PROJECT 2.3: STATISTICAL DOWNSCALING FOR THE NORTH-WEST

Principal Investigator

Steve Charles, CSIRO Land and Water, Private Bag No. 5, Wembley WA 6913
Phone: 08 9333 6795; Email: Steve.Charles@csiro.au

Objectives

- To use statistical downscaling to link circulation drivers to the behaviour of historical rainfall and temperature at spatial and temporal scales suitable for management and planning.
- To compare observed weather state time series with those obtained from downscaling forced GCMs to assess the ability to reproduce the observed changes and trends.
- To assess the abilities of global and regional climate models to reproduce the atmospheric predictors required for statistical downscaling. (The climate models that perform adequately will be used to drive the downscaling model.)
- To produce high-resolution climate change scenarios that can be used for impact and vulnerability assessments by either State agencies alone or in collaboration with IOCI's research providers.

Key Research Findings

- A preliminary network consisting of 57 daily rainfall gauges has been selected. These gauges have substantive records for the period from 1958 to 2008. Of these 57 gauges, 19 are in the Kimberley and 38 in the Pilbara/Gascoyne regions. Records for 10 gauges appear to contain untagged accumulations, particularly over the last 10 years. The percentage of missing data in-filled by interpolation varies considerably between stations and periods. This network will be refined through further consultation with the Bureau of Meteorology.

- Initial fitting of a statistical downscaling model for the summer half-year (November to April) has been undertaken to determine the dominant spatial patterns of rainfall occurrence (states), their seasonality, and trends. The trends found are consistent with the observed rainfall increase in the North-West.
- Work has commenced on the selection of atmospheric predictors for the downscaling scheme.

MILESTONE 2.3.1: REPORT ON SELECTION OF HIGH-QUALITY DAILY RAINFALL GAUGE NETWORKS

Background

Statistical downscaling requires quality-controlled atmospheric fields, as predictors of regional rainfall pattern (‘weather state’) sequencing, that are only available in sufficient quality from 1958 onwards in reanalysis data sets such as that produced by the National Centers for Environmental Prediction/National Center for Atmospheric Research (NNR). Hence the selection of daily rainfall gauge networks was undertaken for the 1958-2008 period. Assessment of all available Bureau of Meteorology (BoM) stations recording daily rainfall in the IOCI 3 North-west study area (Kimberley, Pilbara and parts of the Gascoyne) during this period led to the initial selection of 57 stations with substantive records for this period, separated into two distinct sub-regions: 19 stations in the Kimberley region and 38 stations in the Pilbara/Gascoyne region (Figure 2.3.1; Table 2.3.1).

