Chapter 5 Identifying hydrological persistence

The previous chapter demonstrated that hydrologic data from across Australia is modulated by fluctuations in global climate modes. Various climate indices describe persistence within atmospheric circulations and form an approach to relate this persistence to rainfall and streamflow variability. This chapter investigates persistence within these rainfall data together with the relationships between this persistence and features shown in climate indices. The statistical tests introduced in Chapter 3 that were derived from spells analyses are applied here as a means to identify hydrological persistence. These data represent rainfall across a range of geographical locations that are influenced by a variety of climate conditions, such that persistence is shown to be a significant aspect of Australian hydrology. Importantly differences between hydrological persistence and the Hurst phenomenon are also clearly illustrated.

5.1 Using runs analysis to identify hydrological persistence

The clustering of high and low values is a major component of the Australian hydrological regime and are an artefact of persistent global circulation phenomena that influence the regional climate of this country. The previous sections outlined the significant influence of global climate modes on the magnitude and variability of monthly hydrologic observations across Australia. The inherent quasi-periodicity of these modes provides the physical persistence within the hydrologic regimes of this country. Investigations using the statistical tests of Section 3.1.1 to describe persistence across Australia have not been previously described in the literature, yet can clearly demonstrate the clustering of rainfall data at a monthly time scale. Stronger persistence is characterised by lower LORT exceedance probabilities, higher run skews, longer mean run lengths and higher autorun coefficients. The hydrologic data used in the previous section are analysed here.

5.1.1 Persistence within monthly rainfall

The deseasonalised monthly rainfall for the four selected districts are analysed for statistically significant persistence using the tests described in Section 3.1.1, with results summarised in Table 5.1. To test for persistence in hydrologic data, it is important to evaluate bounds around these runs statistics for independent (or non-persistent) samples that have identical lengths to the observed data. An independent series has a probability of 0.5 of having a value either above or below the median at each time *t*, leading to an average run length of 2 and an expected value of 0.5 for the lag-1 autorun coefficient. Persistent series are characterised by a clustering of values either side of the median, and this produces a lower probability of above-median values following below-median values and vice versa. Average run lengths for persistent series are greater than 2 with lag-1 autorun coefficients being greater than 0.5.

After simulating 10,000 independent samples of length 1080, 95% of series had an average run length of below 2.101 and a lag-1 autorun of less than 0.525. These random sequences also produced a 90% confidence interval for the run skew of (10.48, 15.89). In observed time series persistence is therefore evident (at a 5% level) from run skews being greater than 15.89. Together with the probability measure for the LORT, these three measures are used here to determine the statistical significance of the hydrological persistence. From this Monte Carlo approach, Table 5.1 shows that each of the four series provides evidence for strong persistence when evaluating each of the four runs statistics. Furthermore Table 5.1 includes statistics evaluated from the monthly NINO3 series, which shows strong persistence in the ENSO signal, with the LORT probability for these data being below the scale of computation.

Table 5.1 Persistence statistics for four district-averaged monthly rainfall series and NINO3

District	LORT prob.	Run skew	Mean length	Autorun lag-1
9A	5.143E-05	16.982	2.122	0.528
16	7.258E-30	28.087	2.427	0.589
27	1.511E-30	28.704	2.389	0.582
71	5.951E-33	25.507	2.348	0.575
NINO3	≈ 0	621.761	8.638	0.885

The spatial variability of monthly persistence is demonstrated through analysing deseasonalised monthly rainfall records for all 107 meteorological districts. The LORT is initially calculated for the deseasonalised monthly totals in each district over their common period 1913-2002, and then to series of annual totals (using April-March water years). The meteorological districts that display significant persistence over monthly and annual periods using the LORT are shaded in Figure 5.1 and Figure 5.2 respectively. Figure 5.1 indicates that 98 districts across Australia (92% of all districts) show a tendency for wet and dry spells to persist longer than expected for independent series. By comparing these results to Figure 5.2, in which only 21 districts indicate significant persistence, it appears that persistence is masked when monthly totals are aggregated. This is either due to the dominant pattern of climatic persistence acting over this finer time scale, or that the smaller sample size makes detecting persistence more difficult.

