CHAPTER 5

PARAMETER ESTIMATIONS

Parameters have to be estimated separately for the binary chain, the jitter model, the capping technique and the seasonality incorporated variance relationship of the stochastic disaggregation model. As shown in the previous chapter in section 4.1.3, the statistical moments of the jitter model were derived from the statistical moments of the binary chain. Hence the estimation of the binary chain parameters is sufficient to cater for both the binary chain and the jitter model. As per section 3.1 of chapter 3, the parameters for the stochastic disaggregation model will be estimated for regions 35, 39 and 40 in Queensland, Australia. It was also emphasised in section 4.1.2 in the previous chapter for using a constant regional monthly value of lag-1 autocorrelation, $\rho_{\bar{Y}}(h)$. Therefore its value has to be determined for Queensland.

For parameter calibration, all 65 selected stations in regions 35, 39 and 40 in Queensland from Table 3.1 were used. Each of these stations has at least 10 years of 6-minute BOM observed data for all months of the year. The following Queensland map shows their geographic locations:

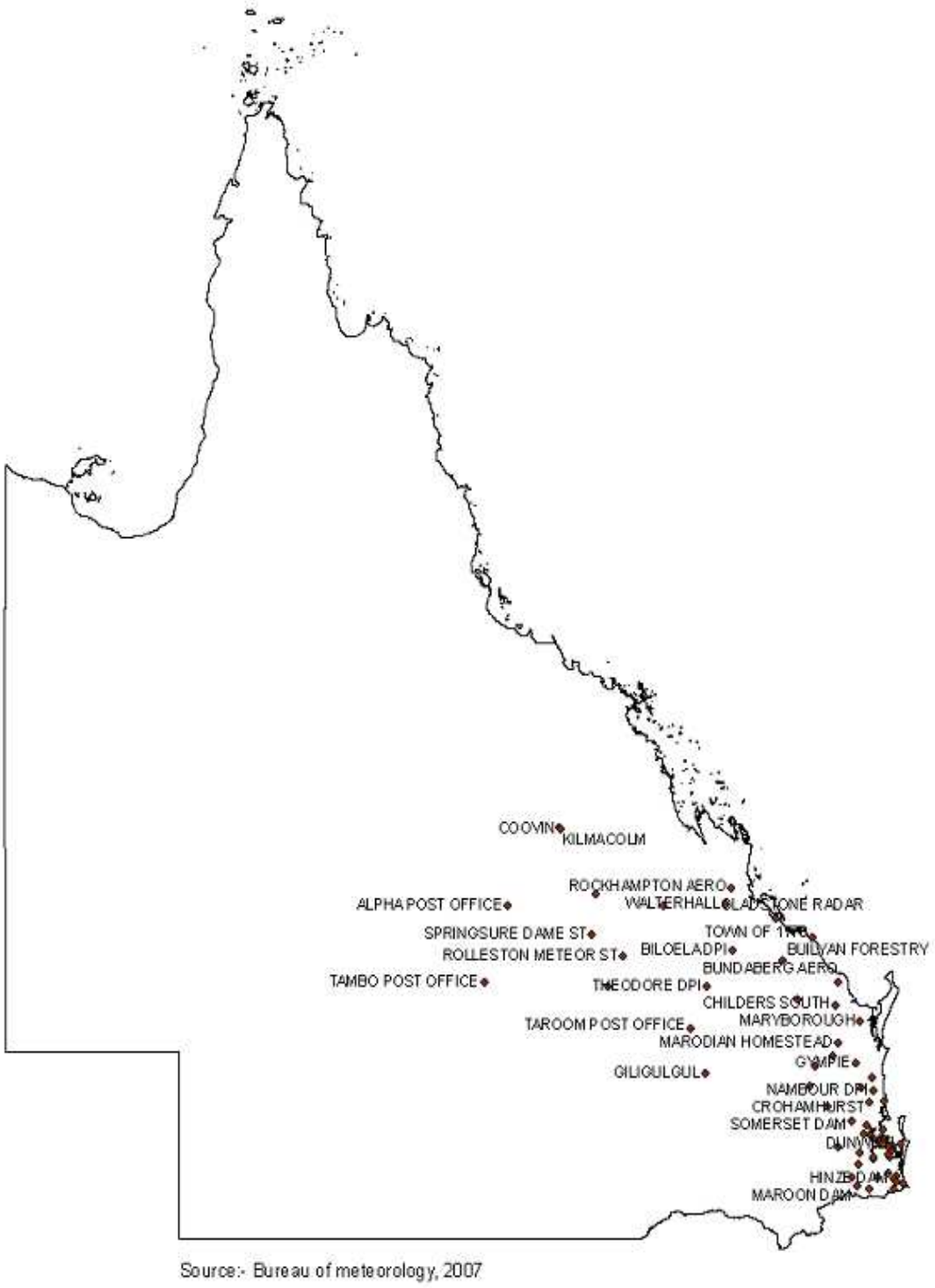

Figure 5.1 Queensland Rainfall Stations

5.1 Binary Model Parameter Estimation for Queensland

6-minute rainfall data of 65 stations obtained from the Bureau of Meteorology (BOM), Australia, were used to develop the Queensland parameters of the binary chain. For a given year, months with missing data were excluded. This resulted in rainfall data of different station-months having a different number of years of record. For the data analysed, the station-month number of years of data varied between 10 and 84 years. The monthly BOM data were analysed at 12 aggregation levels (0.1, 0.2, 0.3, 0.5, 1, 2, 4, 6, 8, 12, 18, and 24 h).

The binary model has seven parameters (β, η, and γ [a_0, a_1, b_1, a_2 and b_2]) to be calibrated. For a given parameter set, values of β, η, and γ, and the daily dry probability, $P(24)$, equation 4.2 was used to estimate parameter λ for each station-month. With these parameters the analytical dry probabilities for the various aggregation levels were obtained. The parameters were calibrated by minimising the objective function, J as:

$$J = \sum_{j=1}^{65} \sum_{i=1}^{12} \sum_{t=1}^{12} [P_{oij}(h_t) - P_{sij}(h_t)]^2 \qquad (5.1)$$