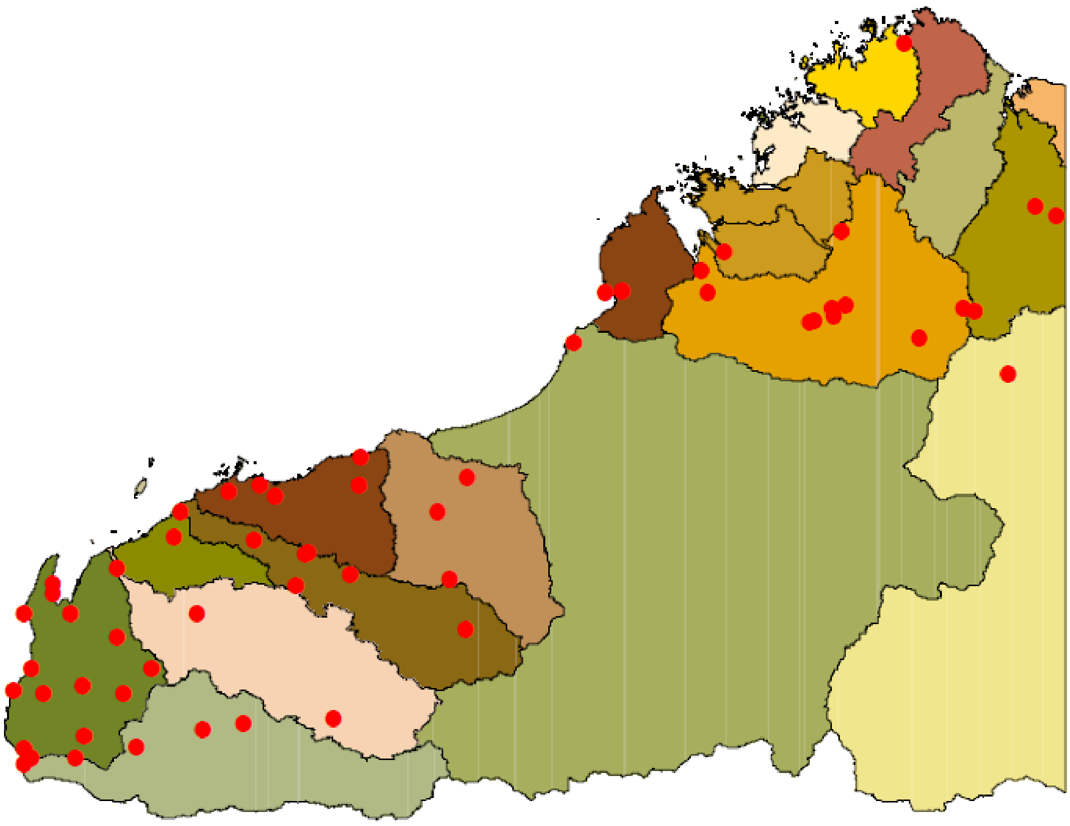


Figure 2.3.1 Location of 19 Kimberley Region and 38 Pilbara Region daily rainfall stations (see Table 2.3.1).

Table 2.3.1 Station details.

No.	BoM number	Station name	Longitude	Latitude	Region	Record quality*	Closed
1	1021	KALUMBURU MISSION	126.64	-14.30	1		2005
2	2012	HALLS CREEK A/P	127.66	-18.23	1	Y	
3	2016	LISSADELL	128.55	-16.67	1		2007
4	2019	MARGARET RIVER	126.86	-18.62	1	Y	
5	2020	MOOLA BULLA	127.50	-18.19	1	Y	
6	2028	SPRING CREEK	128.87	-16.82	1		2007
7	2029	STURT CREEK	128.16	-19.16	1	Y	
8	3003	BROOME AIRPORT	122.23	-17.95	1	Y	
9	3006	FITZROY CROSSING	125.56	-18.19	1		2000
10	3009	JUBILEE DOWNS	125.30	-18.35	1	Y	
11	3014	GOGO STATION	125.59	-18.29	1		2007
12	3015	MEDA	123.99	-17.37	1		2000
13	3017	MOUNT HOUSE STATN	125.70	-17.05	1		2007
14	3022	QUANBUN DOWNS	125.23	-18.38	1	Y	
15	3023	ROEBUCK PLAINS	122.47	-17.93	1	Y	
16	3024	UDIALLA	123.74	-17.95	1		2006
17	3026	YEEDA	123.65	-17.62	1		
18	3027	FOSSIL DOWNS	125.78	-18.14	1	Y	
19	3030	BIDYADANGA	121.78	-18.68	1	Y	
20	4006	BONNEY DOWNS	119.93	-22.18	2	Y	
21	4016	INDEE	118.60	-20.79	2		
22	4020	MARBLE BAR COMP	119.75	-21.18	2		2006
23	4032	PORT HEDLAND A/P	118.63	-20.37	2	Y	
24	4035	ROEBOURNE	117.15	-20.78	2	Y	
25	4039	WARAMBIE	117.37	-20.95	2	Y	
26	4046	YARRIE	120.20	-20.67	2		
27	5001	COOLAWANYAH	117.81	-21.80	2	Y	
28	5003	ETHEL CREEK	120.17	-22.90	2		2003
29	5004	EXMOUTH GULF	114.11	-22.38	2	Y	
30	5005	HAMERSLEY	117.68	-22.28	2		
31	5007	LEARMONTH AIRPORT	114.10	-22.24	2	Y	
32	5008	MARDIE	115.98	-21.19	2	Y	
33	5012	MILLSTREAM	117.07	-21.59	2	Y	
34	5013	MINDEROO	115.05	-22.00	2	Y	2008
35	5014	MOUNT FLORANCE	117.86	-21.79	2		
36	5015	MULGA DOWNS	118.47	-22.10	2	Y	
37	5020	NINGALOO	113.67	-22.70	2	Y	
38	5028	WYLOO	116.23	-22.69	2	Y	2008
39	5032	YARRALOOLA	115.88	-21.57	2	Y	
40	5045	GIRALIA	114.37	-22.68	2		2006
41	5052	KARRATHA STATION	116.68	-20.88	2		

