Res. Lett. Inf. Math. Sci., 2004, Vol.**6**, pp97-108
Available online at http://iims.massey.ac.nz/research/letters/

A model of rainfall based on finite-state cellular automata

D.R. Munroe, B.I. Mills & P.S.P. Cowpertwait

Institute of Information & Mathematical Sciences
Massey University at Albany, Auckland, New Zealand.

The purpose of this paper is to demonstrate that a finite state cellular automata model is suitable for modeling rainfall in the space-time plane. The time-series properties of the simulated series are matched with historical rainfall data gathered from Whenuapai, NZ. The spatial scale of the model cells in related to land-area by optimizing the cross-correlation between sites at lag 0 relative to rainfall data collected from Auckland, NZ. The model is shown to be adequate for simulation in time, but inadequate in spatial dimension for short distances.

---

1 Introduction

The model proposed in this paper for modeling rainfall is a finite state cellular automata Model (FSCAM) inspired by a continuous model known to exhibit chaos. Due to the nature of the FSCAM, rainfall is modeled in the space and time dimension simultaneously. The reduction from a continuous model to a finite state model has resulted in a reduction in 'interesting' behaviour, however it is inherently simpler.

Self-organized criticality, as defined in [2], provides the foundation of the FSCAM for rainfall. Self-organized criticality (SOC) may be an underlying component of rainfall [5], and a rainfall series is highly likely to fit into this class of physical phenomena [1]. Thus a deterministic model which generates stochastic behaviour while demonstrating SOC is expected to be suitable for simulation of rainfall series.

2 Model Formulation

The model conceptually simple, Appendix A, shown in the series of steps below:

1. Acceleration is computed as a function of the surrounding points and 'gravity'.

2. Integrate Acceleration to get Velocity.

3. Integrate Velocity get the new Position.

As we are working with a finite state model, technically we are actually taking finite differences and summing rather than integrating.

3 Simulation

The model used for the simulation of rainfall was originally developed by Bruce Mills, in collaboration with Paul Cowpertwait, for the purpose of simulating cloud behaviour. As cloud density is related to the amount of rainfall [4], a model for emulating cloud behaviour is expected also to be suitable for modeling rainfall.

An image of the cells at different times during the simulation has been reproduced in Figure 1. The dark areas of the images indicate little or no cloud whereas the lighter cells represent dense cloud formations (and therefore heavy rainfall). The initial linear starting position of the model is shown in the top left hand corner of Figure 1.

The starting cells are clearly not random, and the system takes a number of iterations to reach a state of SOC. The top right hand corner of Figure 1 is the same model 1000 time steps later. The linear starting position is still evident in the plot, but the chaotic behaviour of the model is clearly beginning to form. The two lower images in Figure 1 represent the model at 5000 time-steps and 5500 time steps respectively. At this point, little, if any, evidence of the linear starting position remains and there is no pattern in the model.

4 Statistical Analysis

In order to model rainfall using the FCA model, two important conditions must be met. Firstly, the scale of the generated rainfall must be determined relative to the actual rainfall. Secondly, the spatial scale of the model needs to be determined in regard to land-area covered.

4.1 Time Series Statistics

4.1.1 Scale

As rainfall is an extreme value distribution, a scale factor must be applied to the simulated data in order to emulate this. Due to the finite range of the model generator, from 0 to 256, the extreme values are too low and not 'stretched' enough.

Two methods were considered suitable for transforming the simulated data. The first was to use a power transformation (PT), the second to use a logistic inverse transformation (ILT) based on the range of the simulated data.