where h_t is the timescale of aggregation level t, $P_{oij}(h_t)$ and $P_{sij}(h_t)$ are the observed and analytical dry probabilities of month i and station j, respectively. The global optimisation search strategy (Duan et al., 1992) of the Bayesian Non-Linear regression software NLFIT (Kuczera, 1994) is used to calibrate the binary chain parameters. Table 5.1 shows the globally optimised parameters and their corresponding standard deviations for individual regions.

Table 5.1 Regional parameters under global optimization search

Parameters	Region 35	Standard Deviation of Region 35	Region 39	Standard Deviation of Region 39	Region 40	Standard Deviation of Region 40
β	0.400002	0.0389463	0.400000	0.0336551	0.400000	0.0175275
η	0.751466	0.0453324	0.845788	0.0491831	0.728609	0.0225506
a_0	0.146094	0.0032701	0.123936	0.00271617	0.107513	0.0011891
a_1	0.062006	0.0020003	0.030032	0.00142998	0.026937	0.0006632
b_1	-0.01527	0.0017679	-0.01236	0.00169775	-0.01445	0.0007678
a_2	0.012224	0.0016810	-0.00585	0.00146642	-0.00342	0.0006495
b_2	-0.00999	0.0015644	-0.00999	0.00134542	-0.00349	0.0006011

From the above optimisation results a *"t"* statistics testing is formulated to find out the confidence interval of the estimated parameters. The formula of the *"t"* statistics is given by Watson et al. (1993) as follows:

$$t = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} \tag{5.2}$$

Here $\bar{X}$ and $\frac{s}{\sqrt{n}}$ are the sample mean and standard error respectively that change from region to region. The population mean μ has to be estimated from the sample mean. At 95% confidence limit, the level of significance *(α)* is (1-0.95) i.e., 0.05. As the sample size varies, the degrees of freedom differ from sample to sample. Table 5.2 shows the sample sizes and the corresponding degrees of freedom.

Table 5.2 Sample sizes and corresponding degrees of freedom

Sampling criteria	Sample names	Sample size, n	Degrees of Freedom, n-1
Stations having monthly rainfall data $\geq$ 10 years	Region 35	11	10
	Region 39	11	10
	Region 40	43	42

If the left and right bounds of the confidence limit for the population mean μ are denoted as L and R respectively, their expression is (Watson et al. 1993):

$$L = \bar{X} - t_{\frac{\alpha}{2},n-1}\frac{s}{\sqrt{n}} \tag{5.3}$$

$$R = \bar{X} + t_{\frac{\alpha}{2},n-1}\frac{s}{\sqrt{n}} \tag{5.4}$$

Using these boundary equations, the upper and lower bounds for parameters β, η, a_0, a_1, b_1, a_2 and b_2 were calculated and plotted. There are some common ranges for some of the parameters which will be discussed subsequently. The boundary values of parameter β and parameter η are shown in Figures 5.2 and 5.3 respectively.

