Modelling Daily Rain-Gauge Network Measurement Responses Under Changing Climate Scenarios

Report to the
Water Research Commission

by

Geoffrey Pegram and Scott Sinclair
Pegram and Associates (Pty) Ltd
DURBAN

&

András Bárdossy
Universität Stuttgart

WRC Report No. 1964/1/13
ISBN 978-1-4312-0476-2

November 2013

Obtainable from

Water Research Commission
Private Bag X03
GEZINA, 0031

orders@wrc.org.za or download from www.wrc.org.za

Executive summary

What is this project about? The title says it succinctly: Modelling daily rain-gauge network measurement responses under changing climate scenarios. This solicited work falls within the WRC Programme: Developing predictive tools and adaptive measures to global climate change and hydroclimatic variability. We specifically look at gauge rainfall in relation to Regional Circulation Model (RCM) rainfall estimates.

Why are we doing this work? Is there change in the rainfall regime in South Africa? The answer is yes. How fast is the change? We are not sure, as it is still subtle and the data are very noisy.

What is the current status of our rainfall regimes in different parts of the country? This study shows that there was a small change over the country from the decade 1990-2000 to the years 2000-2008. We observe change and can successfully model it, but we are not yet sure that what we observe is not just typical natural fluctuation; therefore monitoring, frequent updates and analyses of data, and most importantly adaptation, are vital. As a result, maintenance of existing data monitoring networks is paramount. We cannot manage on conjecture. Nor can we rely on remote sensing, because those estimates need conditioning on ground-based data. Hence the need for the maintenance of existing and re-establishment of gauge networks measuring both rainfall [and temperature]. However, with what we have available, we are confident in making some suggestions for the way forward, having devised the tools to monitor and model changes in the rainfall estimates offered by RCMs.

This project report is based on two lines of research. The first is a study performed in Germany on 4 Regional Circulation Model (RCM) records and data from over 700 daily raingauges spanning 40 years over a large area [107 000 sq km]. The principal outcome of interest there is that it showed us that linking the rainfall regime to Circulation Patterns (CPs) of daily Sea Level atmospheric pressure, allowed us to effectively model the change in the rainfall regime, but at the 30 km spatial scale of the RCMs. The lessons learned there were applied over South Africa, the difference being that we here downscale the RCM rainfall to point measurements at the gauge-sites and perform direct validation of results in a 'blind' study. The results show that we can measure and model the change in the rainfall regime over 5 climatically different mesoscale regions in South Africa: the Western Cape [winter rainfall], the East London region [coastal summer temperate rainfall], central Free State [summer rainfall], the upper Vaal catchment [heavier summer rainfall] and Mpumalanga [East coastal subtropical].

The report details the answers to the following research questions:

- How does one get a handle on the problem of monitoring and adapting to climate change – what does one look for from the point of view of hydrology? It turns out that when RCM rainfall estimates are downscaled, the amounts and occurrences of rain at gauges should be right. This can only be downscaled RCMs in this context where gauges are available. This is different from the usual idea of downscaling being the generation of a fine scale grid (covering the entire domain), using coarser resolution information.
- How does one ensure that the resulting RCM block-scale (50 km) estimates have the observed spatial interrelationship as measured by their cross correlations so that areal rainfall over large catchments is properly modelled?
- What properties of the data are needed? It turns out that to link RCM and gauge data meaningfully, the gauges should not be too sparse because they have a limited correlation length, beyond which they are not statistically linked; hence the importance of at least maintaining the current network and preferably growing it back to its earlier density
- How can the spatially averaged gauge data be related to the RCM block scale of about 50 km? A special method of spatial Block averaging is devised.
- How to exploit the information offered by the RCMs? It was established that Rainfall regimes are captured by CP dependent RCM conditions – this was shown in Germany and confirmed

in RSA. Learning from the German experience it was identified that the sets of circulation patterns of atmospheric pressure at 700 hPa were better than Sea Level pressures at discriminating rainfall types. This idea was applied part way through the study and found that different CP sets affect the weather patterns in the 5 different regions differently. Thus the CP sets are region dependent; one pattern does not fit all.

- What are we looking for in 'good' gauge data? Which RCM to choose? We chose the PRECIS reanalysis data derived at the Climate Systems Analysis Group (CSAG) of the University of Cape Town. Initial comparison of quantiles and correlations between Block averaged gauge records and PRECIS data, before downscaling, showed that the RCM data were generally wetter than the gauges recording light rainfall and sometimes wetter than the gauge maxima.
- How do we ensure that the PRECIS data are transformed to have the same characteristics as the Block averaged gauge data? Two procedures were required. The first was the decorrelation/recorrelation methodology devised herein. Second, this was followed by the quantile-quantile (QQ) downscaling from the Block scale to the gauges. This combination was shown to be successful and is the core of the methodology used here to decide where and how much the changes happen and under the umbrella of which CPs.
- How practical are these ideas? Of interest to hydrological practitioners and managers, is that the procedure can be adapted to produce a set of stochastically different downscaled rainfall sequences based on the RCM output. This has the benefit of determining the sensitivity of the output to the analysis method, giving realistic confidence in the results.
- How to model the future? We carefully take the 'future' RCM rainfall output which we have in a validation period and downscale it based on the recorrelation procedures and double QQ transforms that we have introduced in this study.
- How do we know that the modelling procedure captures change in the rainfall regime suggested by the RCM? How can we be sure? This is achieved by comparing the downscaled daily rainfall at the networks with actual observations; firstly during a verification period and then using 'blind' data (not used in the downscaling procedure) in a validation period. A range of statistics is checked to ensure that the features are right.

The body of the report relates the methodology in as visual a way as possible. Besides the organisation and cleansing of the data, the technical road to success was this last year's work spent on developing the recorrelation methodology which is *completely original and is the core of the project*. Like many other modelling activities, to conceptualise and achieve the objective depended heavily on computer programming, much of which was rewritten and superseded. This caused considerable frustration and despair in one's inability to design the algorithms to get the right results. Then we found the key to recorrelation by iteration combined with effective quantile transforms. The level of elation at finding a proper solution to the problem and obtaining meaningful results made up for the grief!

There are three major original research highlights that come out of the project:

- The use of historical, mesoscale region specific, atmospheric Circulation Patterns (CPs) on which to condition the downscaling to gauge sites of rainfall estimates provided by RCMs
- The introduction of a recorrelation technique to ensure that the spatial structure of the downscaled rainfall is maintained; this is very important when estimating flows out of large catchments, especially for flood analysis
- The iterative improvement of the recorrelation method to ensure that the dependence structure is not biased by the relatively large number of dry days we experience all over the country; we have between 80% and 95% dry days over the bulk of the country. Without this innovation, the desired level of interdependence would not have been achieved.

Acknowledgements

The links germane to this project include the following:

- a fifteen year association with the research arm of SAWS resulting in a six year old memorandum of understanding between SAWS and the University of KwaZulu-Natal, specifically between Dr Deon Terblanche and his research team and the Satellite Applications and Hydrology Group, which has allowed a fruitful interchange of data, ideas and work, (Pegram et al., 1997, Pegram and Terblanche, 1998, Pegram and Mittermaier, 1998, Terblanche et al., 2001) which is ongoing
- a ten year period of collaboration with Prof Tom McMahon of the University of Melbourne, (McMahon et al., 2008, Peel et al., 2004, 2005, 2009, Pegram et al., 2008, resulting in parallel current research into the effect of climate change on rainfall at the seasonal scale. The Project Leader spent January 2011 working on the same. The rainfall model of Srikanthan and Pegram (2009) was the core of the code to model the daily network rainfall affected by Climate Change
- a ten year period of collaboration with Prof András Bárdossy of the University of Stuttgart, which has already resulted in three papers (Bárdossy et al., 2005; Bárdossy & Pegram, 2009 and Bárdossy & Pegram, 2011), which sets out the methodology for achieving some of the objectives of this WRC project, particularly the link between Circulation Patterns and Rainfall and the recorrelation technique.
- a three year period since the inception of the project allowed a developing working relationship with Prof Bruce Hewitson of CSAG at UCT, who has not only been instrumental in facilitating the data collection vital to the project but who has now included us as partner in a WRC project that started in 2013: ‘Managing limits in skill for seasonal climate forecasting’

We gratefully acknowledge the reference group (and their willing alternatives) for their support and contributions. They are:

Mr C Moseki	:	WRC (Chairperson)
Mr W Nomquphu	:	WRC
Prof BC Hewitson	:	UCT
Prof MJ Savage	:	UKZN
Prof R Schulze	:	UKZN
Dr F Engelbrecht	:	CSIR
Dr GC Green	:	Advisor
Dr E Everson	:	UKZN
Dr E van Wyk	:	DWA
Mr N Kroese	:	SAWS
Mr E Becker	:	SAWS
Mr T Newby	:	ARC then lately CSIR
Mr A Clulow	:	UKZN (alt Dr C Everson)
Mr C Jack	:	UCT (alt Prof BC Hewitson)

Contents

	Executive Summary	iii
	Acknowledgements	v
	List of Tables	viii
	List of Figures	ix
	List of Acronyms	xii
1	**Introduction**	1
2	**Review of Stochastic Rainfall Models**	9
3	**Tools to Detect Change**	24
4	**Assessing the adequacy of the Link between RCMs and Rainfall**	39
5	**Downscaling RCM estimates of rainfall to the gauge scale**	45
6	**Conclusion**	64
	References	67
	Appendix	71

List of Tables

Table	Caption	Page
2.1	The four sets of parameters used in simple re-scaling	10
2.2	The eight sets of parameters used in improved re-scaling	10
2.3	Transition probability matrix for the lag-one Markov chain for the 10 EL CPs of Period 1	21
3.1	Weights calculated at the 9 sites in the interior of the target square in Fig 3.8	34
4.1	labelling of PRECIS blocks in the square containing gauges	40
4.2	Labels, coordinates and calculated weights of gauges in Block 6 of Figure 4.1	41
4.3	Weights of all 48 Mpumalanga gauges in the 12 occupied blocks of Figure 4.1.	41
5.1	Two vectors (20 terms) and their Gaussianisation and recorrelation	49
5.2	List of CPs by region for the first 20 days of Period1	58

List of Figures

Label	Description	Page
1.1	Part of North America including Florida, the Great Lakes and the Rockies, as is evident in the lower right image. GCMs have resolutions intermediate between the top two images, whereas RCMs typically model at scales like the lower left image.	4
1.2	Sea Level Pressure anomalies corresponding to four selected circulation patterns, with the associated precipitation wetness index (wetness of each block conditioned on the CP relative to the areal average of the winter total) patterns for winter, over the Rhine grid during the observation period. The small green trapezium indicates the scale of the SLP data; the orange oval contains the part of the Rhine basin used in the study. (From: Bárdossy and Pegram, 2010).	6
2.1	Comparison of averaged annual ensemble statistics against observed, before (upper row) and after rescaling adjustment to the desired statistics – annual means in the left column and annual standard deviations to the right of the figure. The data used and modelled were 25 stations from the Olifants-Doring district in the Western Cape, all having record lengths of 50 years.	12
2.2	Rainfall averages (mm/day) over 25 km-square blocks in the Rhine region in Germany. Left panel: raingauge observations, averaged over each block. Right panel: estimate obtained from the HadR3M climate model before correction (Bardossy and Pegram, 2011)	13
2.3	The frequency distribution of the occurrences of the 20 CPs during the observation (1961-90) and futures (2021-50) period. The ranking is based on the observations (blue) and the frequencies of the CPs identified from the three candidate RCM in their futures runs are shown in the other colours. (Bardossy and Pegram, 2011)	15
2.4	Entropies calculated on a network of 32 stations in Germany. The entropy is calculated on constrained acute angled triangles sought from the network by search as shown on the right (Bardossy and Pegram, 2009)	19
2.5	Nine Circulation Patterns chosen by matching 700 hPa anomalies to gauge wetness over the East London region in period 1: 1990 to mid-2000; the 10th CP was the 'bucket' of unclassified CPs, not shown here.	21
2.6	The frequencies of the 10 CPs in the two periods plotted coaxially. The vertical axis is the average number of occurrences of each CP per year; the horizontal axis labels the CPs. Also plotted are the quantile bounds based on 100 simulations, with the same length as the first period, using the Markov Chain transition matrix of Table 2.1. The first 7 of the 9 series indicated in the legend are the quantiles from the Markov Chain simulations	22
3.1	Mpumalanga rain gauges. Left: 318 candidate rain gauges. Right 49 gauges remaining after checking their records manually	24
3.2	Mean Annual Precipitation map showing the SAWS numbering of 30′ squares covering the country	26
3.3	6-month moving averages of the spatial daily averages over Mpumalanga, of (i) mean rainfall and (red) (ii) number of wet gauges (blue)	26
3.4	Occurrence of wet days (47, 48 & 49 wet out of 49 stations) (blue dots) and the 6 month moving average of their counts, with linear mean (green sequence) over 10.5 years – Mpumalanga	27
3.5	6-month moving averages of the spatial daily averages over Free State showing negligible negative change in both daily spatial averages and proportion of	27

	wet/dry days over the same 10.5 year period as shown in Figure 3.3	
3.6	The first year (1990) of East London region rainfall on the 57 stations	30
3.7	Sequence of daily Wet# count in the East London region network (blue), with a 3-month moving average (black)	30
3.8	TRMM 3B42RT estimate of daily rainfall over South Africa on 21/10/2008	31
3.9	Daily rainfall estimate made by combining all radar-rainfall images (at 5-minute intervals) observed by the Bethlehem MRL5 radar on a day in 1998	31
3.10	A montage of 12 consecutive days of rain estimated by the Bethlehem MRL5 radar and formed by accumulating the images obtained at 5-minute intervals	32
3.11	Correlogram of spatial ccc versus distance for Vaal daily gauge data. The stars show the temporal (serial) daily correlation for a chosen gauge at lags of 1, 2 & 3 days	32
3.12	Correlogram of spatial ccc versus distance for Vaal daily gauge block average data. The stars show the temporal (serial) daily correlation for a chosen gauge at lags of 1, 2 & 3 days.	33
3.13	Correlogram of spatial ccc versus distance for Vaal daily PRECIS1 data. The stars show the temporal (serial) daily correlation for a chosen gauge at lags of 1, 2 & 3 days	33
3.14	Three maps of gauges in the East London area and some daily time series. Top left: a segment of the MAP map (Figure 3.1) and the SAWS 30′ numbered squares; Top right: the sites of the 57 selected (cleaned) gauges in the SAWS squares; Bottom right: the same sites on the PRECIS grid, with the target square in red, whose border length is 0.44° or 26.4′. Red diamonds will be discussed after the weights are presented in Table 3.1; Bottom left: a 45-day segment of the time series of the nine gauges (black) in the red square and their multiquadric averages (red)	34
3.15	The set of 10 CP anomaly patterns corresponding to different rainfall regimes in the Cape region, based on Sea Level Pressures	35
3.16	The correspondence between three chosen CPs and the rainfall distribution at a particular rain-gauge in the Cape region. Black line: total record; Blue line: Wettest; Red line: Driest; Green line: Heaviest.	36
3.17	A chosen set of relatively highly inter-correlated CPs, each conditioned on the rainfall for a different region. They are arranged according to the smaller table in the top left corner of the figure: Top row: East London and Mpumalanga; bottom row Free State and Upper Vaal regions	36
3.18	Cumulative frequency distribution functions (cfdfs) of rainfall over the years 1990 to 2000 for a chosen site of the Upper Vaal region. The cfdf for the whole year is shown as the Black line. The other 4 lines correspond to the occurrences of rainfall based on the CPs shown in Figure 3.17 for the 4 different regions.	37
3.19	Left: RCM modelled rainfall on a particular day after Q-Q transform [Bárdossy and Pegram (2011)]. Right: the same day's estimate after the whole dataset was subjected to a recorrelation technique. [Bárdossy and Pegram (2012)]	38
4.1	PRECIS grid and rain gauge sites – Mpumalanga. Red square is # 6, (as shown in Table 4.1) whose gauge weights appear in Table 4.2	40
4.2	Computing the location of the average of the tail of an N(0,1) distribution below point [b]. The blue cross is for a humid environment, whereas the red cross is the centroid of a semi-arid environment with a dry probability P[0] = 0.8, typical of South African conditions, whose cut-off is the red line	42
4.3	Block diagram of the recorrelation of PRECIS data to match Gauge block averages	43
4.4	Comparisons of the cccs of (i) raw PRECIS1, (ii) recorrelated PRECIS1 and then (iii) recorrelated PRECIS2 [vertical axes] against those the cccs of the gauge blocks	44

[horizontal axes]

5.1	Block averaged Raw PRECIS1 time series [blue lines] and Downscaled Block-averaged gauge series [red lines] derived from PRECIS1 in periods: March-April 1990 [days 60-120] and September-October 1990 [days 260-320].	47
5.2	Left panel: compares cccs of the Observed and the Downscaled Block daily Averages for the Vaal and Free State regions. Right panel: time series for the Free State similar to Figure 5.1	48
5.3	Example data A & B from Table 1	50
5.4	Precorrelated Gaussianised sets GAUSSA & B with a positive but small ccc	50
5.5	Decorrelated data sets at 1st iteration	51
5.6	Recorrelated data sets from Table 5.1	51
5.7	Recorrelated Gaussian Series after correction of values below p_0 threshold	52
5.8	Recorrelated Gaussian Series after 2nd iteration	52
5.9	The sequence of the original and the first 4 iterations of GAUSSB compared to the 11th iteration which is also shown last	53
5.10	Comparison of Cross Correlation Coefficients for Block Averaged values PRECIS1 in the left column and PRECIS2 on the right	56
5.11	The cccs of the downscaled PRECIS gauge estimates compared to the observations. Top left block to bottom right: Cape, East London, Vaal & Mpumalanga. In each block the three left panels are PRECIS1, the right three are PRECIS2. From top to bottom, the three panels compare uncorrected, decorrelated and recorrelated ccs of gauge estimates in the vertical against the cccs of the observations on the horizontal axes	57
5.12	Three pairs of gauges; top pair Cape, middle pair East London, bottom pair Vaal. The observations in the first and second periods are given by the Green and Purple lines; the downscaled PRECIS values labelled P1 and P2 are coloured Red and Blue respectively	59
5.13	Four fdf comparisons for gauge 228170 in the Free State, from top left to bottom right: the complete set of distributions, CP3, CP5 and CP6 dependent relationships. The legends are all coloured the same: OBS1 and 2 are Green and Purple; P1 and P2 are Red and Blue	60
5.14	Frequency distribution functions of complete records (left) and CP3 based records (right). Top row raw PRECIS1 & 2 Block 10 (OFS). Bottom row a downscaled gauge in Block 10	61
A1	Geometry of distances in triangle formed from a control point and a polygon edge	72

List of Acronyms

ARC-ISCW	Agricultural Research Council - Institute for Soil Climate and Water
ARMA	Auto Regressive Moving Average [time series model]
CCAM	Cubic Conformal Atmospheric Model
CCC	Cross Correlation Coefficient
CORDEX	Coordinated Regional Downscaling Experiment program
CP	Circulation Pattern
CSAG	Climate Systems Analysis Group
DWA	Department of Water Affairs
ECHAM	European Centre Hamburg Model (GCM)
ECMWF	European Centre for Medium-Range Weather Forecasts
EGU	European Geosciences Union
ERA Interim	The latest ECMWF global atmospheric reanalysis from 1979 to present
GCM	General Circulation Model (also known as Global Climate Model)
HadRM3	Met Office Hadley Centre's regional climate model
HMM	Hidden Markov Model
HYLARSMET	Hydrologically Consistent Land Surface Model for Soil Moisture and Evapotranspiration Modelling over Southern Africa using Remote Sensing and Meteorological Data
IAHS	International Association of Hydrological Sciences
KNN	K nearest neighbours
MAHYVA	Multidisciplinary Analysis of Hydroclimatic Variability at the Catchment Scale
NCAR	National Center for Atmospheric Research
NCEP	National Centers for Environmental Prediction
NOAA	National Oceanic and Atmospheric Administration
PEGRAIN	Multisite daily rainfall model designed for DWA by Geoff Pegram
PRECIS	An RCM ported to run on a Linux PC with a simple user interface
PRECIS1 & 2	PRECIS runs from CSAG in periods 1990-2000 and 2000-2008
QQ	Quantile-quantile [transform]
RACMO2	RCM developed by the Royal Netherlands Meteorological Institute
RCM	Regional Climate Model
REMO	RCM developed by the Max-Planck-Institut für Meteorologie, Hamburg, Germany
SAWS	South African Weather Service
SLP	Sea Level Pressure
SVD	Singular Value Decomposition
TPM	Transition Probability Matrix

1. Introduction

In the past, Water Resources design and planning was predicated on the assumption of temporal stationarity in the time series of interest – in particular, daily rainfall measured by networks of gauges – for example in the calculation of risk of flooding in urban areas. Under a changing climate, the statistics of sets of network rainfall are also likely to change; therefore a downscaling methodology should interpret and reflect what the Global Circulation Models (GCMs) and Regional Circulation Models (RCMs) are telling us. Downscaled daily rainfall network estimates must reflect the changes offered by the models combined with meaningful adjustment of their intrinsic characteristics: magnitude, spatial dependence and intermittency. These estimates may then be usefully employed by practitioners for projections and designs. The products we offer are not simulations, although the method can be adapted for use in simulation mode. What they provide, which is novel, are plausible local daily network raingauge measurements at the point scale, capturing the meaningful changes and fluctuations that are found in the large scale GCMs and their embedded RCMs.

The reason that our method succeeds where others possibly do not, is that the method preserves whatever signal of change in climate that the RCMs model, in spite of the well-known fact that these models produce biased estimates (for example, excessive amounts of low intensity drizzle), which are in turn limited to spatial scales which are larger than those gauge-scale data that practitioners find useful. The method developed here specifically removes the bias without damaging the RCMs' signal of relative change; in addition it yields gauge estimates that have realistic statistics of amounts of rainfall and spatial interdependence at all spatial scales, simultaneously capturing and preserving the changes derived from the RCMs while removing the bias.

Our product is different from other studies of future scenario modelling. Some include techniques which estimate trends from observations in the past and then try to interpret the future via a statistical simulation around the trend-line. Others (e.g. Teutschbein and Seibert, 2012) downscale GCM/RCM model predictions by direct statistical comparisons of means and standard deviations – a linear methodology which tends to destroy the signal in the RCMs and perpetuates the historical weather, as described in Chapter 2.

By contrast, we accept the offerings of the designers and keepers of GCMs: they do not claim that they are accurate in scale and bias, but they do claim that they can model differences or changes at the large scale. RCMs interpolate the GCM scale and are more detailed spatially, but they inherit the bias of the GCMs, so are still different from observations. The Engineer still needs an interpreter to be able to use this information.

We claim that we have devised techniques which do not assume stationarity of the time series of the gauge data and the RCM estimates; in particular, if the RCM 'weather' changes between periods, we aim to model the differences, but not preserve the bias. We correct the RCM bias, but try not to destroy the signal - we allow a transition from large to small scale preserving the local statistics and dependence structure. In Bárdossy and Pegram (2011), we achieved this on the RCM scale (25 km square) and produced a future which is unbiased and is valid at aggregated large spatial scales. However, this is still not what the Engineer wants. The bias-corrected and recorrelated RCM product we there devised is too different in character from the local observations provided by gauges, i.e. is typically too smooth at too coarse a scale for common hydrological applications.

By contrast, what we have achieved in this study is to produce, from an RCM, a set of gauge records which capture the changes predicted by the RCM but which have the correct point scale statistics. This is a form of rainfall information which is familiar to the Engineer and from which we have been careful to eliminate the RCMs' inherent bias yet preserve its message of possible future changes. The RCM message we want to preserve is the climate signal which includes shifts in rainfall magnitudes and spatial dependence, which are strongly linked to the RCM signal; we wish to preserve the 'good' things and eliminate the 'bad'. We reckon that our product of reconstructed rainfall gauge network values can be used by Engineers for future design scenarios with the same confidence we have in GCM/RCM predictions. We note that our methodology automatically adapts to the inherent biases in any RCMs output, so further changes/improvements in the RCMs will not affect our procedure, but will improve the final product.

The title and thrust of this research project concern daily raingauge network modelling, taking into account the effect of current climate and future possible changes. To that end, a working understanding of models of climate change is required, although it is not the main aim of the project. Therefore it is important to set out the characteristics of the climate model outputs and their effect on the daily rainfall modelling at the scale of raingauge networks over hydrological catchments.

There are two levels of climate modelling: Global Circulation Models (GCMs) and Regional Circulation Models (RCMs); the latter are typically embedded in the former. Some of these models are run in two modes: reanalysis and simulation. A quote will give the meaning of reanalysis:

> **The term re-analysis is an interesting one and generally needs clear defining. Typically a re-analysis refers to a model that is constrained not only by boundary conditions (SSTs [Sea Surface Temperatures] in the case of a global model) but also by various observational data such as station observations and vertical profiles. … The intent of this is to reproduce as best as possible the historical atmospheric states while maintaining a dynamically consistent system.**
>
> (Chris Jack, CSAG, UCT, personal communication).

Climate simulation, on the other hand, is the process of letting the GCM run freely after initializing the model with the complete set of variables, e.g. temperature, humidity, soil moisture state, pressure systems, wind patterns, etc. These are set at the beginning of a 'run', but during the simulation procedure, the model is conditioned by simultaneously modelling the hydrology to get information on evapotranspiration and albedo, while additionally coupling the GCM with an Ocean model to obtain Sea Surface Temperatures, etc. Thus in simulation mode, the variables are not corrected or conditioned on observations of atmospheric variables.

For the purpose of this study, we are interested in the outputs of the two types of Regional Circulation Model runs: reanalysis and simulation. Reanalysis runs give the historical modelled rainfall and Sea Level Pressures (SLPs) or alternatively the height of the 700 hPa pressure surface which we prefer in this study. Simulation runs give the responses of the model to future possible climates (futures scenarios) under changing greenhouse gas states of the atmosphere. The reanalysis (sometimes called control) runs provide us with information which will enable us to form the link between the observed and modelled precipitation. The futures simulation runs give us the projection of what might happen to the rainfall characteristics under a changing climate. Clearly we need the outputs of each available model running in both historical and futures mode so we can form intelligent links and draw decent conclusions about the possible changes to the rainfall characteristics.

This Chapter records the procedure followed to determine the GCMs and their nested RCMs that are suitable and available for modelling dependent rainfall over selected areas in South Africa. The adjective "selected" is carefully chosen. The reason is that local mesoscale rainfall behaviour in the Western Cape is forced by different Circulation Patterns (CPs) of atmospheric pressure than those driving the rainfall in Gauteng. Not only are there seasonal effects, the responses are also location specific due to the causative weather regime. As examples, cold fronts from the Antarctic typically affect the Cape in winter and the Intertropical Convergence Zone dominates Gauteng's summer weather.

1.1 Searching for suitable RCMs and the GCMs in which they are nested

Preamble

Global Circulation models (GCMs) provide scenarios for the possible future development of climate. Unfortunately their spatial resolution is coarse thus they cannot be used directly for the assessment of regional consequences of climate change. For this purpose downscaling methods have been developed. Regional Circulation Models (RCMs) use the output of the GCMs and provide climate variables, including daily rainfall, at a finer spatial resolution. Unfortunately these models inherit some of the biases of the GCMs (and possibly add their own).

It is the purpose of this WRC study to link daily spatial rainfall to RCM-modelled rainfall produced by reanalysis so that projected future daily rainfall amounts obtained by simulation might be estimated with minimum bias. The first step is to select suitable RCMs (and their parent GCMs), with estimates of the key variables (daily rainfall and CPs) over the Southern African region. A later step is to determine the link between Circulation Patterns (CPs), defined by Sea Level Pressure (SLP) fields obtained by reanalysis, with daily rainfall wetness patterns in a chosen geographical area. This technique is based on the work reported by Bárdossy and Pegram (2011); indeed, familiarity with the GCM/RCM pairs with European emphasis gained in the work that led to that paper allows an informed opinion to be made concerning the choices of suitable models in this study.

In a parallel study, another paper co-authored by the Project Leader (McMahon et al., 2011) deals with the estimation of global responses of rivers and reservoirs (dams) to changes in precipitation and evaporation modelled by a GCM. The time scale treated in that paper is annual, but the experience gained again feeds this study, particularly in the choice of GCM.

A land-mark study supported by the WRC was the report on the effect of Climate Change on SA Water Resources by Schulze (Ed) (2005). There the early reanalysis runs of a selected RCM are reported by Hewitson et al. (2005b), giving guidance and explanations concerning the desirable properties of such model runs. Excerpts from this work will be quoted in the sequel to illuminate the arguments leading to the choices made.

Returning to the focus of this Chapter, it seems appropriate to add to the invaluable collaboration leading to the publications quoted above, by mentioning useful contacts and discussions made at a National Workshop and three International Meetings in 2010, the first year of the project.

The workshops and meetings attended by the Project Leader in 2010, which enrich this study, were:

- The South African Weather Service Climate Change Workshop, held in Pretoria from 16th to 17th March 2010, where local and international researchers working on the subject made presentations. This meeting provided the opportunity for exchange and establishing intellectual links. Dr Deon Terblanche, Prof Bruce Hewitson, Dr Francois Engelbrecht were very helpful in offering suggestions affecting this project and, most valuably, agreed to sit on the WRC reference group guiding the project.
- The CORDEX-Africa meeting organised by the Climate Systems Analysis Group (CSAG) of the University of Cape Town took place from 15th to 16th April 2010. CORDEX stands for Coordinated Regional Downscaling Experiment. The purpose of the meeting was to plan, coordinate and share ideas about RCM studies over Africa, performed by South African and European laboratories. Here links were formed with national and international researchers. A short presentation of the downscaling methodology planned for this project was made and this elicited considerable interest. This was particularly because the effective CP-based method of double quantile-quantile transforms devised by Bárdossy and Pegram (2011) was new to the participants. In a valuable and generous gesture, an offer of a South African calculated set of RCM data was made by Prof Hewitson's team in CSAG at UCT, which has since been realised.

PRECIS data have been downloaded from Cape Town with the kind help of Ms Ruth Cerezo, who was working on these data for her PhD at CSAG.

- A presentation of the downscaling methodology (Bárdossy and Pegram, 2010) was made by the Project Leader at EGU in Vienna in May 2010, which prompted discussion with, among others, John Schaake of NOAA who runs the HEPEX study and Chris Onof of Imperial College, London, well known for his work on precipitation modelling.
- A presentation on a methodology to detect and then rectify bias using observed rainfall network data was made at the IAHS International Workshop: Advances in Statistical Hydrology, Taormina, Italy, in May 2010, entitled "Interpolation of Daily Rainfall Network observations by simulation based on Copulas and Circulation Patterns". This collaborative work with Professor Bárdossy came in useful later in the project.

Before turning to the choice of RCMs nested in GCMs, the key requirements of RCMs are listed in the next section.

1.2 Perspective on the Key Requirements of RCMs

Scale

The first issue is the scale of the models. Figure 1 [from 2007, now a bit out of date!] illustrates the problem. GCMs typically perform their computations on a mesh of points (representing the properties of areas of the same scale) which are from 100 to 400 km apart. The figure illustrates how various models "see" the topography of North America, including Florida, the Great Lakes and the Rockies.

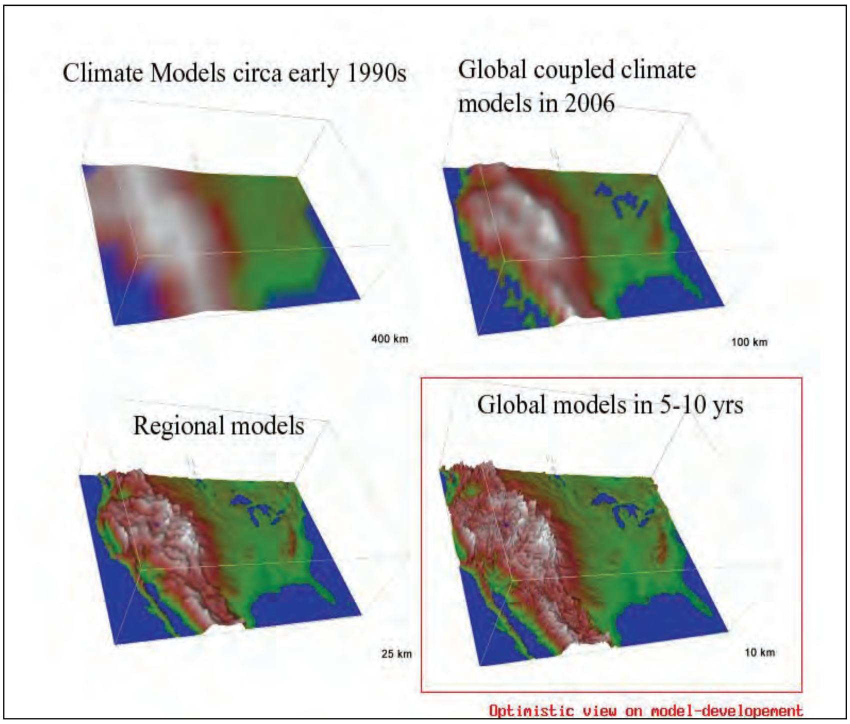


Figure 1.1 Part of North America including Florida, the Great Lakes and the Rockies, as is evident in the lower right image. GCMs have resolutions intermediate between the top two images, whereas RCMs typically model at scales like the lower left image. This figure comes from:
http://www.realclimate.org/index.php/archives/2007/05/climate-models-local-climate/

Evidently in the first panel of Figure 1 (400 km scale), the mountains have been made smooth and Florida has disappeared. The result is that local detail of meteorological variables is smeared out over large (mesoscale) regions and is intrinsically biased compared to the “truth” obtained through local observations. As the scale gets smaller in the figure, more detail becomes apparent. RCMs start to capture the local topographical features although they miss those smaller than 10 km. Even so, RCMs still have local topography problems, because they are embedded in GCMs, from which they inherit their bias, particularly in the vicinity of the interface.