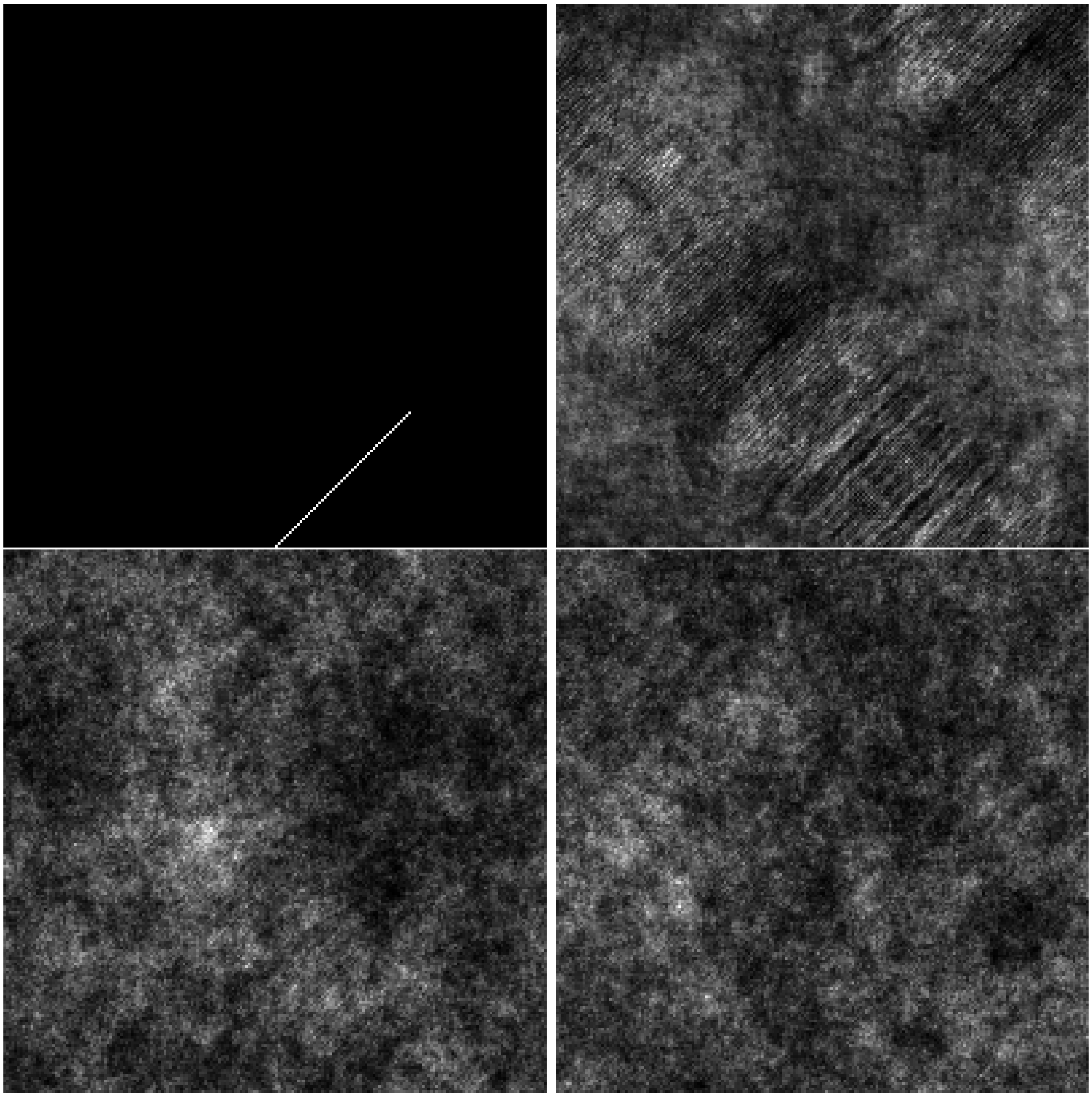

Figure 1: Simulated Cell images and different time steps (from left to right: 0, 1000, 5000, and 5500 respectively)

A simple method of finding the optimal power transformation is to minimise the Kolmogorov-Smirnov test statistic and thus the difference between the empirical CDFs. The result of the optimisation for the standardised series is shown below:

The series collected from the thirteen different 'sites' were checked for consistency against the actual data and the same transformation was suggested in all cases. The effect of several extreme points in the real data series on the transformation was examined but was found to be irrelevant. The transformation suggested was consistent even when the largest 30 observations were removed.

Power-Transformation	D-Statistic
2.763954	0.8767534

The inverse logistic transformation is calculated as in equations 1.

$$Rain_{Transformed} = log(255/(255.0000001 - Rain_{Simulated})) \quad (1)$$
$$Rain_{Standardized} = Rain_{Transformed}/mean(Rain_{Transformed}) \quad (2)$$

The effect of the transformations are best seen in the comparison of the QQ-Plots of the standardized model series versus the standardized rainfall series, Figure 2.