42	6003	BOOLATHANA	113.69	-24.65	2		2005
43	6005	BRICKHOUSE	113.79	-24.82	2	Y	2008
44	6011	CARNARVON AIRPORT	113.67	-24.89	2	Y	
45	6019	DOORAWARRAH	114.43	-24.81	2	Y	
46	6023	GNARALOO	113.52	-23.82	2		2008
47	6030	LYONS RIVER	115.34	-24.63	2	Y	
48	6032	MARDATHUNA	114.56	-24.47	2		2004
49	6033	MAROONAH	115.55	-23.48	2		2006
50	6039	MINILYA	113.97	-23.85	2		2008
51	6050	WANDAGEE	114.55	-23.76	2		2007
52	6052	WILLIAMBURY	115.15	-23.86	2		2008
53	6072	NYANG STATION	115.04	-23.03	2	Y	
54	6084	WARROORA	113.79	-23.48	2		
55	7053	MOUNT AUGUSTUS	116.91	-24.31	2		
56	7058	MOUNT PHILLIP	116.31	-24.40	2		
57	7059	MOUNT VERNON	118.24	-24.23	2	Y	

*No untagged accumulations or extended periods of missing data

Technical Details

Daily records for these 57 stations for 1958 to 2008 were extracted from the Queensland Government and BoM's Patched Point Dataset (PPD, http://www.longpaddock.qld.gov.au/silo/), where missing data are in-filled using interpolation and flagged multi-day recordings are 'deaccumulated'. Investigations into the data quality of the daily records assessed the quantity and periods of in-filled records as well as periods of suspected untagged accumulations. Untagged accumulations are multi-day recordings that are erroneously not flagged as such, and so are unable to be deaccumulated.

Results (not shown) indicate that about 10 of the 57 stations (3 in the Kimberley and 7 in the Pilbara) contain periods suspected of containing untagged accumulations, predominantly in the last decade of the record. The percentage of missing data in-filled by interpolation varies considerably between stations and periods. We aim to use as many stations as possible in the statistical downscaling model (SDM) fitting, to adequately and uniformly cover the spatial extent of the regions. Thus an initial fitting period of 1980 to 1999 has been selected, when untagged accumulations and in-filled data are low (Table 2.3.2). The proportion of in-filled data significantly increases in the most recent decade, particularly given that 8 of the 11 Kimberley stations and 13 of the 38 Pilbara stations close post-2000. Thus this later poor quality period was excluded from the SDM fitting. Ideally we would prefer not to include closed stations; however spatial coverage would be degraded if they were left out.

Table 2.3.2 Station-average in-filled/deaccumulated days per decade

Decade	Kimberley (19 stations)	Pilbara (38 stations)
1958 to 1967	249	112
1968 to 1977	110	145
1978 to 1987	49	164
1988 to 1997	104	45
1998 to 2007	921	748