From the point of view of applying the outputs of RCMs, modern hydrological models typically have spatial resolutions ranging from 1 km to 10 km and urban models have smaller scales still. Therefore the issue facing this study is: How to compare raingauge measurements in networks to the biased RCM outputs averaged on a 50 km grid? Clearly it would not be wise to work from the rainfall offered by the GCMs directly. Hence for this project the choice of GCMs is conditioned by the availability of historical rainfall output of RCMs over Southern Africa.

Bias

Regional Circulation Models in reanalysis mode produce two sets of variables of interest in this study: Sea Level Pressures (SLP) [or 700 hPa levels] and Rainfall. The SLP data obtained from reanalysis, and used to define the CPs, are conditioned on the observed data, and so the individual RCMs are all driven by the same historical SLPs on each day. The RCM data are typically computed on a grid of 0.22° to 0.44°, depending on the choice of RCM. Although they offer data at a scale which is manageable for hydrological applications, the rainfall estimates are usually biased. This is because the RCMs inherit bias in the important variables from the GCMs in which they are embedded, as a direct result of scale, as discussed above. Therefore using the RCM output uncritically is counterproductive - the RCM rainfall data have to be bias corrected using the CP-based quantile-quantile transforms discussed above.

What are needed for this process are two sets of contemporaneous rainfall data over the region of interest. The first is a reasonably large and long set of observed, error-free raingauge rainfall observations on a network. These data are then averaged over the grid of the RCM of choice a day at a time. The second set comprises RCM modelled rainfall data available at a daily frequency over recent history. The raingauge data and the RCM data are sorted according to the CP dominant on each day, in seasons, but split into calibration and validation periods. Their CP-dependent frequency distributions are extracted and stored for downscaling by performing the quantile-quantile transforms against which to compare (validate) the transforms.

Downscaling

It has been found (Bárdossy and Pegram, 2010) that the rainfall at the RCM grid scale within an area of mesoscale size (100 to 200 km square) depends strongly on the Circulation Patterns based on the SLPs (defined by fuzzy rules) optimised using simulated annealing (Bárdossy, 2010). It is therefore crucial to match local rainfall to RCM estimated rainfall conditioned on CPs. Bárdossy and Pegram (2011) found that the frequency distributions of rainfall were dependent on the CPs as shown in Figure 1.2. True, this work was done in Germany, but the ideas translate well to Southern Africa.

The panels in Figure 1.2 show, from top to bottom, that the wind directions driven by the circulation patterns (CPs) affect the rainfall patterns. On the right of each panel, the average daily rainfall during winter over the Rhine region (within the orange oval on the CP plot) on each 25 km square, depends uniquely on the CP. To obtain this kind of match, it is essential that RCMs produce plausible historical rainfall sequences that can be matched in quantile-quantile plots to perform the required adjustment to remove the bias. Once this has been done with the historically based RCM outputs, then the RCM rainfall produced in the futures scenarios can be downscaled using the modelled CPs. If the historical data are not available (i.e. if the offerings are limited to simulations) there is no link to work from.

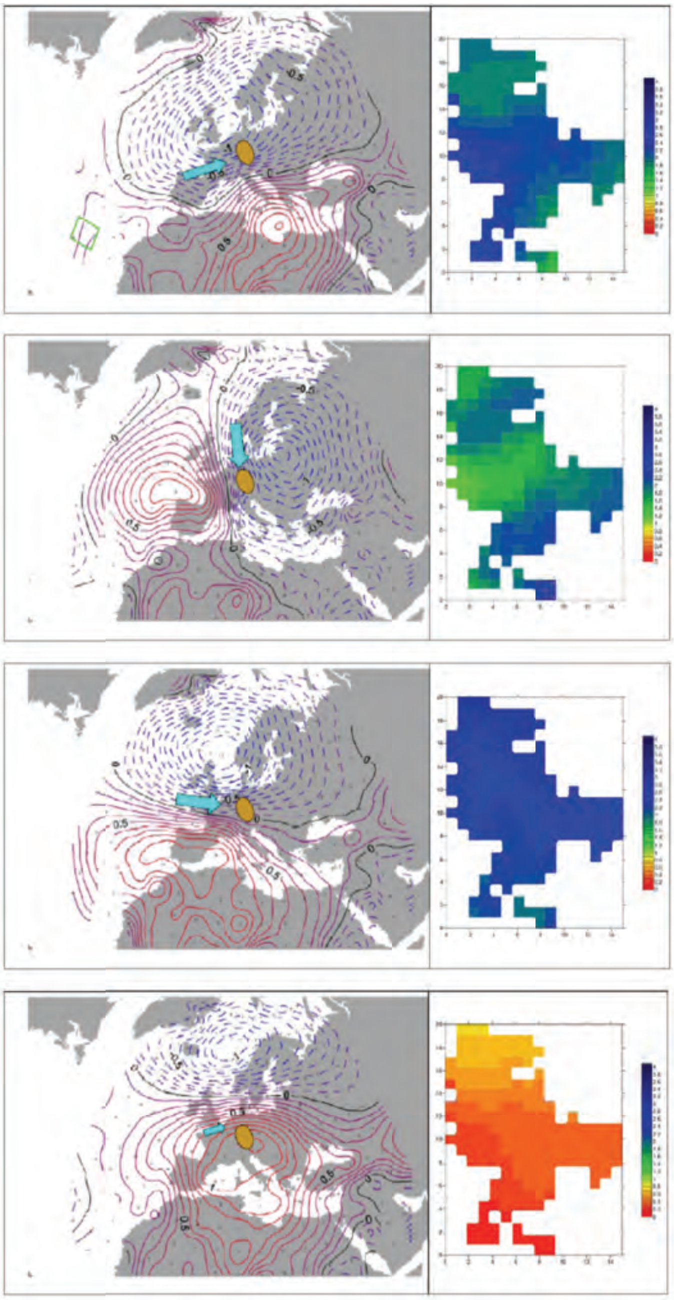


Figure1.2. Sea Level Pressure anomalies corresponding to four selected circulation patterns, with the associated precipitation wetness index (wetness of each block conditioned on the CP relative to the areal average of the winter total) patterns for winter, over the Rhine grid during the observation period. The small green trapezium indicates the scale of the SLP data; the orange oval contains the part of the Rhine basin used in the study. (From: Bárdossy and Pegram, 2010).

1.3 Choice of GCMs and their embedded RCM combinations

In Europe, RCMs available for CP-downscaling work include the following three, which were utilized by Bárdossy and Pegram, (2010):

- HadRM3:- developed by Hadley centre for Climate prediction and research, Met Office UK and embedded in the GCM: HadAM3h
- RACMO2:- developed by the Royal Netherlands Meteorological Institute, KNMI and embedded in the GCM: ECMWF cycle 23 release 4
- REMO:- developed by the Max-Planck-Institute fur Meteorologie, MPI, embedded in the GCM: ECHAM5/MPIOM.

For RCM control runs contemporaneous with the observed rainfall data, these had a common time period of 1961-1990 while for transient futures runs, time periods during 2001-2099 were considered. It is worth noting the genealogy of the GCMs. ECHAM is a 'child' of ECMWF and the MPI group's forcing variables, because it is their adaptation of ECMWF at Hamburg. This became evident in the analysis of the Rhine basin. The outputs of the two RCMs' (REMO & RACMO) which were dependent on ECMWF were much closer to each other than to HadRM3, both numerically and in their patterns.

To continue the list:

- The HADCAM GCM (Jones et al., 2006) was the only one used by McMahon et al. (2011), as a proof of concept. It is the successor to HadAM, but to our knowledge, has not yet been downscaled over Africa.
- PRECIS was used by Hewitson (2005b) to "downscale HadAM3 GCM, the Hadley Centre global atmospheric model. The lower boundary forcing was provided by observed SSTs for the control and future SSTs are calculated as the HadCM3 (coupled model)."

The RCM reanalysis data available over the Southern African region is the RCM output based on the ERA-Interim GCM, using the PRECIS software developed by the Hadley Centre. The work is performed by CSAG at UCT. According to the Hadley Centre website: http://www.ecmwf.int/research/era/do/get/era-interim , "ERA_Interim ... is an 'interim' reanalysis of the period 1989-present …", although, as cautioned by Chris Jack (personal communication): "PRECIS is not what we would call a re-analysis model in that it is not constrained by anything other than the lateral boundary conditions. The interior of the model is 'free' in the sense that the model dynamics are free to develop with no restrictions other than remaining dynamically consistent with the lateral boundaries." This is important to the study, because there is no conditioning on rainfall data in PRECIS.

At the early stage of the project, to enable us to get a start, the collected wisdom of the project's reference group and others consulted about RCMs suitable for application in Southern Africa, resulted in the following list of 3 RCMs: **NCEP, PRECIS** & **CCAM**.

- The **NCEP**/NCAR reanalysis over Southern Africa was chosen for identifying the CPs. It produces Sea Level Pressures on a daily basis during the period 1949-2008, at a grid scale of 2.5° and has (in a preliminary analysis in this project) already been used to identify selected CPs associated with daily rainfall over Mhlatuze. To recapitulate the analysis, this set of CPs is crucial for identifying the historical rainfall scenarios over selected mesoscale areas. Once the CPs were identified by their fuzzy rules, these rules were stored for CP identification purposes, both in the historical rainfall matching and in the futures scenarios. It is thus essential that RCMs are run in free mode so we can compare their daily rainfall estimates against the observed patterns on the same days, conditioned on the same CPs; this requirement limited appropriate candidates.
- As mentioned in the previous paragraph **PRECIS** is an RCM, embedded within the ERA-Interim Reanalysis GCM from ECMWF. The RCM was created at CSAG in UCT and data were made available to this study. The period available was 1990-2008, giving Sea Level Pressures &

Rainfall on a daily basis at a grid scale of 0.44° [we subsequently asked for and obtained 700 hPa data for the same period].

- Output from **CCAM**, the Conformal-Cubic Atmospheric Model with a grid of 0.5° is available from Dr Engelbrecht of CSIR. He mentioned at the project's Reference Group meeting in August 2010 that CCAM can be run in Ensemble mode, which is valuable for estimating precision. A follow-up visit to Dr Engelbrecht by the Project Leader to Pretoria in September 2010 elicited the information that the Ensemble would comprise CCAM being run in RCM mode driven by 5 (possibly 6) separate GCMs. Unfortunately CCAM outputs are *simulations* which are run to mimic historical climate, not match daily weather, therefore the Ensemble cannot be used in CP-rainfall analysis mode. However, it might be useful to use the CCAM output as simulations of the historical and future climates and compare the resulting rainfall statistics against the PRECIS downscaling. The individual Ensemble runs which come from different GCMs, although downscaled using CCAM as the RCM, are likely to provide a range of results, which will provide a measure of confidence in the model outputs, in a probabilistic sense. In retrospect, this data set was not obtained.

These (**NCEP & PRECIS**) were the models selected for this study. As noted by Hewitson (2005a) in section 2.6:

> "Regional climate change projections can now readily be produced at a range of spatial and temporal detail, down to the point scale on sub-daily temporal resolutions. However, despite the detail of the projections, as a definitive *forecast* of climate change these projections are, at best, probabilistic indicators of change. At present, it is not possible to regard the magnitude of change from any one projection as more likely than another. There remain notable uncertainties around forcing, understanding of the science, and the limitations of the regionalisation tools."

Therefore it would have been useful to employ more than one RCM for making useful projected or futures comparisons, as was done by Bárdossy and Pegram (2011), where three were used. Only then is one likely to have a sense of the precision of the results in a probabilistic sense. However, this was not done here as it was outside the specification of the project.

1.4 Summary

The identification of suitable RCM candidates for CP-downscaling was achieved. The ingredients for the CP-rainfall identifications and comparisons were available: NCAR data was stored on the Project Leader's computer with the PRECIS data. As a result of a visit of Prof Bárdossy to Durban in early 2011, CPs were identified for each of the 5 chosen regions in South Africa, matching CPs with rainfall regimes.

2. Review of Stochastic Rainfall Models

Preamble

Future Climate behaviour is modelled by General Circulation Models (GCMs) at scales of the order of 100 km. To obtain more detail, Regional Circulation Models (RCMs) are embedded in GCMs to estimate meteorological variables at scales as small as 25 km. The RCM output of variables of interest to hydrologists and water resources planners (rainfall, temperature, humidity, etc.) are presented as averages over the 25 km blocks. In contrast, daily rainfall measured by networks of raingauges (usually spaced from 10 to 20 km apart) gives more detail of intermittency and extremes than do the RCM estimates.

When RCM block estimates of rainfall are compared with contemporaneous rainfall network averages over the same blocks, they typically differ. In addition, RCMs tend to model more drizzle. Statistical measures of the spatial estimates averaged over months or seasons, like means and variances, also differ for GCMs and observations during the observation period. In futures scenarios modelled by GCMs and RCMs we do not have any measured data to check against, so it is important to do two sets of scaling. The first is to find the statistical relationships between the GCM rainfall estimates and those of the Network during the observation period over a given area and secondly, to use these links to downscale GCM/RCM futures rainfall products to the Network scale. How do we proceed to model futures rainfall once this link is available?

To produce reasonable simulations of future rainfall, a stochastic daily rainfall network model must be able to do two things: (i) replicate the *observed rainfall* behaviour faithfully in terms of intermittency, frequency of amounts of rain across the spectrum and in addition, temporal and spatial dependence, (ii) be able to be adapted to model *future rainfall* scenarios and capture the essential behaviour that it does in the observed period; the model must be amenable to having its parameters altered at will to suit a given rainfall regime.

It is the purpose of this chapter is to document the description of suitable stochastic daily rainfall network models which can be adapted to partner the GCMs and RCMs in estimating network rainfall in the future. Not only must they be able to model observed rainfall faithfully, they must be able to be configured to model future rainfall with confidence. To determine which models are suitable for this partnership, it is necessary to develop the desiderata from simple initial steps through to the appropriate level of complexity. Section 2.2 lays out the problem. Section 2.3 develops the solution based on an historical perspective obtained from the technical literature. Section 2.4 gives more detail of the mathematical constructs which show the way to the candidate rainfall models. Section 2.5 uses these ideas to suggest the way forward to futures rainfall modelling.

2.2. Initial steps at rescaling

To start with, consider information of rainfall observed at raingauges and estimated by RCMs gathered on an annual basis. Simple scaling of a simulated historical sequence using a naïve approach would proceed as follows, based on the information summarised in Table 2.1 which lists four sets of parameters: m_H, s_H, m_F, s_F; e.g., the sets of areal values of historical annual mean are given collectively as m_H. We start by simulating the observed gauge data with the chosen rainfall model using the historical parameters and check that the model can replicate the observations adequately. Only then would we rescale the model output using the GCM futures parameters.

Table 2.1. The four sets of parameters used in simple re-scaling

Historical data		Future GCM values	
Mean	Standard Deviation	Mean	Standard Deviation
m_H	s_H	m_F	s_F

The simple rescaling would involve the following two-step operation:

(i) Simulate the historical sequence then standardize a sequence of the ensembles:
  Standardized simulated historical sequence = [Simulated historical record - m_H]/ s_H
(ii) Rescale the simulated historical sequence by futures statistics:
  Futures rainfall estimates = [Standardized simulated historical sequence].$s_F + m_F$

This simple shift and scale procedure would ignore an important missing link – there is a mismatch of scales. What is missing is the initial adjustment of the simulated historical sequence by the GCM control-run (observation period) estimated rainfall outputs. This extra step is necessary because the spatial scales of the Network and GCMs are markedly different. To be fair, the raw GCM estimates are not used directly, as intermediate Regional Climate Models (RCMs) are frequently used to refine the estimates. For the purpose of discussion in this section, we use the acronym GCM as a catch-all implying numerical physically based climate modelling. Thus to make the necessary improvement there should be another two steps inserted in Table 2.1. These are shown in Table 2.2, symbolically following the flow of the arrows.

Table 2.2. The eight sets of parameters used in improved re-scaling

	Historical data		Future values (ensembles)	
	Mean	Standard Deviation	Mean	Standard Deviation
Network	m_{NH} $\Downarrow$ para	s_{NH}	m_{NF}	s_{NF} para $\Uparrow$
GCM	m_{GH}	s_{GH} physics $\Rightarrow$	m_{GF}	s_{GF}

The operation is as follows: take the historical data, and determine its parameters (moments, etc.). Do the same with the GCM 'data' during the historical period and determine their correspondence: symbolically, determine T such that $G_H T \Rightarrow N_H$, which translates to "transform the historical GCM values G_H to Network values N_H, using the transform T". Technically this is downscaling.

Next, trusting the physically-based computational link between the Historical and Future GCM calculations, determine the new parameters from the futures GCM and call the set G_F. Determine the future Network parameters from the future GCM set using the ***original*** transform T:

$$G_F T \Rightarrow N_F,$$

yielding m_{NF} & s_{NF}.

In a proper application of downscaling the future GCM outputs, there are more parameters than simple annual means and variances that need to be dealt with; in particular, intermittency and spatial dependence structures have not been included here and would need to be parameterized. Nevertheless, this methodology can be extended to more detail by using seasonal or monthly statistics (in addition to annual ones), or as we prefer CP-based parameters, to rescale. These will have to be transformed into parameters that drive the rainfall network models; in order to understand what is required we proceed to examine stochastic daily rainfall network models in some detail.

2.2.1 Candidate multisite daily rainfall models

Srikanthan and McMahon (2001) gave a thorough review of rainfall models, from annual down to daily and for single and multiple sites. For modelling rainfall at individual sites, they listed four types of model:

- two-part models (a Markov chain model for the occurrence of dry and wet days and a serially correlated model for the generation of rainfall amount on wet days)
- transition probability matrix models (a multi-state Markov chain model with a Uniform distribution for each of the wet states except for the last, for which an Exponential distribution was used)
- resampling models (marginal, joint and conditional probability densities of interest such as dry spell length, wet spell length, precipitation amount and wet spell length given prior to dry spell length, are estimated non-parametrically using at-site data and kernel probability density estimators)
- time series models of the ARMA type (a first-order Autoregressive model generates standardised daily rainfall data and uses a threshold to truncate the low values to dry days)

For multisite modelling they listed three techniques:

- conditional models such as the hidden Markov chain
- extension of two part Markov chain Autoregressive models
- random cascade models.

and came to the conclusion that the extended Markov chain models seemed to have merit but were rather complex; they tended to favour the random cascade models.

Following on from that review, Mehrotra et al. (2005) later compared up-to-date examples of the three multisite models: the parametric Hidden Markov Model (HMM), a first order multi-site Markov stochastic precipitation generation model (proposed by Wilks, 1998) and a nonparametric *k*-nearest neighbour (KNN) random cascade model. The HMM aims to preserve the spatio-temporal precipitation distribution by conditionally generating the response; this in turn depends on a hidden discrete weather state that is Markovian in nature. The Wilks-type model simulates precipitation at multiple sites through the assumption of a first order Markov chain at each location; the spatial dependence is maintained by using binary random innovations that are dependent across space and the amounts process is modelled by a multivariate Autoregressive process. The k-nearest neighbour approach reproduces the spatial precipitation distribution structure by simulating precipitation occurrences jointly at multiple locations; temporal persistence is preserved through Markovian assumptions on the rainfall occurrence process. The three methods were evaluated by Mehrotra et al. (2005) for their ability to model spatial and temporal dependence in the rainfall field. Results indicated that all three approaches were successful in reproducing the spatial pattern of the multi-site rainfall occurrence field. However, overall the Wilks model was found to produce better results in comparison to the KNN and HMM models, and the authors selected the extended Markov chain model of Wilks (1998) adapted by Srikanthan (2005) as having the most promise.

Srikanthan's (2005) model extended Wilks's model by using a technique of rescaling, which is the important key in this study. Srikanthan's concept of rescaling was done by standardization of monthly and annual aggregations of the daily rainfall. The model captured the first two moments of the amounts process well and also reproduced the temporal structure faithfully. However it was not designed to preserve the spatial dependence at the monthly and annual scales after aggregation of the daily values. The model was subsequently adapted by Pegram in Srikanthan and Pegram (2009) to enable post-conditioning of the spatial as well as the temporal covariance structure. This modification puts it well ahead of the other two types of model tested by Mehrotra et al. (2005). Because it can in addition be easily adapted to accept futures statistics in its parameterization, it is thus the covariance model of choice in this study. To indicate the efficacy of the procedure, here follow four images in Figure 2.1, not appearing in Srikanthan and Pegram (2009).

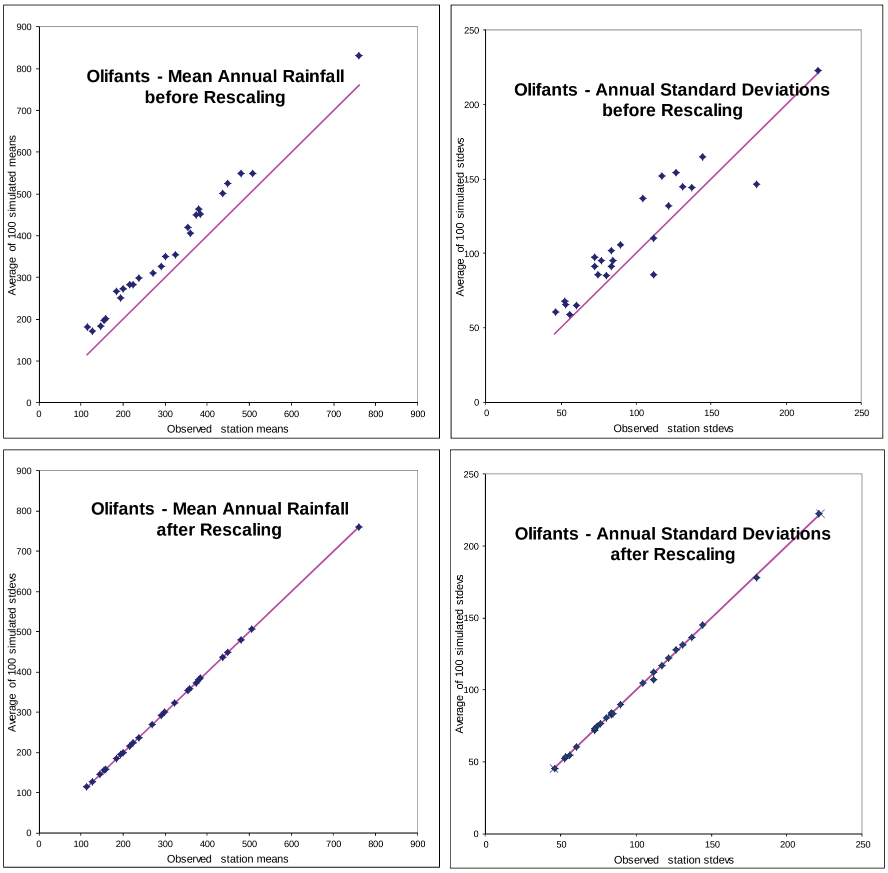


Figure 2.1. Comparison of averaged annual ensemble statistics against observed, before (upper row) and after rescaling adjustment to the desired statistics – annual means in the left column and annual standard deviations to the right of the figure. The data used and modelled were 25 stations from the Olifants-Doring district in the Western Cape, all having record lengths of 50 years.

As a South African application of the model, the four panels show the results of generating and rescaling the annual totals of 25 stations in the Olifants-Doring basin in the Western Cape – the upper row shows the statistics of the unscaled ensembles, the lower row shows the result of rescaling on annual parameters.

In each panel the horizontal axis gives the statistic of the observations (annual mean to the left or standard deviation in the right panels); the vertical axes are the means of the 100 ensemble members' relevant statistic, obtained from generated sets with the same record length as the observed – in this case 50 years long.

It will be observed that the rescaling methodology is entirely successful. Srikanthan and Pegram (2009) give a full set of results for 30 stations with record lengths of 43 years sited around Sydney, Australia. The Sydney data exhibit less seasonality, but were primarily chosen because of the very careful screening and error correction of the data by the Bureau of Meteorology, for whom Dr Srikanthan works.

2.2.2 Application of these models to climate change studies

When it comes to modelling *future* network rainfall, it is now "common knowledge" (based on the outputs of several suites of GCMs) that there are going to be changes in the occurrences and amounts of rainfall in different locations. This means that there may be longer dry (or wet) spells in some regions and there is the added threat of more severe and more frequent convective storms when it rains in the drier parts, increasing erosion and the risk of flash floods. Thus blanket rescaling (shifting means and changing the spread of accumulated rainfalls) as suggested in Table 2.1 will be unable to change the modelled rainfalls in the correct way. Something else has to be done.

As a point of departure, we start with the downscaling method using the T-transform as outlined in the discussion on Table 2.2, changing the parameters of the distributions of the simulated future daily network rainfall amounts in such a way as to give the correct monthly averaged daily means and variances. This will be legitimate if the occurrence process (wet/dry sequencing) is the same in the future as the past. It is also acceptable as a strategy if the relative spatial distribution of very wet days and very dry days (as measured by the proportion of the number of wet gauges in the network on a given day) is maintained. A recent study suggests that current wetness will not be maintained in future (Bardossy and Pegram, 2011) and that RCMs do not succeed in maintaining the correct spatial patterns, as shown in the example in Figure 2.2.

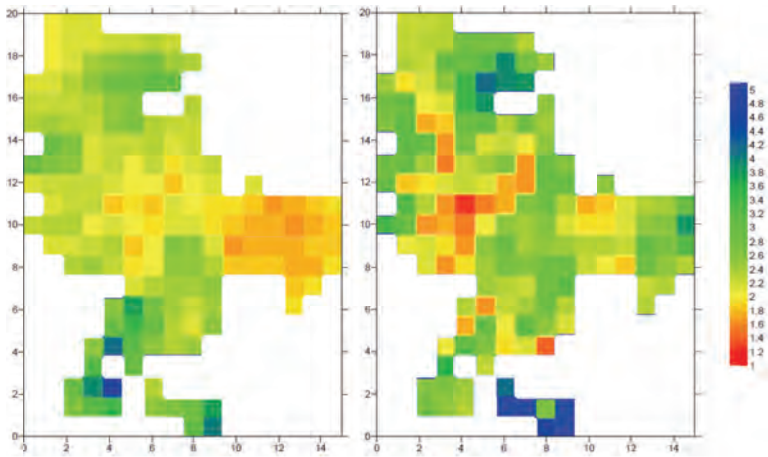


Figure 2.2. Rainfall averages (mm/day) over 25 km-square blocks in the Rhine region in Germany. Left panel: raingauge observations, averaged over each block. Right panel: estimate obtained from the HadR3M climate model before correction (Bardossy and Pegram, 2011).

Figure 2.2 has two panels showing rainfall averages over 25 km-square blocks in the Rhine region in Germany, taken from Bardossy and Pegram (2011). They are mean annual rainfall estimates during the 30 year period 1961 to 1990. The one on the left is calculated from raingauge observations, averaged over each block. The panel on the right shows the estimate obtained from the HadR3M climate model. The model gets parts right in general, but not in detail. The scale is in mm/day. What do we need to do to the rainfall network models to enable them to model the rainfall correctly? Three things have to happen in modelling the future:

1. the proportion and sequencing of wet and dry days at a gauge in a season/month have to be right
2. the resultant marginal distributions of the amount of rain falling on wet days at a site in a season/month has to be right
3. the proper amount of isolated convective events, frontal systems and widespread stratiform events, the proportion and clustering of the wet gauges on the network on the raining days has to be right.

How are these objectives to be achieved?

The link to modelling the future daily network rainfall is through the GCMs, via the RCMs to the networks, as already emphasised. The underlying assumption is that, unlike the statistics of the observations and estimates, the *physics* governing the generation of rainfall in the future are immutable – they will be the same as they always have been. The adjustment has to be made to the parameters of the rainfall models after downscaling the relevant statistics.

To perform downscaling, some modellers have concentrated on the probability distributions of the amounts of rain at each gauge to adjust the mean and variance of the accumulated (monthly/seasonal) amounts. In those cases, the models of future intermittency have not been adjusted – the rainfall amounts processes alone are given the statistics of the futures GCM/RCM outputs. The historical proportion and sequencing of wet and dry spells is not altered.

This is a weakness that we feel can be addressed, as will be explained in Section 2.5. What we will show is that, with recent research results, we are now in a position to adjust not only the marginal distributions of the amounts but also the statistics of the occurrence processes which are strongly linked to the prevailing weather, so that we can adjust the accumulated rainfalls to suit.

2.3. The way forward, based on an historical perspective

It has been known for at least 40 years (Lamb, 1972) that there is a strong link between atmospheric pressure systems (as indicated by maps of Sea Level Pressures) on the one hand and the character of local (regional) rainfall, on the other. Recent work by many researchers has exploited this relationship to systematise the link between rainfall behaviour and Sea Level Pressures (SLPs). In particular, the automatic (non-subjective) association between rainfall behaviour in a region and SLP patterns defined by fuzzy rules and simulated annealing has been successfully applied in Europe to downscale futures rainfall (Bardossy and Pegram, 2011). An example of the downscaling is given in Figure 2.3, where the annual frequency of the 20 CPs during the observation and future period are compared. The ordering is on the increasing rank of the frequency of the CPs occurring in the observation period. It is seen that only three RCM future CPs (CP01, CP11, indicated by red arrows and, to a lesser extent, CP19) show material differences from their frequencies during the observation period.

What that study demonstrated is that the frequency of the various SLP patterns during *historical* and *future modelled periods* is substantially different. This means that the frequency, thus arrival rates, of the various weather types are different, suggesting differences in the rainfall occurrences processes; this is in addition to the amounts and spatial distributions. It is true that meteorological variables other than rainfall (e.g. temperature, humidity and energy) conditioned on SLPs will not necessarily be the same in the two periods. However, the modelling of the physical processes producing rain are believed to be consistent in the past, current and future periods.

What Bardossy and Pegram (2011) were led to do was to define a double quantile-quantile transform of rainfall marginal distributions, on each pixel of the RCM's rainfall output, to modify the future rainfall on the target region. The intermittency was not addressed in that work.

Why is all this important? In addition to defining the rainfall marginal distributions of rainfall *amounts* on each of the 25 km-square areas, we consider that there is a useful link between (i) the frequencies and sequencing of the individual identified CPs and (ii) the local *occurrences* behaviour on the ground; it remains to determine this linkage and exploit it. Once we have the correct temporal and spatial distribution of wet spells as suggested by the CPs, we will be able to modify the rainfall occurrence parameters in the PEGRAIN model, simultaneously adjusting the parameters of the marginal distributions of the raining amounts, to match the rescaled mean and variance produced by the GCM through the RCM of choice.

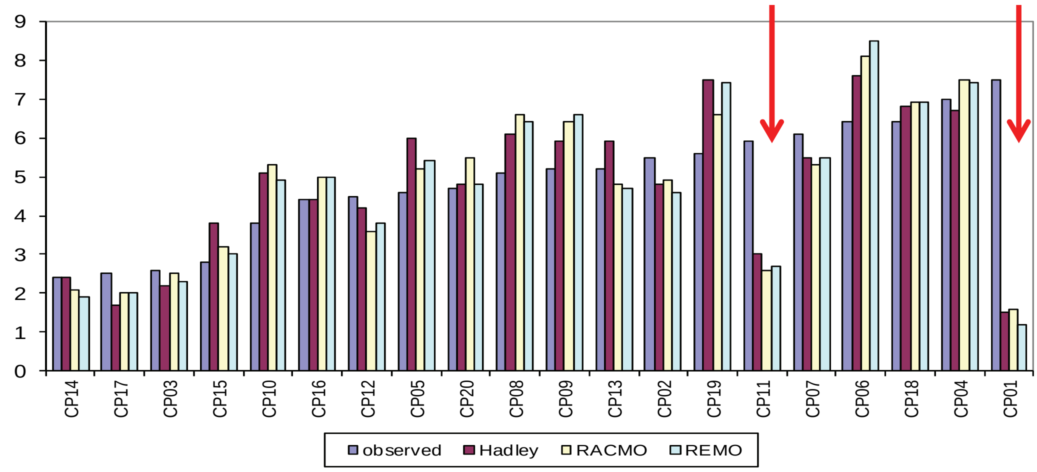


Figure 2.3. The frequency distribution of the occurrences of the 20 CPs during the observation (1961-90) and futures (2021-50) period. The ranking is based on the observations (blue) and the frequencies of the CPs identified from the three candidate RCM in their futures runs are shown in the other colours. (Bardossy and Pegram, 2011).

2.4. The mathematics linking CPs to rainfall occurrences

2.4.1 The occurrence process

In order to produce a stochastic rainfall model of the future we need, for each pair of raingauges (say 1 and 2), the following information, following the notation of Katz (1983):

$p_1(1)$, $p_1(0)$, d_1; $p_2(1)$, $p_2(0)$, d_2 and ρ_{12},

where the subscripts refer to the gauges; $p(1)$ and $p(0)$ are the probabilities that the current day is wet given that the previous day was respectively wet or dry and

$$d = p(1) - p(0) \qquad [1]$$

is the binary serial correlation or temporal dependence parameter. ρ_{12} is the spatial binary cross-correlation coefficient between the gauge records.

The unconditional probability of rain on a day is then derived as:

$$P = p(0)/(1-d) \qquad [2]$$

2.4.2 Extending the ideas to the amounts process including the occurrences.

Typically, but not exclusively, the literature suggests that the marginal distribution of daily rainfall Y (excluding dry days) has a Gamma distribution, specified by the shape and scale parameters α and β, so for the rest of the development, we will continue with that choice. However it is interesting to note in passing that Wilks (1992) found that the double exponential model suggested by Woolhiser and Pegram (1979) was a superior model in his application. If we turn to the Gamma distribution (because it appears in the key equations [8] and [9] to come, chosen by Srikanthan, 2009), its mean and variance, in terms of the parameters, are $\mu = \alpha\beta$ and $\sigma^2 = \alpha\beta^2$. If $\alpha = 1$, the Gamma distribution specializes to the Exponential distribution.