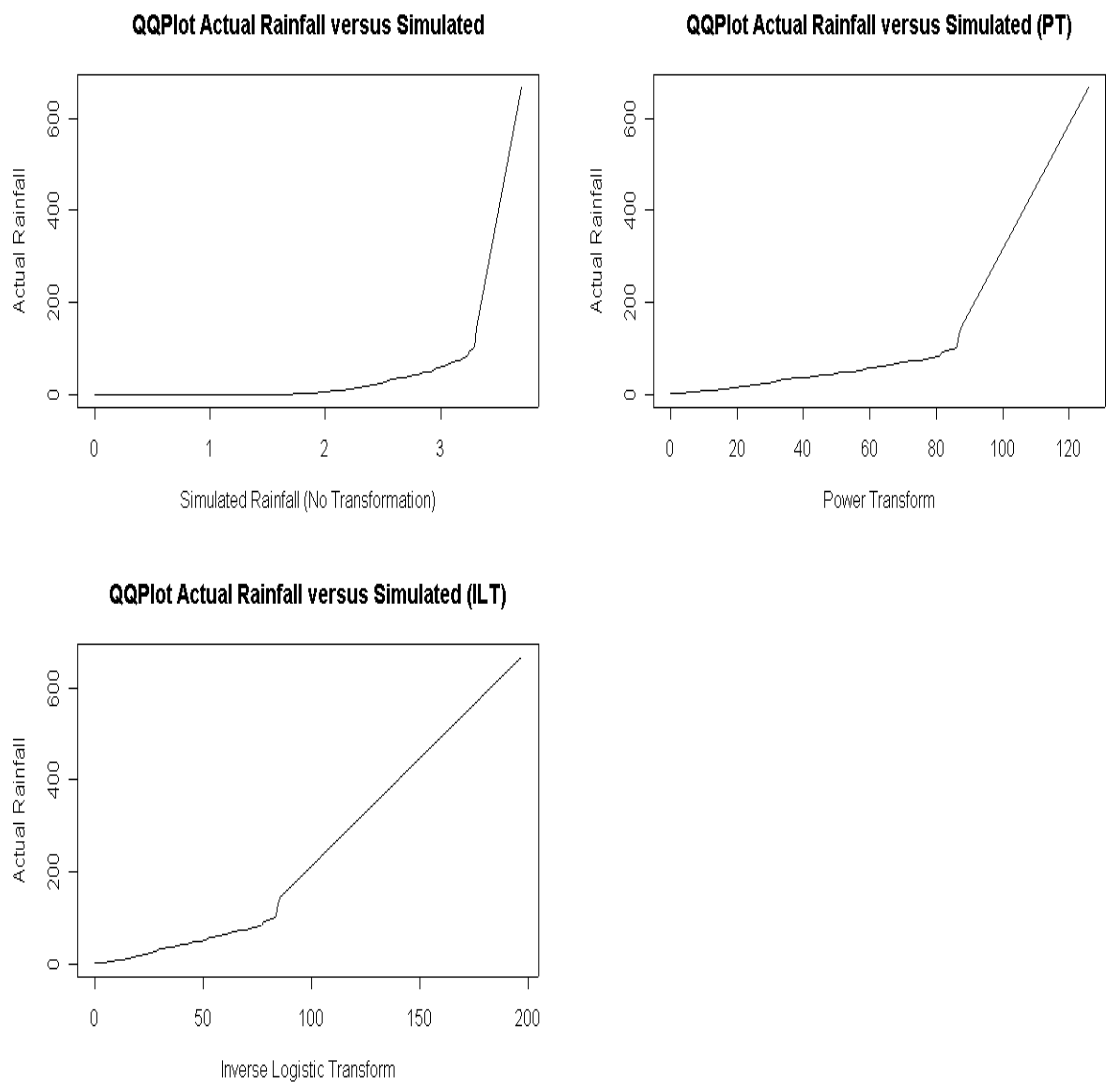

Figure 2: Quantile-Quantile Plots: Simulated and Actual (Hourly)

It is evident that while the series are still very different, the transforms have improved the situation significantly. In addition, it is obvious that the second transformation (ILT) is superior to the power transformation and thus all subsequent analysis relates to this transformation.

4.1.2 Cloud Density and rainfall

As the simulator actually generates cloud intensity, the next step is to determine at what level cloud intensity generates rainfall. As we are interested in a simple model, the probability of rainfall is assumed to be absolute, in that, above a particular cut-off point, the probability of rain is 1 and below that point, the probability of rain is 0. The distance between the cutoff point and the simulated value is taken to be the amount of rain falling. Clouds with greater density generate more rain than clouds of lower density.