MILESTONE 2.3.2: REPORT ON DEVELOPMENT AND TESTING OF DOWNSCALING MODELS

Background

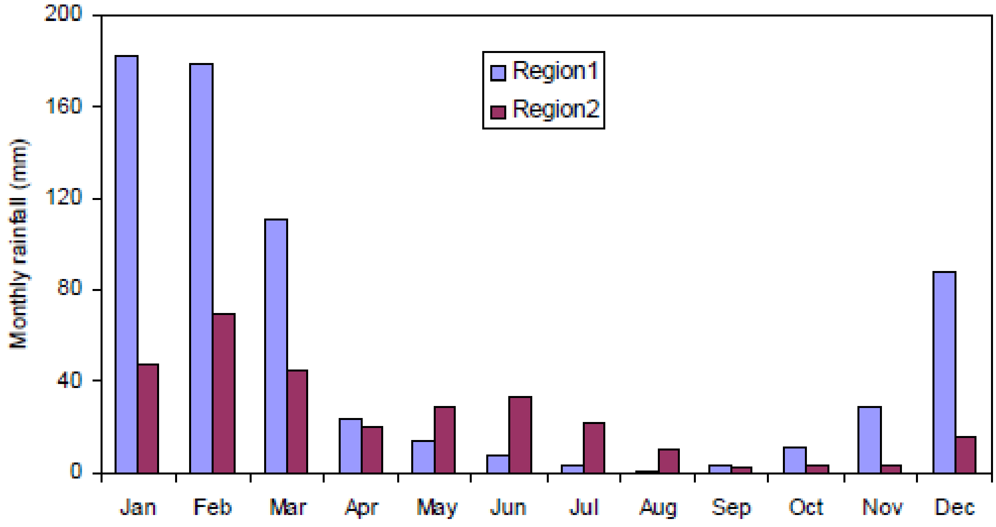
The annual cycle in monthly rainfall for the Kimberley and Pilbara regions is shown for monthly totals (Figure 2.3.2a), and as percentages of annual total rainfall for each region (Figure 2.3.2b) and each station (Figure 2.3.3). Rainfall in the Kimberley region is dominated by the strong monsoon signal, 'the wet', during December to March. Over 50% of annual rainfall occurs in just two months, January and February (Figure 2.3.2b). Thus, initially, a November to April half-year has been selected for SDM fitting for the Kimberley region. For the majority of selected stations, the annual cycle of monthly rainfall for the Pilbara region exhibits a distinct double peak (Figure 2.3.2 and 2.3.3b). The summer peak, centred on February, is a feature of the monsoon/cyclone season. A small number of stations do not show the summer peak, i.e. their rainfall is winter dominated. These stations are concentrated in the extreme southwest of the study region. The second winter peak is centred on June. Interactions between the subtropics and midlatitude westerlies contribute to this season's rainfall. As different meteorological processes can produce rainfall in these two seasons, separate SDMs are being fitted for each season.

Technical Details

The SDM used is the nonhomogeneous hidden Markov model (NHMM), as used previously in IOCI and widely across Australia and internationally. The NHMM defines a small number of distinct, discrete 'weather states' that represent the dominant rainfall occurrence patterns of the station network. For a fitted NHMM, the daily sequence of states is determined by the Viterbi algorithm which assigns each day to its most likely state based on the spatial pattern of rainfall occurrence across the station network. Simulations of daily weather state sequences, and rainfall, are also NHMM outputs. A detailed description can be found in previous reports for IOCI Stages 1 and 2.

Initial fitting of the NHMM has been undertaken to determine the dominant rainfall patterns (states), their seasonality, and trends. An example of preliminary results for the 6-state NHMM fitted to the 19 Kimberley stations for the November-April half-year are shown in Figures 2.3.4 to 2.3.6. Two SDM calibrations were undertaken, one that includes all the deaccumulated and in-filled data, and one that marks such suspect data as missing. An important component of this initial research is the assessment of the impact of station data quality on (1) the weather state patterns and frequencies for the 1980 to 1999 fitting period, and (2) the resultant weather state time-series, using the full period from 1958 to 2008.

