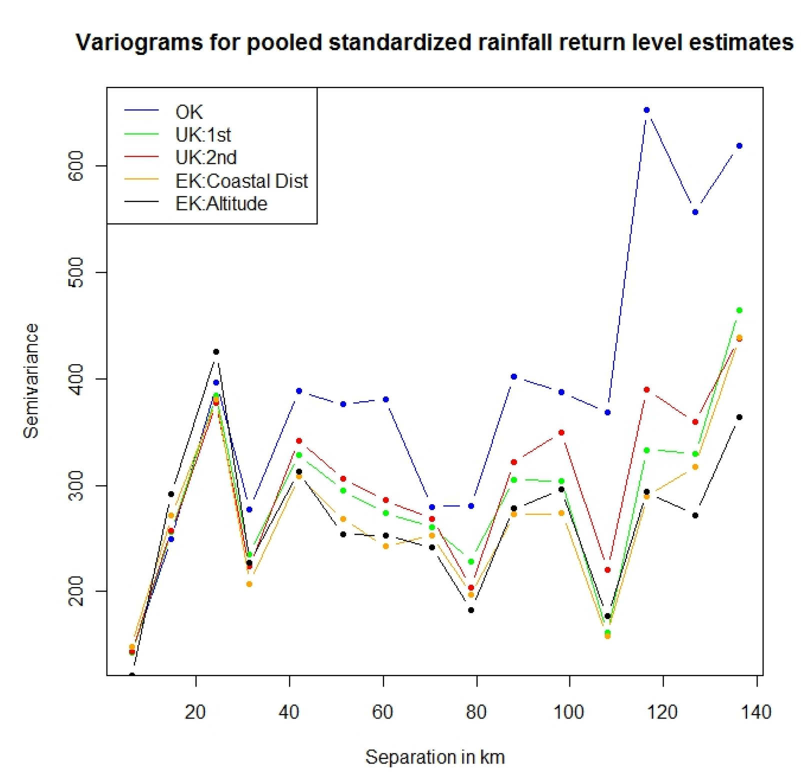

Figure 3: *A comparison of variogram models*

Once the variogram parameter estimates are obtained these are re-scaled for use as input in prediction at unsampled sites, i.e. resulting in a kriging map of an estimated 50-year 24-hour design rainfall surface. Prediction performance for each of the five cases is evaluated through a leave-one-out-cross-validation procedure, obtaining the root mean squared prediction error (RMSPE) for comparison. Low RMSPE values were obtained for OK and KED Alt (Tab 1).

Table 1: *Range parameter (in* km*) for the variogram models and LOCCV RMSPE*

	OK	UK 1^{st}	UK 2^{nd}	KED Coast	KED Alt
R^2	0.28	0.019	0.36	0.27	0.29
Range	37	27	27	14	25
RMSE	36.17	33.99	60.42	38.42	35.40

The range of prediction differences shows that the trend surface model with distance to the coast as an external covariate tends to predict lower values than the model with altitude as a covariate as well as OK values. This bias can clearly be seen close to sampled locations (Fig 4). Differences in predicted values for the distance to the coast trend surface and OK are nearly 3 times those of the altitude trend surface and OK.