We now turn to the occurrences process. At a given gauge site, or station, S_n is the number of wet days in a sequence of n days (n = 28, 29, 30 or 31 for months). S_n has mean and variance given by

$$E[S_n] = nP \quad [3]$$

$$\text{var}[S_n] \approx nP(1-P)(1+d)/(1-d) \quad [4]$$

This is a result that seems to have first been given by Markov (1924), who gave his name to the discrete state model with limited temporal dependence, of which this lag-one binary model is a special case. We will use the label 'Markov chain' throughout the discussion which follows to specify the lag-one binary chain unless otherwise qualified. The Gaussian approximation to the variance presented here is an application of the Central Limit theorem. Gabriel (1959) showed that equation [4] was adequate for most cases of practical situations in daily rainfall modelling, in that it was close to the exact expression that he derived for a stationary Markov chain:

$$\text{var}[S_n] = P(1-P)(1-2P)[n(1+d)f - 2df^2(1-d^n) \quad [5]$$

where $f = 1/(1-d)$.

Incidentally, drawing from the work producing these results, Gabriel and Neumann (1962) [Neumann is frequently given incorrectly as Newman in references] published their seminal paper on applications of Markov chains to daily rainfall in Tel Aviv. They showed that, although there is a weak dependence on days prior to the previous, the model was sufficient (statistically) and yielded a variety of derived statistics from the two parameters $p(1)$ and $p(0)$. Woolhiser and Pegram (1979), amongst others, also demonstrated this notion and developed a methodology for fitting daily varying parameters governed by Fourier series. In spite of the fact that Pegram (1980) showed that there were legitimate cases where a longer lag (up to 3) was appropriate, the Fourier based lag-1 model was adopted by Zucchini and Adamson (1984) in their study of South African rainfall for the Water Research Commission.

The lag-1 Markov chain with seasonally varying parameters has been adopted as the default, frequently without checking, by many researchers since. A notable exception was that of Katz and Parlange (1998) who examined the efficacy of Markov chains with lags up to 4 in an attempt to counter the effect of the model simulations underestimating the variance of the modelled rainfall, with some success. Srikanthan and Pegram (2009) used a lag-4 Markov chain to ensure that the distribution of the dry spell durations was properly modelled in a multisite stochastic daily rainfall model – a lag-1 model can only generate wet and dry spells with geometrically distributed lengths – they found that some dry spells had a hypergeometric distribution.

To summarise this subsection, we specify the construct of the covariance/Markov model of choice. Its description of the rainfall recorded at individual stations is parameterised by (i) the Gamma distribution for amounts and (ii) the Markov chain for occurrences.

2.4.3 The methodologies used for rescaling

In the early 1980s Climate Change researchers started looking for a link between the statistics of daily rainfall and those of monthly totals. Katz (1983) came up with an elegant solution based on Gabriel's (1959) results, which provided modellers with the armoury to generate daily rainfalls at individual stations with chosen monthly rainfall characteristics. He put it in the following way.

Following the definition above in Section 2.4.2, let Y be the rainfall depth occurring on a wet day, a process assumed to be stationary in a month or possibly a season. Let the mean and variance of Y be $\mu = E[Y]$ and $\sigma^2 = \text{var}[Y]$. Let the total observed precipitation during n days, of the season or month, be

W_n accumulated during the S_n wet days. Based on a result by Feller (1968), Katz (1983) used Gabriel's (1959) equations [3] and [4] to show that:

$$E[W_n] = nP\mu \quad [6]$$

$$\text{var}[W_n] \approx nP[\sigma^2 + \mu^2(1\text{-}P)(1+d)/(1\text{-}d)] \quad [7]$$

It seems we have a link between the parameters of the rainfall model (at least of a single gauge) and the n-day rainfall depth statistics. We do, but there are 4 of the former parameters (μ, σ^2, P and d) and only 2 of the latter ($E[W_n]$ and $\text{var}[W_n]$) in 2 equations. Assumptions have to be made about 2 of the 4 model parameters.

We turn to the seminal paper by Richardson (1981), who seems to have been the first to jointly model the three important daily meteorological variables: rainfall, temperature and radiation. [Incidentally, Richardson extensively used the Fourier series model-fitting strategy of seasonally varying Markov chains suggested by Woolhiser and Pegram (1979).] Richardson used a combination of Markov chains and multivariate Autoregressive (lag-1) model based on the seminal work of Matalas (1967) to generate his ensembles and successfully recaptured the serial and spatial covariance structure of the observations. This he did by generating the rainfall at the chosen site and then generated the other variables conditioned on the wet or dry status of the day. He showed that the serial and cross correlation structure of temperature (max & min) and radiation were strongly dependent on the wet/dry process.

Based on Richardson's (1981) work and Katz's (1983) result in equations [6] and [7], Wilks (1992) suggested the following strategy: fix the occurrence probability P and dependence d at the historically observed values, which means that the rescaled generated rainfall occurrences will behave the same as the observed ones. Using this ploy, it is straightforward to solve for μ and σ^2 using equations [6] and [7], which fixes the parameters of the marginal distribution of the rainfall amounts. It remains to find a way to specify P and d independently, using the statistics of the GCMs.

2.4.4 The development of multisite daily rainfall models

During a sabbatical with Prof Ezio Todini in Bologna, Pegram (1996) derived a good approximate relationship between the binary wet/dry occurrence processes of m gauges in a network. To elaborate: an m-site binary switching process is defined by 2^m parameters, which means that a network of m=10 gauges needs 1024 and a 30-gauge network requires a princely 1 073 741 824 parameters to define the high order joint probabilities of the different combinations of gauges being wet or dry. Pegram showed that the multinormal distribution, used as a "hidden covariance" model, could replicate these relationships in rainfall fields to a high degree of accuracy with only $m(m+1)/2$ parameters so that the 10 and 30 gauge networks only need 55 and 465 parameters respectively – much more manageable! The theoretical advantage in using the multinormal distribution is that all joint probabilities in m-dimensional space are defined by only $m(m+1)/2$ parameters, because the higher moments of the Normal distribution are completely defined by the second order moments. This result was published in conference proceedings by Srikanthan and Pegram (2006).

In the meantime, Wilks (1998) produced a multisite stochastic daily rainfall generating model based on the same idea of the hidden covariance model, not implemented in exactly the same way, but there were difficulties in capturing the correct mean and variance of the modelled monthly and annual values. The well-known problem in the process of stochastically modelling short term hydrometeorological variables is that the variances and the correlations of accumulations in the original series are not recovered unless special steps are taken. Based on Wilks's (1998) paper, Srikanthan (2004) devised a method to post-dress the generated daily values to recover the moments and serial correlations of the accumulations, but he did not manage to recover the spatial correlations.

After an exchange of ideas with Pegram, a joint paper was published which solved this problem: Srikanthan and Pegram (2009). Pegram (2010) made subsequent modifications to the model, adapting it to South African conditions for the Department of Water Affairs, and called it PEGRAIN. This model uses post-dressing techniques to ensure that all the historical moments up to second order are recovered as well as possible. In this model, it is straightforward to modify the daily *occurrences* process, which the model successfully reproduces. Effectively, PEGRAIN accomplishes what Wilks (1992) suggested for individual gauges, but extends the process to the whole network, maintaining the serial and spatial dependence structures at all levels of accumulation.

In addition to being a quantum improvement to the model of choice of Mehrotra et al. (2005), PEGRAIN is the model chosen for application in this study, as it is driven by multisite Markov chains and the wet amounts are distributed as Gamma, both specified by the parameters in equations [8] and [9] appearing in Section 2.5 following. Most importantly, besides being the best model for the observation period, it has the facility to be reconfigured to perform temporal and spatial rescaling to ensure the correct first and second order moments of the futures rainfall.

During the development of the PEGRAIN model, an important feature of network rainfall was determined, namely the wetness of the network on individual days (Pegram, 2010). It was found that the wetness (proportion of network gauges recording rain on a particular day) was an important neglected parameter. The wetness is a function of the local climate and season. It is a function of the physical rainfall generation mechanism; convective, frontal or stratiform rainfall systems. These in turn are dependent (admittedly not exclusively) on the CPs, which should give an indication of what weather type is prevailing locally. Pegram (2010) has already done work to adapt the PEGRAIN model to be driven by wetness indices with considerable initial success. It is competitive with the standard model of Srikanthan and Pegram (2009) in terms of its ability to recapture observed statistics.

To elaborate, conditioned on the wetness, the mean, standard deviation and interstation correlation length all increase. This indicates that the distribution of rainfall at an individual station on drier days is different to those on wetter days. This suggests that the cross correlation between stations is not independent of magnitude, as suggested by Gaussian models. This led to the development of a method of modelling the correct wetness by use of a copula model (Bardossy and Pegram, 2009). More difficult to devise (and possibly to adapt) to changes in occurrences as well as amounts probabilities, this copula-based model supports the need for modelling the clustering behaviour observed in the random fields of rainfall. The difference between the fixed Gaussian version of PEGRAIN and the copula model was demonstrated by computing the spatial entropy of triples of gauges as Figure 2.4 shows. The version of PEGRAIN that is based on wetness accommodates this difference and is therefore a candidate worth considering for downscaling.

We can describe clustering using entropy, which we measure using the joint occurrences of wetness on a collection of sets of 3 gauges. Entropy is given by $\varepsilon = \{(\pi-1).\ln(1-\pi) - \pi.\ln\pi\}$, where π is the joint probability of three gauges on the vertices of acute triangle being wet, determined by counting these joint occurrences. The square root of the Area (h) of a triangle is a one-dimensional measure of its size. Entropies for the network are plotted against h on the left of Figure 2.4, comparing the outputs of the covariance version of PEGRAIN and the copula model to data.

The comparisons in the left panel in Figure 2.4 show that it is to some extent important to take note of the spatial structure. This is because ε-points of the copula model (correlation sensitive to magnitude) are closer to the data values than those of the covariance-based model.

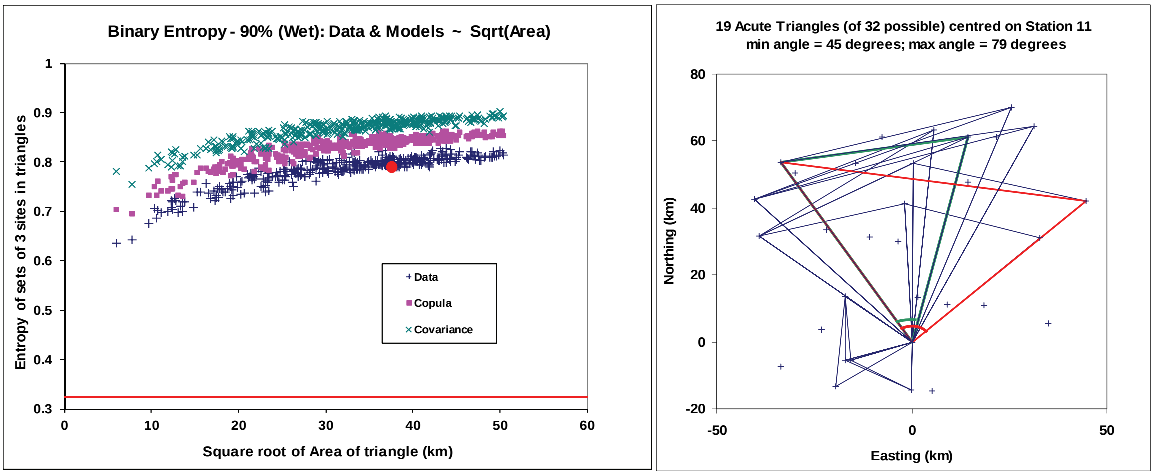


Figure 2.4. Entropies calculated on a network of 32 stations in Germany. The entropy is calculated on constrained acute angled triangles sought from the network by search as shown on the right (Bardossy and Pegram, 2009).

This refinement, which indicates that covariance-based models do not completely capture the clustering behaviour of network wetness, will be kept in mind in the futures rainfall network modelling and accommodated by the wetness-based version of PEGRAIN.

2.5. Modifying the PEGRAIN model to adapt the stochastic output to reflect change

Srikanthan (2009) published the partial solution to the problem of scaling future GCM/RCM model products, based on the joint work of Srikanthan and Pegram (2009). He suggested a rescaling methodology for climate change studies which involves adjusting the statistics of daily amounts, using Wilks's (1992) equations [6] and [7], which were based in turn on Gabriel's (1959) results [3] and [4].

If we again use the Gamma distribution to describe the daily rainfall amounts Y (when wet) then the mean and variance are $\mu = \alpha\beta$ and $\sigma^2 = \alpha\beta^2$. In addition, recall that P = probability of rain on a day and d = the serial correlation of the wet/dry process. Wilks (1992) suggested the following scaling equations [8] and [9] based on [6] and [7], where a prime (') indicates a futures (desired) parameter and the lack of prime indicates the present known value. Because the calculation is done by comparison for the same n days, the n appearing in [6] and [7] cancel in the numerators and denominators of [8] and [9]:

$$\mathrm{E}'[W_n]/\mathrm{E}[W_n] = P'\alpha'\beta'/\{P\alpha\beta\} \qquad [8]$$

$$\mathrm{var}'[W_n]/\mathrm{var}[W_n] \approx P'\alpha'\beta'^2[1 + \alpha'(1-P')(1+d')/(1-d')]/\{P\alpha\beta^2[1 + \alpha(1-P)(1+d)/(1-d)]\} \qquad [9]$$

The idea of Srikanthan (2009) appears to solve the modelling problem, because it was applied by him in the following way based on these two equations, effectively using the idea of Table 2.1:

- generate a multisite set (ensemble) of rainfall using the model of Srikanthan and Pegram (2009)
- post-dress the model's daily amounts to ensure that the future GCM monthly and annual mean and variance ($\mathrm{E}'[W_n]$ and $\mathrm{var}'[W_n]$) are captured.

However the intermittency problem remains: there are 4 unknowns on the right hand sides of two equations. Srikanthan (2009) decided to let $P' = P$ and $d' = d$, as Wilks (1992) did, but this forces the occurrences process in the future to be the same as in the past, which we understand is unlikely to happen.

Another issue is ignored by the simple scaling equation, subtle though it is. In the introduction, we showed that a naïve two-step rescaling procedure in Table 1.1, applied to annual (which could also be monthly) rainfall accumulations, just does not work. This is because the GCM parameters cannot be used directly "as is" without transforming them into the Network space because the statistics of the GCMs do not match those of the Networks (primarily due to the differences in the spatial scale). True, Srikanthan (2009) was able to show that the GCM statistics had been recaptured, but that is not enough, because they are different from the gauge-based statistics. In Table 2.2 we drew attention to the necessity of at least going through the three step process of scaling from GCM to network in both past and future. How does this knowledge help this study? Can we do better?

2.5.1 The suggested solution

Bardossy and Pegram (2011) found that the daily rainfall type/climate in a local region down to 25 km scale is highly dependent on the Circulation Patterns (CPs), identified in the historical period and identified in the futures period, as demonstrated in the discussion on Figure 1.3. The frequency and sequencing of these CPs was found to be different in historical and futures periods as shown in figure 2.3. True, we devised a methodology to downscale rainfall estimated by RCMs over the 25 km scale Rhine region and show that daily rainfall differs substantially between past and future. However the procedure we adopted was more a proof of concept than a practical operational hydrological tool; it now has to be made practical in the current context.

Therefore, to produce a rainfall model which can be used routinely, based on the CP-based research of Bardossy and Pegram (2011), it is necessary to define the remaining two of the 4 unknown parameters via the CP sequences: P' and d' (recall that P' = probability of rain on a future day and d' = the future serial correlation of the wet/dry process). The necessary intermediate step to finally close the information loop is to find the scaling link between:

- the CP-based estimates of the spatial average precipitation over the RCM pixels (in the case of this WRC study 0.44 ° degrees on each side – about 50 km square) and
- the spatial average of the observed/modelled gauge rainfall, over the same areas.

That should be relatively straightforwardly achieved by spatially averaging observed network rainfall and making comparisons conditioned on CPs. In addition, it seems that the added information available in the identified CP sequencing in past and future RCM runs will help to define the numbers required for the occurrence process. In fact the sequencing of the CPs can be modelled by a Markov Chain, as indicated in sub-section 2.5.2, which will define the wet/dry process.

The previous paragraph focuses on the P' and d' values defining future individual gauge behaviour. There is one more step that has to be taken: spatial wetness needs to be modelled. Not yet addressed specifically by any published multisite rainfall model, is the count of the number of gauges in an area which are wet on a given day and in more detail, their clustering behaviour. True, attention has been given to the issue by Bardossy and Pegram (2009), but we think it is important to model the number of gauges that are wet on a particular day, i.e. the wetness index or ratio, as indicated in Section 2.4.4.

2.5.2 Exploiting the time series of CPs to model rainfall

Early on in this study we compared the discrimination of the following numbers of classifications of CPs, based on Sea Level Pressure anomalies: 8, 10 and 12, each set with a 'rubbish bin' of days with no particular signal. There are not enough data in the short period of PRECIS (and the even shorter period of rainfall) data available to us to make more classifications than 12. When the 700 hPa data set from

PRECIS was made available, the classification was re-run using 9 CPs, with the 10th being the set with no particular signal and these are used in Chapters 4 and 5 for the downscaling. Their time series are useful for explaining how they can be exploited to generate plausible sequences of daily CPs by simulation. As an example, we take the CP set for East London appearing in Figure 2.5.

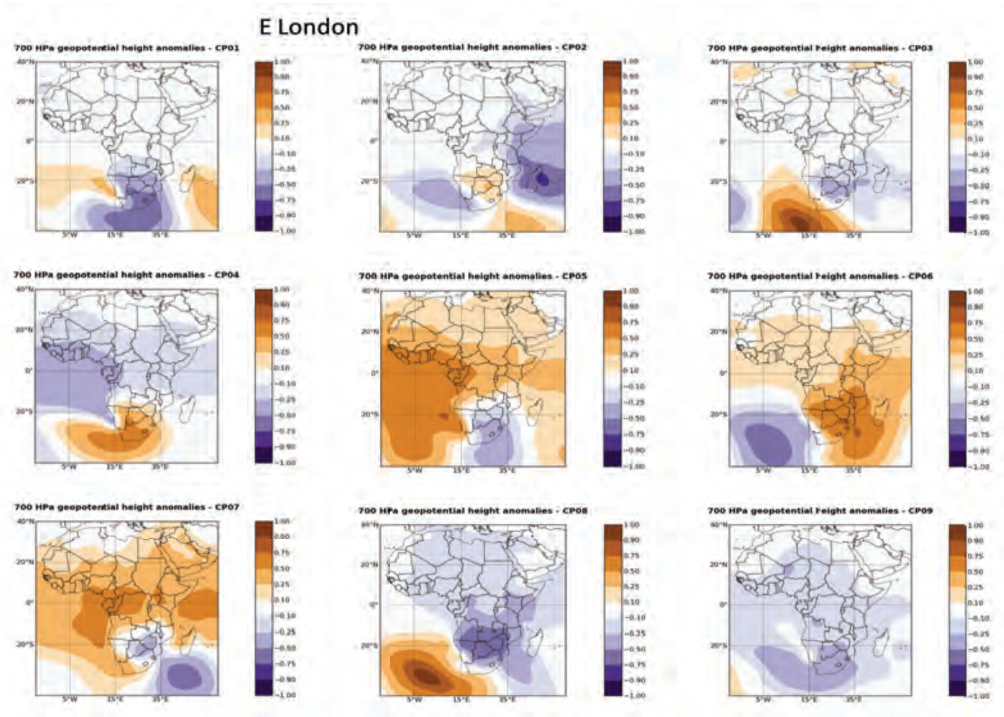


Figure 2.5 Nine Circulation Patterns chosen by matching 700 hPa anomalies to gauge wetness over the East London region in period 1: 1990 to mid-2000; the 10th CP was the 'bucket' of unclassified CPs, not shown here.

Their transition counts (how often a CP is followed by one of the others) produces a frequency table, which, on division by each corresponding column sum, produces a transition probability matrix for the lag-one Markov chain shown in Table 2.3.

Table 2.3 Transition probability matrix for the lag-one Markov chain for the 10 EL CPs of Period 1

	From									
To	1	2	3	4	5	6	7	8	9	10
1	0.213	0.033	0.013	0.127	0.024	0.045	0.059	0.068	0.038	0.014
2	0.176	0.165	0.013	0.106	0.171	0.053	0.140	0.023	0.013	0.021
3	0.102	0.285	0.226	0.010	0.061	0.053	0.135	0.027	0.002	0.031
4	0.037	0.041	0.077	0.365	0.159	0.122	0.056	0.163	0.405	0.127
5	0.019	0.045	0.030	0.027	0.268	0.012	0.005	0.018	0.023	0.014
6	0.065	0.128	0.141	0.050	0.098	0.362	0.051	0.050	0.015	0.024
7	0.213	0.062	0.103	0.158	0.024	0.077	0.306	0.095	0.058	0.120
8	0.000	0.029	0.026	0.072	0.061	0.069	0.029	0.367	0.058	0.065
9	0.074	0.041	0.171	0.045	0.073	0.126	0.081	0.163	0.370	0.329
10	0.102	0.169	0.201	0.041	0.061	0.081	0.137	0.027	0.019	0.257

Generating one hundred sequences each 10 ½ years long from this Markov chain, and determining the frequencies of the each of the CPs occurring, gives us a sampling distribution of each CP based on the data of the first period. If the frequencies of occurrence of the CPs of the second period are close to the median of the generated first set, then the 'climate' (as sampled by the CPs) will not have changed.

In Figure 2.6, the frequencies (average occurrence per year) of the 10 CPs in the two periods are plotted coaxially. Also plotted are the quantile bounds based on 100 simulations, each of the same length as the first period, using the Markov Chain transition matrix of Table 2.1.

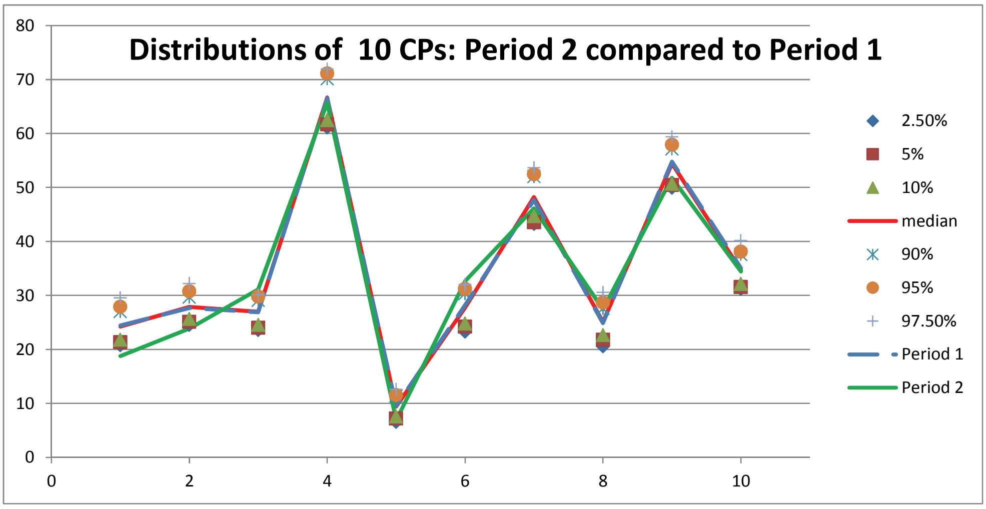


Figure 2.6 The frequencies of the 10 CPs in the two periods plotted coaxially. The vertical axis is the average number of occurrences of each CP per year; the horizontal axis labels the CPs. Also plotted are the quantile bounds based on 100 simulations, with the same length as the first period, using the Markov Chain transition matrix of Table 2.1. The first 7 of the 9 series indicated in the legend are the quantiles from the Markov Chain simulations.

What will be seen from Figure 2.6 is that there is excellent reproduction of the CP frequencies during the first 10½ year period. This is indicated by the complete overlay of the set of frequencies sampled in the first period, shown in a dashed blue line, by the red line, which in turn links the 50th percentiles (the medians) of each CP frequency. However the frequencies of the CPs in the second period, joined by the green line, show dissimilar distributions when compared to those in the first period (indicated by the red line), as well as the 100 sets of 10 ½ years generated from the data in the first period. This difference is demonstrated by the positions of the green line relative to the 90% quantile range indicated by the brown squares and the biscuit circles, as listed in the legend. That information tells us that things have materially changed over the two periods.

This change may well be due natural fluctuations, but is difficult to be certain based on a short set of ten or so years. Nevertheless, Figure 2.6 is therefore an indication of another significant result we owe to the rainfall classification based on the CP discrimination.

This result is interesting to us from the point of view of simulation. Generating plausible sequences of CPs, each set different from the other, provides the core for filling the information gap highlighted in sub-section 2.5.1. The CPs define two items of particular interest to us: the wetness and therefore marginal distribution of the rainfall for the day, or in symbolic terms, P = probability of rain on a day and d = the serial correlation of the wet/dry process. This work has not been undertaken in this project, but the above maps out a very plausible solution to the information closure problem we have highlighted and is likely to find its way into the literature, in the not too distant future.

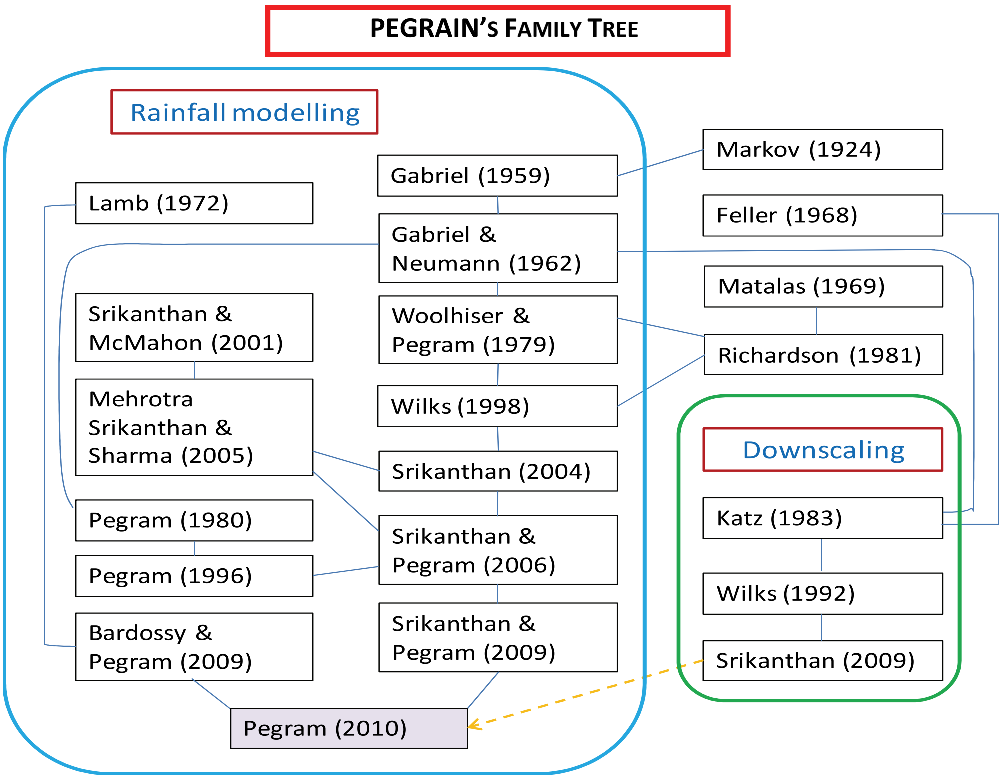
2.7. Summary

As a first step, the PEGRAIN model (Pegram, 2010) with minor modifications suggested by Srikanthan (2009), can model futures generated daily rainfall, based only on monthly accumulation statistics. This would ignore any change to the occurrences process and would assume that the spatial dependence on any day was the same in both past and future. That is a start, and is (we believe) an improvement to the current state of the art.

To improve this model we suggest three steps, using the techniques developed by Bardossy and Pegram (2011):

1. obtain the link between CP-dependent and observed occurrences processes to obtain the intermittency parameters
2. determine the parameters of the amounts process (to define the distributions of rainfall) using the Wilks (1992) transform of equations [8] and [9]
3. determine the strength of the relationship between the rainfall network wetness index and the CPs, and if successful, extend the analysis procedure in order to assist in defining the covariance transformations between GCM (through RCM) to network rainfall.

Use steps 1 and 2 in the current covariance-based version of PEGRAIN to model the future rainfall. Use the wetness information in step 3 to guide the wetness-driven version of PEGRAIN and compare results.



3. Tools to Detect Change

Preamble

How does one assess the ability of observation networks to detect and follow change in rainfall patterns, possibly due to Climate Change, and measure the concomitant uncertainty? How does one best learn about the rate and direction of changing rainfall regimes, by combining information streams emanating from observation networks and Global Circulation Models (GCMs) and their embedded Regional Circulation Models (RCMs)? How does one link these meaningfully when gauge observations and GCM/RCM outputs are at vastly different scales? What tools, models and strategies are appropriate to harness in answering these questions?

The contents of this Chapter will achieve this end. A summary follows:

- We identified 5 regions in different climate zones of RSA containing large numbers of gauge data. For the data collected by gauge networks to be useful in this context, relevant statistics were extracted, such as: proportion of dry and wet sequences by season; marginal distributions of wet amounts; spatial interdependence and network wetness, among others.
- We selected, refined and used modelling tools which incorporate these statistics and whose outputs reflect them. These modelling tools include (i) descriptive and simulation tools of a statistical/stochastic nature (ii) GCMs and RCMs which are largely based on physics, operated in both reanalysis and simulation modes.
- The goal achieved is the information transfer stream between GCM/RCM outputs and the network-focussed statistically based modelling tools. To achieve this end, the identification and choice of relevant statistical measures is important.

3.1 Description of the data

Five networks with different but representative regimes were selected to cover South Africa, satisfying, as far as possible, the following criteria to enable a gauge network to be able to detect and to gauge changing rainfall patterns associated with climate change. Each should have long, intact, error-free time series of rainfall. All gauges in a network should belong to the same climatic region, affected by-and-large by the same weather drivers, so as to ensure consistency and coherence of the measurement properties. There should be a sufficiently dense population of gauges in each of the networks to capture small spatial scale features (e.g. thunderstorm rainfall) as well as those from large or persistent (frontal or stratiform) systems. To achieve this, each gauge record during the period 1990 – 2000 was examined for evidence of infilling and coherence with its neighbours, reducing the original set of 1931 stations to 302. These were cross-correlated and examined for outliers and mismatches and either rectified or rejected. Figure 3.1 illustrates the decimation in the Mpumalanga region.

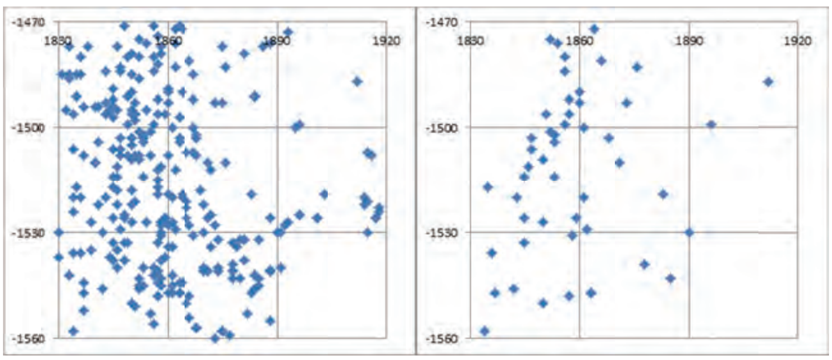


Figure 3.1 Mpumalanga rain gauges. Left: 318 candidate rain gauges. Right 49 gauges remaining after checking their records manually.

What statistics were obtained from these sets? Rain does not fall from clear skies. Therefore the primary basic statistic to gather from a gauge network is *the sequence of wet and dry days* over the network. The next important statistic was to determine *the proportion of wet gauges on a given wet day* over the network, having determined a minimum threshold of rain detection, set at 2 mm per day. The reason for the importance of this statistic is that local afternoon thunderstorms may wet only a small proportion of the network (although very heavily and causing flash floods), while frontal or stratiform systems wet a large proportion and are responsible for large floods on large catchments. The wetness proportion statistic was found to be particularly helpful in identifying those days in the RCM outputs which have similar Circulation Patterns (CPs) so we could group rainfall responses meaningfully.

Next, we determined *the frequency distribution of the amount of rain that falls at each site on each day in a given season* (or month), because the rainfall characteristics change with the weather and therefore the CPs. This set of statistics gave us the marginal probability distribution of wetness and the probability of dryness at each site. The spatial statistics of the rainfall are important from the point of view of determining how wet gauges cluster and how the amounts are spatially related. *The spatial binary correlations of wet/dry processes* as well as the interdependence of rainfall as measured by *the spatial cross-correlation coefficients of the amounts* are crucial statistics to gather.

Thus, to summarise the vital statistics, we determined the following from each network to characterise the rainfall regime under different climates:

- the stochastic structure of the sequences of wet and dry days of individual gauges
- the proportion of wet gauges on a given wet day
- the frequency distribution of the amount of rain that falls at each site in a season
- the spatial binary correlations of wet/dry patterns
- the spatial cross-correlation coefficients of the amounts in the rainfall patterns.