(a)



(b)

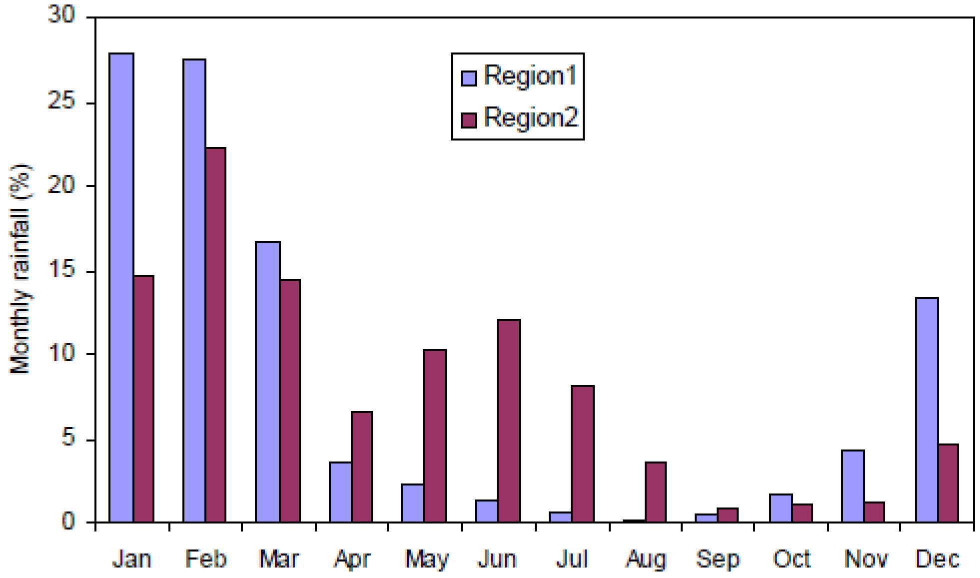


Figure 2.3.2 Regionally averaged (a) monthly rainfall totals and (b) monthly rainfall as a percentage of annual total for Kimberley (Region 1) and Pilbara (Region 2) station networks.

Figure 2.3.4 presents the spatial patterns of rainfall occurrence probabilities for the weather states. A 6-state model was selected based on assessment of the parsimony of 2 to 8 state models. A 6-state model gives a good fit to the data without over-fitting, as determined by the Bayes Information Criterion, a metric that compares NHMMs with different numbers of states (not shown). The 6 states can be briefly summarised as:

State 1. Wet throughout

State 2. Dry throughout

State 3. Wet predominantly along coast

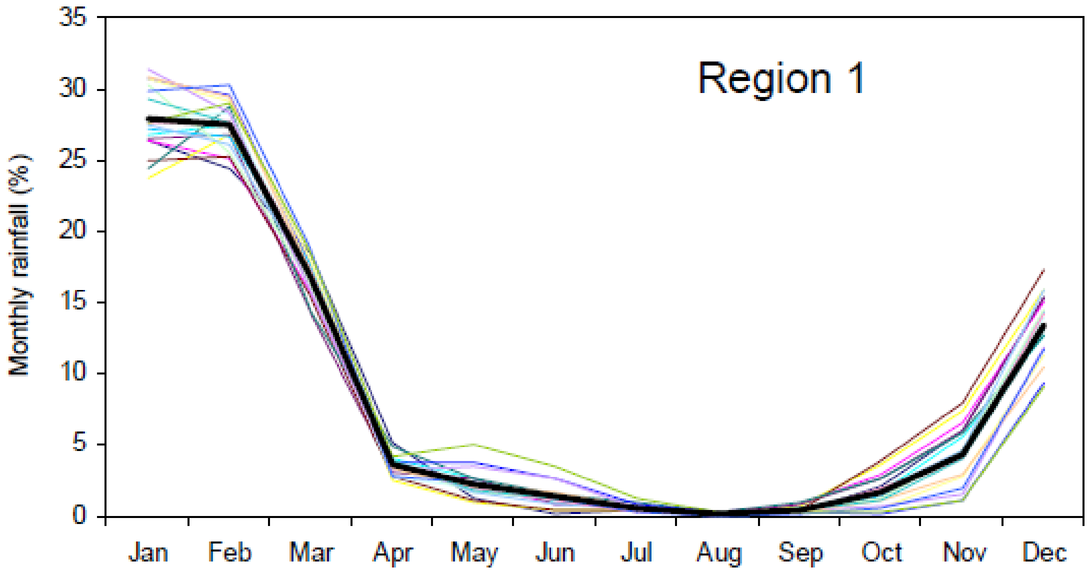
State 4. Wet predominantly inland

State 5. Mostly dry, some rain in north and west

State 6. Dry south-west, wet north and east

Further research is investigating the mean atmospheric conditions coincident with these rainfall patterns, and this will guide predictor selection (described in brief below).