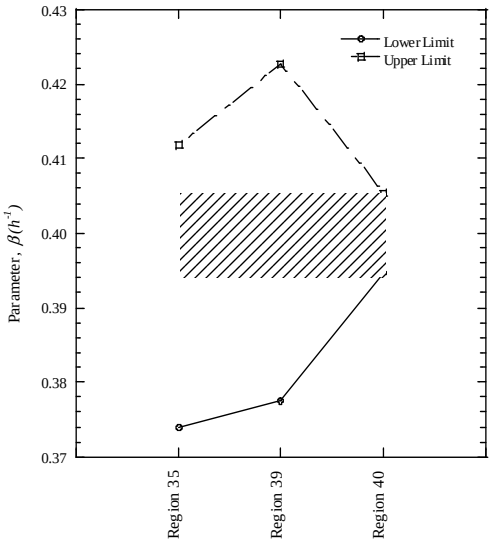

Figure 5.2 Confidence Intervals of parameter β under global optimisation search with shaded area representing overlapping values for all three regions

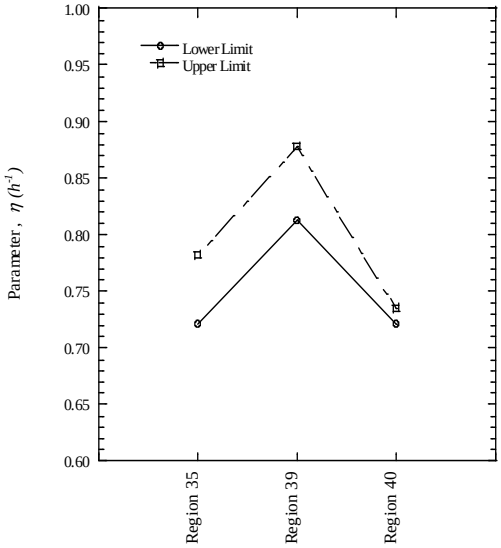

Figure 5.3 Confidence Intervals of parameter η under global optimisation search

The shaded area in Figure 5.2 represents the overlapping values of β for all the three regions under consideration. With 95% confidence it can be estimated that the mean population value of parameter β lies within the range of 0.395 to 0.405. No such overlapping values were found for parameter η in Figure 5.3. However, there are overlapping values for parameter η between regions 35 and 40 ranging from 0.722 to 0.736. To find out the possibility of fixing these parameters as constants over the three regions, correlation analyses of these parameters will be done later.

As stated earlier, parameter γ is estimated for its monthly variations from a Second Harmonic Fourier series (Gyasi-Agyei, 1999). The globally optimised values of parameters a_0, a_1, b_1, a_2 and b_2 are used to parameterise γ. To find out the values of these parameters, the same *"t"* statistics testing with a 95% confidence limit was performed again. Figures 5.4 to 5.8 depict the ranges for these parameters.