We now move on to the selection of RCM outputs. The available RCM with the necessary outputs matching recent historical observations is the PRECIS model run by CSAG in UCT. These data, specifically Sea Level Pressure and 700 hPa Circulation Patterns (CPs), together with matching daily rainfall coverage over Africa, were kindly made available to this project. To be able to compare the gauge network statistics to the outputs of this RCM, we need to match the sub-areas of the observation networks to the pixels of the climate models. The RCM pixels are 0.44° [of the order of 50 km] square. Therefore, the daily gauge network rainfall over these matching large areas needs to be estimated before direct comparison with RCM outputs. However, a caution is necessary when it comes to comparisons. Current RCMs (such as PRECIS) provide rainfall at a reasonably high spatial and temporal resolution. Nevertheless, RCM estimates of rainfall suffer from deficiencies which trouble hydrologists: the RCM precipitation distribution at a given location differs from the gauge observation; for example, in the RCM output, there is too much drizzle (with too little dryness). Even long term means are not reproduced well. The systematic biases of these models can be corrected using bias correction methods, such as Q-Q transformations with or without dependence on circulation patterns (CPs), as shown by Bárdossy and Pegram (2012), with CP-based methods showing better results.

Once the links between these areal averages of rain (RCM and gauges) have been obtained as an intermediary, only then can the exercise of downscaling RCM rainfall estimates to gauge data commence. Thus, all five of the classes of statistics mentioned above were obtained for both model and matching aggregated observations, so that they can be linked in order to perform the following two important functions:

- “understand how gauge networks can detect and gauge changing rainfall patterns associated with climate change” [Hewitson (2005b)],
- then model the future possible changes (with concomitant uncertainties) to rainfall affected by climate

3.2. Background – Time series examples

To set the stage before moving on to the technical details, some preliminary investigations were carried out on South African regions. South Africa possesses a multitudinous variation of local climate. We chose five subregions for the purpose of obtaining a reasonable span of climate diversity. For the purpose of the analysis reported here, the following were chosen: Mpumalanga, Upper Vaal, Free State, East London and Cape. They appear on the Mean Annual Precipitation map showing the SAWS numbering of 30′ squares covering the country in Figure 3.2. There are from 48 to 67 rainfall stations in each of the subregions spanning the period from Jan 1990 to July 2000, to match the outputs from the RCM PRECIS runs performed by CSAG of UCT, previously referred to in Chapter 2.

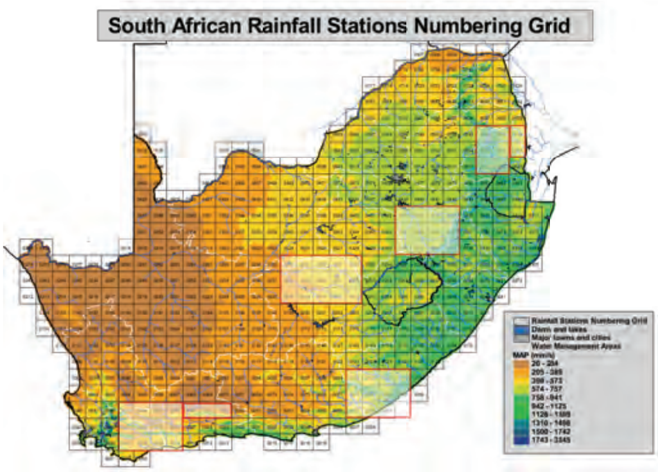


Figure 3.2. Mean Annual Precipitation map showing the SAWS numbering of 30′ squares covering the country [by kind favour of Bennie Haasbroek]

Taking the Mpumalanga subregion in the Northeast (49 gauges in the network over 18 000 km^2) shown in Figure 3.1, we obtained in Figure 3.3, the 6-month moving averages of the spatial daily averages of (i) mean rainfall (red) and (ii) number of wet gauges (blue). The figure shows a (linear) increase of daily spatial average from 1.96 to 3.61 mm and an increase of probability of wet days from 0.66 to 0.71 over 10.5 years:

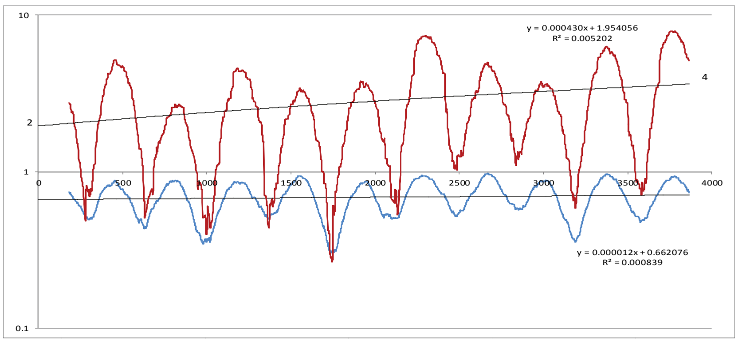


Figure 3.3. 6-month moving averages of the spatial daily averages over Mpumalanga, of (i) mean rainfall and (red) (ii) number of wet gauges (blue).

However, taking only those very wet days in the 10 years limited to when only 47, 48 or 49 gauges were wet, we found that there is a negative trend in very wet days, as shown in Figure 3.4. It is also significant that the periods with very wet days coincide with periods of high average rainfall shown in Figure 3.3. What is important here? - amount of rain? frequency of wet periods? occurrence of very wide coverage? The answer is: All of these, and more. These questions will be addressed in the sequel.

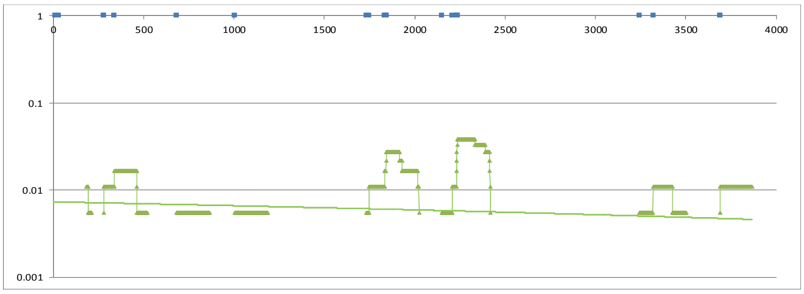


Figure 3.4. Occurrence of wet days (47, 48 & 49 wet out of 49 stations) (blue dots) and the 6 month moving average of their counts, with linear mean (green sequence) over 10.5 years – Mpumalanga.

Turning to the Free State region (the central block in Figure 3.1, with 57 gauges over 38 000 km^2) we obtained Figure 3.5, showing negligible change in both the daily spatial averages and the proportion of wet/dry days over the same 10.5 year period. Because this behaviour is different from the upward trend of the Mpumalanga set over the contemporaneous period in Figure 3.2, we might conclude that the behaviour of gauge observations is region specific, a feature that is exploited in the analysis to follow.

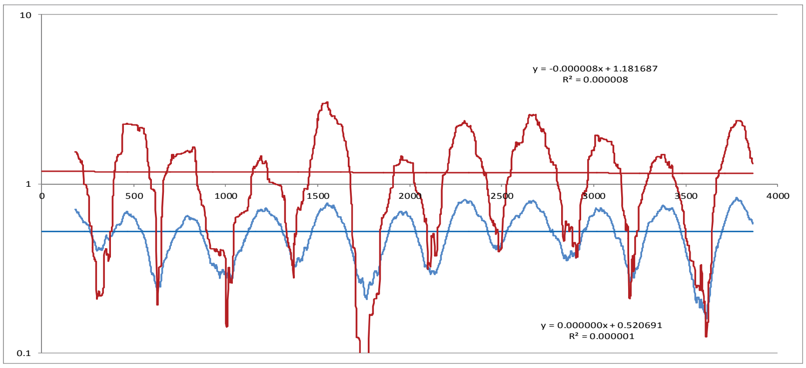


Figure 3.5. 6-month moving averages of the spatial daily averages over Free State showing negligible negative change in both daily spatial averages and proportion of wet/dry days over the same 10.5 year period as shown in Figure 3.3.

It must be emphasized that these sequences are far too short to detect change due to climate and that putting a linear trend-line through the spatial mean statistics of such a short period is not very meaningful, as will be demonstrated in what follows from a different perspective. Nevertheless, the above statistics (the time series of network wetness and means) are valuable in characterising rainfall regimes under changing climate scenarios.

Perhaps one should offer a warning – a blast from the past, as it were. Thirty-five years ago, Dyer and Tyson (1977) reported as follows in their abstract:

> "By fitting a trigonometrical regression model to a regionally averaged time series for the period 1910-72, and by extrapolation to the year 2000, an estimate of future extended wet and dry spells has been made for the summer rainfall region of South Africa. It is suggested that the periods 1972-81 and 1991-2000 will experience above normal rainfall and the period 1981-90 will be drier than normal. Results to date show that mean regional rainfall for the hydrological years 1972-73 to 1975-76 has conformed to the suggested pattern."

In the same document, they produced the following figure, which is based on annual rainfall totals over the Summer rainfall region of South Africa. It caused a huge disturbance in the South African Hydrology community in the early 1980s, with people hunting for links with sunspots and Milankovitch cycles, etc. It actually had farmers and DWA planning according to this message, which was (in a sense, unfortunately) reinforced by the worst drought over the hinterland since 1933-39, occurring in the 1980s, as Dyer & Tyson (1977) had suggested.

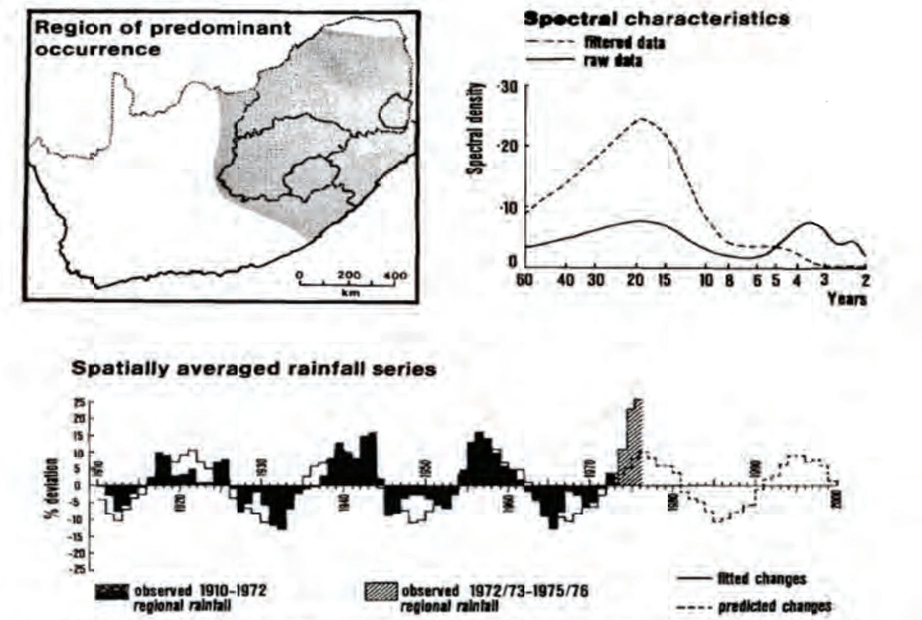


FIG. 1. The region over which the quasi 20-year oscillation in rainfall occurs predominantly and the region for which the mean time series has been determined (upper left); the spectral characteristics of the spatially averaged regional time series (upper right); and observed, calculated and predicted regional rainfall changes (lower).

The following plot of the river flows in the summer months of the Mokolo River in the North of the country, excerpted from the MAHYVA project (Rouault et al., 2010) reinforces the 'pseudo-cyclicity' observed above. However the periodicity predicted by Dyer & Tyson (1977) disappeared part-way through the 1990s – the first five years were dry, followed by a five year wet period, then back to five years dry at the turn of the millennium, breaking the 20-year pseudo-cycle.

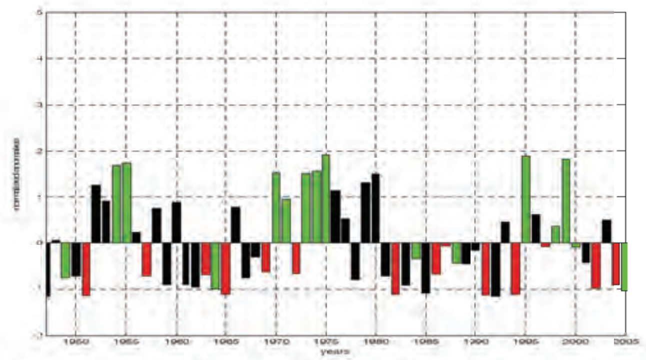


Figure 3.10 Normalised anomalies (mean discharge of January February and March (JFM) for each year minus the JFM climatological mean divided by the JFM standard deviation) for the Mokolo River at Zandrivier (drainage area 1777 km^2).. The positive phase (El Niño) of the El Niño Southern Oscillation (ENSO) is in red and the negative phase (La Niña) in green, while the neutral phase is in black. Years plotted along the abscissa correspond to the start of summer: i.e. 1982 corresponds to the 1982/983 summer.

Admittedly this streamflow example is drawn from a particular catchment (the response of others mentioned in the MAHYVA report is not as dramatic), but it goes to show (besides the historically interesting excitement over periodicity) that short records of rainfall or streamflow (much less than 40 years in length) are not suitable for detecting change. This is because such change tends to be masked by many other factors, including natural climate variability.

Nevertheless, catchments are (nonlinear) integrators of rainfall (and temperature, wind-speed, humidity and other climate variables). Thus, if there was a change in the climate signal, one would expect to see some sort of trend over the last few years, as indicated by the natural filter of the streamflow records. This would likely appear, even in the seasonal streamflow totals, if there was a genuine signal in the noisy rainfall. Evidently, such change is difficult to detect visually in the above images, so is unlikely to appear in a statistical analysis. It should however be mentioned in passing, that the stochastic streamflow models used by DWA for their water resources management purposes (devised by Pegram some 27 years ago) currently appear to embrace most of the variability in the observed data. This suggests that the current models we are using are still relevant. However, we need to develop models that adapt to uncertainty and change, which is what this project is about.

To enable such change to be detected (so we can model the new environment), we have established links between network rainfall and the climate model outputs based on observations, so that we can estimate the future rainfall regime to ensure responsible contingency planning under a changing environment. The way this was achieved was to harness the skill of the climate modellers' GCMs and RCMs, which are based on physics, and interpret the correspondence between model and observation. The downside is that RCMs are embedded in GCMs and therefore inherit their biases. This observation applies not only to rain falling on a given area or gauge site, but also to the spatial dependence, commonly measured by cross-correlations. However, on the up-beat side, the signal that the RCMs contain are likely to reflect the relative change in rainfall behaviour under a changing climate, so the task is to find the relationship between RCM and network rainfall and exploit it. Table 2.2 illustrates this information transfer path, so germane to this report.

In that table, only two sets of parameters (means and standard deviations) were mentioned. We have now enlarged and generalised this portfolio of parameters to 5, so the information flow is enlarged to include:

i. the stochastic structure of the sequences of wet and dry days of individual gauges
ii. the proportion of wet gauges on a given wet day
iii. the frequency distribution of the amount of rain that falls at each site in a season
iv. the spatial binary correlations of wet/dry patterns
v. the spatial cross-correlation coefficients of the amounts in the rainfall patterns

3.3. Obtaining the Important Statistics

From the time series plots offered in the examples in Section 3.2 above, it is clearly difficult to use short period rainfall (especially with a pseudo-cycle of 20 years and given that one could find a decent set of raingauges) to comfortably detect climate change. What we have done instead, is to harness the skill of the climate modellers who have performed reanalysis of the weather variables, including rainfall, over the recent past. The set of 20 years of PRECIS data in itself is not enough to detect change, which (in light of the observations made in Section 3.2) may well be masked by natural, approximately decadal, variability, but the futures simulations will incorporate the estimated trends. However we can compare details, as shown later in the report.

To get there, the first task performed was to obtain the sequence of daily wetness counts over the networks. An idea of the clustering nature of rainfall is given by Figure 3.5, which is the first year (1990) of East London region rainfall on the 57 stations arranged by their SAWS numbers. The maximum daily rainfall depth in that year was 283 mm. Spatial and temporal clustering is easily seen.

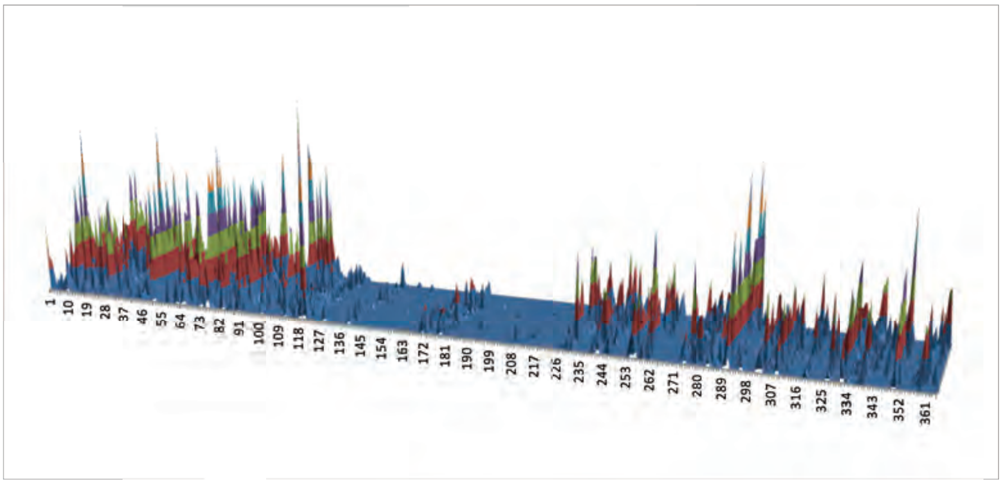


Figure 3.6. The first year (1990) of East London region rainfall on the 57 stations.

An example time series of the gauge wetness count from the East London region appears in blue in Figure 3.7, with a 3-month moving average in black. The rainfall behaviour on these days is strongly associated with the prevailing CPs, which are captured by the PRECIS RCM, to be demonstrated later in this section.

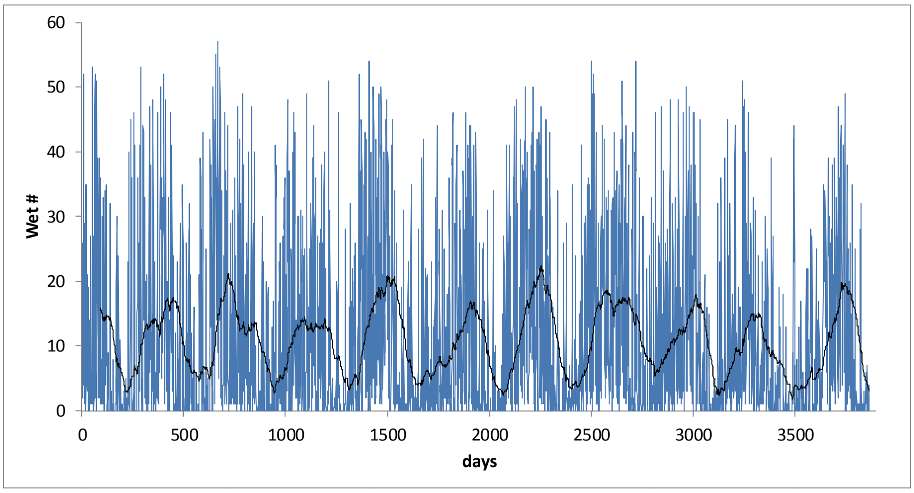


Figure 3.7. Sequence of daily Wet# count in the East London region network (blue), with a 3-month moving average (black).

The second task in treating the gauge networks was to obtain the average rainfall over the areas of the RCM pixels, so we can compare spatially congruent values. In any correlation-based technique like Kriging (in contrast to splines which only fit smooth surfaces to points), if the points are too sparse there is no information transfer between them. The following discussion examines the linkages.

3.3.1 Spatial Structure of rainfall estimates

The following image in Figure 3.8 is the TRMM 3B42RT estimate of daily rainfall over South Africa composed by summing 8 images obtained at 3-hour intervals in a parallel WRC project K5/2024: HYLARSMET. The scale of the blocks is about 25 km, half that of PRECIS. Note the random clustered nature of the rainfall. The centre of the small red ring locates the MRL5 S-band radar at Bethlehem. A day's rainfall estimated by the radar appears in Figure 3.9. Note the structure and preferred dependence in the image, caused by a fast moving convective front.

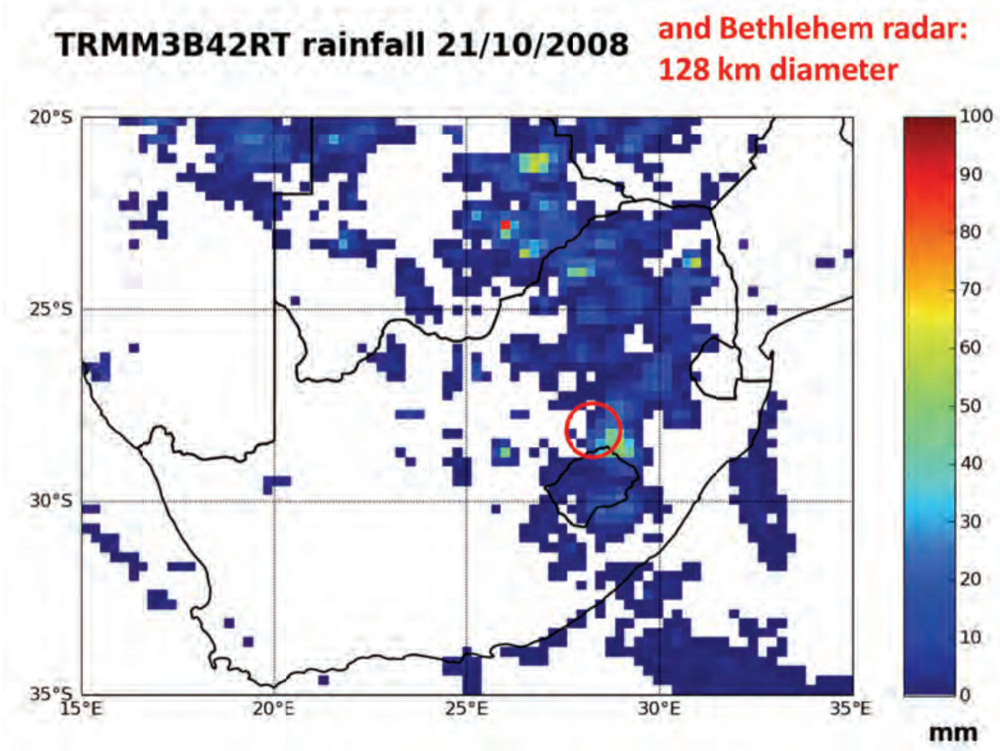


Figure 3.8 TRMM 3B42RT estimate of daily rainfall over South Africa on 21/10/2008

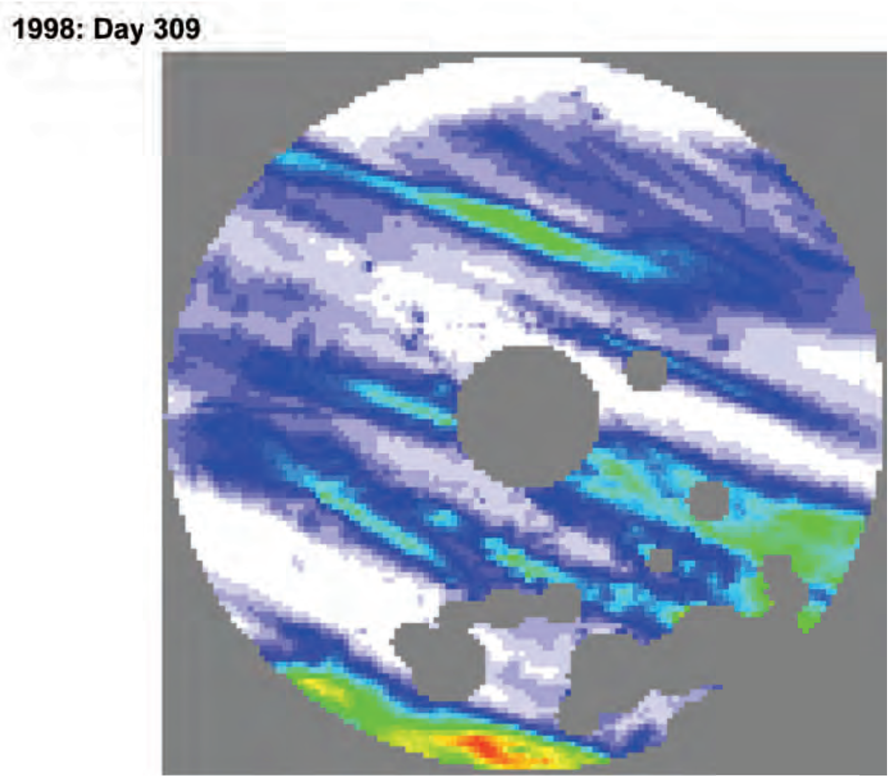


Figure 3.9 Daily rainfall estimate made by combining all radar-rainfall images (at 5-minute intervals) observed by the Bethlehem MRL5 radar on a day in 1998

Figure 3.10 is a montage of 12 consecutive days preceding that shown in Figure 3.9. Again, note the structure of the wide-spread rain and the scattered rainfall with less structure.

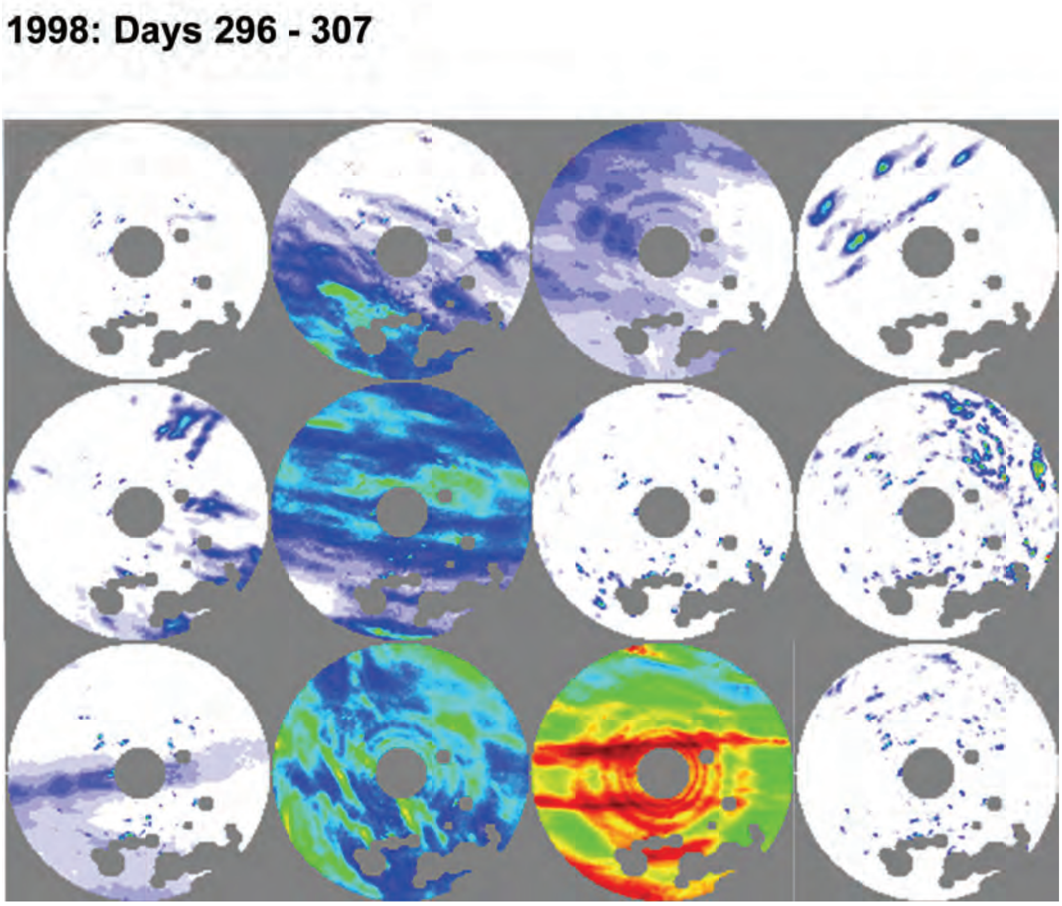


Figure 3.10 A montage of 12 consecutive days of rain estimated by the Bethlehem MRL5 radar and formed by accumulating the images obtained at 5-minute intervals.

In the three figures that follow, we have computed the cross correlations of the Vaal region raw daily rainfall data (including the zeros) at three scales, all for Period1 [1990-2000]. Figures 3.11, 3.12 and 3.13 show (i) the observed gauge values, (ii) the block averaged gauge values [over the same blocks as defined by PRECIS and (iii) the PRECIS1 estimates of rainfall, with all coordinates in minutes of arc. Superimposed on the ccc versus distance plots are the trend-line (red) and the 1/*e* threshold level (green) whose intersection with the trend-line defines the correlation length – a measure of spatial dependence.

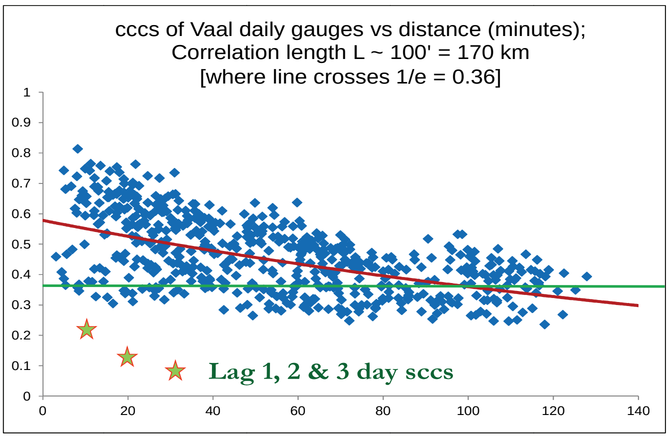


Figure 3.11 Correlogram of spatial ccc versus distance for Vaal daily gauge data. The stars show the temporal (serial) daily correlation for a chosen gauge at lags of 1, 2 & 3 days.

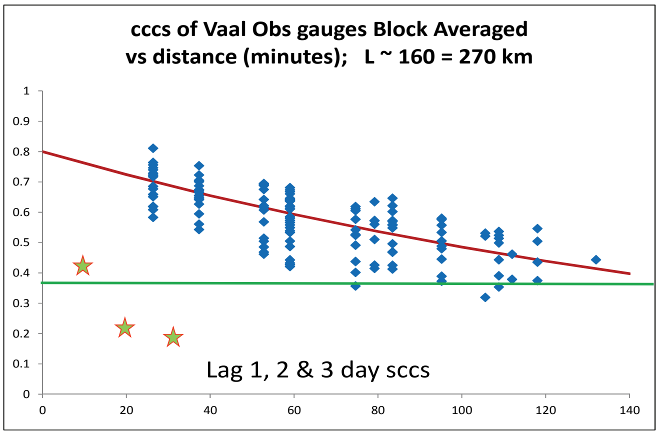


Figure 3.12 Correlogram of spatial ccc versus distance for Vaal daily gauge block average data. The stars show the temporal (serial) daily correlation for a chosen gauge at lags of 1, 2 & 3 days.

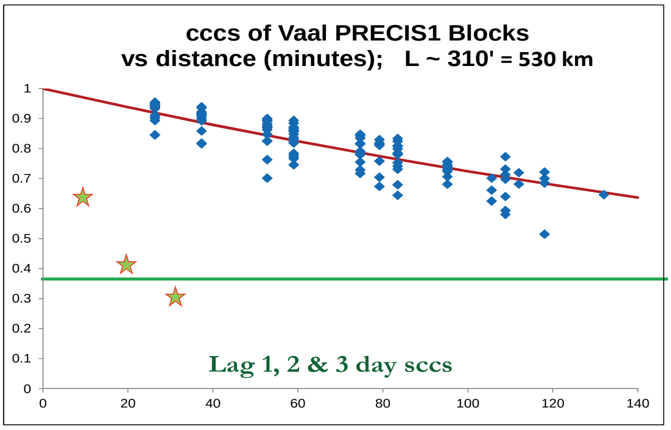


Figure 3.13 Correlogram of spatial ccc versus distance for Vaal daily PRECIS1 data. The stars show the temporal (serial) daily correlation for a chosen gauge at lags of 1, 2 & 3 days.

Three things are worth noting from these figures. First, the spatial dependence structure of the PRECIS1 data is considerably stronger than that of the gauge Block averages, even though they both have the same spatial scale. Secondly (and not surprisingly) the gauge data are much less spatially correlated than either of the block scale data; in comparison to the gauge Block averages, the spatial smoothing has reduced the noise inherent in the gauges. This must be dealt with in any downscaling work. On cannot simply arithmetically scale RCM data to have the same spatial means as the observations – the correlations need to be addressed as well.

3.3.2 The Block Averaging Algorithm

The technique used for computing weighted average rainfall is based on integration of multiquadric interpolated surfaces over target areas as described by Pegram and Pegram (1993); the mathematics was developed in that paper. The method of computing the surfaces is almost identical to Kriging with a semivariogram of rectangular hyperbolic shape, defined by the offset parameter of the rectangular hyperbola (equating to the nugget of the variogram). Why the method is valuable is that the volume integral under the multiquadric surface can be calculated analytically piecewise over the target area, which is approximated, as closely as desired, by an enclosing polygon. The calculation is very quick and after scaling, the weights associated with each gauge can be calculated. These weights are then used to determine the average depth over the horizontal surface of the target area for each rainfall occurrence, be it daily or over another interval.

The benefit of the method is that it is somewhat similar to the method of Thiessen polygons which ascribes weights to each gauge based on tessellation of the area, so that those gauges which are crowded together have less weight than those sparsely distributed. The difference here is that the method is easy to automate, once the coordinates of the gauges and vertices of the polygon enclosing the surface have been determined. The offset parameter (nugget) of the variogram (or basis function) of the multiquadric method should be set to zero, as it was shown by Pegram & Pegram (1993) that this is possibly the best averaging method, followed closely by Thiessen polygons. There is no benefit to be gained by including gauges outside the neighbourhood of the target area, because multicollinearity soon reduces the weights to oscillate meaninglessly about zero. This is a manifestation of the typical 'shielding effect' found in Kriging (Wesson and Pegram, 2004). An example using East London data appears below in Figure 3.14, with the weights for various offsets listed in the accompanying Table 3.1. In that table, the variable 'param' is the offset of the centre of the hyperboloid from the vertical origin. A value of zero yields an inverted cone for the basis function.