(a)



(b)

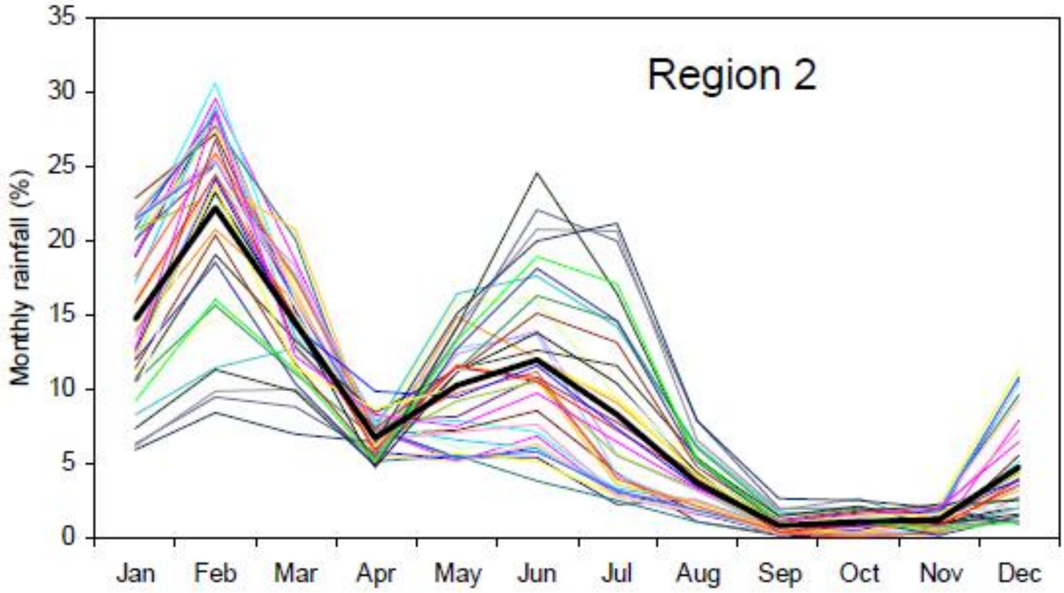


Figure 2.3.3 Station monthly rainfall as a percentage of annual total for (a) 19 Kimberley (Region 1) and (b) 38 Pilbara (Region 2) stations. Thick black line is mean.

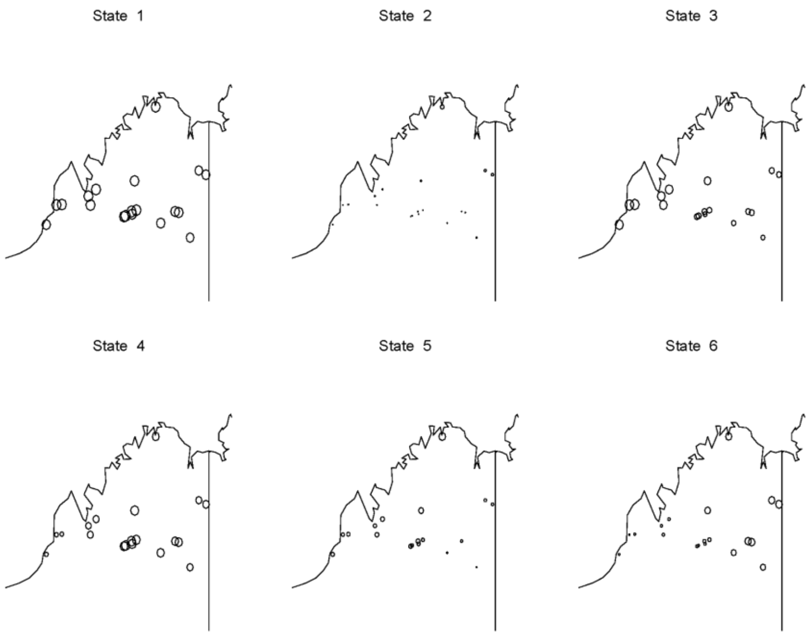


Figure 2.3.4 Mean precipitation occurrence patterns of Kimberley region 6-state SDM fitted to Nov-Apr 1980-1999 data. Circle diameters are proportional to the probability of rainfall.