In order to achieve the correct behaviour between simulated rainfall, three statistics were simultaneously optimized. Specifically, the lag 1 autocorrelation, the covariance, and the proportion of time units with no rainfall. As the data was standardized to have unit mean, the optimization of the covariance is equivalent to optimizing the standard deviation.

The simulated data for the optimized values and actual rainfall over an equivalent time scale is shown in Figure 3.

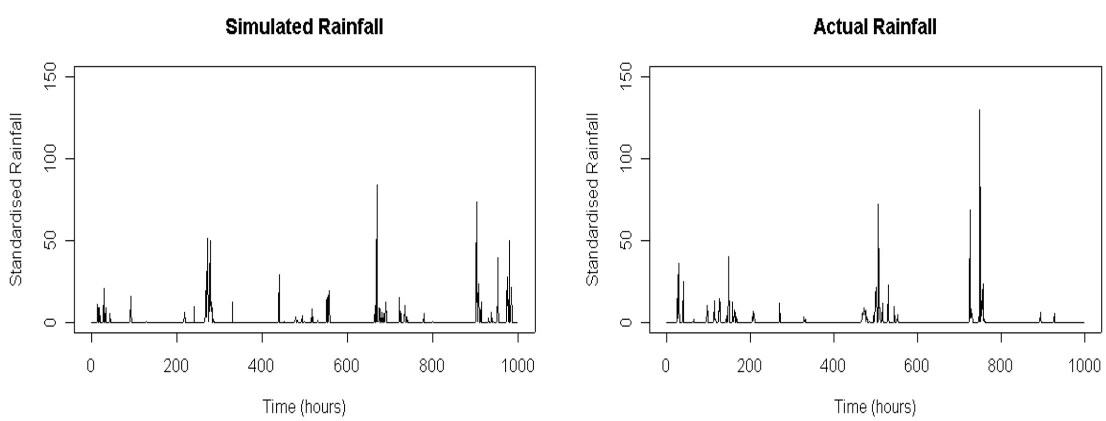

Figure 3: Rainfall (hours): Simulated and Actual

In order to determine whether there was any significant difference between the actual series and the simulated series (ILT), a series of 500 samples of sizes 100, 200, and 400, were taken from the simulated series (ILT) and compared to a sample of the same size taken from the rainfall series. A Kolmogorov-Smirnov test was then computed for each sample, Figure 4. In each case, the test was computed 500 times. As we can see from the results in Figure 4, the simulation is consistently similar for the small samples but, as the sample size increases, the dissimilarity