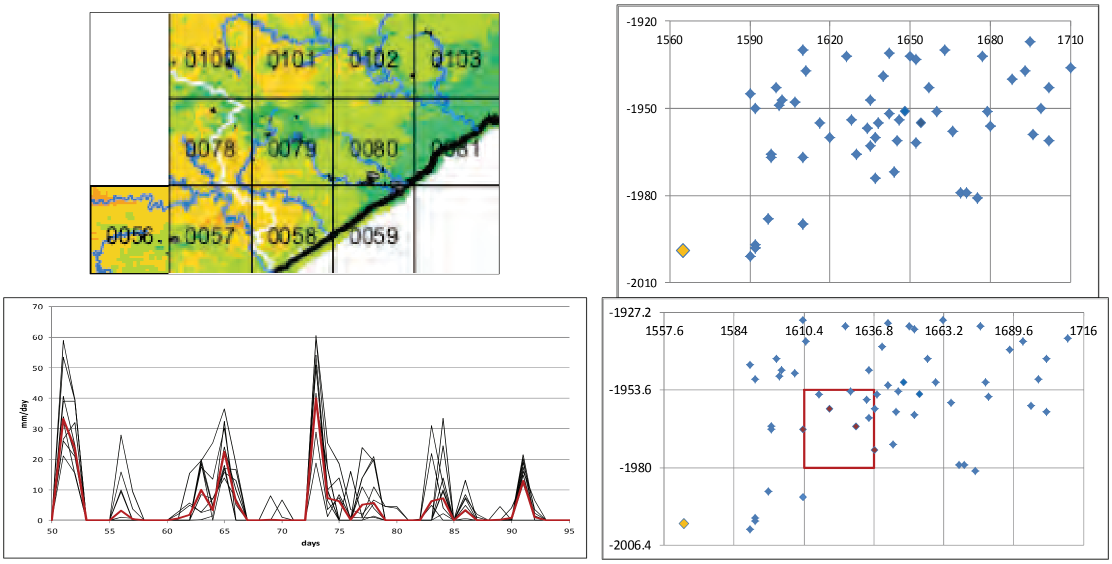


Figure 3.14. Three maps of gauges in the East London area and some daily time series. Top left: a segment of the MAP map (Figure 3.1) and the SAWS 30′ numbered squares; Top right: the sites of the 57 selected (cleaned) gauges in the SAWS squares; Bottom right: the same sites on the PRECIS grid, with the target square in red, whose border length is 0.44° or 26.4′. Red diamonds will be discussed after the weights are presented in Table 3.1; Bottom left: a 45-day subsequence of the time series of the nine gauges (black) in the red square and their multiquadric averages (red).

Table 3.1. Weights calculated at the 9 sites in the interior of the target square in Fig 3.8.

param	Station number								
	10	**11**	**12**	**13**	**14**	**15**	**16**	**18**	**19**
0	***0.245***	0.054	***0.174***	0.046	***0.221***	0.041	0.026	0.002	***0.191***
9	0.222	0.076	0.173	0.042	0.200	0.027	0.078	0.020	0.162
100	0.119	0.111	0.117	0.106	0.117	0.105	0.110	0.105	0.111

The weights in Table 3.1 corresponding to the nugget offset param = 100 are nearly all equal to 1/9, the average weight, and due to the very flat interpolating function that results. Those with the greatest variability (due to param = 0) are in the first row, estimated with a conical interpolation function. The 4 largest weights (bold italic), corresponding to param = 0, are associated with the four sites which have been highlighted maroon in the bottom panel of Figure 3.14. These are relatively widely spaced sites across the middle of the square and have no sites south of them (in the square). Their weights total 0.831; the cluster of 5 sites in the North and East of the square share the remaining 0.169 of the weight, indicating that they have considerably lower influence than that in an arithmetic mean. Not shown here, is the observation that if points external to the square were used to determine its average rainfall, the weights of those external points oscillated about zero and many of them were mildly negative due to multicollinearity.

It was found that the best solution, in spatial averaging using this technique, is to use only those gauges internal to the square with param = 0. The time series appears in the bottom left panel of Figure 3.14 and it will be seen that this weighted mean does not always follow what might be expected by eye from the arithmetic mean.

3.4. Links between RCM and Network rainfall

The key, to understanding and detecting possible climate change signals, is the sequence of daily Circulation Patterns CPs associated with locally varying rainfall regimes, as described in Chapter 1 and reported by Bárdossy and Pegram (2011). The CPs associated with different levels of wetness were identified for all five regions using a method of classification described in Bárdossy (2010). An example set based on the Cape network appears in Figure 3.15. Notice that the pressure anomalies surrounding Southern Africa are deliberately more defined than the remainder North of the equator, although data are available over the images shown. Each CP is associated with a different rainfall regime in the Cape.

Note that the 'pixel' size of the CPs is 5° - about 550 km (NS) and 480 km (WE) across at 30° latitude – just over 11 times larger than the pixel scale (26.4') of the PRECIS RCM rainfall output. In the figures which follow later, this tessellation is disguised by smooth interpolation and we will use fewer CPs based on 700 hPa heights.

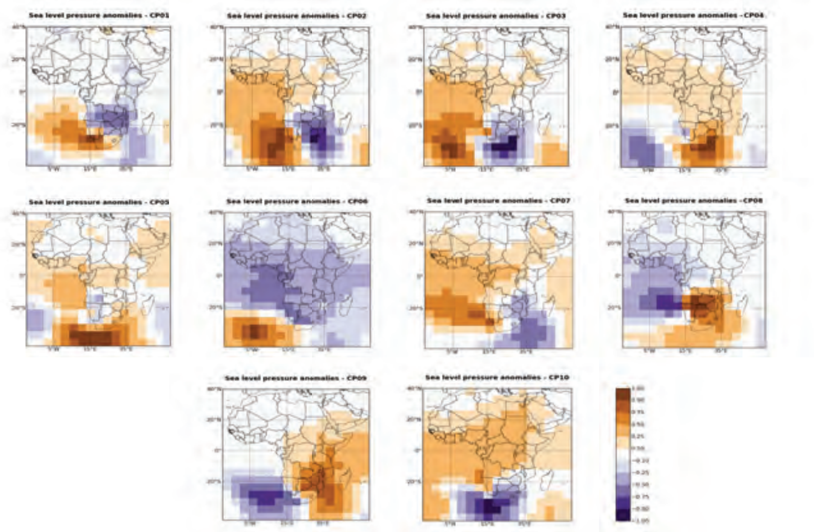


Figure 3.15. The set of 10 CP anomaly patterns corresponding to different rainfall regimes in the Cape region, based on Sea Level Pressures.

Figure 3.16 shows the correspondence between (i) three CPs chosen from Figure 3.15 and (ii) the rainfall distribution at a particular rain-gauge in the Cape region. The black line in Figure 3.16 is the overall behaviour for Season 1 (the wet season in the Cape) and the others are explained in the caption. What is useful about the CP dependence is its power to discriminate between rainfall types.

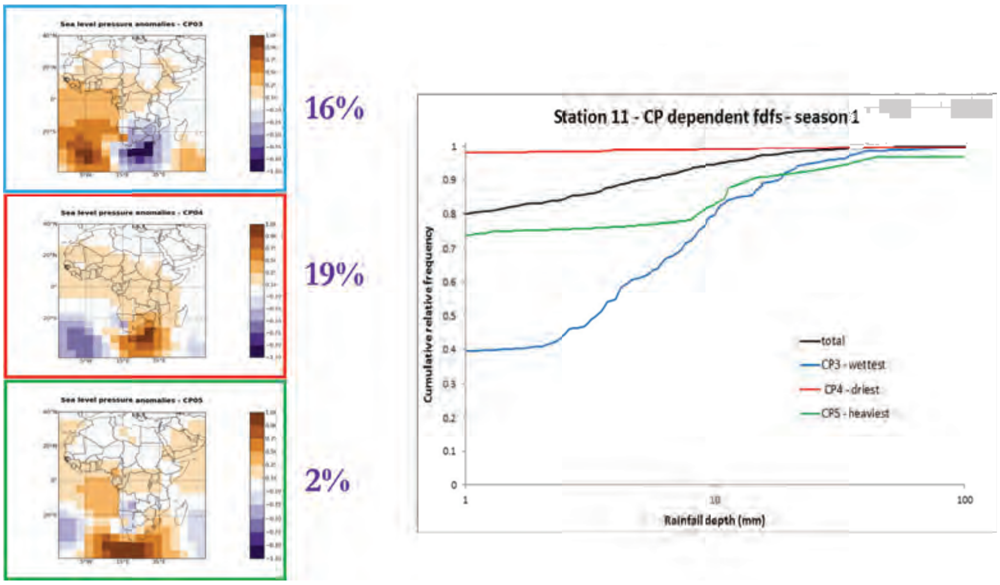


Figure 3.16. The correspondence between three chosen CPs and the rainfall distribution at a particular rain-gauge in the Cape region. Black line: total record; Blue line: Wettest; Red line: Driest; Green line: Heaviest.

When we applied the same methodology to the other 4 regions (the Cape was omitted as its drivers are Winter rainfall), we found that there was a difference between the 4 sets of CPs modelled on each sub-region's rainfall behaviour, although there are similarities. By searching through all the possible pairs and computing their Spearman cross correlations based on the binary incidences over the historical record, we came up with the set of relatively highly correlated CPs, shown in Figure 3.17. They all possess a low pressure system lying off-shore to the West and a high pressure system on the continent (not all in the same place). The cross-correlations between them appear in the table at the top of the figure, together with figures indicating the frequency of the pairs occurring on the same calendar day in 10 years. The number 11 in EAL11 refers to its CP – the orderings don't necessarily match in the four different sets, because they are patterns and not intrinsically ordered.

	MPU09	OFS10	VAA11
EAL11	200 (0.935)	110 (0.675)	196 (0.791)
MPU09		103 (0.623)	157 (0.690)
OFS10			173 (0.770)

EAL11	**MPU09**
OFS10	**VAA11**

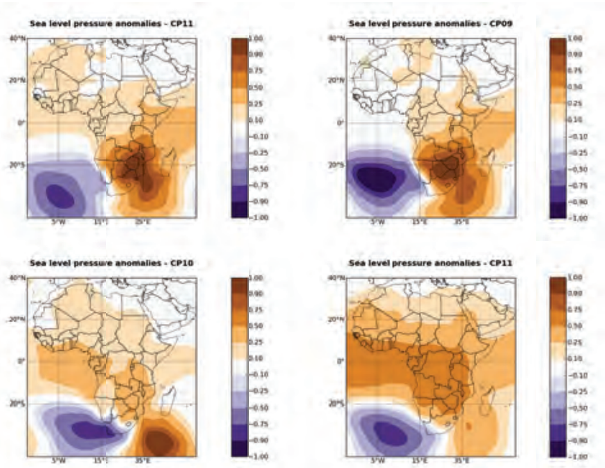


Figure 3.17. A chosen set of relatively highly inter-correlated CPs, each conditioned on the rainfall for a different region. The cccs are in the 4x4 table top-left. They are arranged according to the smaller table top right: Top row: East London and Mpumalanga; bottom row Free State and Upper Vaal regions.

We selected a rain gauge station in the Upper Vaal sub-region, as an exemplar, to plot its cumulative frequency distribution function (cfdf) for the whole season, shown as the Black line in Figure 3.18 below.

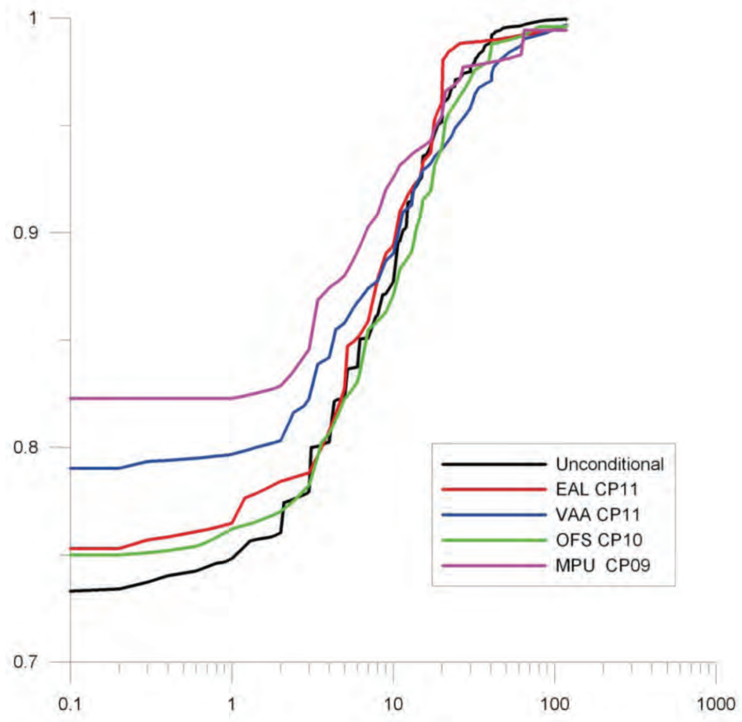


Figure 3.18. Cumulative frequency distribution functions (cfdfs) of rainfall over the years 1990 to 2000 for a chosen site of the Upper Vaal region. The cfdf for the whole year is shown as the Black line. The other 4 lines correspond to the occurrences of rainfall based on the CPs shown in Figure 3.17 for the 4 different regions.

The vertical axis in Figure 3.18 is the probability that the rain falling on a given day chosen at random will be below the corresponding value in millimetres per day given on the horizontal axis – we call it a cumulative frequency distribution function (cfdf). When we conditioned the rain on the days corresponding to the 4 similar CPs shown in Figure 3.17, based on the 4 different sub-regions, we obtained somewhat different results, not only regarding the total rainfall in black, but between themselves. Why are these curves so valuable? The message offered by this cfdf diagram is that: *small changes in the atmosphere cause significant differences on the ground at different places*. It is these differences which we can exploit in detecting change in climate, assisted by the CPs modelled by the RCMs.

3.5. Summary

Gauge site rainfall frequency distributions are sensitive to the prevailing Circulation Pattern of the day, whether Historical or projected into the Future. Both sets may be captured from Regional Circulation models driven by reanalysis or in simulation mode. In spite of the RCMs exhibiting bias in the areal rainfalls, both locally and spatially, they faithfully capture the flavour of the prevailing climate. The rainfall outputs of the RCMs were related to observed rainfall conditioned on CPs and their local bias corrected. Once done, the RCMs become good indicators of possible change because of their modelling consistency, based on physics.

The Circulation Patterns obtained from Sea Level Pressures (or more usefully by 700 hPa levels as we used later in the project) and defined by local regional rainfall behaviour, are very capable in discriminating between rainfall days, both in terms of wetness and also in terms of distribution frequency. These can be exploited to show how modelled climate change is likely to affect rainfall. The paper by Bardossy and Pegram (2011) sets out the theory behind this methodology and indicates its successful application in the data-rich German Rhine basin. These ideas have been shown, in the above discussion, to be transferrable to South Africa.

But, there is always more to be done, as we discovered part way through the project! Uneasy about the spatial statistics of the downscaled RCM rainfall reported in Bárdossy and Pegram (2011), we calculated the spatial cross correlation coefficients (cccs) of daily rainfalls of the same high quality German set. We found that the cccs of the RCM precipitations were significantly lower than those of the observations (the converse of the conclusions offered on the South African data in sub-section 3.3.1). CP-based downscaling led to a slight increase of the cccs but they remained below the observed cccs. This underestimation has potentially deleterious consequences for flood calculations over large areas based on RCM outputs, even after full CP-based bias elimination at the 25 km scale. A recorrelation treatment, a significant innovation to the set of modelling tools, was devised to correct the cccs of the RCM estimates back to the observed set, before undertaking the final Q-Q transform.

This Section ends with a pair of images in Figure 3.19, from Bárdossy and Pegram (2012). On the left is the rainfall modelled on a particular day as estimated by an RCM after quantile-quantile transform on each 25 km square pixel, using the methodology of Bárdossy and Pegram (2011). The image on the right is the same day's estimate after the whole dataset was subjected to the recorrelation technique we have since devised. The one on the right is an improvement because it models the clustering behaviour over large regions better than before recorrelation, determined by aggregating RCM data over large areas and computing the cccs between the aggregated blocks – of recorrelated blocks versus precorrelated. This method can also be applied to futures scenarios.

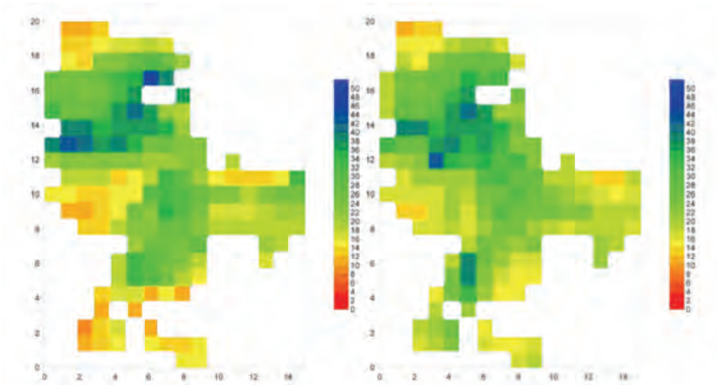


Figure 3.19. Left: RCM modelled rainfall on a particular day after Q-Q transform [Bárdossy and Pegram (2011)]. Right: the same day's estimate after the whole dataset was subjected to a recorrelation technique. [Bárdossy and Pegram (2012)]

There is thus an extra step which needs to be taken: Exploit new knowledge and estimate the pairwise spatial correlation coefficients of (i) the network grid and (ii) the RCM daily rainfalls

4. Assessing the adequacy of the Link between RCMs and Rainfall

Preamble

In this Chapter, we not only assess the adequacy of the link between RCM output and gauge rainfall, we go further and offer a technique for successfully performing the downscaling. Herein we will deal with the Block scale; Chapter 5 will treat the gauge scale.

4.1 Setting the stage for the downscaling methodology

Five regions were selected for downscaling PRECIS data – the same as had been used for determining the Circulation Patterns related to rainfall behaviour described in Chapter 3: Cape, East London, Free State, Upper Vaal and Mpumalanga.

The data sources available are:

- PRECIS RCM daily rainfall and SLP data from 1990 to 2008 inclusive, split into two periods PRECIS1 [Jan 1990 to Jul 2000], & PRECIS2 [Aug 2000 to Dec 2008], to match the SAWS rainfall data availability at the time
- SAWS daily rainfall data for selected sites was initially available only from 1990 to 2000; this set was later augmented to 2008, the latter set being used as 'blind' data for validation, as if we were able to see the 'future', after the event of downscaling the PRECIS2 data.

The PRECIS RCM modelled rainfall estimates are at the 50 km scale. We have devised techniques to modify these to match the gauge block averages at the same scale in the 5 RSA regions in the same way as we did in the German set. The remainder of this section lays out the procedures. In Chapter 5 that follows this one, we extend and modify this thinking and methodology in order to downscale PRECIS data to the gauge scale.

In outline, the procedure takes the following steps:

1. the first activity was to average the observed daily gauge rainfall for Period 1 over the PRECIS 0.44° blocks, to match the space scales for comparison purposes
2. the gauge block averages were QQ transformed to censored Gaussian variates to accommodate dry days
3. the spatial cross correlation coefficients between the QQ'd gauge block averages were obtained
4. the PRECIS daily rainfall estimates in Period 1 (1990-2000) were QQ transformed to censored Gaussian variates. These were decorrelated then recorrelated to match the gauge block averages following the method described in Bardossy and Pegram (2012)

It was initially thought that the way to proceed to downscale the PRECIS estimates to gauge values would follow the next steps:

5. each of the rainfall estimates on each day in these recorrelated block averages were then shared among the member gauges in the blocks and perturbed by a small amount of added noise. In a second stage these PRECIS gauge estimates were decorrelated then, using gauge network (point) estimated statistics, recorrelated using the cccs of the gauge observations
6. all the gauge values based on recorrelated PRECIS point values were then given the corrected marginal distributions using an appropriate QQ transform
7. these RCM-based point rainfall estimates were then pairwise compared to the sequences of the original gauges in a verification procedure.

Unfortunately, although not too bad, these estimates were not good enough in our opinion, particularly when the downscaled gauge data were aggregated back to the block scale, so a new method based on the procedures detailed in this Chapter were developed for downscaling to gauges and is detailed in Chapter 5. In this Chapter 4, we will confine our discussion to the block scale.

In the sub-sections that follow this, we deal with the following ideas and their realisation:

- Averaging Daily Gauge Observations over RCM Blocks
- Censored Gaussian Transformation of Rainfall
- Recorrelation of PRECIS data to match the Gauge Block Averages

4.2. Averaging Daily Gauge Observations over RCM Blocks

The data

We take as example the arrangement in Mpumalanga (used previously in Section 3.3.2) and present the following image in Figure 4.1. The green dots are the SW corners of each of the blocks containing the gauges are offset by 0.01′ so as not to dichotomise the gauges; the lilac crosses are the coordinates of the PRECIS grid where the RCM rainfall estimates are found; the blue diamonds are the sites of the gauges. The coordinates are in minutes East of the Prime Meridian and North of the Equator (hence the negative values on the vertical axis). The box labels count across and then up from the SW corner as shown in Table 4.1, following SAWS numbering convention, with the difference that their blocks are 30′ square, whereas the PRECIS blocks are 26.4′ square. This mismatch required some careful organisation of coordinates.

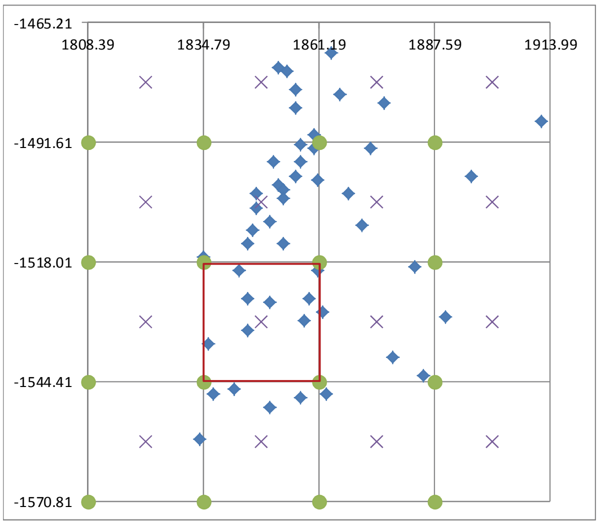


Figure 4.1 PRECIS grid and rain gauge sites – Mpumalanga. Red square is # 6, (as shown in Table 4.1) whose gauge weights appear in Table 4.2

Table 4.1: labelling of PRECIS blocks in the square containing gauges

13	14	15	16
9	10	11	12
5	6	7	8
1	2	3	4

The block occupations are as follows with empty blocks ringed blue:

BOX	1	2	3	0	0	6	7	8	0	10	11	12	0	14	15	16
# GAUGES	1	4	1	0	0	8	4	1	0	16	3	1	0	5	3	1

The gauge weights are all calculated by integrated Multiquadrics, as outlined in Section 3.3.2, restricted to those gauges which lie in each box, because inclusion of sites outside the box induces negative weights, which we want to avoid. Also, it seems sensible to restrict the box averages to the gauges inside the box; otherwise one is importing information from distance. The weights sum to 1.0. An example of the output of such a computation was performed on Box 6 which contains 8 gauges. Their labels, coordinates and weights appear in Table 4.2. The closer the gauges cluster, the lower their weights, which is clear from the details of Tables 4.2 and 4.3.

Table 4.2. Labels, coordinates and calculated weights of gauges in Block 6 of Figure 4.1.

Gauge	X	Y	Weight
2	1836	-1536	0.2186
5	1845	-1533	0.2021
13	1843	-1520	0.1141
14	1845	-1526	0.0549
18	1850	-1527	0.0759
24	1858	-1531	0.2094
25	1859	-1526	0.0471
26	1861	-1520	0.0780

Once all weights have been calculated for each box, they are arranged in a matrix relating gauge weights in each block to individual gauges. This matrix ensures that the correct information is collected in each column of box averages; the set for Mpumalanga appears in Table 4.3.

Table 4.3. Weights of all 48 Mpumalanga gauges in the 12 occupied blocks of Figure 4.1.

		active blocks											
block	gauge	1	2	3	6	7	8	10	11	12	14	15	16
1	1	1	0	0	0	0	0	0	0	0	0	0	0
6	2	0	0	0	0.2186	0	0	0	0	0	0	0	0
2	3	0	0.3691	0	0	0	0	0	0	0	0	0	0
2	4	0	0.0093	0	0	0	0	0	0	0	0	0	0
6	5	0	0	0	0.2021	0	0	0	0	0	0	0	0
2	6	0	0.2805	0	0	0	0	0	0	0	0	0	0
2	7	0	0.3410	0	0	0	0	0	0	0	0	0	0
3	8	0	0	1	0	0	0	0	0	0	0	0	0
censored		0	0	0	0	0	0	0	0	0	0	0	0
7	9	0	0	0	0	0.3115	0	0	0	0	0	0	0
7	10	0	0	0	0	0.1117	0	0	0	0	0	0	0
10	11	0	0	0	0	0	0	0.1431	0	0	0	0	0
10	12	0	0	0	0	0	0	0.0745	0	0	0	0	0
6	13	0	0	0	0.1141	0	0	0	0	0	0	0	0
6	14	0	0	0	0.0549	0	0	0	0	0	0	0	0
10	15	0	0	0	0	0	0	0.0636	0	0	0	0	0
10	16	0	0	0	0	0	0	0.1482	0	0	0	0	0
10	17	0	0	0	0	0	0	0.0619	0	0	0	0	0
6	18	0	0	0	0.0759	0	0	0	0	0	0	0	0
10	19	0	0	0	0	0	0	0.0198	0	0	0	0	0
10	20	0	0	0	0	0	0	0.0122	0	0	0	0	0
10	21	0	0	0	0	0	0	0.0071	0	0	0	0	0
10	22	0	0	0	0	0	0	0.0385	0	0	0	0	0
10	23	0	0	0	0	0	0	0.1111	0	0	0	0	0
6	24	0	0	0	0.2094	0	0	0	0	0	0	0	0
6	25	0	0	0	0.0471	0	0	0	0	0	0	0	0
6	26	0	0	0	0.0780	0	0	0	0	0	0	0	0
7	27	0	0	0	0	0.3209	0	0	0	0	0	0	0
11	28	0	0	0	0	0	0	0	0.1604	0	0	0	0
11	29	0	0	0	0	0	0	0	0.4673	0	0	0	0
7	30	0	0	0	0	0.256	0	0	0	0	0	0	0
8	31	0	0	0	0	0	1	0	0	0	0	0	0
10	32	0	0	0	0	0	0	0.1462	0	0	0	0	0
14	33	0	0	0	0	0	0	0	0	0	0.5110	0	0
14	34	0	0	0	0	0	0	0	0	0	0.0551	0	0
14	35	0	0	0	0	0	0	0	0	0	0.019	0	0
14	36	0	0	0	0	0	0	0	0	0	0.1289	0	0
10	37	0	0	0	0	0	0	0.0177	0	0	0	0	0
10	38	0	0	0	0	0	0	0.0637	0	0	0	0	0
10	39	0	0	0	0	0	0	0.0019	0	0	0	0	0
14	40	0	0	0	0	0	0	0	0	0	0.2859	0	0
10	41	0	0	0	0	0	0	0.0298	0	0	0	0	0
10	42	0	0	0	0	0	0	0.0607	0	0	0	0	0
15	43	0	0	0	0	0	0	0	0	0	0	0.3365	0
15	44	0	0	0	0	0	0	0	0	0	0	0.1597	0
11	45	0	0	0	0	0	0	0	0.3723	0	0	0	0
15	46	0	0	0	0	0	0	0	0	0	0	0.5038	0
12	47	0	0	0	0	0	0	0	0	1	0	0	0
16	48	0	0	0	0	0	0	0	0	0	0	0	1

To obtain the daily block averages of the observed gauge data, the column matrix of gauge observations (48 in the case of Mpumalanga) is post-multiplied by the weight table in Table 4.3.

4.3. Censored Gaussian Transformation of Rainfall

What to do with the zeros in a rainfall record? How do we find the correlations between pairs of gauge records? If we are to recorrelate these records to match the cccs of another set, how do we proceed? These questions led to an investigation which took the better part of 6 months of intense work and ended in the publication of Bárdossy and Pegram (2012), where the detail of the argument is put. We will here confine ourselves to the main points of the algorithm.

When we compare the spatial structure of the block averaged gauge data (derived using the methods of Section 4.2 above) against that of the PRECIS RCM values on their blocks, we find they are quite different. To ensure that the RCM is downscaled properly, we have to recorrelate their values so they have the right clustering behaviour. To measure the correlation, we have a choice between the methods devised by the following people: Pearson, Spearman and Kendall. However, in all the rainfall data there is a large quantity of zero values – typically of the order of 80% and above. These will yield misleadingly high cross correlation coefficients (cccs) between the individual time series. We performed some experiments, taking pairs of Gaussian distributed series that we constructed with ranges of given cccs. The most useful substitute for zeros in the individual series was the mean of the section of the Gaussian distribution below the threshold defining the zeros cut-off, an idea due to Rosenbleuth (1975). Figure 4.2 will give the idea – the blue circled cross locates the mean of this lower segment below [b] – if the Gaussian variable is labelled {y} we call the mean of the segment $\acute{y}_0$; it always has a negative value, given by the formula:

$$\acute{y}_0 = -\,\phi(b)/\Phi(b)$$

where:

b is the N(0,1) variate corresponding to zero in the rainfall set

$\phi(b)$ and $\Phi(b)$ are the density and cdf of the standard N(0,1) distribution.

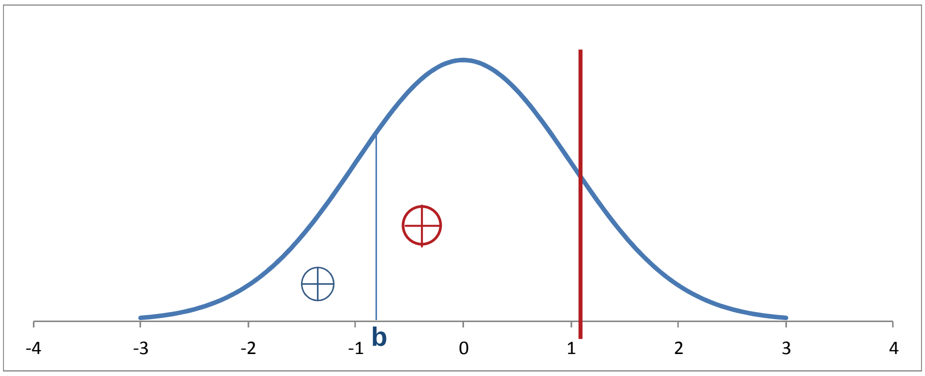


Figure 4.2. Computing the location of the average of the tail of an N(0,1) distribution below point [b]. The blue cross is for a humid environment, whereas the red cross is the centroid of a semi-arid environment with a dry probability P[0] = 0.8, typical of South African conditions, whose cut-off is the red line.

We also tried out the median of the tail as a candidate instead of the mean. We even examined the behaviour of measures other than Pearson (which assumes Gaussianity of the data) but they were not as good.

The upshot was that we ranked each time series (be they gauge, Block averaged gauge or PRECIS data) and transformed the rainfall values to the censored Gaussian distribution above {y}, setting the zeros to the $\acute{y}_0$ value (we call this operation: to Gaussianise). We then had a collection of transformed series in a matrix G, from which we could derive C (the cross correlation matrix of the set) in the conventional way.

4.4. Recorrelation of PRECIS data to match the Gauge Block Averages

Recall that there are two periods, the first (1 January 1990 to 31 July 2000) with raingauge records, the second (1 August 2000 to 31 December 2008) without gauge records (at least for the purpose of verification and validation). Contrary to our German experience, we found that the values in the cross correlation (cc) matrix of the RCM data were much larger than those of the Gauge block averages, but the methodology we used was the same as in Bárdossy and Pegram (2012). The first step was to recorrelate the Gaussianised RCM set to have the same cc-matrix as the Gauge block averages. That was done in the following way. Firstly, we found that the number of dry days in the PRECIS values was much less than those of the gauge block averages in all cases and also that the blocks in the second period of PRECIS data were wetter than the first. A sketch will help to understand the procedure.

In Figure 4.3, matrix G is the set of Gaussianised Gauge Block Averages, matrix Y1 is the set of Gaussianised raw PRECIS data for period 1 and PR1 is the resulting recorrelated PRECIS data-set for this first period.

In more detail, from matrix G we calculate its cross correlation matrix C. We derive the "square root" matrix of C by Singular Value Decomposition (SVD) in the following way. SVD of G yields 3 matrices:

$$C = U_c W_c V_c^T$$

where W_c is a diagonal matrix of singular values. If we are fortunate and all the Eigenvalues of C are non-negative, then $U_c = V_c$ and it turns out that $U_c V_c^T = I$, the identity matrix. We can then take the square root of each diagonal element of W and write the resulting matrix as $W_c^{1/2}$. The square root matrix S of C is then:

$$S = U_c\, W_c^{1/2} U_c^T$$

Similarly, we can obtain the cc matrix of Y1 and call it T':

$$T' = U_T\, W_T^{1/2} U_T^T$$

and obtain its inverse square root matrix by taking the reciprocal of W:

$$T = U_T\, W_T^{-1/2} U_T^T$$

where the very small values in W_T are put to zero in $W_T^{-1/2}$, if necessary.