The statistical significance of run skew, average run lengths and lag-1 autorun coefficients for each district is identified using credibility intervals from independent samples of length 1080. Statistics that exceed the 95th percentile from samples of independent variates are interpreted as significant persistence. From this analysis, the 79 districts that are shaded in Figure 5.3 show significant persistence at a 5% level using each of the four different statistics. These districts, which include the four districts analysed previously, provide strong evidence for significant persistence in the time series of monthly rainfall. Seven of the unshaded districts fail to show significant persistence with any of the four statistical measures.

Figure 5.1 Districts showing statistically significant persistence in deseasonalised monthly rainfall using the length of runs test (LORT)

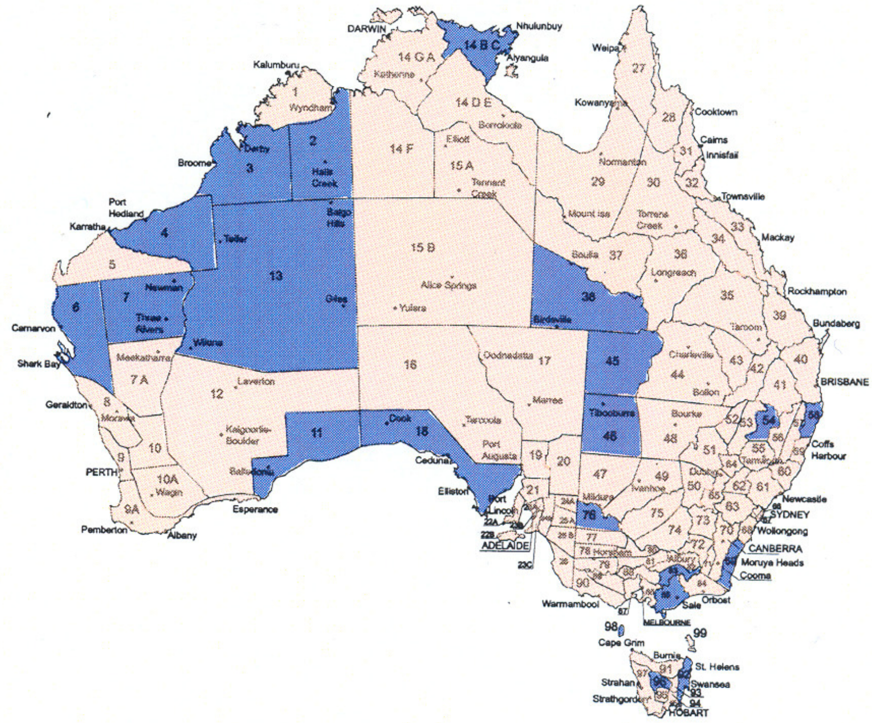

Figure 5.2 Districts showing statistically significant persistence in annual rainfall using the length of runs test (LORT)

Figure 5.3 Districts showing statistically significant persistence in deseasonalised monthly rainfall using each of the four runs statistics

The importance of monthly hydrological persistence as suggested in Figure 5.3 is reinforced by relationships between the different persistence statistics. Exceedance probabilities associated with the LORT statistic are compared to the mean run lengths, run skew and lag-1 autorun coefficient across each district. By first ranking the test statistics for each of these characteristics, the four series of 107 rank values are then correlated. This produces a non-parametric rank correlation between the four separate characteristics of persistence across Australian districts, and the results are presented in Table 5.2. Each correlation is highly significant ($p < 0.001$ in each case), indicating that the four spell statistics describe monthly persistence in the district-averaged rainfall across Australia in a similar manner.