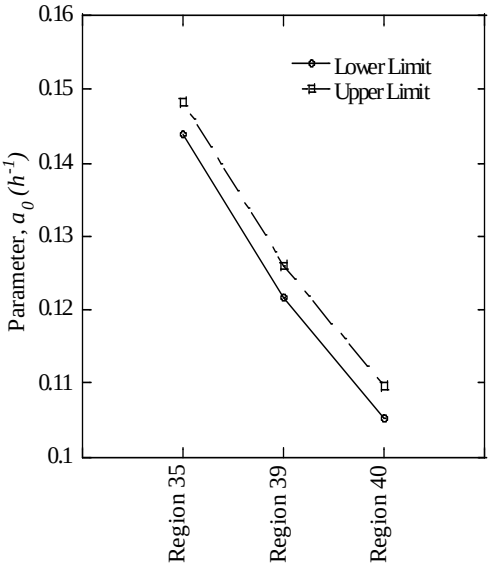

Figure 5.4 Confidence intervals of a_0

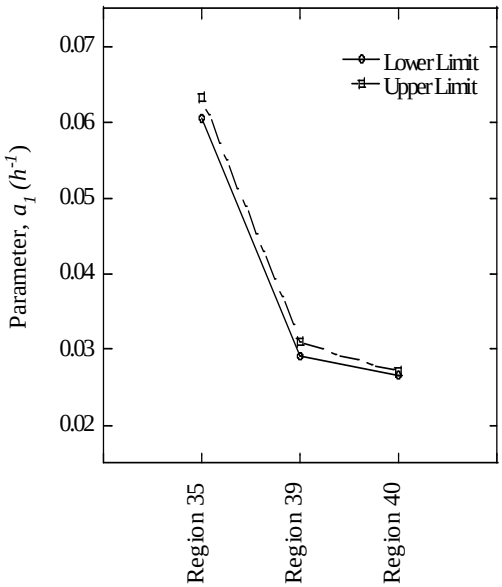

Figure 5.5 Confidence intervals of a_1

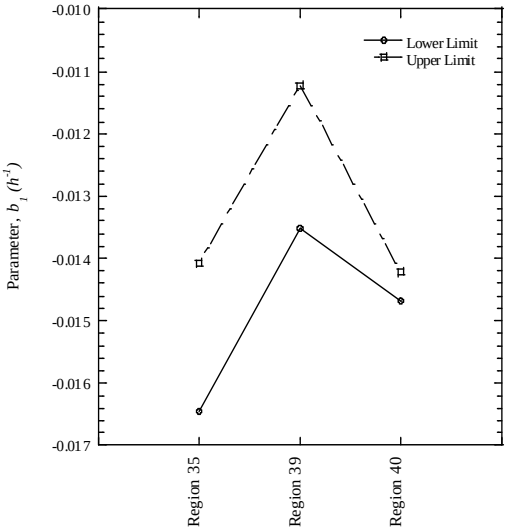

Figure 5.6 Confidence intervals of b_1

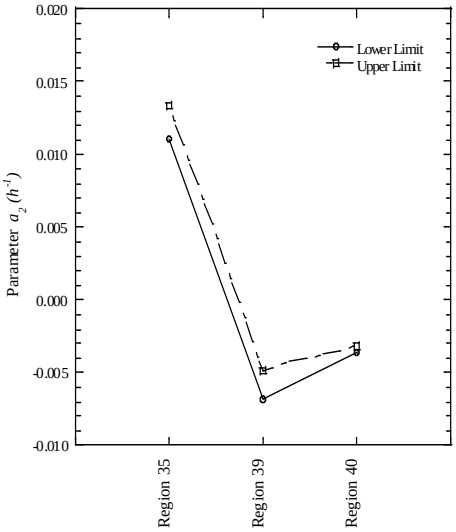

Figure 5.7 Confidence intervals of a_2

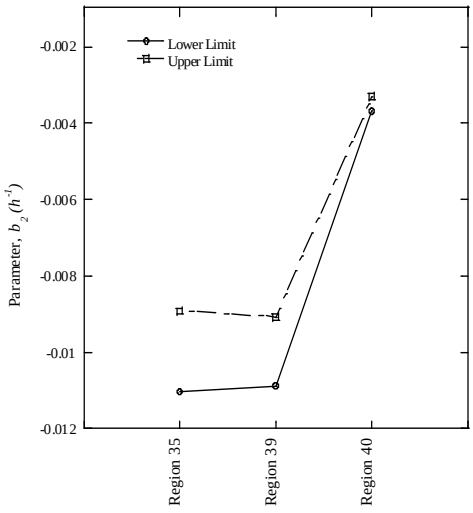

Figure 5.8 Confidence intervals of b_2

It is obvious from Figures 5.4 to 5.8 that there are no commonly identifiable ranges for parameters a_0, a_1, b_1, a_2 and b_2. Hence the generalisation of these parameters is not

possible. Rather, we have to use the regionalised values for these parameters to determine the parameter γ.

To establish the above assertion on parameters, a detailed step-by-step parameter separation technique by correlation analyses has been adopted. In step 1 all the parameters were considered as variables and the correlation coefficients between them were determined. Mathematically, the correlation coefficient, usually designated by "ρ" for the population and "r" for the sample, is defined as follows:

$$r = \frac{\text{Covariance between the two variables}}{(\text{Standard deviation of one variable})(\text{Standard deviation of other variable})}$$

That is, $$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}} \qquad (5.5)$$

The correlation coefficient can be either positive or negative within a range of -1 to +1. When the absolute correlation coefficient of two random variables is very close to unity then there is a strong correlation between them. If it is close to "0" then there is no correlation between the two random variables. During step 1, it was observed that there were strong correlations between $\beta - \eta$, $\beta - a_0$ and $\eta - a_0$. This means that if any of these parameters were considered as constant then the loss of accuracy in parameter estimation would be catered for by the other correlated parameters. The correlations between the other parameters were very small and therefore ignored.