Lastly, we postmultiply Y1 by T and then S, in turn, to give the recorrelated set of PRECIS1 data:

$$PR1 = Y1\ T\ S$$

which first decorrelates Y1, then recorrelates the product to the Gauge block cc relationship.

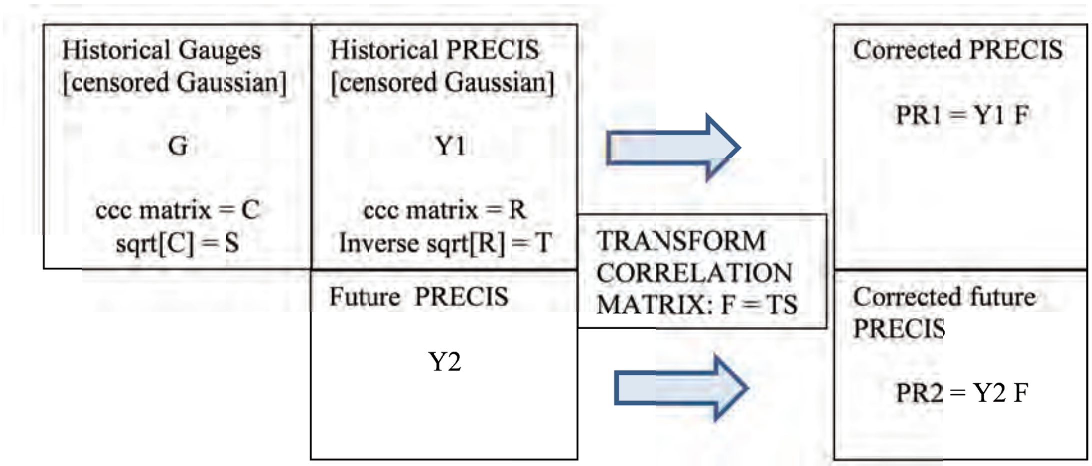


Figure 4.3. Block diagram of the recorrelation of PRECIS data to match Gauge block averages

The lower two boxes in the sketch, labelled 'Future PRECIS', are treated as follows. Y2 is the matrix set of the Gaussianised PRECIS block estimates in the "future" period [Aug 2000 to Dec 2008]. We call it the future, because at the time of starting we did not have the raingauge records for this period. Only in 2011 were we able to unpack them and then 'hide' them from the analysis until they were needed for validation of the downscaling method.

We want to preserve the differences in the rainfall distributions and correlations between the two sets of PRECIS values. We therefore treat the Y2 data in exactly the same way with the same de- and re-correlation matrices as we used on the Y1 data, and see if there is a difference. Thus we postmultiply Y2 by F=TS, as we did Y1. It turns out that there are mild but insignificant differences between the resulting cccs.

Figure 4.4 has three panels showing comparisons of cccs of PRECIS values (vertical axes) against those of the Gauge Block averages (horizontal axes) over the Mpumalanga region. In the left, middle and right panels are the cccs of (i) raw PRECIS1, (ii) recorrelated PRECIS1 and then (iii) recorrelated PRECIS2 against those of the gauge block averages, all Gaussianised.

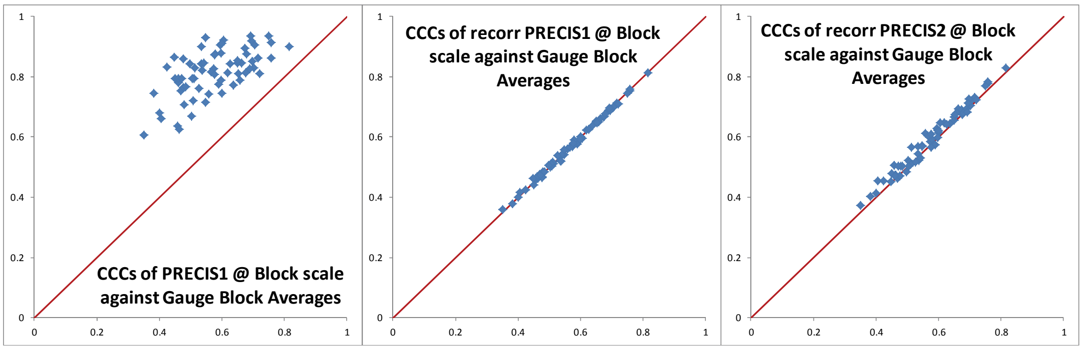


Figure 4.4. Comparisons of the cccs of (i) raw PRECIS1, (ii) recorrelated PRECIS1 and then (iii) recorrelated PRECIS2 [vertical axes] against those the cccs of the gauge blocks [horizontal axes].

Notice the initially considerably higher cccs of PRECIS1, compared to the gauge block averages, in the left panel, following the differences shown in Figures 3.12 and 3.13. The recorrelation process appears to completely eliminate the bias in P1 as seen in the middle panel and importantly the procedure successfully retains the minor differences of the cccs of PRECIS2 vs. PRECIS1 as shown in the right panel. Results for the other 4 regions differ in minor detail from the above so are redundant and not included here. Unfortunately these results are optimistic and are only truly resolved in Chapter 5.

4.5. Summary

The recorrelation of PRECIS data to match gauge Block averages is interesting and sets out the methodology, but is not intrinsically useful for hydrological applications, which are the eventual target of this research work. The reason is that most hydrological studies are based on raingauge data, not on areal averages over 50 km square blocks. In the next Chapter we adapt the above methodologies to gauge-scale downscaling of PRECIS RCM rainfall estimates.

5. Downscaling RCM estimates of rainfall to the gauge scale

Preamble

The purpose of this Chapter is to assess the changing stochastic (space-time) properties of rainfall associated with plausible climate change projections. The work presented here, together with the earlier work anticipating it in Chapter 4, provides the tools to make this assessment possible. Examples of our ability to model the changing stochastic (space-time) properties of rainfall during a validation period, having found the linkages between RCMs and gauge rainfall in a calibration period, are demonstrated here. We also resolve a difficult paradox, hinted at in the Summary in Section 4.5.

The techniques for ensuring that correct spatial dependence, at the gauge and large scales (25, 50, 100 km) simultaneously, have only recently been identified by us (Bárdossy and Pegram, 2009, 2011 and 2012). In the most recent paper a new practical methodology of recorrelation was laid out and summarised in Chapter 4 of this report. The large-scale correlations are of paramount importance in the estimation of extreme rainfall over large areas, thus the ability of this methodology to provide a solution is a boon. It seems that few modellers have noted, let alone addressed, this areal property of rainfall in a downscaling context.

Unfortunately and worryingly, since May of 2012, it was found that the method suggested and reported in the papers and Chapter 4 did not go quite far enough. To define and solve the problem and suggest a complete solution required a prolonged research thrust, requiring experimentation in various directions many of which were blind alleys, an exercise which involved a large commitment of thought, analysis and computer programming. As a result of this endeavour, the steps of a novel method of performing effective recorrelation by means of iteration and correction are presented here, making it possible to downscale RCM data, virtually capturing not only the space-time properties of daily raingauge networks but also their large integrated spatial averages. This activity resolves the paradox referred to above.

It is the main purpose of this Chapter to show that we can properly transform sets of RCM time series to have desired cccs and marginal distributions matching those of rainfall networks, in particular making downscaled RCM data in the image of the observed rainfall. The relatively dense explanation, which is more mathematical than usual, is required to define the subtlety of the corrective procedure. Not to present it properly would hide the fact that it has taken several man-months of computer programming and exploratory calculations to define the problem and solve it. Nevertheless, the explanation will be kept concise and to the point.

Thus, in this report, we present a methodology which is designed to ensure that the downscaled gauge records have the correct spatial interdependence structure as measured by the cross correlation coefficients (cccs), as well as the correct marginal distributions of amounts. In recorrelating and quantile transforming the downscaled sets of PRECIS data for the periods 1990-2000 and 2000-2008, we are able to generate sets of daily gauge network rainfall ensembles, which have the correct marginal distributions of amounts of rainfall over chosen mesoscale regions of RSA. The new method of iterative recorrelation we present here promises that the spatial properties are also preserved.

5.1 Outline of Recorrelation procedure

The easiest way to define the problem is to start by describing the recorrelation procedure in a verbal form of 'pseudo-code'. Having obtained the weights associated with each gauge in a block, using the algorithm explained previously in Chapter 4, we take the following programming steps to perform the initial downscaling of some RCM data over a selected mesoscale region, which will be one of the five zones: Cape, East London, Free State, Upper Vaal and Mpumalanga.

We work in two steps. First recorrelate (using an innovative iterative approach) and downscale the PRECIS1 rainfall estimates using quantile-quantile (QQ) transforms via the gauge data during period 1 [Jan 1990 to July 2000]. The second step is to use the ***same*** recorrelation matrices derived in the first stage (as indicated in Figure 4.3), together with the observed gauge rainfalls in Period 1, to recorrelate and downscale the PRECIS2 rainfall estimates in Period 2.

The algorithm as pseudo-code:

1. Obtain the raw PRECIS1 and PRECIS2 data and the rainfall data for the region
2. Obtain the table of the gauges' associated block numbers, indicating which gauges belong to each PRECIS block
3. In each observed gauge record in Period 1, where there are zeros (dry days), put $\acute{y}_0 = -\phi(b)/\Phi(b)$ which is the average of the tail of the Gaussian distribution below $b = \Phi^{-1}[p_0]$, where p_0 is the proportion of the dry days in the record and ϕ and Φ are respectively the probability density function and cumulative distribution function of the Normal distribution – see Figure 4.2.
4. Having ranked all the positive rainfall values for each gauge in turn above its p_0, as $[r_1, r_2, \ldots, r_n]$, Gaussianise them as $z_j = \Phi^{-1}[r_j]$ and collect them in matrix Z.
5. Compute the ccc matrix C of the set Z of these Gaussianised observed gauge values, in the usual way using Pearson correlations; C is the target ccc matrix with the qualities we need to impart to the downscaled RCM data
6. Compute C's symmetrical square root S such that $SS = C$. This step is performed using Singular Value Decomposition, as explained in Chapter 4.

That first set of calculations gives us our target defined by the observations at the gauges. The next stage is the setting up of the Y matrix of initial gauge values obtained from the PRECIS1 data set. We choose only those PRECIS blocks where there are one or more gauge sites in a block, each with a label of its block membership, as noted in step 2 above; we do not try to fill blocks empty of gauges.

7. Fill matrix Y with Gaussianised PRECIS1 data; start by giving the gauges on that day the parent Gaussian PRECIS rainfall estimate. Where the data rank is below the p_0 of the gauge concerned, put $\acute{y}_0$ which is the average of the Gaussian tail below b for that gauge determined in step 3.
8. Add some noise to all Gaussianised PRECIS values above the threshold b (corresponding to those PRECIS values whose wet quantiles are above the gauges' p_0); after tests, we found that Gaussian noise with a standard deviation of 1/3 was a good choice.
9. Calculate R the ccc matrix of the initial downscaled gauge data in Y using Pearson.

Next perform the recorrelation step to get the target gauges' Gaussian values

10. Compute the square root matrix T of R^{-1}, the inverse of R, such that $TT = R^{-1}$.
11. Form the Transform matrix $F = TS$
12. Recorrelate the Gaussianised PRECIS1 gauge-scale values to $V = YF$

This is the procedure based on that described in Chapter 4. It produces a matrix of recorrelated gauge data in matrix V, examples of which time series can be seen in Figure 5.1.

The important conclusion made in Chapter 4 about the methodology at this stage was that it appears to successfully retain the relative minor differences of the cccs of PRECIS2 vs. PRECIS1 as can be seen when comparing the middle and the right panels of Figure 4.4; there it can be seen that there has been a slight increase of the cccs in the second compared to the first period.

The next step was to check the cccs between (i) the downscaled gauges against the corresponding observations and (ii) the downscaled blocks versus the observed block averages, in the rainfall space, after quantile-quantile (QQ) transformation. Starting with PRECIS1 (the record in the first 10-year period), what this involved was to match the quantile of each Gaussianised downscaled gauge record against the observed gauge quantile and assign the observed rainfall amount to the downscaled gauge. Of course, any downscaled gauge value below a particular gauge's p_0 was set to zero. The series would not be exactly matched in time, but would match in distribution. Then two sets of time series were created: PRECIS1 downscaled to the gauge locations and their block averages obtained from the weights calculated by the Multiquadric method explained in Chapter 4.

Here, plotted in Figure 5.1, follow some selected time series of rainfall from the Upper Vaal region after recorrelation as described in the above algorithm.

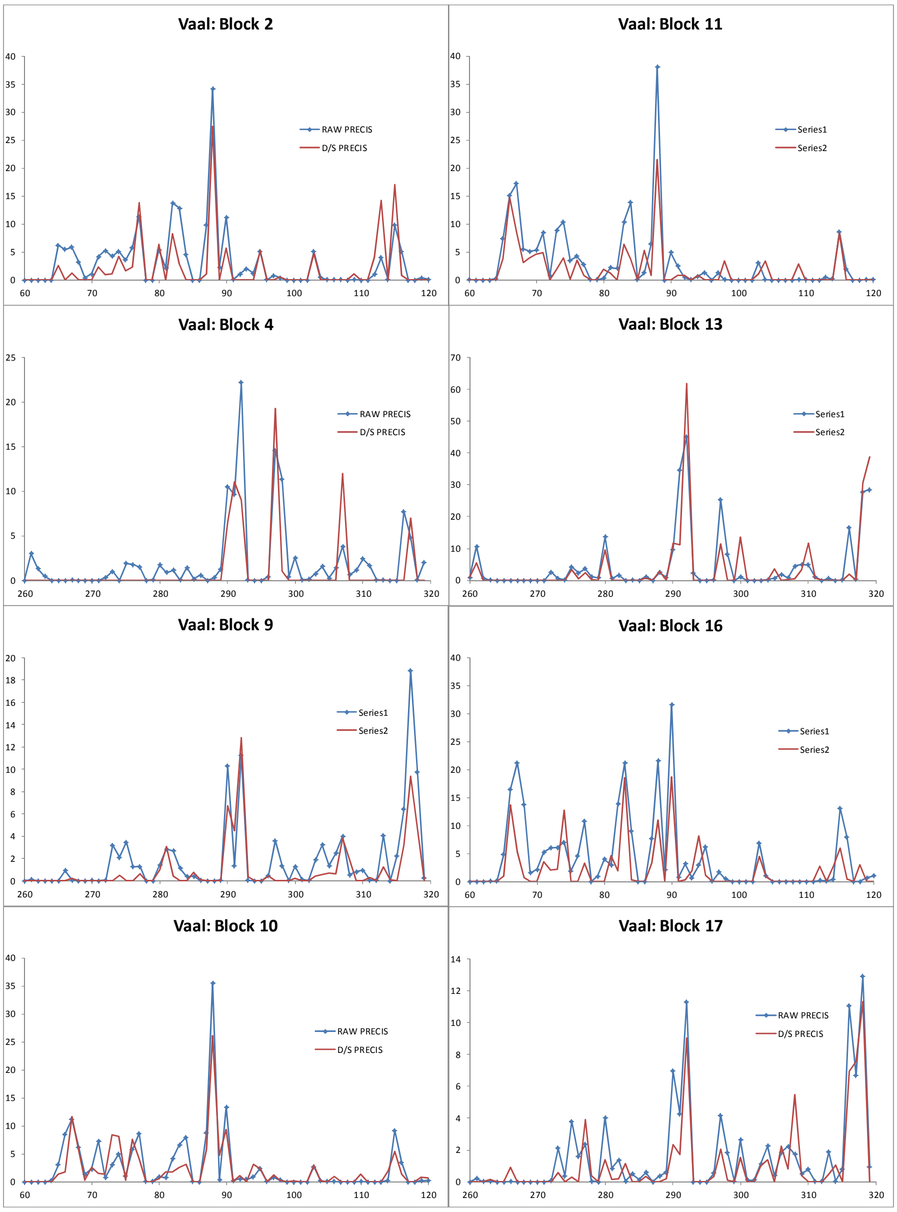


Figure 5.1. Block averaged Raw PRECIS1 time series [blue lines] and Downscaled Block-averaged gauge series [red lines] derived from PRECIS1 in periods: March-April 1990 [days 60-120] and September-October 1990 [days 260-320].

They are Block averaged Raw PRECIS1 and Downscaled Block-averaged gauge values derived from PRECIS1. There are two short periods: March-April 1990 [days 60-120] and September-October 1990 [days 260-320]. They seem to be relatively well matched in timing and values. The individual downscaled raingauges (not shown here) do not match the observed ones well, because of the noise added to the PRECIS values before recorrelation.

To compute the cccs of the recorrelated Downscaled PRECIS block data, the daily rainfall totals are ranked and then Gaussianised as explained in steps 3 and 4 of the pseudo-code, with all the dry days' zero values transformed to the individual stations' $\acute{y}_0$ value. Figure 5.2 compares the cccs of the Gaussianised Observed and Downscaled Block daily Averages for the Vaal and Free State regions and includes some time series for the Free State similar to Figure 5.1.

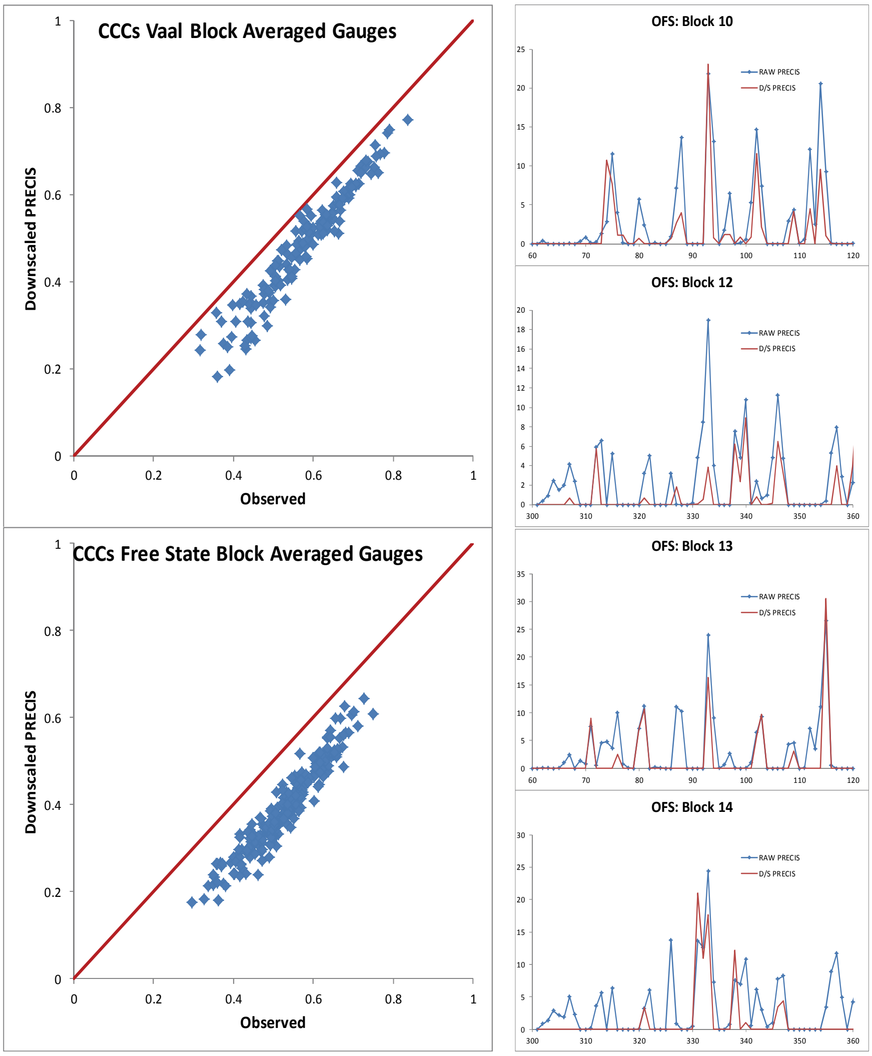


Figure 5.2 Left panel: compares cccs of the Observed and the Downscaled Block daily Averages for the Vaal and Free State regions. Right panel: time series for the Free State similar to Figure 5.1.

Immediately (and disappointingly) obvious is that the cccs of the Downscaled block averages are shifted about 0.15 lower than the observed, in both regions. With the above methodology, we have not been able to fully recover the observed correlations. Why? In many months of searching for a solution, we examined methods of QQ transforms, rewrote the code of the programs and tried many other ideas, too sad to relate. Then it dawned on us.

The values of the QQ'd dry values (below p_0) of the Gaussianised data that produced Figure 4.4 are not the same as those that produced Figure 5.2, even though the (wet) values above the p_0 threshold are the same in both sets. The reason for this disparity is that the recorrelation procedure alters the values below the threshold as well as above whereas, after Gaussianisation, the downscaled rainfall maintains the same values above p_0, but causes their below-threshold values to be to ***reset*** to $\acute{y}_0$. This was not picked up for a long time because of the complexity and subtlety of the procedure. How to remedy this? After some experimentation, we came up with the following idea: iteration. However, the programming problem to be solved was how to do this and how to determine when convergence was achieved on 60+ stations each with 3000+ days of record. The next subsection gives a small dimensional example to explain the methodology of the solution.

5.2 A Two-Dimensional Example of Recorrelation by Iteration

Assume we have two stations where we have obtained estimates of rainfall from downscaling PRECIS data. We consider 20 days to keep the example manageable. These data are listed in Table 5.1 as A and B. Their dry probability (p_0) values are respectively 0.40 and 0.55. Their ranks r_j (the zeros are tied) are in RANKA & B, from which the quantiles are calculated as $r_j/(n+1)$, using the Weibull formula for n = 20 data, appearing as PROBA & B. From these probabilities the Normal transform yields GAUSSA & B, the last pair of columns under the PRECORR block, where the original zero values are given Gaussian $\acute{y}_0$ values of -0.842 and -0.598 for sequence A & B respectively. Below the table are the cccs calculated between the columns above them. The decorrelated data and the recorrelated sets result from the following computation.

Table 5.1 Two vectors (20 terms) and their Gaussianisation and recorrelation

Day	ORIGINAL		PRECORR						DECORR		RECORR	
	A	B	RANKA	RANKB	PROBA	PROBB	GAUSSA	GAUSSB	GAUSSA	GAUSSB	GAUSS A	GAUSS B
1	0	0	13	10	0.381	0.524	-0.842	-0.598	-0.781	-0.495	-0.848	-0.672
2	1.92	0	4	10	0.810	0.524	0.876	-0.598	0.987	-0.740	0.390	-0.439
3	0.89	1.48	6	5	0.714	0.762	0.566	0.712	0.481	0.653	0.775	0.831
4	0	0	13	10	0.381	0.524	-0.842	-0.598	-0.781	-0.495	-0.848	-0.672
5	3.01	2.34	2	3	0.905	0.857	1.309	1.068	1.195	0.912	1.538	1.447
6	1.27	1.51	5	4	0.762	0.810	0.712	0.876	0.609	0.800	0.967	1.029
7	0	4.59	13	2	0.381	0.905	-0.842	1.309	-1.053	1.467	-0.607	0.607
8	0.12	0	11	10	0.476	0.524	-0.060	-0.598	0.024	-0.607	-0.322	-0.573
9	5.88	0.2	1	8	0.952	0.619	1.668	0.303	1.674	0.074	1.496	0.980
10	0.02	0.02	12	9	0.429	0.571	-0.180	0.180	-0.211	0.211	-0.068	0.068
11	0	0.51	13	6	0.381	0.714	-0.842	0.566	-0.947	0.702	-0.713	0.041
12	0	0	13	10	0.381	0.524	-0.842	-0.598	-0.781	-0.495	-0.848	-0.672
13	0.26	0	7	10	0.667	0.524	0.431	-0.598	0.528	-0.677	0.051	-0.502
14	0.21	0.26	8	7	0.619	0.667	0.303	0.431	0.251	0.400	0.437	0.485
15	0	0	13	10	0.381	0.524	-0.842	-0.598	-0.781	-0.495	-0.848	-0.672
16	0	0	13	10	0.381	0.524	-0.842	-0.598	-0.781	-0.495	-0.848	-0.672
17	0	0	13	10	0.381	0.524	-0.842	-0.598	-0.781	-0.495	-0.848	-0.672
18	0.18	0	10	10	0.524	0.524	0.060	-0.598	0.147	-0.624	-0.231	-0.556
19	0.19	8.17	9	1	0.571	0.952	0.180	1.668	-0.052	1.692	0.880	1.442
20	2.8	0	3	10	0.857	0.524	1.068	-0.598	1.184	-0.767	0.536	-0.411
CCCs	-0.0494						0.2714		-0.0008		0.8005	

The cccs of the original data [-0.0494] and the Gaussianised Precorrelated data [0.2714] differ so much because the raw data have a large number of zeros and the positive values are strongly skewed as can be seen in Figure 5.3. Figure 5.4 shows the Precorrelated Gaussianised sets with a positive but small ccc of 0.2714. Note their constant $\acute{y}_0$ values.

Because the sample dimension is small, we can determine the recorrelation matrices by direct computation. If the ccc is r, then the 2x2 decorrelation matrix T has the form:

on the main diagonal, (1/(SQRT(1+r))+1/(SQRT(1-r)))/2 = 1.0292
on the off-diagonal, (1/(SQRT(1+r))-1/(SQRT(1-r)))/2 = -0.1424

The T and S matrices have the following values in this example:

T =	1.0292	-0.1424		S =	0.9487	0.3162
	-0.1424	1.0292			0.3162	0.9487

Post-multiplication of the matrix Y (containing Precorr GAUSSA & B) by T yields $Q = YT$ (containing columns Decorr GAUSSA & B) with r = -0.0008, which is close to zero, as desired. The Decorr data sets appear in Figure 5.5.

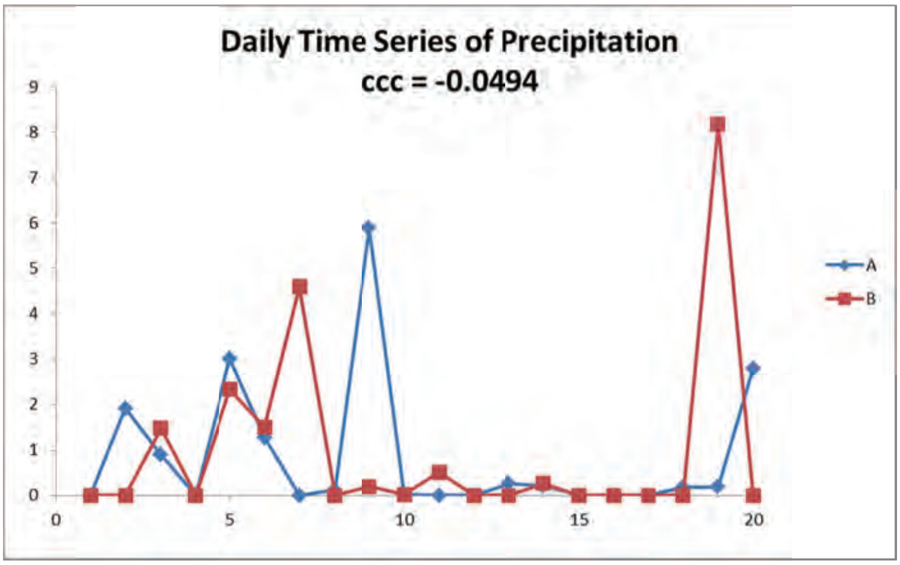


Figure 5.3 Example data A & B from Table 1.

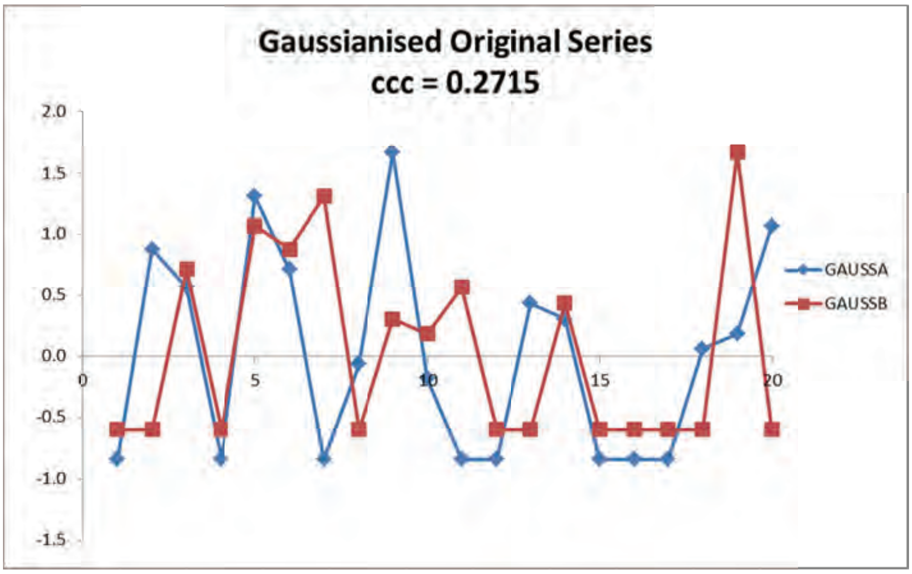


Figure 5.4 Precorrelated Gaussianised sets GAUSSA & B with a positive but small ccc.

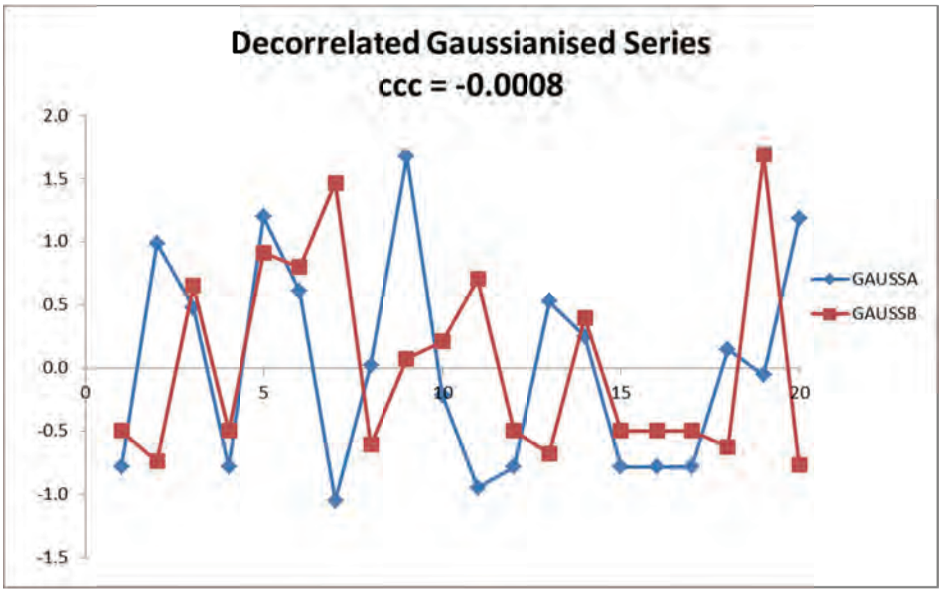


Figure 5.5 Decorrelated data sets at 1st iteration

To recorrelate the series in Figure 5.5 we need the S matrix with the following terms:

on the main diagonal, (SQRT(1+r)+SQRT(1-r))/2 = 0.8944
on the off-diagonal, (SQRT(1+r)- SQRT(1-r))/2 = 0.4472

then the Recorrelated set is matrix $V = QS$ appearing in the last 2 columns of Table 5.1 and depicted in Figure 5.6.

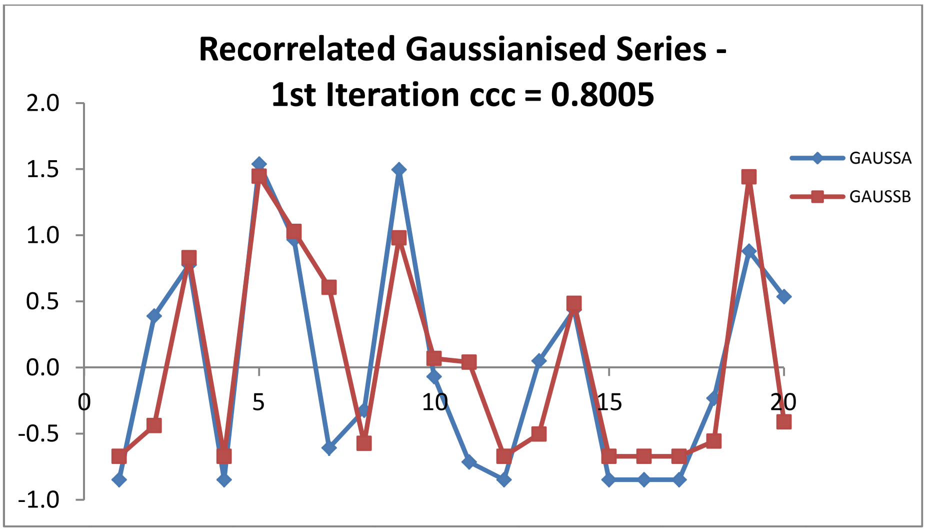


Figure 5.6 Recorrelated data sets from Table 5.1

However, the values matching the zeros in Figure 5.6 are not constant after recorrelation - note the GAUSSA values at 7 & 11 days (-0.607 and -0.713, compared to -0.848 for days 1, 4 & 12 as seen in Table 5.1 – 2nd last column) and GAUSSB at 2, 13 & 18 days. If we set them back to constant $\acute{y}_0$ values without altering the values above the respective p_0 threshold, then we get the result shown in Figure 5.7, with a ccc of 0.6151, much less than the 0.8, the value we got before adjusting the thresholded values. Paradox solved!! To solve the problem, this set is now iteratively treated as a Precorrelation set and it goes through the same procedure as before: calculate a new decorrelation matrix T_1, calculate $Q_1 = Y_1T_1$ and recorrelate that with the desired recorrelation matrix (fixed) S. We get the series in Figure 5.8, which, after correction has an improved ccc of 0.7311, as desired. This figure shows the second iteration after zero correction. Note the mild changes between Figures 5.7 and 5.8.