Table 5.2 Rank correlations between the four runs statistics for deseasonalised monthly rainfall of the 107 districts across Australia

	LORT	Run skew	Mean run length
Run skew	0.923	–	–
Mean run length	0.653	0.746	–
Lag-1 autorun	0.670	0.768	0.980

The relationship between monthly hydrological persistence and ENSO is now investigated by comparing correlations between monthly rainfalls and monthly NINO3 values with spells statistics. Once again the series of 107 rainfall-NINO3 correlations are compared with respective LORT probabilities, run skews, mean run lengths and lag-1 autorun correlations through non-parametric rank correlations. These results are presented in Table 5.3.

Table 5.3 Rank correlations between the four runs statistics and the linear correlations between deseasonalised monthly rainfall in the 107 districts across Australia and monthly NINO3 values

	LORT	Run skew	Mean lengths	Lag-1 autorun	Max lengths
Correlation to NINO3	**0.193** (0.047)	**0.244** (0.011)	**0.373** (<0.001)	**0.397** (<0.001)	**-0.005** (0.962)

These results indicate that districts in which monthly rainfall is strongly related to the NINO3 series have a tendency towards stronger persistence, providing evidence for the role of ENSO in the clustering of high and low monthly anomalies. This result appears to contradict the results of Peel *et al.* (2004b), who found that within annual runoff records, ENSO had minimal influence upon the length of runs either side of median values. It is proposed that the use of annual totals in this previous study obscures the major source of ENSO variability, and that monthly values more clearly demonstrate the hydroclimatic relationships. Although Table 5.3 shows strong relationships, the longest runs in each district remain unrelated to the strength of the ENSO signal, suggesting that ENSO variability fails to influence hydrologic extremes, instead modulating other values such as the mean as was identified in the previous section.

These results demonstrate significance persistence in the time series of deseasonalised monthly rainfall for a majority of Australian meteorological districts. Persistence is interpreted as an important aspect of the hydrological regime of this country and it is vital that monthly rainfall simulations capture these main persistence characteristics. Monthly streamflows from across Australia are now analysed for evidence of significant hydrological persistence. It is expected that these series are more persistent than rainfall due to the effects of catchment storage.

5.1.2 Persistence within monthly streamflows

The five statistics used in the previous section are utilised again to identify monthly persistence in flows from the seven rivers described in Section 4.1. Persistence in the monthly flows of arid zone rivers cannot be assessed in the same manner due to the high proportion of months having zero flow. Furthermore, with only 29 years of data available for the Omatako, 22 years for the Omaruru and 37 years for the Todd, there may be insufficient information to observe statistically significant spells either side of a long-term threshold.

The temporal dependence in the deseasonalised monthly flow record of the seven rivers analysed are shown in Figure 5.4 and Figure 5.5. Although autocorrelations in the monthly flows of the Murray and Darling, in particular, remain significant for many lags, neither flow record appears to have the slow decay indicative of mathematical persistence.