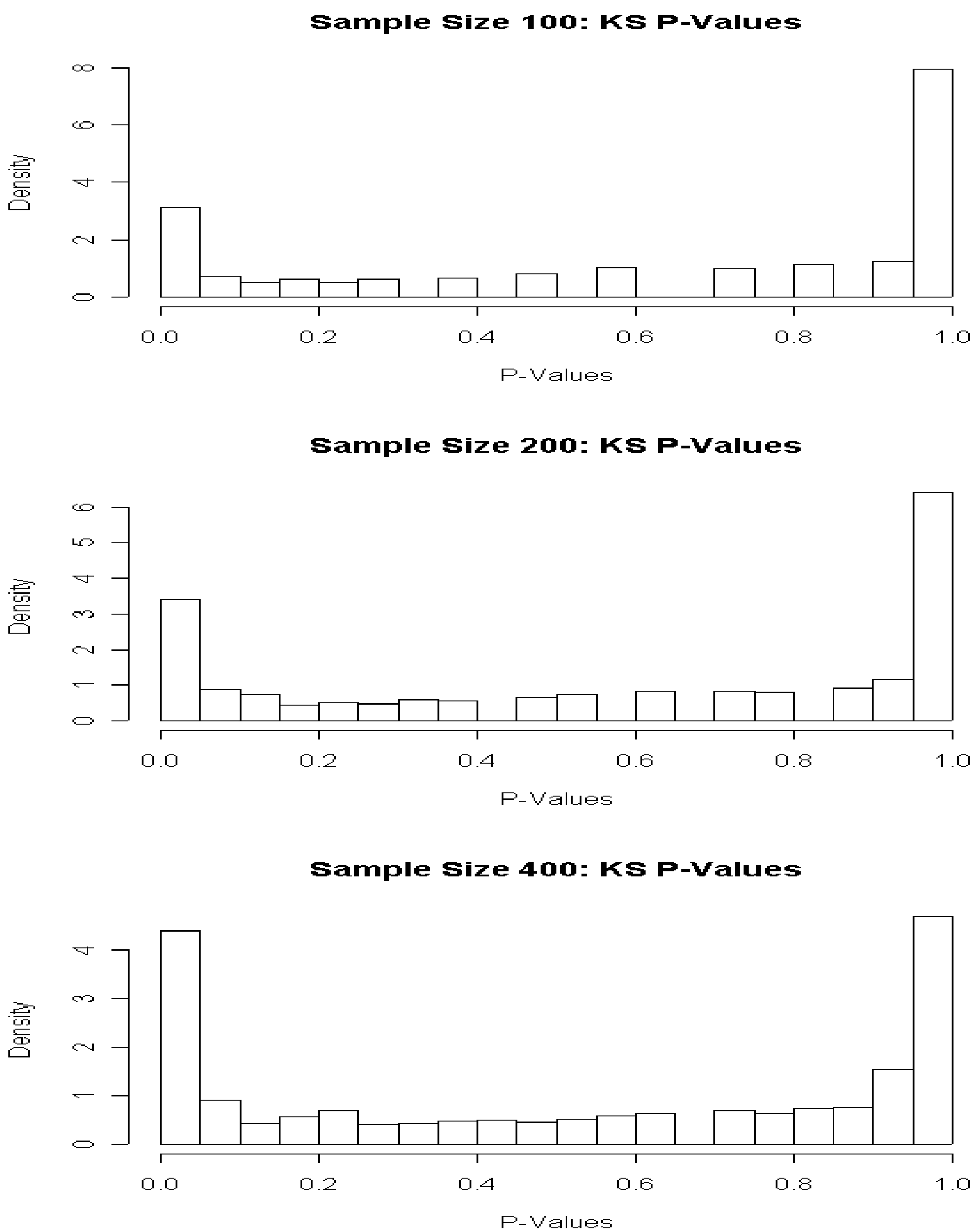

Figure 4: Rainfall (hours): KS-Tests

also increases. The likely cause of this behaviour is the probability of the samples containing extreme values increases as the sample size increases. As was shown previously in Figure 2, the simulation does not match extreme values very well.

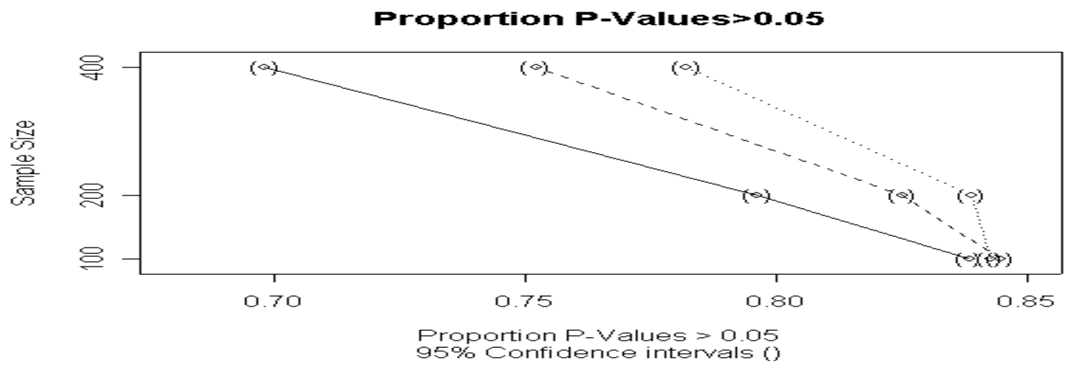

Figure 5: KS P-Values > 5%: Scaled and Unscaled with CI's

4.1.3 Effect of Scaling

The effect of the scaling transformation on the series is not immediately obvious. Visually, the results from the transformed transformation seem identical to the untransformed version. Therefore, the K-S test for each of the series for the sample sizes described above was computed and the proportion of P-Values > 5% were stored. The results for the scaled models are a significant improvement of the unscaled model as shown in Figure 5: where the solid line is the unscaled model, the dashed line represents the PT model, and the dotted line represents the ILT model.