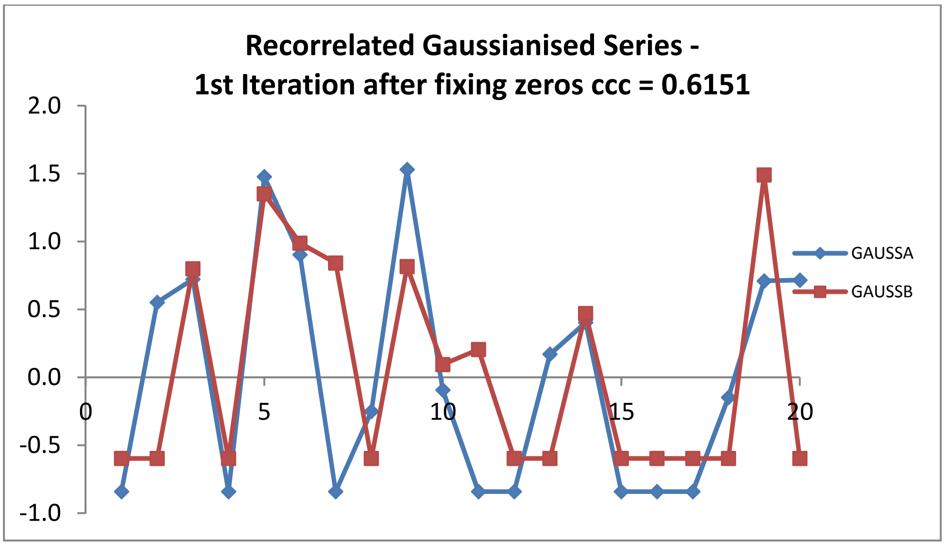


Figure 5.7 Recorrelated Gaussian Series after correction of values below p_0 threshold

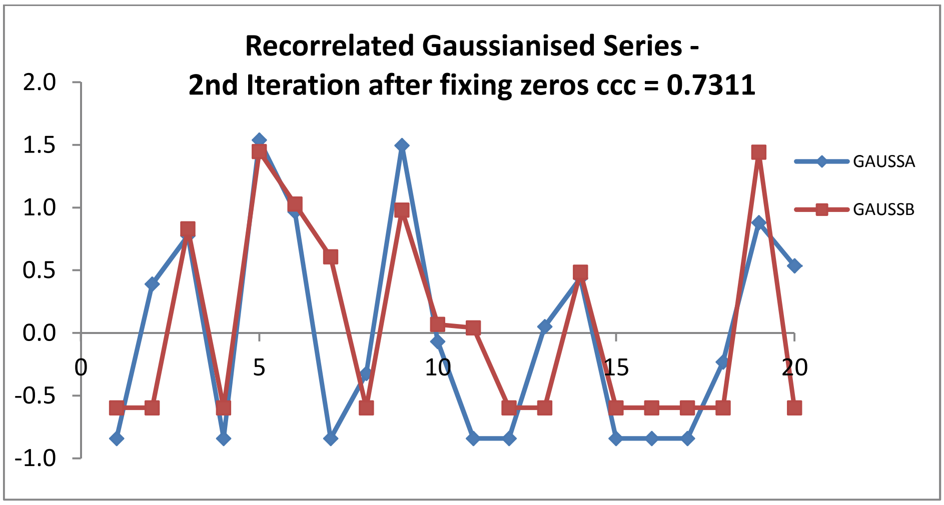


Figure 5.8 Recorrelated Gaussian Series after 2nd iteration

Figure 5.9 shows the sequence of iterations of GAUSSB up to, and compared to, the 11th iteration. Reading from left to right and top to bottom, the first panel shows the original precorrelated GAUSSB relative to the 11th iteration (the orange rings), before zero correction. The sequence is then 1st, 2nd 3rd and 4th iterations compared to the last iteration 11. The corresponding cccs are: 0.6151, 0.7311, 0.7727, 0.7889 and 0.8000. Even after 4 iterations it is evident that there is still some correction required at points 2, 13 and 18.

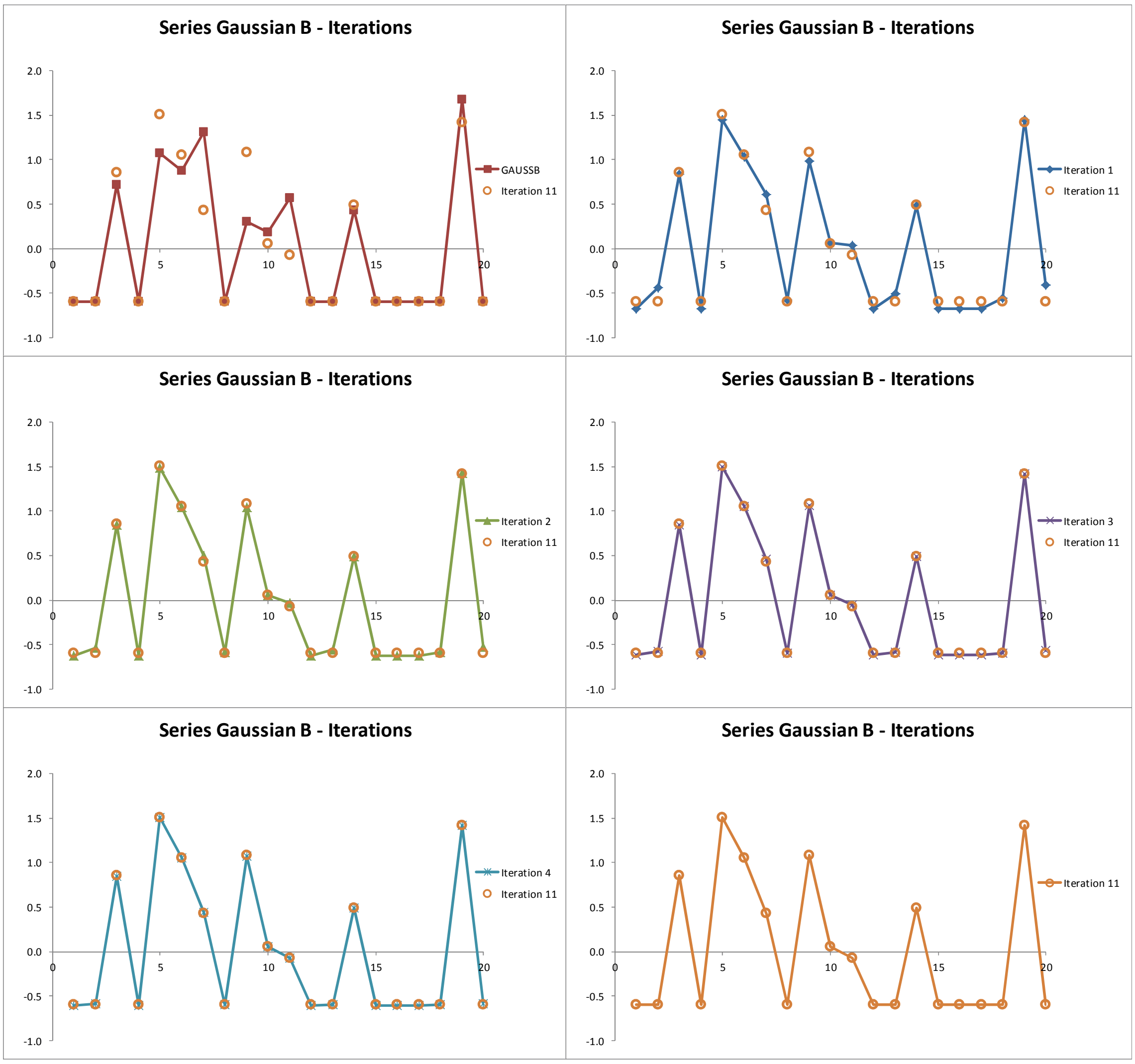


Figure 5.9 The sequence of the original and the first 4 iterations of GAUSSB compared to the 11th iteration which is also shown last.

In summary, the iterative recorrelation procedure works well on a small dimensional example, as shown in this section. It remains to apply it to a real problem.

What it involves is that the Algorithm at the beginning of the section is modified after point 6 by replacing that text with the following:

> The next stage is the setting up of the Y matrix of initial gauge values obtained from the PRECIS1 data set. We choose only those PRECIS blocks where there are one or more gauge sites in a block, each with a label of its block membership, as noted in step 2 above.
> Set J = 1

7. Fill matrix Y_j with Gaussianised PRECIS1 data, i.e. in each block, start by giving the gauges on that day the parent Gaussian PRECIS rainfall estimate. Where the data rank is below the p_0 of the gauge concerned, put $\acute{y}_0$ which is the average of the Gaussian tail below b for that gauge determined in step 3.
8. Add some noise to all Gaussianised PRECIS values above the threshold b (corresponding to those PRECIS values whose wet quantiles are above the gauges' p_0); after tests, we found that Gaussian noise with a standard deviation of 1/3 was a good choice.
9. Calculate Rj the ccc matrix of the initial downscaled gauge data in Y_j using Pearson.

Next perform the recorrelation step to get the target gauges' Gaussian values

10. Compute the square root matrix Tj of Rj^{-1}, the inverse of Rj, such that $TjTj = Rj^{-1}$.
11. Form the Transform matrix $Fj = TjS$
12. Recorrelate the Gaussianised PRECIS1 gauge-scale values to $V_j = Y_jFj$

    New step – set $Y_{j+1} = V_j$, put J = J + 1 and go back to step 13 a few times, storing each iteration's new version of *Tj*; all else remaining the same.

In downscaling PRECIS2 data, do not go back to recalculating the *S* matrix of the Gaussianised gauge data, but keep the original from PRECIS1. Then perform the replicate steps of the above procedure omitting the re-estimation of the sequence of *Tj* matrices, so that the steps are:

Set J = 1

13' Fill matrix Y_j with Gaussianised PRECIS1 data, i.e. in each block, start by giving the gauges on that day the parent Gaussian PRECIS rainfall estimate. Where the data rank is below the p_0 of the gauge concerned, put $\acute{y}_0$ which is the average of the Gaussian tail below *b* for that gauge determined in step 3.

14' Add some noise to all Gaussianised PRECIS values above the threshold *b* (corresponding to those PRECIS values whose wet quantiles are above the gauges' p_0); after tests, we found that Gaussian noise with a standard deviation of 1/3 was a good choice.

18' Recorrelate the Gaussianised PRECIS1 gauge-scale values to $V_j = Y_jFj$

Set $Y_{j+1} = V_j$ & J = J + 1 ; Return to 13' until convergence

5.3 Recorrelation by Iteration – Application to Downscaling RCM data

The iterative recorrelation algorithm was coded in FORTRAN and the results follow. Raw PRECIS data from the first decade were downscaled to gauge sites and then QQ'd to rainfall. It is used in 2 phases.

The core QQ routine works as follows:

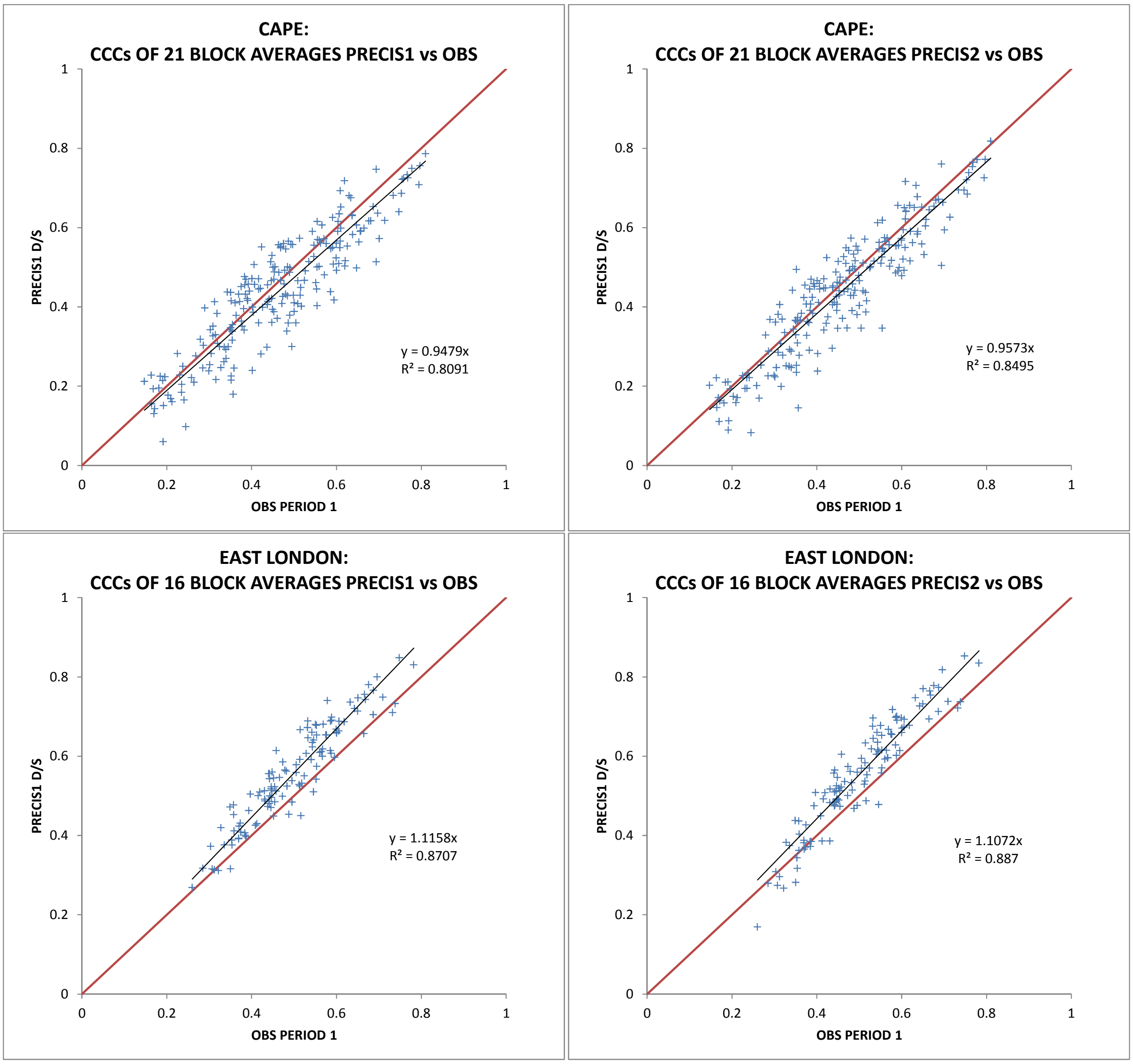
It is designed to get an individual target *x* from Gaussian probabilities p_x via ranked observed rainfall data. It uses an exponential curve fitted to the top two points of the observed data curve as an approximation to the Extreme value Type 1 distribution, when extrapolation is required; see section 18.2.2 of Handbook of Hydrology (1993).

The first phase is in Period1 [1990/2000] when we have an observed gauge record and matching Gaussianised PRECIS1 data, respectively. In this case, because the dates match in both sets, there is no need to condition on CPs as all days have concurrent CPs and match. The resulting quantile/rainfall plots for the observed and downscaled values should be nearly the same, except at the extremes, thus we could use one vector for each set.

When it comes to the second period, the downscaling has to be CP dependent. This split is organised outside the QQ subroutine, which accommodates vectors of different lengths. The routine returns the downscaled rainfall values, which is the responsibility of the calling routine to allocate to the correct CP -dependent days.

5.4 The results

The PRECIS data downscaled to raingauge sites were block averaged over their parent blocks. The cccs for the two sets of recreated block-scale data were calculated and are shown in Figure 5.10 (occupying this page and the next), for the 5 regions. The legends are self-explanatory. The results are a considerable improvement over those in Figure 5.2 – the scatter is a natural result of the iterative recorrelation process and introduces some uncertainty. New sequences with the same overall behaviour may be generated by changing the random number seed. East London and Mpumalanga both present modelled cccs which are about 10% higher than the observed, but the remainder, which sport a larger number of gauges give ***very*** satisfactory results!



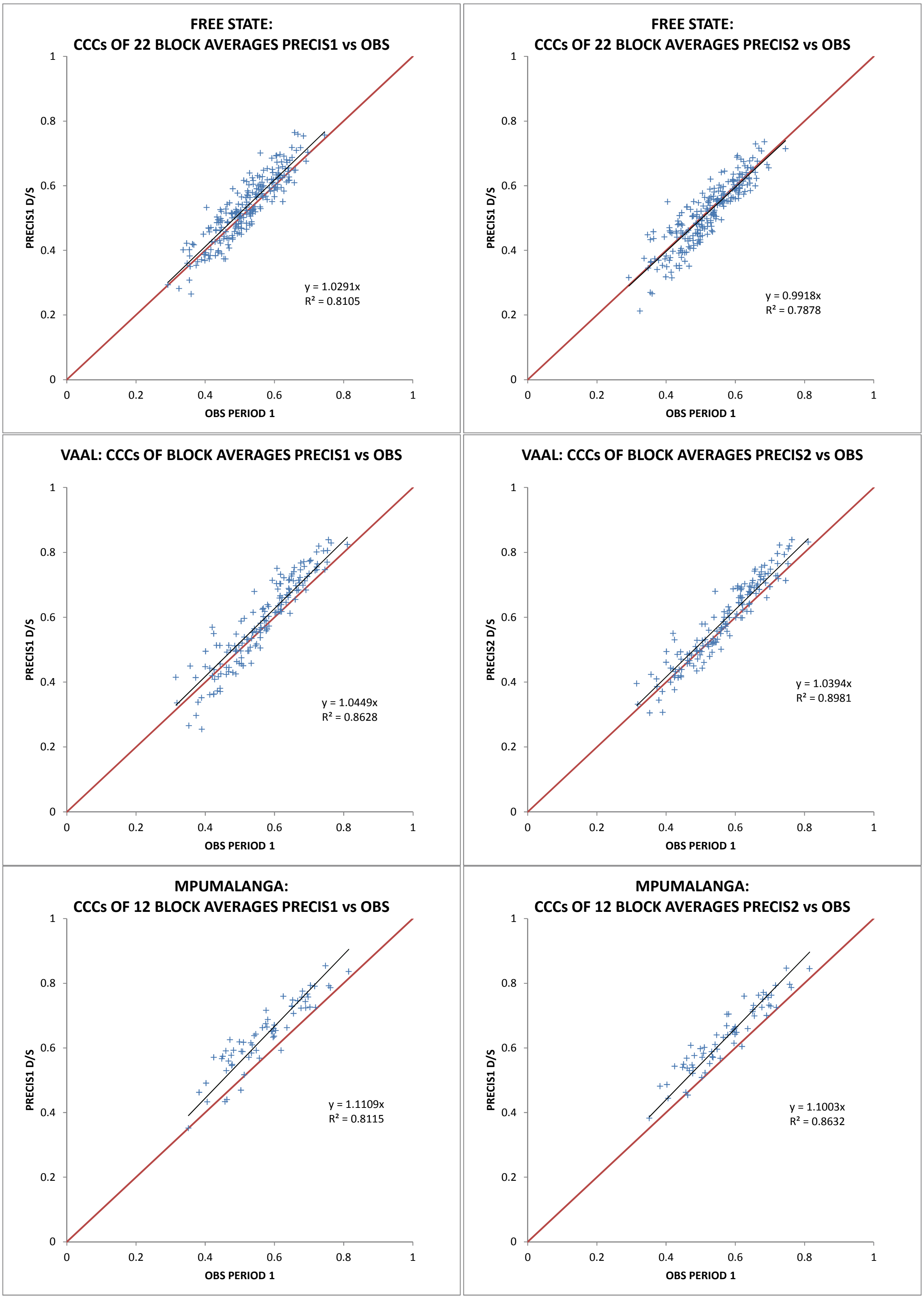


Figure 5.10 Comparison of Cross Correlation Coefficients for Block Averaged values PRECIS1 in the left column and PRECIS2 on the right

We follow the block-averaged results with those of the cccs of the downscaled PRECIS gauge estimates compared to the observations in Period 1 for 4 out of the 5 regions; the results for the Free State are very similar to the Vaal and to include the 5th panel would have spoilt the display!

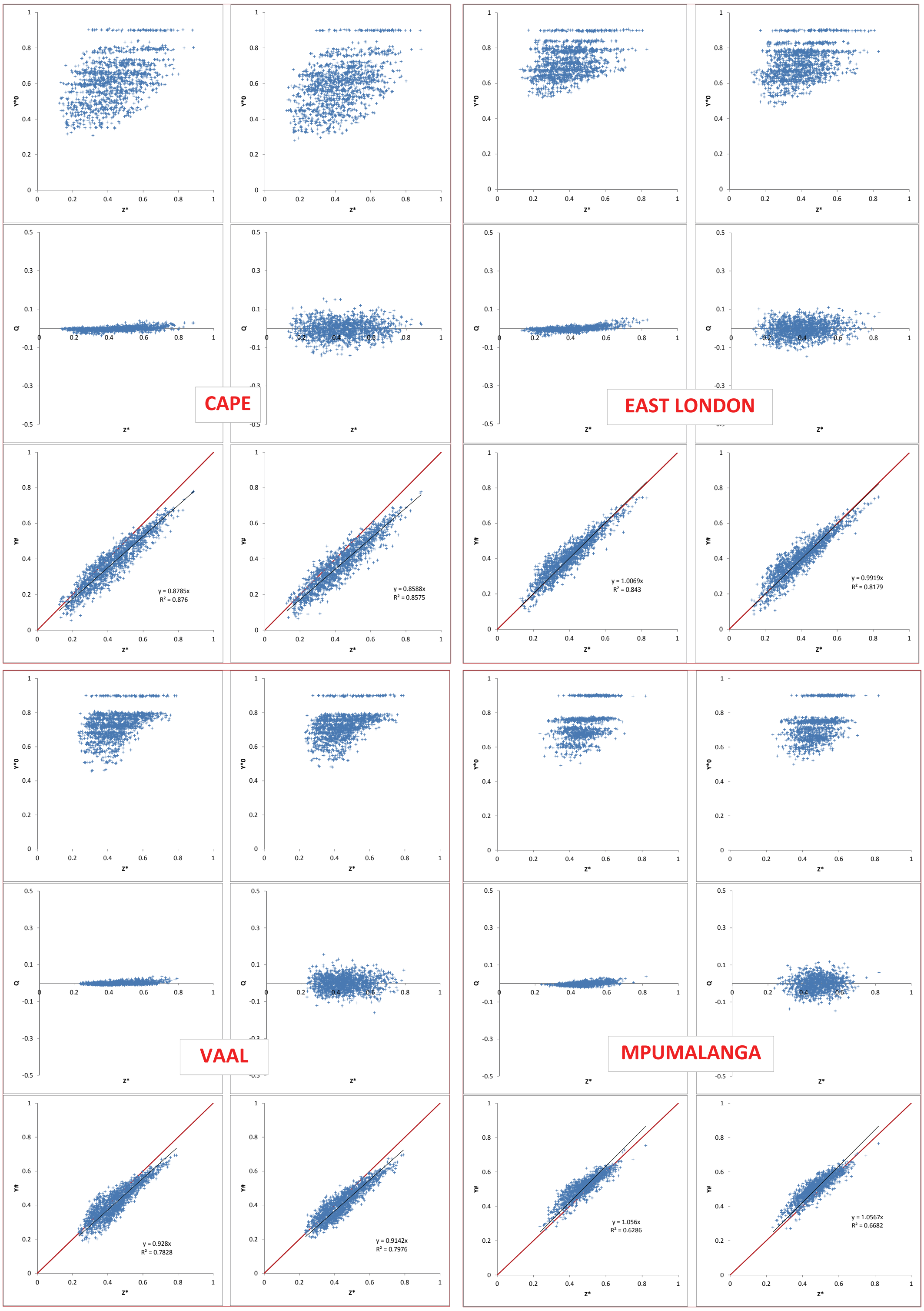


Figure 5.11 The cccs of the downscaled PRECIS gauge estimates compared to the observations. Top left block to bottom right: Cape, East London, Vaal & Mpumalanga. In each block the three left panels are PRECIS1, the right three are PRECIS2. From top to bottom, the three panels compare uncorrected, decorrelated and recorrelated ccs of gauge estimates in the vertical against the cccs of the observations on the horizontal axes

Some of the recorrelated cccs are off a bit; for the Cape the offset ratios are 0.87 and 0.85, For East London they are exactly 1.00, for the Vaal they are 0.93 an 0.91 and for Mpumalanga, they are both 1.06. Overall, this is a very satisfactory result.

5.5 Validation of the results of the downscaling

'The proof of the pudding is in the eating', so besides getting the cccs right, it remains to check the marginal probability distributions of some of the gauges in each region, both conditioned on CPs and also over the full period. We want to check the values obtained from the downscaled PRECIS data in both periods [1990 to 2000 and 2000 to 2008], against the observed rain gauge data in each period. We note again that the gauge data in Period1 was used to train the modelling procedure and was indeed used for downscaling each of the PRECIS estimates, so we were apprehensive about the correspondence of the downscaled PRECIS2 values with the observations in the second period, where we had hidden the data. We note that although the CPs are patterns that are fixed on each day, the membership or collection of CPs that affect the weather in the different regions are different. Thus we cannot find a given labelled CPn in more than one of the regions which has the same pattern – they cannot be numerically ordered in that way, even though they can be associated as shown in Figure 3.17. To emphasise the point, here is a list of the first few CPs in the first period by region. They do not even change from one to another on the same day, from the point of view of influencing the weather in the different regions, although the pattern on each day covering the country is common to all.

Table 5.2 List of CPs by region for the first 20 days of Period1.

year	day	CAP	EAL	OFS	VAL	MPU
1990	**1**	5	8	1	1	2
1990	**2**	5	8	1	1	8
1990	**3**	5	4	9	1	2
1990	**4**	5	4	4	1	2
1990	**5**	4	4	4	4	2
1990	**6**	4	1	6	8	2
1990	**7**	2	7	2	8	4
1990	**8**	1	2	1	8	2
1990	**9**	6	5	2	6	7
1990	**10**	8	5	1	3	5
1990	**11**	8	5	1	1	8
1990	**12**	5	5	9	1	8
1990	**13**	8	8	1	1	8
1990	**14**	2	6	1	1	1
1990	**15**	2	6	1	1	1
1990	**16**	2	6	1	1	1
1990	**17**	2	6	1	1	9
1990	**18**	2	6	1	1	3
1990	**19**	5	8	3	6	3
1990	**20**	8	8	5	3	7

In what follows, we show some selected fdf plots for full station results, pooling all CPs for a chosen gauge. In all these cases (it is not always so clear as shown here) the predicted shift and the observed shift between the two periods follow each other. This is strong affirmation that the downscaling preserves the message of the RCM. The way these gauges were selected was to take the first two, and the 23rd and 24th stations in each set in Period1. This gave us 20 stations from Period 1 with their associated P1 and P2 values. When it came to find matching ‘hidden’ observed data for period 2, only 12 of the 20 stations had survived, some with a fair amount of missing data. The comparisons plotted here are records with complete data – there had to be proper correspondence for meaningful comparative validation.

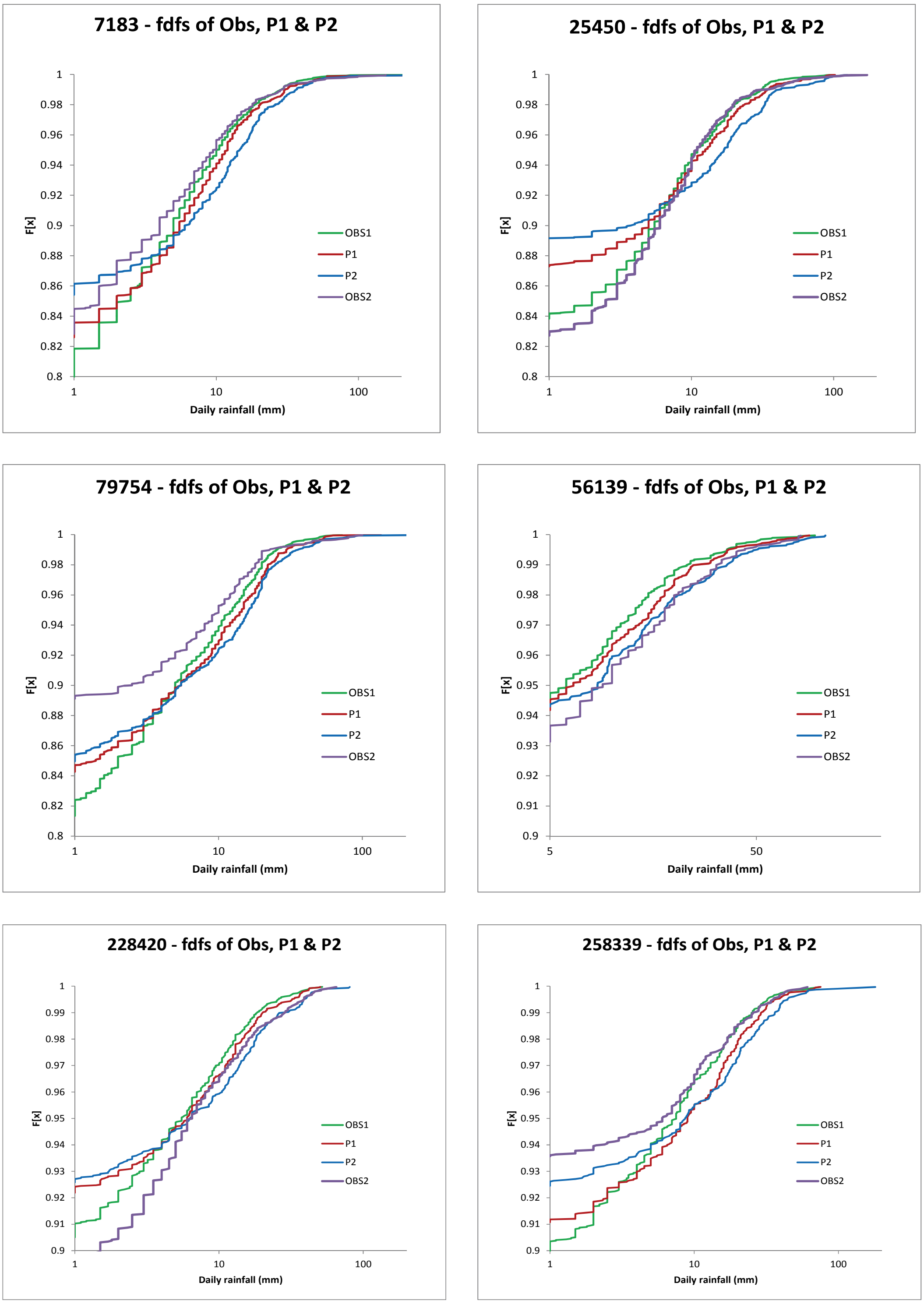


Figure 5.12 Three pairs of gauges; top pair Cape, middle pair East London, bottom pair Vaal. The observations in the first and second periods are given by the Green and Purple lines; the downscaled PRECIS values labelled P1 and P2 are coloured Red and Blue respectively.

Examining the images in Figure 5.12, there is crossing over of some of the lines from 10 mm/day and below. If we therefore confine comments to the wetter part of the image, the following observations can be made, based on the expectation that if the observations get wetter from one period to the next, then so should the downscaled PRECIS lines. This translates to the following: the shift from Green to Purple should be in the same direction as the shift from Red to Blue; this means that both observations and downscaled estimates move in the same direction. Numbering the panels from 1 to 6 for the images from top left to lower right and applying this rule, 1 fails, 2 is indeterminate, 3 fails, 4 is OK, 5 is OK, 6 is indeterminate. There is no conclusion to be made here as there is a tie. One feature that is evident from the plots is that the differentiation is not great. The range of probabilities is of the order of 0.01 to 0.02. A simple Kolmogorov-Smirnov test is that at the 95% level the criterion is $1.36/n^{1/2}$, which for our maximum length of record (of n = 3850) is equal to 0.022. Thus these fdfs cannot be separated statistically. It remains to see whether the CP-dependent plots give us better discrimination.

Here follows a selection of CP-dependent plots, limited to those CPs which had a reasonable proportion of the rainfall. Some were so dry that the number of wet days over the 3072 day periods counted only in the tens and twenties. This is as a result of the almost universal probability of dryness in the 80% range and above – this dryness is clearly evident from the six images in Figure 5.12.

In Figure 5.13, we show the results for the first of our listed stations for the Free State; number 228170. The discrimination on the full fdf in the top-left is small, and discriminating between them does not make sense because the outer range above 5 mm/day fails the KS test.