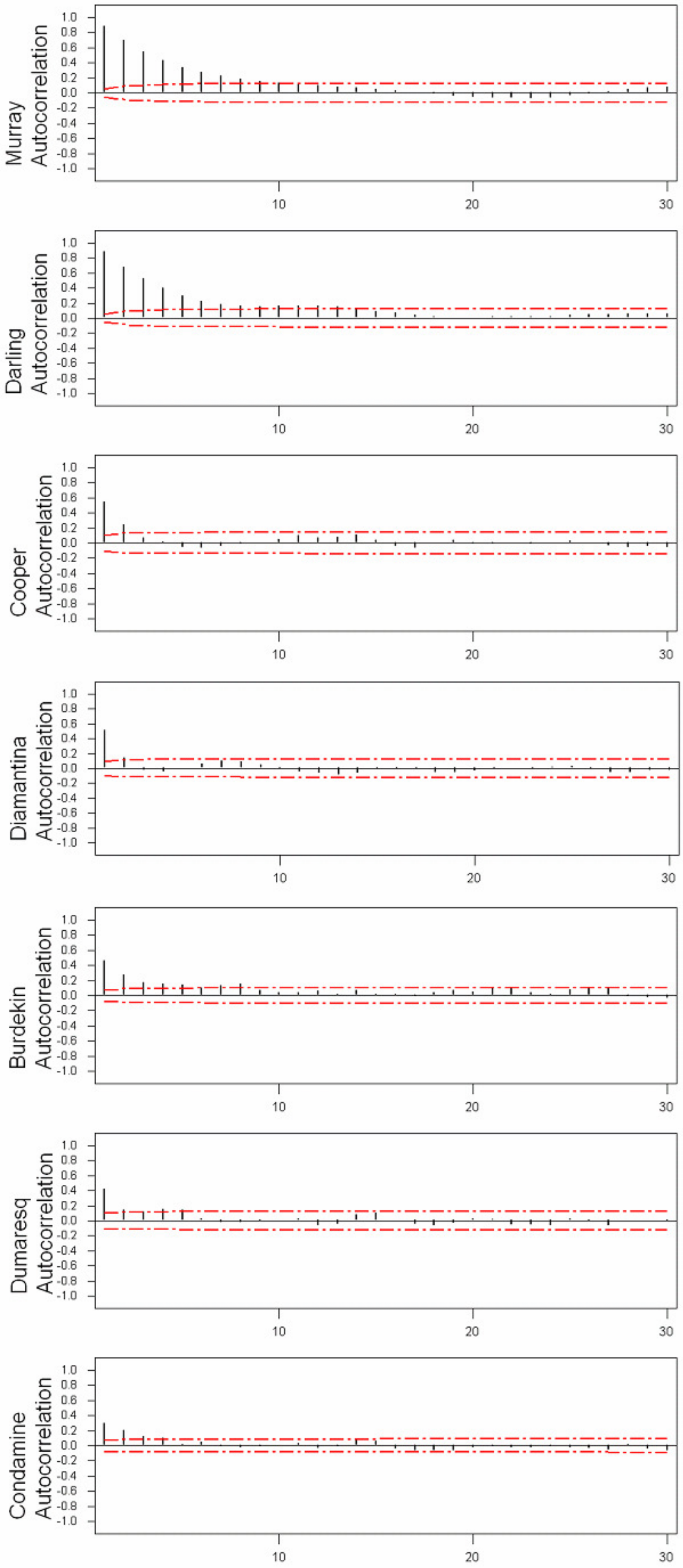

Figure 5.4 Correlograms for deseasonalised monthly flows in the seven rivers analysed