It is clear from the results that the scaled models are more successful at emulating rainfall than the unscaled model 1.

Table 1: KS Average Proportion < 5%

	Sample100	Sample200	Sample400
Unscaled	0.8384067	0.7961933	0.6981067
PT	0.8446867	0.8252067	0.7520133
ILT	0.84264	0.8390867	0.78214

The result of the the scaling is to reduce the likelihood the simulation being significantly different from the actual rainfall at the 5% level. Larger samples are less likely to be significantly different for either transformed series than for the untransformed series.

4.1.4 Aggregation to Daily Rainfall

The simulated ILT series and the actual rainfall series were then aggregated to daily values and the two series compared as shown in Figure 6. The result of a

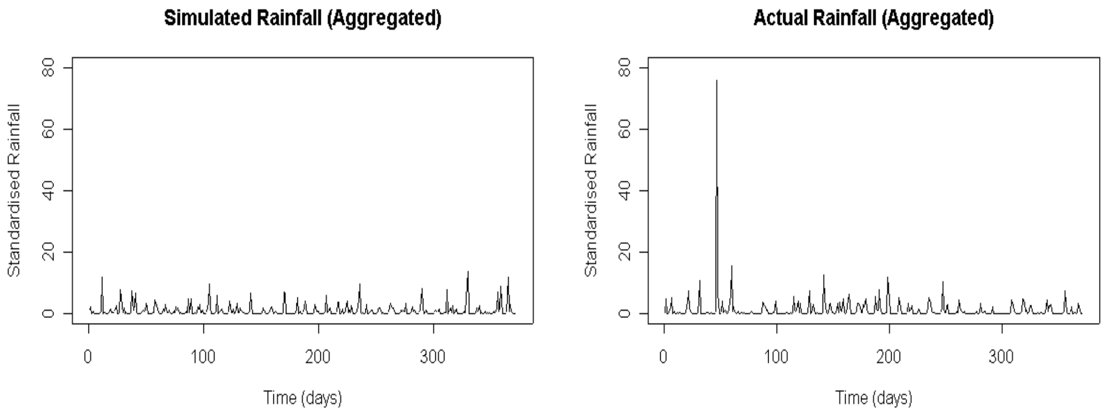

Figure 6: Rainfall (days): Simulated and Actual

Kolmogorov-Smirnov Test comparing the two series is as follows:

Thus, we cannot conclude that the series come from a different distribution at

D-Statistic	P-Value
D = 0.0761,	p-value = 0.03326

the 1% significance level. It is self-evident that the cause of difference between the two series is the lack of 'extreme values' for the aggregated rainfall. If the actual aggregated series is examined over a longer time period as in Figure 7, it is evident that large values of rainfall are not uncommon.

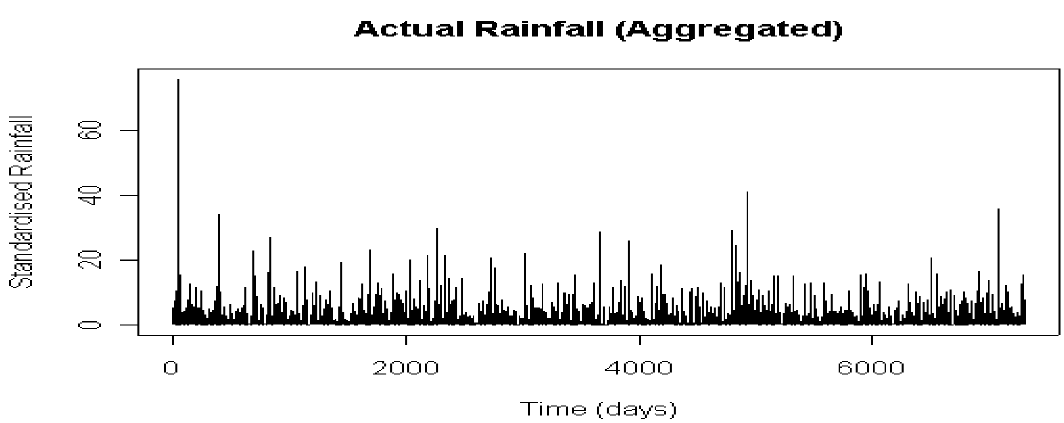

Figure 7: Rainfall (days): Actual

4.2 Spatial Statistics

The simulated rainfall from the finite cellular automata model was collected at a number of 'sites' within the model. In order to appreciate the scale of the model, the cross-correlation at lag 0 was compared against 'distance' for the simulated data and the Auckland City data. The latter, however, covers a relatively small land-area with the maximum distance between sites being only 17.4km.