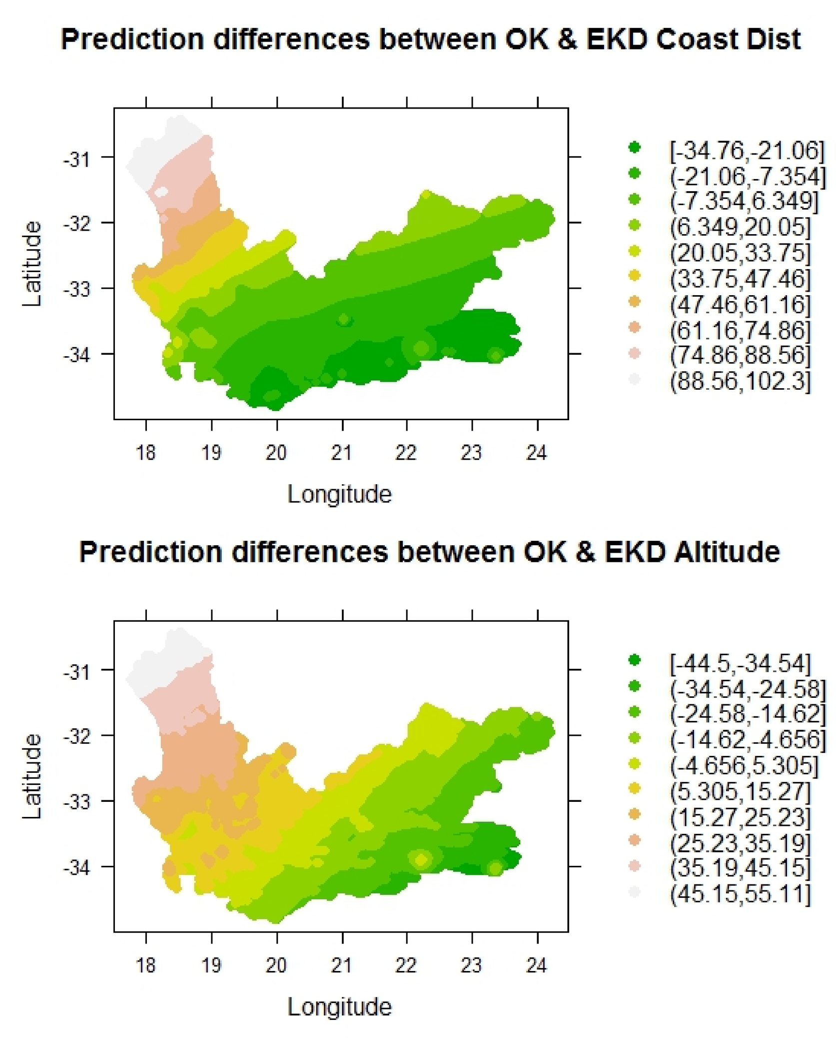

Figure 4: *Differences between ordinary and external drift kriging predictions of 50-yr 24-hr design rainfall values*

4. Conclusion

Extending the sample pseudo-temporally and spatially (Stein and Sterk, 1999) to estimate the variogram model parameters provided a way to implement kriging despite the sparseness of the data, rather than taking an average over an area as large as the Western Cape. KED is useful when the sample is sparse, as strength in prediction is gained by using a more densely sampled covariates to model the regional trend. However, it is important that a highly correlated covariate be used to seethe gains of this approach. In this study elevation was poorly correlated to the design rainfall values, but better than the measure of distance to the coast. Due to poor correlations and sparseness of the sample, KED error variances were much larger than OK variances. The trend surface with distances to the coast shows bias for lower predicted values than those obtained by OK kriging. Given more sampled locations KED with altitude as an external covariate maybe useful to capture the effect of topography of design rainfall surface estimation in the Western Cape.

Acknowledgements

The data used in this study was obtained from South African Weather Service. Dr P. Schmitz for assistance in obtaining point elevations for the Western Cape. TATA Africa Scholarship and funding from the CSIR and ITC - University of Twente for making possible my research visit to ITC, during which part of this work was done. Mark Dowdeswell for his supervision.

References

[1] S. G. Coles. *An Introduction to Statistical Modeling of Extreme Values.* Statistics. Springer Verlag, London, 2001.

[2] N. A. C. Cressie. *Statistics for spatial data.* Wiley, 1993.

[3] C. A. Ferro and J. Segers. Inference for clusters of extreme values. *Journal of the Royal statistical Society*, 65:545–556, 2003.

[4] R. A. Preston-Whyte and P. D. Tyson. *The atmosphere and weather of southern Africa.* Oxford University Press, 1988.

[5] A. Stein and G. Sterk. Modeling space and time dependence in environmental studies. *JAG*, 1(2):109–121, 1999.