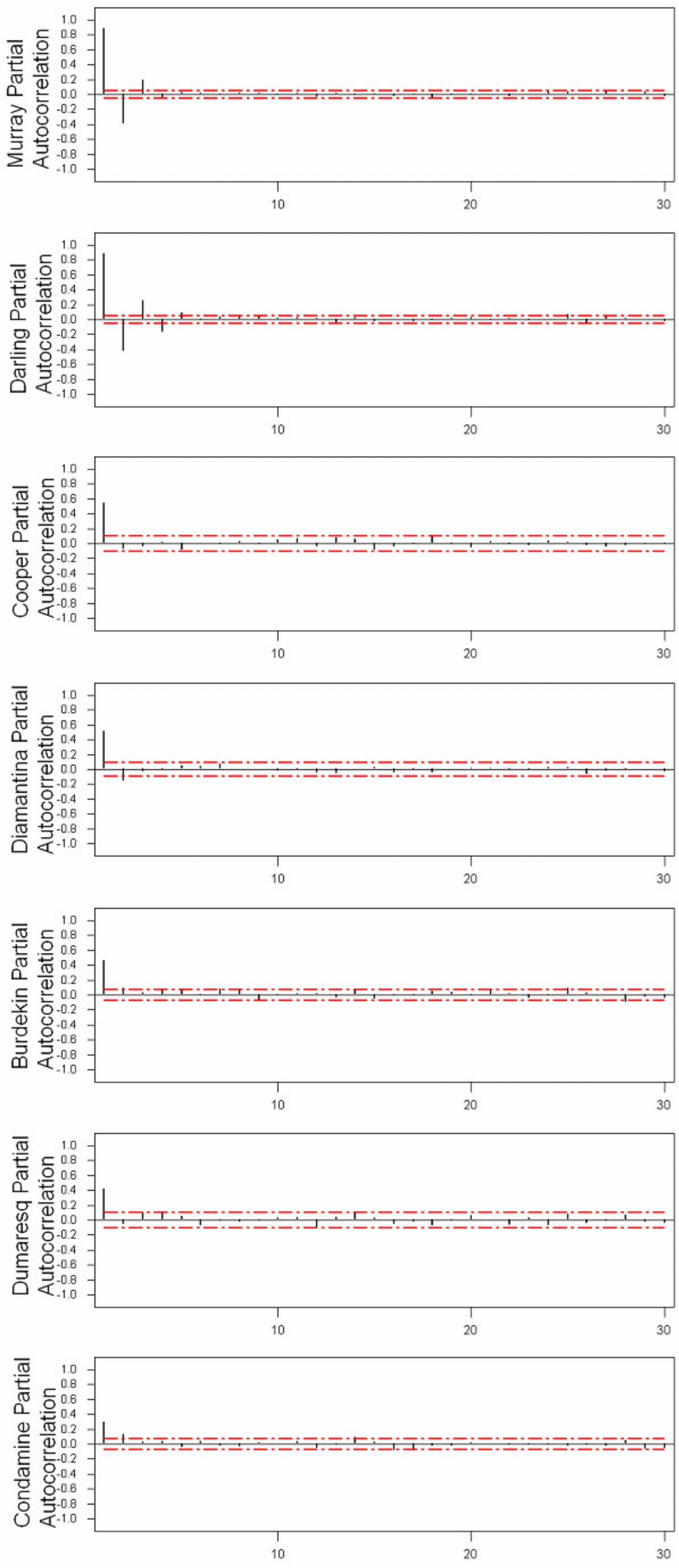

Figure 5.5 Partial correlograms for deseasonalised monthly flows in the seven rivers analysed

Persistence statistics for the deseasonalised monthly flow records are shown in Table 5.4, with each series demonstrating much stronger persistence than the rainfall series in Table 5.1. Persistence is strongest within Murray flows, having an average run length of 7.62 months that well exceeds the longest average run length from the district rainfall series of 2.52 months in District 30. The Murray series also has a run skew of 247.9 (as opposed to the largest skew of the rainfall records that is 32.32 in District 28) and a maximum run length of 39 months. With the Murray having the largest catchment area of the rivers investigated, hence the largest soil storage capacity, catchment size is likely related to the magnitude of streamflow persistence. The LORT probabilities are not shown in Table 5.4 as these are below the lower bound of computation. These results suggest that variability within rainfall that is observed across a catchment may be amplified through catchment storage to produce more evident cycles in streamflow observations. The run skew and maximum run length statistics show an identical ranked order for these seven rivers, as do the average run lengths and autorun statistics. The rank correlations between run skew or maximum run length and average run length or lag-1 autorun are 0.964, which demonstrates that these four statistics are identifying hydrological persistence in a similar fashion.

Table 5.4 Persistence statistics for deseasonalised monthly flows for the seven rivers

	Run skew	Max lengths	Mean length	Autorun lag-1
Murray	247.988	39	7.624	0.870
Darling	212.312	37	6.384	0.843
Cooper	33.095	10	2.924	0.663
Diamantina	36.111	13	2.938	0.664
Burdekin	63.294	21	3.196	0.688
Dumaresq	59.013	19	3.356	0.704
Condamine	69.720	24	3.377	0.706

5.2 Identifying evidence for the Hurst phenomenon

The Hurst phenomenon is the final aspect of monthly-scale persistence to be investigated in these rainfall and streamflow data. This behaviour of the rescaled range statistic is analysed here in order to estimate the Hurst exponent for hydrologic time series, which in turn is related to statistics from runs analyses described earlier.