In step 2, parameter β was considered as a constant by taking the mean population values as determined in the earlier "*t*" testing for all three regions. At the end of this step the correlation between $\eta - a_0$ got stronger than before for all regions indicating similar simulation results from a wide range of parameter space if one is taken out as a constant.

In step 3, both β and η were considered as constants by taking their mean population value from the standard "*t*" tests. At the end of this step the correlations between a_0-a_1, a_0-b_1, a_0-a_2, a_0-b_2, a_1-b_1, a_1-a_2, a_1-b_2, b_1-a_2, b_1-b_2, a_2-b_2 were insignificant. This suggested no further parameters could be isolated from the parameter space. Table 5.3 shows the correlation coefficients at different steps of the parameter identification process.

Table 5.3 Various steps of parameter Identification by correlation analyses

Region	***Variability Criteria***	$\beta - \eta$	$\beta - a_0$	$\eta - a_0$	a_0-a_1	a_0-b_1	a_0-a_2	a_0-b_2
Region 35	All Variable	0.9526	0.7384	0.5317	0.2863	-0.2072	-0.0161	-0.0695
	β fixed	---------	---------	-0.8377	0.5660	-0.3792	0.0958	-0.2014
	β, η fixed	---------	---------	----------	0.2046	-0.4521	-0.0879	-0.1549
Region 39	All Variable	0.9509	0.7789	0.5933	0.1917	-0.2593	-0.1460	-0.0903
	β fixed	----------	----------	-0.7591	0.2837	-0.4480	-0.2108	-0.1607
	β, η fixed	----------	----------	----------	0.1175	-0.5873	-0.2775	-0.1261
Region 40	All Variable	0.9557	0.7434	0.5577	0.1806	-0.2860	-0.1273	-0.0177
	β fixed	----------	----------	-0.7762	0.3085	-0.4718	-0.1729	-0.0611
	β, η fixed	----------	----------	----------	0.1281	-0.6051	-0.2379	-0.0293
All Region	All Variable	0.9543	0.7849	0.6044	0.2014	-0.2653	-0.1167	-0.0337
	β fixed	----------	----------	-0.779	0.346	-0.47	-0.158	-0.093
	β, η fixed	----------	----------	----------	0.1554	-0.6014	-0.2305	-0.0614

As the numbers of parameters set fixed were increased in the step-by-step parameter identification process, it was found that there was little or no change in the correlation

among a_0, a_1, b_1, a_2 and b_2. Even though parameter b_1 showed a slight gradual improvement in its correlation with parameter a_0 during the steps of parameter identification, this improvement was far less than expected. The highest correlation between b_1 and a_0 was achieved as -0.6051 as shown in the Table 5.3 for region 40. These findings validate the decision of considering parameters β and η fixed and a_0, a_1, b_1, a_2, b_2 as regional variables to calculate parameter γ. Finally, the fitted Second Harmonic Fourier Series (Gyasi-Agyei, 1999) of equation 4.1 was used to calculate γ. Figure 5.9 shows the patterns of variations of the parameter γ for the regions of 35, 39, 40 and all three regions together. Since region 40 has more than twice the sample size of regions 35 and 39 combined, the "all region" curve is closer to that of region 40.