The two calibrations were compared in terms of the Viterbi and simulated mean weather state frequencies for the fitting period (Table 2.3.3). This assesses the sensitivity of the NHMM parameters to the presence of in-filled data in the fitting period. Whilst most of the weather state mean frequencies are close when comparing NHMMs fitted to (a) all data or (b) higher-quality data (i.e. in-filled marked as missing), State 2 and 5 appear sensitive to the presence of interpolated data. Further research will assess NHMMs calibrated to smaller sets of stations or shorter periods with less infilling. Preliminary SDM calibration for the Pilbara region has also been undertaken and results are currently being analysed.

Table 2.3.3 Mean weather state frequencies from Viterbi algorithm and simulations for NHMMs fitted to (a) all data ('All': includes in-filled data) and (b) only high-quality data ('HQ': in-filled data marked as missing)

	Viterbi		**Simulations**	
NHMM State	**(a) All**	**(b) HQ**	**(a) All**	**(b) HQ**
1	0.096	0.098	0.092	0.095
2	0.365	0.335	0.368	0.347
3	0.072	0.077	0.073	0.075
4	0.137	0.138	0.132	0.131
5	0.172	0.209	0.170	0.203
6	0.158	0.143	0.165	0.150

The long-term weather state time-series are obtained from the 1980-1999 fitted NHMMs using all or only higher quality data for the full 1958 to 2008 period. Figure 2.3.5 shows the Viterbi sequence obtained using all the data (i.e. accepting the PPD deaccumulated and in-filled data as observations).

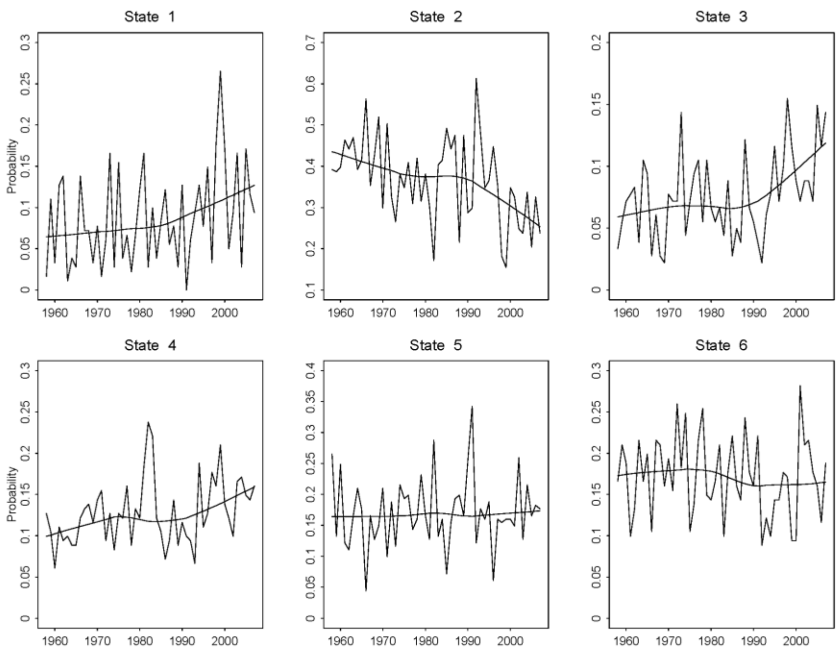


Figure 2.3.5 Weather state time-series of Kimberley region 6-state SDM fitted to Nov-Apr 1980-1999 data using all data.