5.2.1 Sampling bounds for estimates of the Hurst exponent

Identification of the Hurst phenomenon is impeded by the requirement for accurate estimation of the exponent H. The sampling error associated with the rescaled adjusted range statistic is illustrated in Figure 5.6, which shows the Hurst exponent evaluated for multiple time series of independent Gaussian variates. Time series of lengths 1080 and 90 are chosen for analysis, as these are consistent with the length of spatially-averaged rainfall series aggregated over monthly and annual periods. Hurst exponents are estimated for each of 1000 simulations at both lengths, and represented in Figure 5.6 as histograms. These histograms show wide sampling distributions for series lacking any evidence of long-memory. The 1000 samples of length 1080 have a median value of 0.512, consistent with the asymptotic value of 0.5 for non-persistent data, and a 90% credibility interval of (0.376, 0.675). For independent series of length 90, which reflects the length of annual totals for the district-averaged rainfall series, the median estimate of H increases to 0.628, with a wider 90% credibility interval of (0.492, 0.771). Independent trials such as these are useful for identifying districts that display evidence for the Hurst phenomenon.

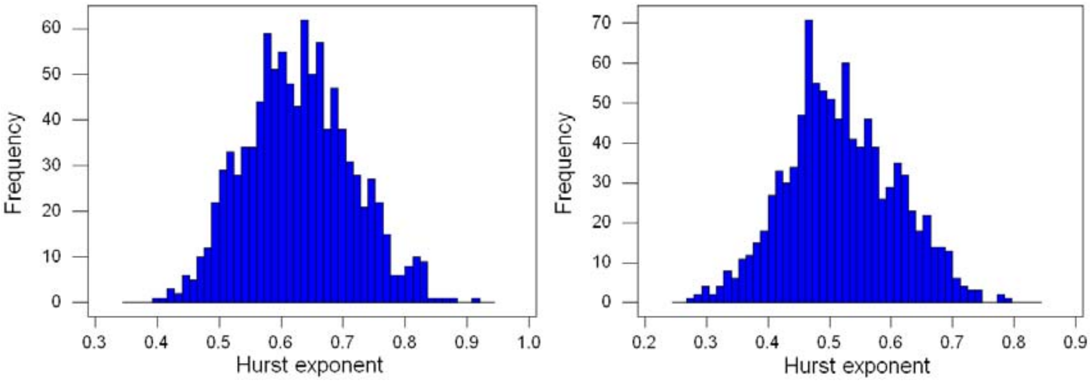

Figure 5.6 Sampling distribution of estimates for Hurst exponents for 1000 simulations of independent series of length 90 (left) and length 1080 (right)

5.2.2 Estimating the Hurst exponent in monthly rainfall data

The rescaled adjusted range statistic is estimated from the deseasonalised monthly rainfall in each district of Australia, with districts having an estimate for H exceeding the 90% credibility bound for independent series being shaded in Figure 5.7. This figure shows that 23 districts across Australia (21% of all districts) have Hurst exponents that exceed the 95% credibility limit for a non-persistent series of length 1080. These 23 districts appear to be located across all rainfall regimes, with no spatial bias towards any single climatic regime. Using the rescaled range statistic to estimate Hurst exponents for annual rainfall in each district, 100 districts in Australia have H estimates that exceed their monthly estimates. However with wider sampling bounds associated with the shorter annual series, only 15 of these districts have Hurst exponents exceeding the 95% credibility limit for independent samples. These 15 districts are shaded in Figure 5.8, and each of these also displays significant Hurst behaviour at a monthly scale.