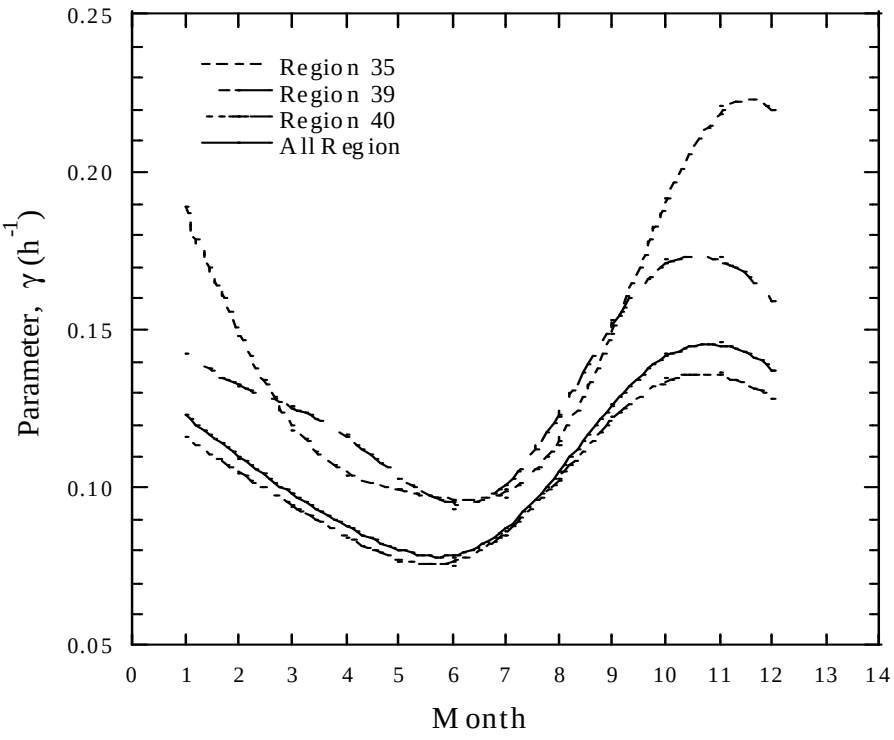

Figure 5.9 Seasonal variations of parameter γ for all regions

The tabulated γ values are given in Table 5.4.

Table 5.4 γ values for all months

Months	Region 35	Region 39	Region 40	All Regions
Jan	0.188866	0.142512	0.116179	0.123322
Feb	0.148543	0.132113	0.104758	0.10955
Mar	0.118221	0.125873	0.09437	0.098014
Apr	0.104135	0.116746	0.084436	0.087905
May	0.099269	0.103217	0.076689	0.079829
Jun	0.096026	0.093668	0.075638	0.077747
Jul	0.097164	0.099375	0.084929	0.086117
Aug	0.11317	0.122771	0.103	0.1048
Sep	0.148654	0.15283	0.122576	0.12699
Oct	0.19222	0.172778	0.135048	0.142844
Nov	0.221403	0.173311	0.136145	0.14601
Dec	0.219483	0.159042	0.128008	0.137437

Since the prediction of the dry probability is the driving factor for the estimation of all the analytical model parameters as per equations 4.2 through 4.5, the corresponding dry probabilities for all months and all aggregation levels are calculated. There are two error statistics: the average absolute percentage error E_1 and maximum absolute percentage error E_2, which are defined as:

$$E_1(\%) = \frac{1}{m \times n} \sum_{h=1}^{m \times n} \left| \frac{P_0(h) - P_a(h)}{P_0(h)} \right| 100 \tag{5.6}$$

$$E_2(\%) = \max_{h=1, m \times n} \left| \frac{P_0(h) - P_a(h)}{P_0(h)} \right| 100 \tag{5.7}$$

There are also two different criteria for calculating E_1 and E_2 statistics: on monthly basis and on aggregation level basis. If these are calculated for months, then *"n"* is the number of aggregation levels. On the other hand, if the error statistics are calculated for aggregation levels, then *"n"* is the number of months. *"m"* is the total number of stations in the region. $P_0(h)$ and $P_a(h)$ are observed and analytical dry probabilities, respectively. $P_0(h)$ is calculated from the equation 3.6 and $P_a(h)$ is calculated from equation 4.2. Figure 5.10 shows the comparative error statistics E_1 and E_2 amongst regions 35, 39, 40 and all regions for all aggregation levels and all months. In each case β

and η were fixed and γ was calculated from the Second Harmonic Fourier Series (Gyasi-Agyei,1999).