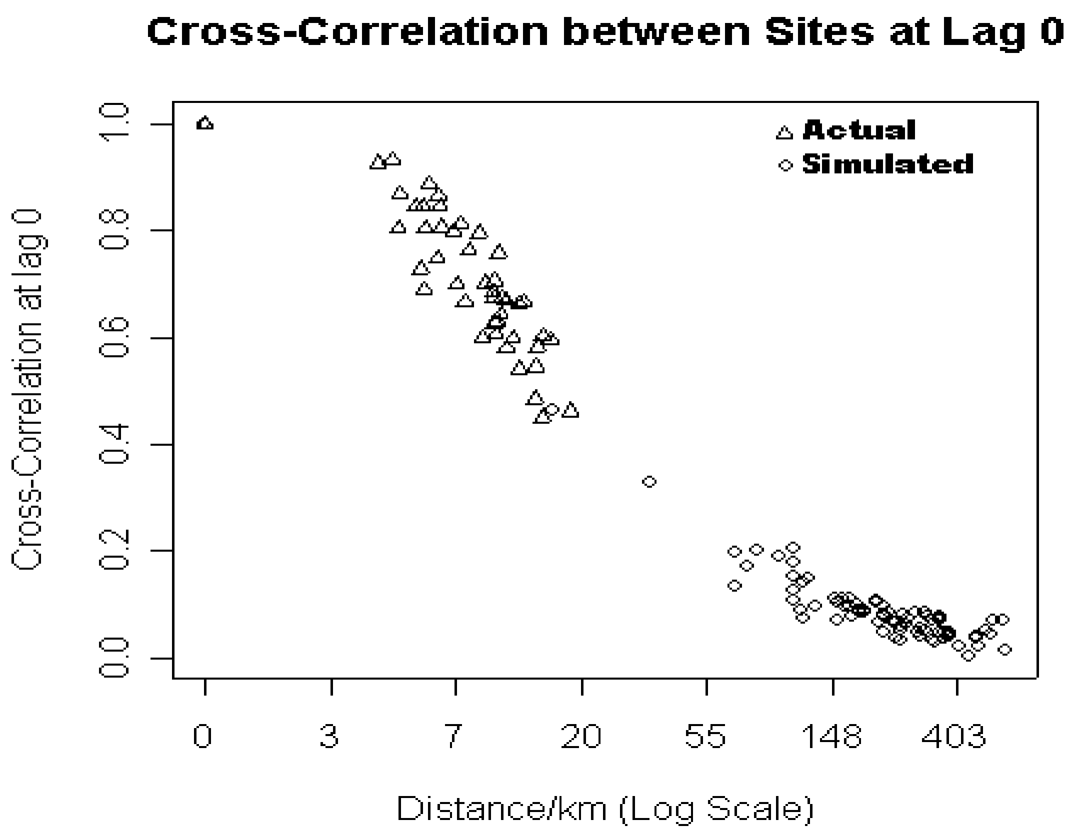

Figure 8: Hourly Rainfall Simulated and Actual: Lag 0 Site Cross-Correlation

The cross-correlation plot in Figure 8 clearly shows the correlation decreasing with distance. The simulated data has had distances assigned to it based on the best match up of the minimum distance. The optimal level for the scale of the FSCAM was found to be $\sim 15km$ per cell. Although the distance covered by the actual data in Figure 8 is clearly not large enough to cover that for the simulated series, the latter is still behaving as is expected. A cross-correlation plot between sites in the Thames catchment [3, Figure 4, page 174] shows that the correlation over a greater distance behaves similarly to that of the simulated series.

The two respective series, simulated and actual, were then aggregated and the cross-correlation at lag 0 compared as before, Figure 9. The distance scale was kept

Figure 9: Daily Rainfall Simulated and Actual: Lag 0 Site Cross-Correlation

at $\sim 15km$ in Figure 9, however, it is evident that the scale factor is not correct. The cross-correlation at lag 0 is not decreasing fast enough between sites as distance increases.