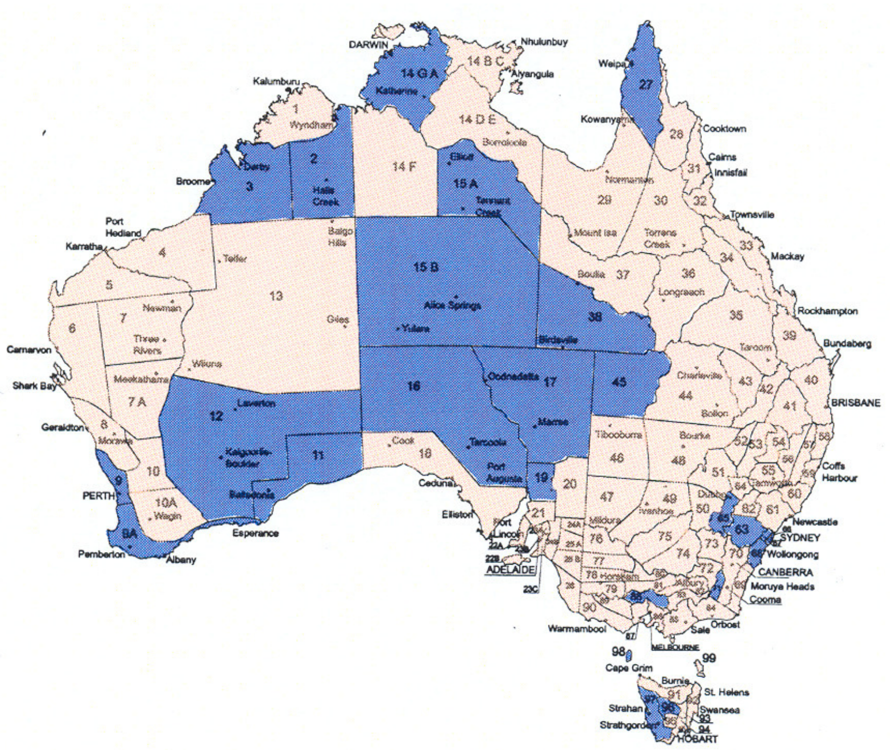

Figure 5.7 Districts showing estimates of Hurst exponents for deseasonalised monthly rainfall series that exceed the 95th percentile of multiple simulations of independent series

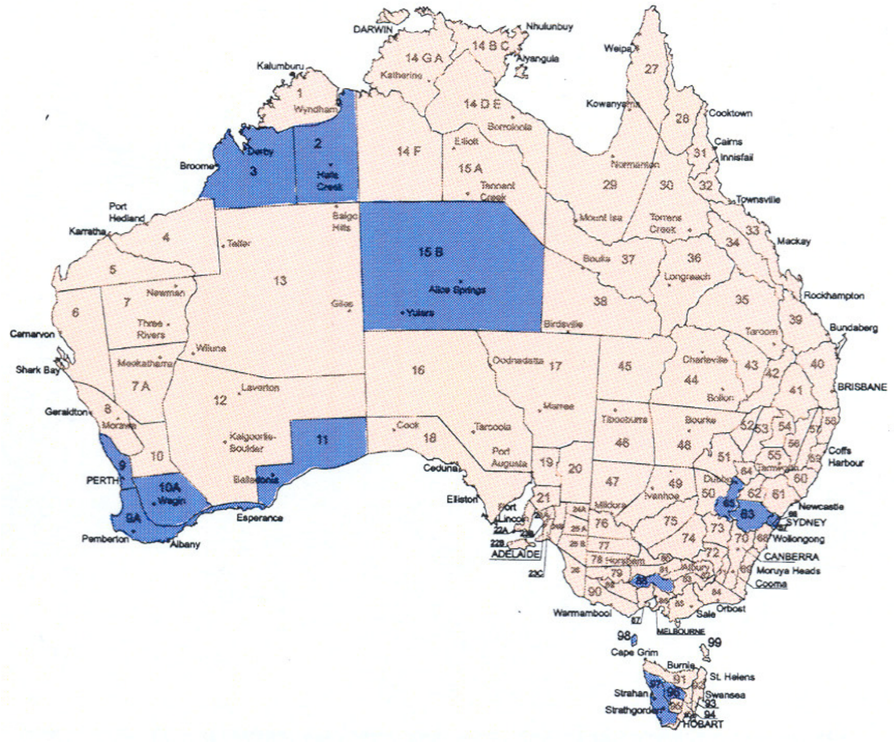

Figure 5.8 Districts showing estimates of Hurst exponents for annual rainfall that exceed the 95th percentile of multiple simulations of independent series