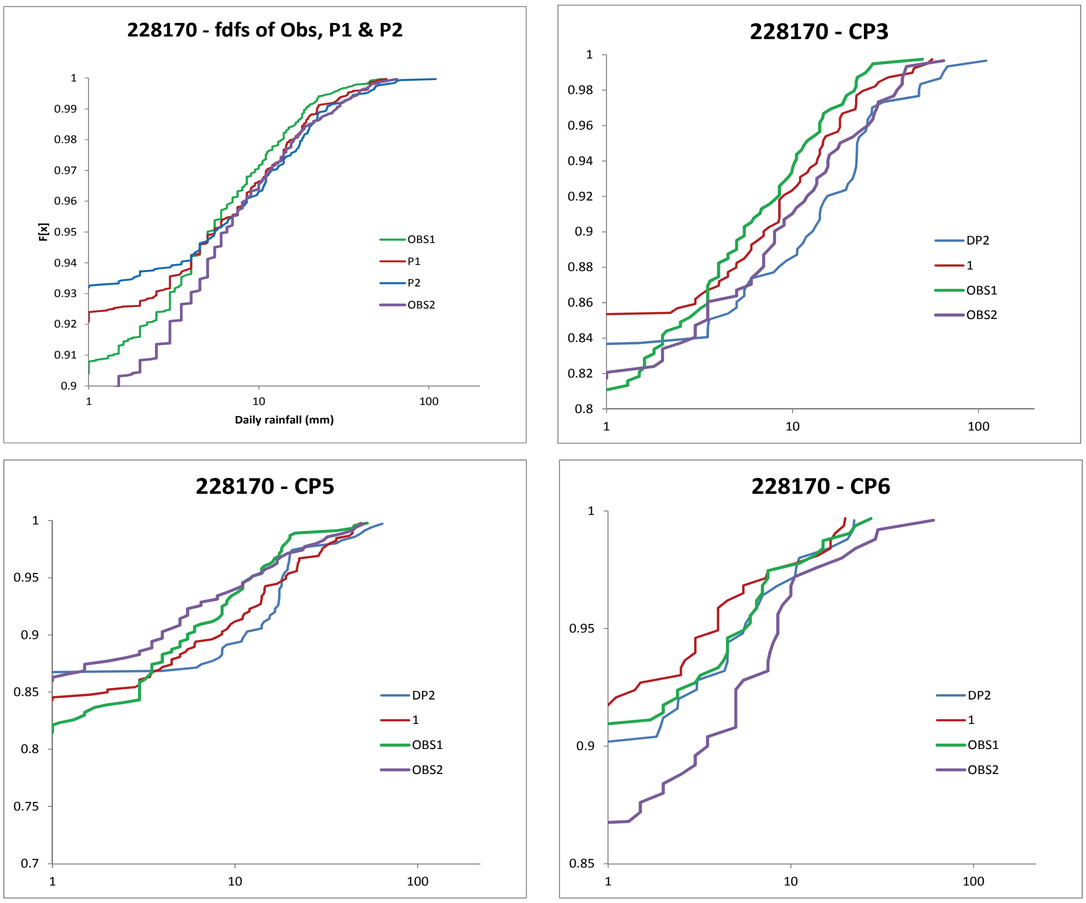


Figure 5.13 Four fdf comparisons for gauge 228170 in the Free State, from top left to bottom right: the complete set of distributions, CP3, CP5 and CP6 dependent relationships. The legends are all coloured the same: OBS1 and 2 are Green and Purple; P1 and P2 are Red and Blue.

However, to make a qualitative comment, P1 is wetter than OBS1 above 10 mm/day and P2 is not much different from OBS2 over most of the range. At the lower rainfall levels, the messages are consistent; the P[0] values (where the curves cut the axis at 1 mm/day) increase from green to red (OBS1 to P1), as does the purple to blue (OBS2 to P2). The legends for the CP-dependent fdfs are differently ordered from the full annual one, but the colour codes are the same. We note that the percentage of total rainfall in Period1 by CP3, 5 & 6 are respectively 33%, 29% and 6%, the remaining 6 CPs are responsible for the remaining 32 % of the rainfall. The 95% KS statistic for CPs 3, 5 & 6 are respectively 0.038, 0.041 and 0.089, wider in all cases than the pairs of plotted fdfs.

Turning to CP3 (top-right), both sets of downscaled values are wetter than the observed, if we only look at 5mm/day and above. In this image, we note that P2 (blue and labelled DP2) is generally wetter than P1 (red and labelled 1). What is pleasing is that OBS2 is generally wetter than OBS1, confirming the shift. In CP5 (bottom-left), they swop over in concert at about 12 mm/day, in the dryer CP6 (bottom-right), the order is recovered.

This consistency is interesting, but it is important that the shift in observations is in the same direction as the shift in the downscaled values. Again we use the rule: the shift from Green to Purple should be in the same direction as the shift from Red to Blue. In the top left panel, which shows the results for the full set (not CP-dependent) the Period2 fdfs are not as definite as the Period1 fdfs, but they tend to go in the same direction. The CP3 shows a match over nearly the full range. CP5 fails. CP6 is consistent.

To help reinforce the message more cleanly, the images in Figure 5.14 following show the capability of the method.

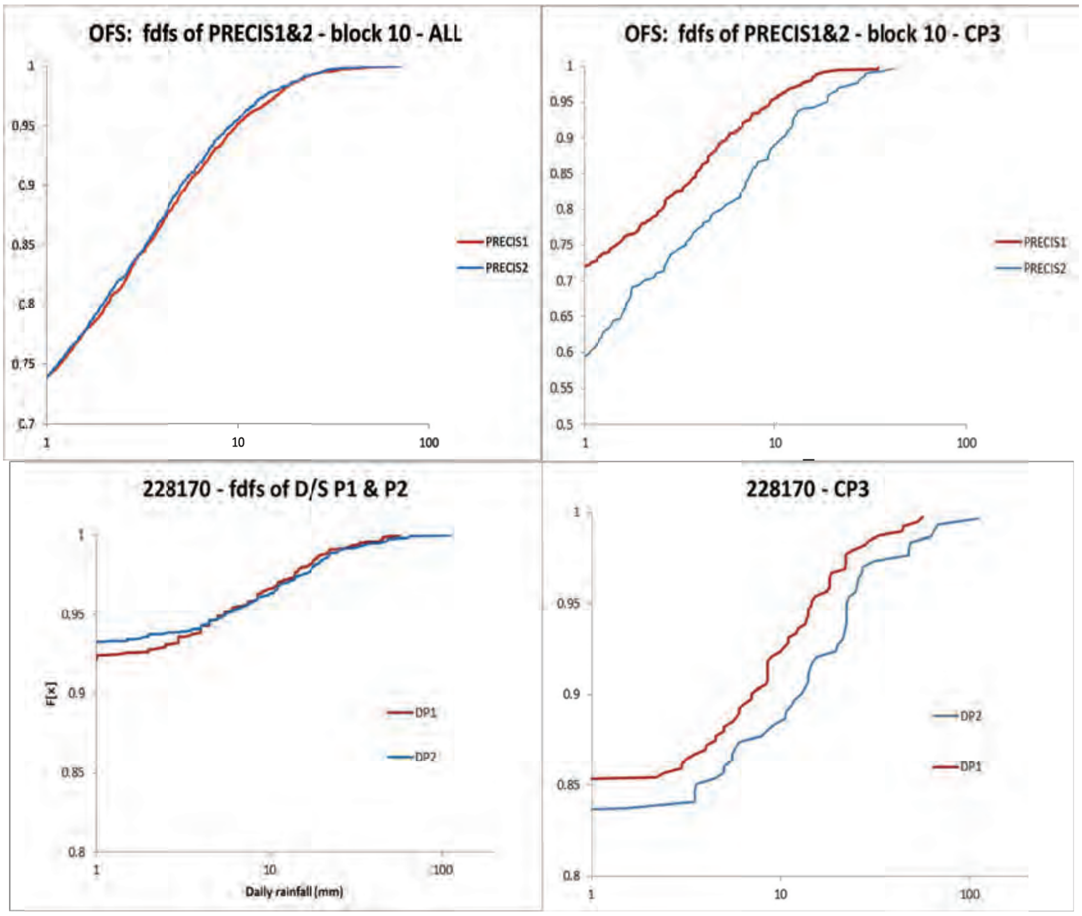


Figure 5.14 Frequency distribution functions of complete records (left) and CP3 based records (right). Top row raw PRECIS1 & 2 Block 10 (OFS). Bottom row a downscaled gauge in Block 10.

The top left panel in Figure 5.14 shows that in Block 10 in the Free State, the PRECI1 & 2 distributions are almost identical. This behaviour is mimicked in the bottom left panel of the figure, by the fdfs of the downscaled PRECIS values to gauge 228170, which lies within Block 10. This indicates that there is negligible difference in the overall rainfall distributions in the two periods in either the PRECIS or downscaled values. However, when we look at the two right hand panels which are conditioned on CP3, the PRECIS and gauge estimates are all in concert. This indicates that when this CP3 occurs, the PRECIS rainfall in the second period tends to be wetter and this is echoed by the downscaled gauge fdf. It is the power of the CPs to perform this discrimination that is the strength of the method. Note that the two panels in the bottom row of Figure 5.14 are the same as the red and blue lines in the top left and right panels of Figure 5.13 – they are the PRECIS fdfs and the downscaled block data to the gauge sites. What is exciting about this set of images is that the signal in the difference of the PRECIS rainfall estimates is preserved in the downscaled gauge values, even though their distributions, means and variances are different. That, in a nutshell, is the message of this project.

5.6 Summary

The contract governing this project contains the following sentence, which appears in the Motivation section of Annexure A:

> The rationale for this project is to explore ways in which systematic changes in the spatial patterns of rainfall at the daily or event scales can be (i) detected with greater certainty, (ii) linked to climate change, (iii) predicted more confidently and, (iv) translated more appropriately into changes in runoff patterns.

Our new modelling procedure satisfies these desirable goals and gives context to a short description of the work which bore this Chapter:

> Assessment of adequacy/inadequacy, and the limitations in interpretation and application, of current climate prediction tools in relation to rainfall modelling requirements

which can be expanded and interpreted as follows.

The "current climate prediction tools" are RCMs nested in GCMs. In particular there are three products of interest to this project which come out of PRECIS. They are the daily maps of (i) precipitation totals (ii) Sea Level Pressure (SLP) (iii) and as an alternative to SLP, 700 hPa geopotential heights, all mapped to the same 0.44° grid over the whole of Africa and surrounding Oceans.

In previous Chapters (1, 2 & 3) and demonstrated here, it was shown that the key to relating a specific gauge network's rainfall to RCM outputs is to determine what happens on days with similar Circulation Patterns (CPs) based on either SLPs or 700 hPa surfaces. Also, importantly, that the rainfall signature used to identify the appropriate CP on a given day is the state of wetness of the network. The gauge network wetness is a good signature for the rainfall intensity and the intercorrelation between the gauges when they experience precipitation. It is also a key to characterising the behaviour of the prevailing climate's sequencing and length of dry and wet periods.

These ideas feed into the specification of the phrase "rainfall modelling requirements" which appears in the description quoted above. From our point of view, the "modelling requirements" of the rainfall are embedded in their parameterization – the statistics defining the stochastic rainfall model. As detailed in Chapter 2, the PEGRAIN model is defined by these fundamental rainfall network parameters. Furthermore, PEGRAIN is adaptable to variations in dry spell length and consequent changes in rainfall regime signature, by the CPs produced in futures scenarios by the RCMs.

It was the task of this Chapter to assess *inter alia* “the adequacy/inadequacy, and the limitations in interpretation and application” of the RCMs in relation to adequately transferring climate information to the network scale. Besides the model parameters, we have now expanded this phrase to include the creation of sets of raingauge network sequences derived from an RCM.

The main question to be answered can be put as follows: How does the set of ensemble outputs from the downscaling model of choice, when driven by the RCM outputs, compare with measured rainfall data not used in its specification? This is a form of model validation. To obtain this result, one needs two periods of rainfall data. The first to calibrate the model on the RCM outputs, the second to check whether a second set of RCM outputs, driving the same model (with possible climate changes) produces similar rainfall to the recorded data of the second (blind) set. This has been done and is a qualified success as far as the amounts go, but a complete success from the point of view of capturing spatial correlation.

6 Conclusion

The rainfall downscaling method described here results from a series of three studies:

- one of which is already published (Bárdossy & Pegram, 2011)
- the second (Bárdossy & Pegram, 2012) has also been published
- the third is reported here, supported by a presentation at a Workshop at SANCIAHS 2012, and is in preparation for publication.

The first study in RCM downscaling (Bárdossy & Pegram, 2011) was over a 107 000 sq km area of the Rhine basin in Germany, where estimates from three RCMs were downscaled over the 25 km square blocks making up the region. Our product was daily average rainfall over the blocks which were interpolated from gauge values during the two periods: observation and verification. The first innovation and main feature of this first study was the conditioning of the downscaling procedure on Circulation Patterns (CPs) associated with various wetness indices of mesoscale networks of daily raingauges in the region. The second innovation in that work was the use of the double quantile-quantile (QQ) transformation of future block RCM rainfall frequencies to gauge averaged block values achieved in two steps: Future RCM to (i) Observed RCM to (ii) Observed gauge block values. The important and original feature of this methodology was to preserve the relative changes in behaviour of the rainfall modelled by the RCM between future and observed periods and then remove the model bias by shifting the distribution to the gauge average domain. This method was tested in the Verification period and found to be effective. What this innovation means is that the signal/message of the RCM is not destroyed even though the bias inherent in the model is removed. This is not simulation. It is a multi-step transformation of a model's information to the real world.

The main feature of the second study (Bárdossy & Pegram, 2012) was to develop a method to recorrelate the estimates obtained in the first study so that they would reflect proper cross correlation coefficients (cccs) between the daily block rainfall estimates. The need for this recorrelation emerged from the realisation that the cccs between the QQ'd RCM block estimates, produced in the first study, during the Observation period were lower than the corresponding cccs between the observed gauge-interpolated 25 km block values. The lower cccs between the downscaled block values (before recorrelation) meant that heavy extreme events over large aggregated areas would be fragmented whereas, in nature at the mesoscale, higher cccs produce large areas of high rainfall with low variability between blocks.

Hydrologically, the effect of this limitation of the method would be that floods estimated using RCM outputs over large areas (even QQ'd correctly locally) would be dangerously underestimated. We had to find a way of readjusting the QQ'd RCM estimates of the first study, to reflect the correct spatial interdependence, without destroying the distribution of the amounts of rainfall averaged over the blocks at both original and larger scales. The methods of recorrelation that we devised achieved this and more importantly preserved the relative changes in cccs between the future and observed periods. Again, we were able to preserve the signal of the RCMs while removing the bias, but this time not only in the marginal distributions produced in the first study, but also in the spatial dependence, albeit at the block scale.

In the third study (first reported in this project report) we take the results of stages one and two and, going to point scale instead of large scale, downscale via recorrelation and QQ transforms the RCM estimates to networks of individual raingauges. The study area was applied to 5 regions in South Africa, varying in size from 10 000 to 30 000 sq km. This downscaling to a pattern of points is achieved while still preserving all the block-scale and large-scale marginal distribution statistics and dependence structures we had previously devised. The downscaled gauge site estimates of the daily rainfall based on the RCM are shown to have the same statistics as those of the Observation period. More excitingly, and quite fortuitously, we found that in many cases the methodology also captured the behaviour of the individual gauges during a Validation period, which means that we indeed have a tool that faithfully replicates the gauge-scale characteristics of rainfall measuring networks, not only now but in the modelled future. This observation will hold so long as the RCMs capture the climate correctly.

This validation came about by default. During the first part of the work reported in this Chapter, started in October 2011, the only available daily rainfall data had been obtained from the Lynch data-base with end-dates in mid-2000. Nevertheless, we had been able to obtain from CSAG in UCT a complete set of PRECIS RCM data comprising Seal Level Pressures and Rainfall modelled on 0.44° blocks over the whole of Africa and surrounding Oceans from 1990 to 2008. Therefore all downscaling work started during the contemporaneous period: 1990 to 2000. Only as late as November 2011, after gaining permission from SAWS, were we gratefully able to obtain an augmented rainfall data-base from CSAG at UCT whose end-date is early 2010. It is this augmented set of gauge rainfall data which has been used to validate the downscaling procedure – it is as though we fortuitously had a Time machine to glimpse a view of the future in some of its detail!

The result is very exciting. The downscaled recorrelated estimates of rainfall at the gauge sites during the first period match the observations acceptably well, indicating that the technique has been verified – in other words it works as it should when we have the gauge data. More importantly, the PRECIS estimates of Block rainfall during the second period, when "differenced" from the PRECIS estimates in the first period, appear to be different in frequency of occurrence and marginal distribution but not appreciably in spatial correlation. These characteristics of the RCM rainfall (more likely due to inter-decadal variation than genuine climate change - the sequences are too short to tell), were preserved in the downscaling procedure, successfully estimating rainfall at the gauge sites during the second period. These readjusted, recorrelated, downscaled raingauge estimates in the second period, when compared with the "hidden" raingauge records, show a remarkably good match – in other words, the method has been validated.

Apart from this technical success, there are several implications which are even more beneficial. The message that this unique validation procedure sends is:

- to the climate modellers: you are providing reasonable rainfall estimates, through capturing the changes in the rainfall and therefore of the climate. What your models produce is not phantasy nor useless to the hydrometeorologist, but their outputs need transformation and interpretation, and hopefully continued improvement. If your models faithfully reproduce the *relative change* in weather and climate, we here have a method of unpacking the signal and making it valuable in other communities.
- to the hydrometeorological modellers: we have devised a method which can capture the subtle changes in spatial structure and amounts of rainfall modelled by the RCM PRECIS, even if the modelled values and cross correlations are very different from those of observed rainfall. This can be done using QQ transforms and matrix recorrelation techniques routinely and should be readily transferable to other RCMs.
- to the practitioner: we think we have a winner for modelling daily rainfall over observation networks, not only in the present, but also in the modelled future. We will need to make more experiments to be confident in our prognostications. Something else comes out of our methodology – we can routinely produce ensembles of rainfall from this procedure, not the single downscaled offering made by linear scaling of the RCM outputs. This means that the method is viable for assisting in risk analysis in plausible futures.

Despite the excitement, there is a limitation to this method of downscaling RCM block values to gauge sites: the blocks must contain gauge records which are contemporaneous with the sites: [No gauges] = [No downscaling]. But be aware that our raingauge records are dying off. In the validation period after mid-2000, we found that out of the 279 gauges which existed in the first period, only 180 survived until 2010 – a tragic, irreparable loss of 36% in ten years.

There is a possible solution to this problem of diminishing raingauges, so long as they don't all become extinct. That solution is to produce realistic interpolated gauge values, following on from the excellent work reported in WRC Report No 305/1/94 by McNeill, Brandão, Zucchini and Joubert, but improving their method by preserving the local spatial cross correlations which were absent in that study. This fourth step in our current research (conducted by Bárdossy and Pegram), which is running in parallel to this study, is already far advanced and will soon be published. It is going to be used in a new WRC directed/solicited project starting April 2013: Revision of the Mean Annual Precipitation (MAP) estimates over Southern Africa.

Finally, there are three major original research highlights that come out of the project:

- The use of historical, mesoscale region specific, atmospheric Circulation Patterns (CPs) on which to condition the downscaling to gauge sites of rainfall estimates provided by RCMs
- The introduction of a recorrelation technique to ensure that the spatial structure of the downscaled rainfall is maintained; this is very important when estimating flows out of large catchments, especially for flood analysis
- The iterative improvement of the recorrelation method to ensure that the dependence structure is not biased by the relatively large number of dry days we experience all over the country; we have between 80% and 95% dry days over the bulk of the country. Without this innovation, the desired level of interdependence would not have been achieved.

REFERENCES

Bárdossy A, GGS Pegram and L Samaniego, (2005). Modeling data relationships with a local variance reducing technique: Applications in hydrology, Water Resour. Res., 41, W08404, doi:10.1029/2004WR003851

Bárdossy A and GGS Pegram (2009). Copula based multisite model for daily precipitation simulation. Hydrol. Earth Syst. Sci., 13, 2299–2314

Bárdossy, A. (2010), Atmospheric circulation pattern classification for south-west Germany using hydrological variables., *Physics and Chemistry of the Earth*, *12*, doi: 10.1016/j.pce.2010.02.007.

Bárdossy, A., and G. Pegram (2011), Downscaling precipitation using regional climate models and circulation patterns toward hydrology., *Water Resources Research*, *47(W04505)*, doi:10.1029/2010WR009,689.

Bárdossy, A., and G. Pegram (2012), Multiscale spatial recorrelation of RCM precipitation to produce unbiased climate change scenarios over large areas and small, *Water Resources Research*, 48, W09502, doi:10.1029/2011WR011524 [Publication Date:1 September 2012]

Dyer, T.G.J., and P.D. Tyson (1977), Estimating above and below rainfall periods over South Africa, 1972-2000., *Journal of Applied Meteorology*, pp 145-7.

Feller W (1968) *An Introduction to Probability Theory and its Applications: Volume 1*. (3rd Ed.) Wiley.

Gabriel K R (1959) The Distribution of the Number of Successes in a Sequence of Dependent Trials: *Biometrika*, Vol 46, No. 3/4, pages 454-460.

Gabriel KR and J Neumann (1962) A Markov chain model for daily rainfall occurrence at Tel Aviv. *Quarterly Journal of the Royal Meteorological Society*, Vol 88, pages 90-95.

Hewitson, B., Tadross, M. and Jack, C. (2005a). Climate Change Scenarios: Conceptual Foundations, Large Scale Forcing, Uncertainty and the Climate Context. *In*: Schulze, R.E. (Ed) *Climate Change and Water Resources in Southern Africa: Studies on Scenarios, Impacts, Vulnerabilities and Adaptation*. Water Research Commission, Pretoria, RSA, WRC Report 1430/1/05. Chapter 2, 21 – 38

Hewitson, B., Tadross, M. and Jack, C. (2005b). Scenarios from the University of Cape Town. *In*: Schulze, R.E. (Ed) *Climate Change and Water Resources in Southern Africa: Studies on Scenarios, Impacts, Vulnerabilities and Adaptation*. Water Research Commission, Pretoria, RSA, WRC Report 1430/1/05. Chapter 3, 39 - 56.

Johns, T.C., C.F. Durman, H.T. Banks, M.J. Roberts, A.J. McLaren, J.K. Ridley, C.A. Senior, K.D. Williams, A. Jones, G.J. Rickard, S. Cusack, W.J. Ingram, M. Crucifix, D.M.H. Sexton, M.M. Joshi, B.-W. Dong, H. Spencer, R.S.R. Hill, J.M. Gregory, A.B. Keen, A.K. Pardaens, J.A. Lowe, A. Bodas-Salcedo, S. Stark, and Y. Searl, (2006). The New Hadley Centre Climate Model (HadGEM1): Evaluation of Coupled Simulations. *J. Climate*, **19**, 1327-1353.

Katz RW (1983) Statistical Procedures for making inference about precipitation changes simulated by an atmospheric general circulation model. *Journal of the Atmospheric Sciences,* Vol 40, pages 2193–2201.

Katz, RW and MB Parlange (1998) Overdispersion phenomenon in stochastic modeling of precipitation. *Journal of Climate,* Vol 11, pages 591–601.

Lamb HH (1972) British Isles weather types and a register of daily sequence of circulation patterns, 1861-1971, *Geophys. Mem.*, no. 110, 85 Meteor. Office, London.

Handbook of Hydrology (1993) David Maidment (Editor in Chief), McGraw-Hill

Markov AA (1924) *Probability Theory* (4th ed.). Moscow (in Russian).

Matalas NC (1967) Mathematical assessment of synthetic hydrology. *Water Resources Research*, Vol 3 (4), pages 937–945.

McMahon, TA, AS Kiem, MC Peel, PW Jordan & GGS Pegram, (2008). A New Approach to Stochastically Generating Six-Monthly Rainfall Sequences Based on Empirical Mode Decomposition, *J Hydrometeorology*, DOI: 10.1175/2008JHM991.1, December, Vol. 9, 1377-89.

McMahon, Thomas A., Murray C. Peel, Geoffrey G.S. Pegram and Ian N. Smith, (2011), A simple methodology for estimating mean and variability of annual runoff and reservoir yield under present and future climates. *Journal of Hydrometeorology,* Vol. **12**, pp 135–146
doi: http://dx.doi.org/10.1175/2010JHM1288.1

Mehrotra, R, Srikanthan, R and A Sharma (2005) Comparison of three approaches for stochastic simulation of multi-site precipitation occurrence. *29th Hydrology and Water Resources Symposium,* Canberra.

Peel MC, McMahon TA and Pegram GGS (2005).Global Analysis of Runs of Annual Precipitation and Runoff Equal to or Below the Median: Run Magnitude and Severity. *International Journal of Climatology* vol **25**: pp 549–568

Peel MC, McMahon TA & Pegram GGS, (2009). Assessing the performance of rational spline-based Empirical Mode Decomposition using a global annual precipitation dataset. *Proceedings of the Royal Society London Series A*. 465: 1919-1937.

Peel MC, Pegram GGS and McMahon TA, (2004). Global analysis of runs of annual precipitation and runoff equal to or below the median: Run Length. *International Journal of Climatology*, Vol 24: pp. 807–822.

Pegram, GGS (1980) An Auto-regressive Model for Multi-lag Markov Chains. *Journal of Applied Probability*, Vol.17, pages.350-362.

Pegram GGS (1996) A hidden covariance model for spatial binary processes. Invited seminar at *National Technical University of Athens*, Greece, hosted by Demetris Koutsoyiannis and Efi Foufoula-Georgiu.

Pegram GGS (2010) PEGRAIN - a complete daily rainfall network simulator. Report by Pegram & Associates (Pty) Ltd to the *RSA Department of Water Affairs and Forestry*, January, 18 pages.

Pegram GGS and MP Mittermaier (1998). Catchment Management. *Workshop On The Utilization Of Weather Radar In Southern Africa*, Water Research Commission and SAWB, Centurion Park, November

Pegram GGS and Terblanche DE (1998). Radar Hydrology developments in South Africa. *Proceedings of an Advanced Study Course: Radar Hydrology for Real time Flood Forecasting,* sponsored by Wat Man

Res Centre, Univ of Bristol & European Commission Env and Climate Prog, Bristol, June, published as EUR 19888, ISBN 92-894-1640-8, pp. 245-262

Pegram GGS, Mackenzie RS, van Biljon S and Terblanche D, (1997). Towards a flood forecasting system for the Vaal Dam. *8th South African National Hydrology Symposium,* Pretoria, November

Pegram, G.C. and G.G.S. Pegram, (1993). Estimating Areal Rainfall by Integration of Multi-Quadric Surfaces over Polygons. Journal of Hydraulic Engineering, American Soc. Civ. Eng. Vol.119, pp.151-163

Pegram, GGS, MC Peel & TA McMahon, (2008). Empirical Mode Decomposition using rational splines: an application to rainfall time series. *Proceedings of the Royal Society A.* Vol. 464, 1483–1501 doi:10.1098/rspa.2007.0311

Richardson, C W (1981) Stochastic simulation of daily precipitation, temperature, and solar radiation. *Water Resources Research*, Vol 17, pages 182–190.

Rosenblueth, E. (1975) Point estimates for probability moments. *Proceedings of the National Academy of Science* U.S.A., 72: 3812-3814

Rouault, M., N. Fauchereau, B. Pohl, P. Penven, Y. Richard, C.J.R. Reason, G.G.S. Pegram, N. Phillippon, G. Siedler and A. Murgia, (2010), Multidisciplinary Analysis of Hydroclimatic Variability at the Catchment Scale, *WRC Report No*. 1747/1/10, ISBN 978-1-77005-955-9

Schulze R.E. (Ed) (2005). Climate Change and Water Resources in Southern Africa; Studies on Scenarios, Impacts, Vulnerabilities and Adaptation. WRC Report No : 1430/1/05, ISBN No : 1-77005-365-4

Srikanthan, R (2004) Stochastic generation of daily rainfall data using a nested model. In: *57th Canadian Water Resources Association Annual Congress*, 16–18 June 2004, Montreal, Canada.

Srikanthan R (2005) Stochastic Generation of Daily Rainfall Data at a Number of Sites. Technical Report 05/7, *CRC for Catchment Hydrology*. Monash University, 66 pages.

Srikanthan R (2009) A multisite daily rainfall data generation model for climate change conditions. *18th World IMACS / MODSIM Congress*, Cairns, Australia 13-17 July 2009.

Srikanthan, R and GGS Pegram (2006) Stochastic generation of multi-site rainfall occurrences, Advances in Geosciences, Vol 6: *Hydrological Science*, World Scientific Publishing Company, p 1-10.

Srikanthan, R and GGS Pegram (2009) A Nested Multisite Daily Rainfall Stochastic Generation Model, Journal of Hydrology, Vol 371 pp 142-153. DOI: 10.1016/j.jhydrol.2009.03.025

Srikanthan, R and TA McMahon (2001) Stochastic generation of annual, monthly and daily climate data: A review. *Hydrology and Earth System Sciences*, Vol 5(4), pages 653–670.

Terblanche, D.E., G.G.S. Pegram and M.P. Mittermaier, (2001). The development of weather radar as a research and operational tool for hydrology in South Africa, *Journal of Hydrology* (241)1-2, pp. 3-25

Teutschbein, Claudia and Jan Seibert (2012). Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods, *Journal of Hydrology*, Vols 456 & 457 pp12–29

Wesson SM and GGS Pegram, (2004). Radar Rainfall Image Repair Techniques. *Hydrological and Earth Systems Sciences*, European Geophysical Society, 8(2), 220-234

Wilks, DS (1992) Adapting stochastic weather generation algorithms for climate change studies. *Climatic Change,* Vol 22, pages 67–84.

Wilks, DS (1998) Multisite generalisation of a daily stochastic precipitation generation model. *Journal of Hydrology* Vol 210, pages 178–191.

Woolhiser DA and GGS Pegram (1979) Maximum Likelihood Estimation of Fourier Coefficients to Describe Seasonal Variation of Parameters in Stochastic Daily Precipitation Models. *Journal of Applied Meteorology*, Vol 5, No 1, pages 34-42.

Zucchini W and PT Adamson (1984) The occurrence and severity of droughts in South Africa, Report WRC 91/1/84 to the *Water Research Commission*, Pretoria,

Appendix

Expansion of methodology described in Pegram and Pegram, (1993)

The gauge weights are all calculated by integrated Multiquadrics, restricted to those gauges which lie in each box, because inclusion of sites outside the box induces negative weights, which we want to avoid. Also, it seems sensible to restrict the box averages to the gauges inside the box; otherwise one is importing information from distance and inducing false correlations between block averages. The weights sum to 1.0. The method is fully described (with mathematical development) by Pegram and Pegram (1993); the method is essentially Simple Kriging with a hyperbolic variogram, but with the important addition of analytically computed weighting functions and the restriction to local interpolation. A mathematical description to give the core ideas follows.

In Simple Kriging with a hyperbolic variogram (in other words, in Multiquadrics) the standard equations are:

$$z = F\,c$$

$$z_0 = f_0^T c$$

where z is the n-vector of known values of the variable at control points with coordinate vector x, each element f_{ij} is the variogram value at range $|x_i - x_j|$, where x_i is the position vector of control point i = 1, 2, …, n. The vector c is unknown and is found as:

$$c = F^{-1} z$$

To determine the volume under the surface defined by the range of z_0 over the region of interest, with a depth z_i at each location x_i, we proceed analytically:

$$\mathrm{Vol} = \int_R z_0 dx_0 = \int_R \{ f_0^T c \} dx_0$$

which, because of linearity, can be written as the scalar product of two vectors:

$$v^T c = [\int_R f_{01}\, dx_0 \quad \int_R f_{02}\, dx_0 \quad \dots \quad \int_R f_{0n}\, dx_0]\ c$$

where $f_{0i} = f(|x_i - x_0|)$ and x_0 ranges over the region.

Each of the individual integrals over the region of interest, centred on each control point, is obtained as follows. The boundary of the region is approximated by a closed polygon with m edges and nodes. The integration is done piecewise over triangles with one apex at the site x_i of the control station, the other two at the contiguous nodes x_k and x_{k+1} of the k^{th} edge: k = 1, 2, …, m, where it is understood that m+1 = 0.

In Figure A1, the locations of the vertices are given as x_i, x_k and x_{k+1}; the length of the perpendicular to the line produced along the k^{th} edge is p_k, meeting the line at x_k^*, from which the distances to nodes x_k and x_{k+1} are s_{k1} and s_{k2} respectively. These distances may all be calculated from simple coordinate geometry.

If we choose to set the nugget of the variogram to zero, the f_{ij} becomes the distance between x_i and x_j. This choice yields the simplest formula [the full set appears in Pegram and Pegram (1993)] for the elemental volume below the inverted cone covering the right-angled triangle [x_i, x_k, x_k^*]:

$$I_{k1} = p_k^3 \sinh^{-1}[s_{k1}/p_k]/6 + p_k\, s_{k1}\, f_{ik}/6$$

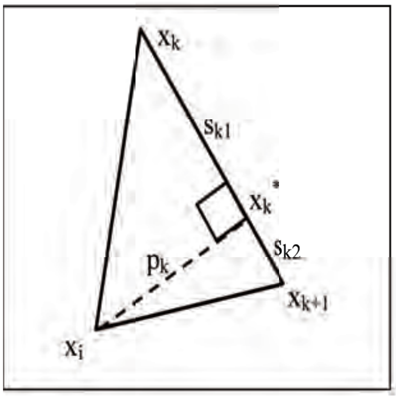


Figure A1. Geometry of distances in triangle formed from a control point and a polygon edge.

where in this special case, f_{ik} is the distance $|x_i - x_k|$, an element of the matrix F. I_{k2} is the volume over the other right-angled triangle with vertex at x_{k+1}. Their signs depend on the angle swept out by the direction of travel from x_k^* (which needs care) and the sum of I_{k1} and I_{k2} yields the volume v_{ki} associated with the k^{th} edge and with x_i as vertex. Each of these v_{ki} can be summed over k = 1, 2, … m edges to give the volume over each control point v_i

The volumes v_i can be arranged in an n-vector v^T. Then the complete volume under the closed z_0 surface over the closed polygon, when the depth at each site x_i is z_i, is:

$$\begin{aligned} \text{Vol} &= v^T c \\ &= v^T F^{-1} z \\ &= w^T z \end{aligned}$$

where now w is a constant n-vector of weights associated with each control site and depends only on the geometry of the region and the controls. It is the n-vector z which varies between time-steps.

The mean depth over the region of area A experiencing rainfall z_t at the gauge sites on day {t} is simply given as:

$$\text{Mean depth} = w^T z_t / A$$