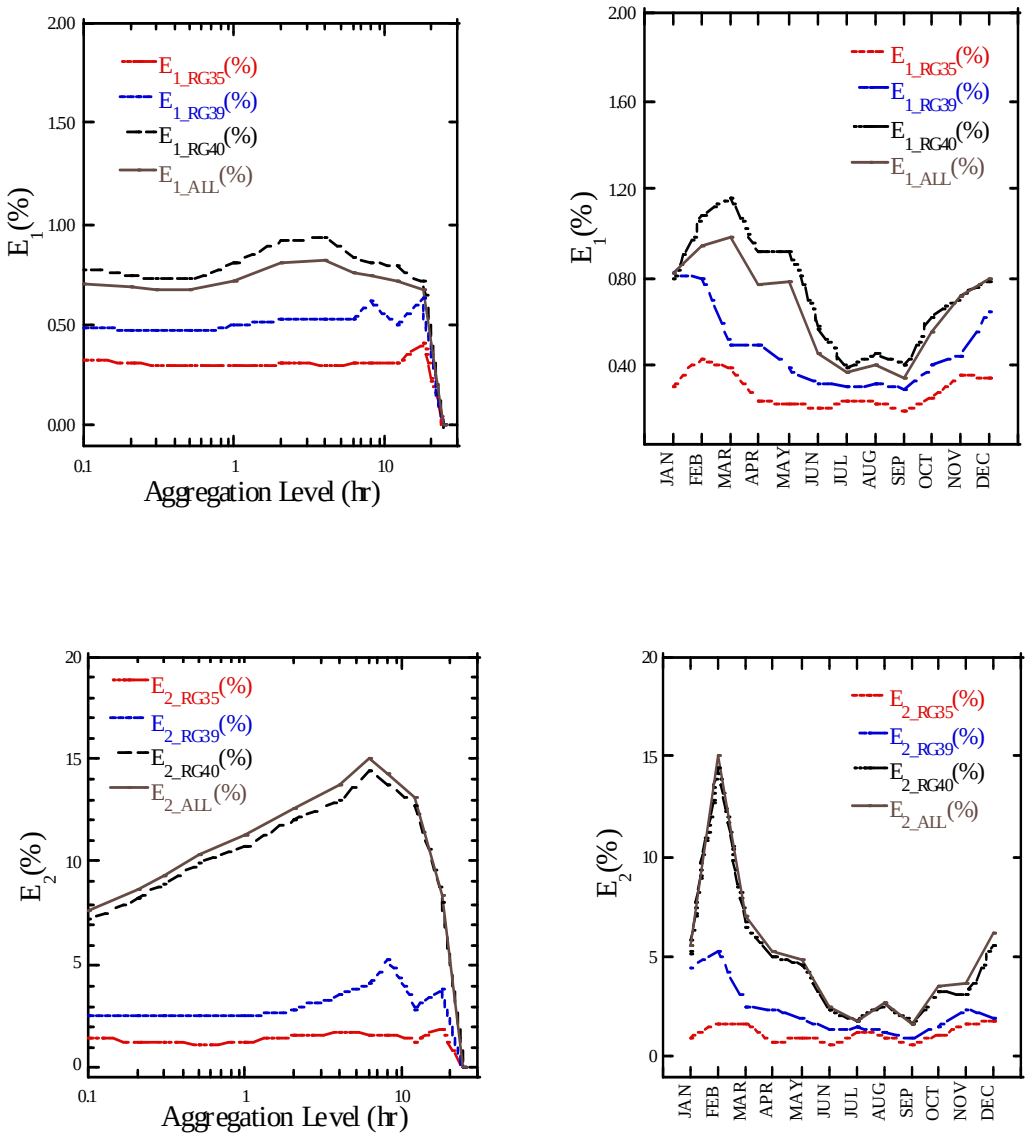

Figure 5.10 Comparison of E_1 and E_2 statistics for all months and all aggregation levels

From the comparison it is evident that the average error statistic E_1 is less than 1.2% for both criteria when all the stations are taken together. The maximum error statistic E_2 is highest (15%) for the 6-hour aggregation level and during the month of February. A

further investigation of the histograms of each of the maximum error statistic E_2 for all stations at the 6 hour aggregation level and for the month of February has been carried out. Figure 5.11 shows both histograms.