A visual inspection of Figure 5.8 shows little homogeneity between districts that demonstrate significant Hurst effects and the various rainfall regimes of this country. This suggests that the rescaled adjusted range statistic cannot provide a consistent description of relationships between climatic processes and rainfall observations across the country.

It is important to now investigate the relationship between the Hurst phenomenon and the hydrological persistence in district rainfall that was previously outlined. Table 5.5 displays the rank correlations between the series of Hurst exponents for the 107 deseasonalised monthly rainfall series and statistics derived from runs analyses, with none of these correlations being significant at a 5% level. The estimates of H are significantly related to the series of linear correlations between monthly observations and NINO3 values, although the R^2 associated with this rank correlation is less than 6%, proving that this relationship is not strong. These results suggest that values of the Hurst exponent in the range $0.5 < H < 1$ fail to describe the same characteristics as those identified by runs analyses. This provides evidence that the Hurst phenomenon is distinct from the anomalous run behaviour that has been termed hydrological persistence, and it is proposed that the array of persistence statistics presented here provide a superior description of hydrological persistent processes than Hurst statistics such as the rescaled range.

Table 5.5 Rank correlations between estimates of the Hurst exponents, runs statistics and linear correlations between deseasonalised monthly rainfall and monthly NINO3 values for the 107 districts across Australia

	LORT	Run skew	Mean lengths	Lag-1 autorun	NINO3 corr.
Hurst exponent	**0.188** (0.053)	**0.142** (0.144)	**0.173** (0.075)	**0.156** (0.109)	**0.237** (0.014)

Correlations between Hurst exponents and persistence statistics for monthly streamflow records are similar to those for monthly rainfall. The magnitudes of the Hurst coefficients shown in Table 5.6 are inconsistent with the magnitude of persistence as defined by runs statistics. Although the Murray shows strong hydrological persistence, its low Hurst exponent provides further evidence that the Hurst phenomenon identifies a different form of persistent behaviour than that described by runs statistics. Furthermore, the ranked order of these exponents is uncorrelated to the order of the runs statistics shown in Table 5.4. As a consequence, stochastic models designed to preserve sequences of values either side of a long-term threshold are more suitable for hydrologic data than those designed to replicate estimates of the Hurst exponent.

Table 5.6 Hurst exponents for deseasonalised monthly flows for the seven rivers analysed

Murray	Darling	Cooper	Diamantina	Burdekin	Dumaresq	Condamine
0.610	0.744	0.692	0.559	0.667	0.635	0.508

5.3 Summary of chapter

Evidence for significant hydrological persistence at a monthly scale was identified in this chapter within various hydrologic records from across Australia. Persistence is interpreted as extended periods during which observed values remain either side of a long-term threshold, and is identified through a range of statistical tests that are based on the theory of runs analysis.

Spatially-averaged rainfall data from defined meteorological districts showed much of the Australian rainfall regime to be persistent. Importantly, much of this persistence was concealed when time series of monthly values were aggregated to annual totals. Furthermore, time series of deseasonalised monthly streamflow anomalies displayed much stronger persistence than time series of rainfall observations. The Hurst phenomenon has been previously discussed as a measure of long-term persistence; however results in this chapter indicate that this characteristic is dissimilar to fluctuations between extended wet and dry spells that has been termed persistence in the present work.

Hydrological persistence is likely to have vital importance upon water resource management across Australia given the extent of regions that demonstrate this feature. Stochastic models for hydrologic time series will be improved by accounting for fluctuations between wet and dry spells. The following chapters investigate hidden Markov models (HMMs) as a method to incorporate the hydrological persistence shown in this chapter.