Figure 2.3.6 is the corresponding set of plots using only the observed data. As noted in Table 2 the quantity of infilling significantly increases in the post-fitting period. The effect of this is evident when comparing Figures 5 and 6, for example State 1 has an increasing trend using all the data and a recent decreasing trend if not, State 2 has a decreasing trend using all data not evident when only using observations. On-going analysis is assessing individual station rainfall trends in the presence and absence of in-filled data, to complement this analysis, to get a more confident picture of rainfall pattern trends in the two regions. This preliminary analysis emphasises the significant impact station data quality has on trend analysis.

All this analysis is predicated by the assumption that the BoM station rainfall data entering the PPD database are correct. It is beyond the scope of this project to undertake a full quality control, as this is the expertise of the BoM and concurrently being undertaken in Project 1.4. Thus the results highlighted here are preliminary, as station networks for the two regions will change based on advice from Project 1.4 as to which stations to include or remove.

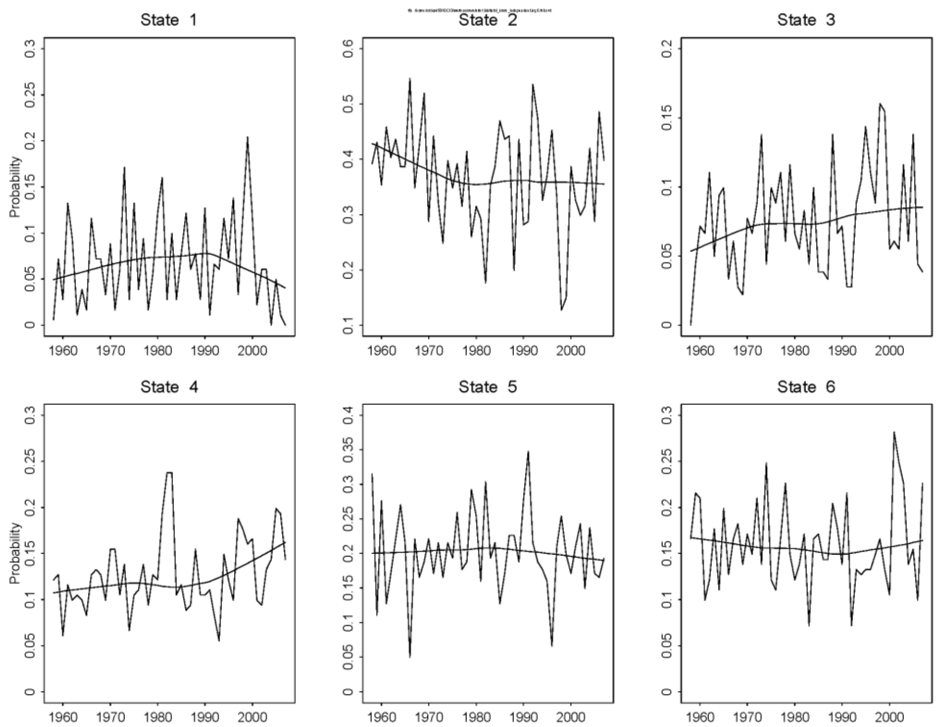


Figure 2.3.6 Weather state time-series of Kimberley region 6-state SDM fitted to Nov-Apr 1980-1999 data using only high-quality data.

Research has also commenced on determining candidate atmospheric predictors for the two regions. NNR daily mean sea level pressure (MSLP) and winds, humidity and temperature (at the 850, 700 and 500 hPa pressure levels) for 1958 to 2008 are being extracted on a 2.5^{o} by 2.5^{o} grid for a region encompassing 107.5^{o} to 132.5^{o} East and 10^{o} to 30^{o} South (11 by 9 NNR grid points). Choice of predictors has to be undertaken with reference to availability of GCM IPCC AR4 daily data, for which only limited daily fields are archived for current (1961-2000), mid-century (2046-2065) and end-of-century (2081-2100) time slices.

Initial predictor screening will be undertaken using correlation analysis between daily series of station rainfall and each candidate atmospheric predictor for each grid and adjacent grid differences (i.e. gradients). More advanced statistical predictor selection techniques are currently under development, and when ready for application they will be applied to determine which combinations of predictors (means and gradients) relate to observed daily rainfall variability (see report for project milestone 2.4.3).

LIST OF TABLES

LIST OF FIGURES