The cross-correlation plot with a 'correct' scale factor of $\sim 7.5km$ is shown in Figure 10. Clearly, this second plot is more correct as the correlations are behaving in accordance with what is expected. By increasing the aggregation level to daily data rather than hourly data, the effect on the spatial dimension is to reduce the scale factor by 1/2.

5 Conclusions

The finite state cellular automata model is suitable for simulating rainfall - especially over short time periods. Simulating rainfall over longer periods requires the simulated data to be transformed before the optimal cutoff is found. The 'best' transformation found was the inverse logistic transform, which generated extreme values more readily than the power transformation. Some further model adjustment may need to be applied so that extreme points can be generated as commonly found in actual rainfall.

Although the model was applied to hourly data, when the simulated data and the Whenuapai rainfall data were aggregated the time-series were still similar. This property of time-scale invariance, if held for higher aggregation levels, may be useful

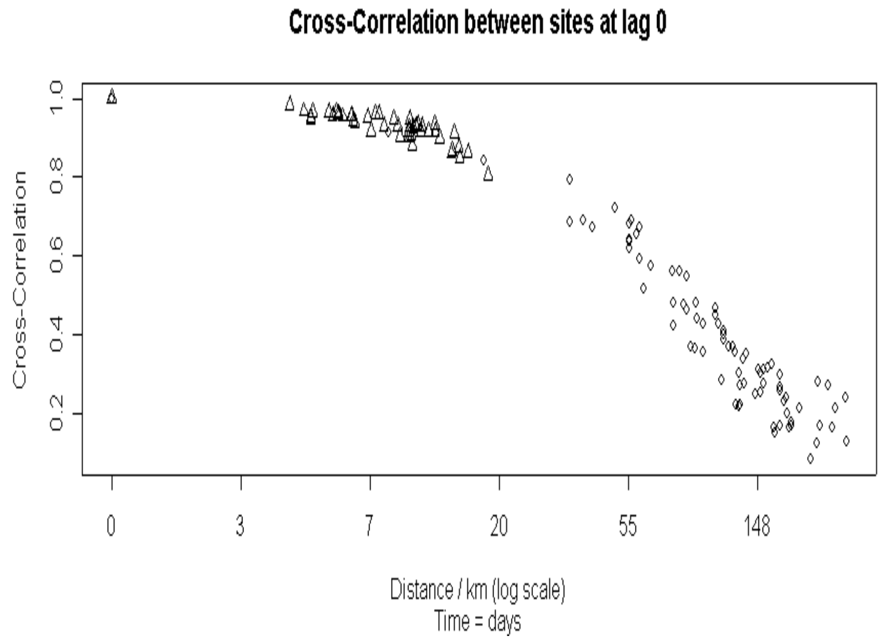

Figure 10: Daily Rainfall Simulated and Actual: Lag 0 Site Cross-Correlation

as a means of disaggregating series where only high aggregation levels are available.

The spatial correlation between cells within the FSCAM is consistent with the model covering a large land area. However, the actual scale of the model should be checked more thoroughly against actual rainfall collected from a wider catchment area. In addition, while aggregation produces sensible simulation models for the time series, the scale of the spatial model is reduced by some magnitude. In order to ensure the correct scale, the spatial dimension must be re-optimized when the simulated rainfall series is aggregated.

6 Acknowledgments

We are grateful to the Auckland Regional Council and Metrowater Ltd for granting permission to use the Auckland City rainfall data. We also thank Scott Whineray for providing a useful reference.

References

[1] R.F.S. Andrade, H.J. Schellnhuber, and M. Claussen. Analysis of rainfall records: possible relation to self-organised criticality. *Physica A*, 254:557–568, 1998.

[2] Per Bak, Chao Tang, and Kurt Wiesenfeld. Self-organised criticality: An explanation of 1/f noise. *Physical Review Letters*, 59:381–384, 1987.

[3] Paul S. P. Cowpertwait. A generalized spatial-temporal model of rainfall based on a clustered point process. *The Royal Society*, 450:163–175, 1995.

[4] Ignacio Rodrigues-Iturbe Ed Waymire, Vijay K. Gupta. A spectral theory of rainfall intensity at the meso-β scale. *Water Resources Research*, 20:1453–1465, 1984.

[5] Ole Peters and Kim Christensen. Rain: Relaxations in the sky. *Physical Review E*, 66, 2002.