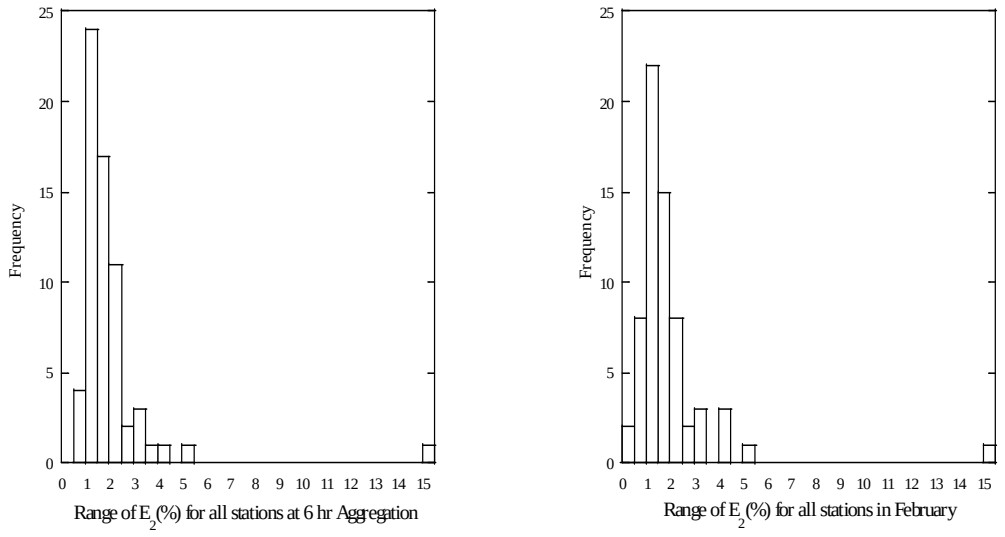

Figure 5.11 Histograms showing range of E_2 for all stations at 6 hour aggregation and in the month of February

From the histograms of Figure 5.11 it is evident that the maximum errors E_2 of 15% at the two worst case scenarios constitute only 1% of the total frequency. Hence, it is necessary to see the individual absolute errors for all data points and find their distribution in a histogram format for all stations to see whether this is an isolated phenomenon.

The individual absolute error is defined as:

$$E' = \left| \frac{P_0(h) - Pa(h)}{P_0(h)} \right| \times 100 \qquad (5.8)$$

where $P_0(h)$ and $P_a(h)$ are observed and analytical dry probabilities, respectively.

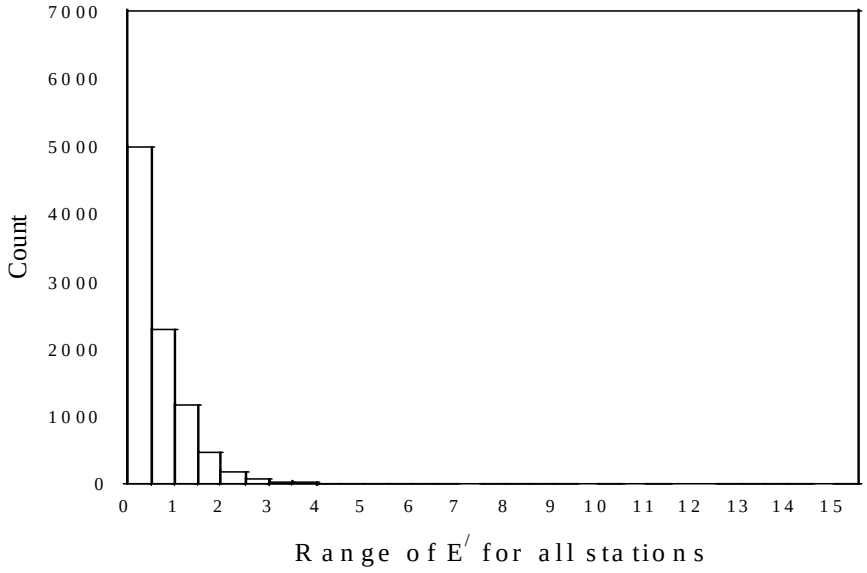

Figure 5.12 Histogram of the individual error statistic E' for all stations

The distribution of absolute error E' in Figure 5.12 is clearly exponential which indicates the maximum error statistics E_2 of 15% at the two worst case scenarios is not a driving factor in the model prediction of the dry probability. These are just some isolated cases caused by one or two isolated stations which require more observations. Since this research is limited to the rainfall data from the Bureau of Meteorology, our observation is also restricted to the same. As the bulk of the absolute errors ranges from 0 to 1.5, the model prediction of the dry probability is rather controlled by the average error statistics E_1.

Up to this stage model parameters β and η are considered fixed and a_0, a_1, b_1, a_2 and b_2 are considered variable. The E_1, E_2 and E' statistics are based on this arrangement of model parameters. Since we are looking to reduce the error statistics of the dry probability prediction as well as trying to reduce the number of parameters, three more

arrangements of parameters have been chosen. The error statistics are calculated again for the following arrangements:

- All the parameters β, η, a_0, a_1, b_1, a_2 and b_2 are taken as variables for the joint calibration of regions 35, 39 and 40.
- Parameters a_2 and b_2 are taken out of the parameter space and the remaining parameters β, η, a_0, a_1, b_1 are taken as variables for the joint calibration of regions 35, 39 and 40.
- Parameters a_1 and b_1 are taken out of the parameter space and the remaining parameters β, η, a_0, a_2, b_2 are taken as variables for the joint calibration of regions 35, 39 and 40.

Figure 5.13 shows the comparison of the E_1, E_2 error statistics for all of the abovementioned parameter arrangements.

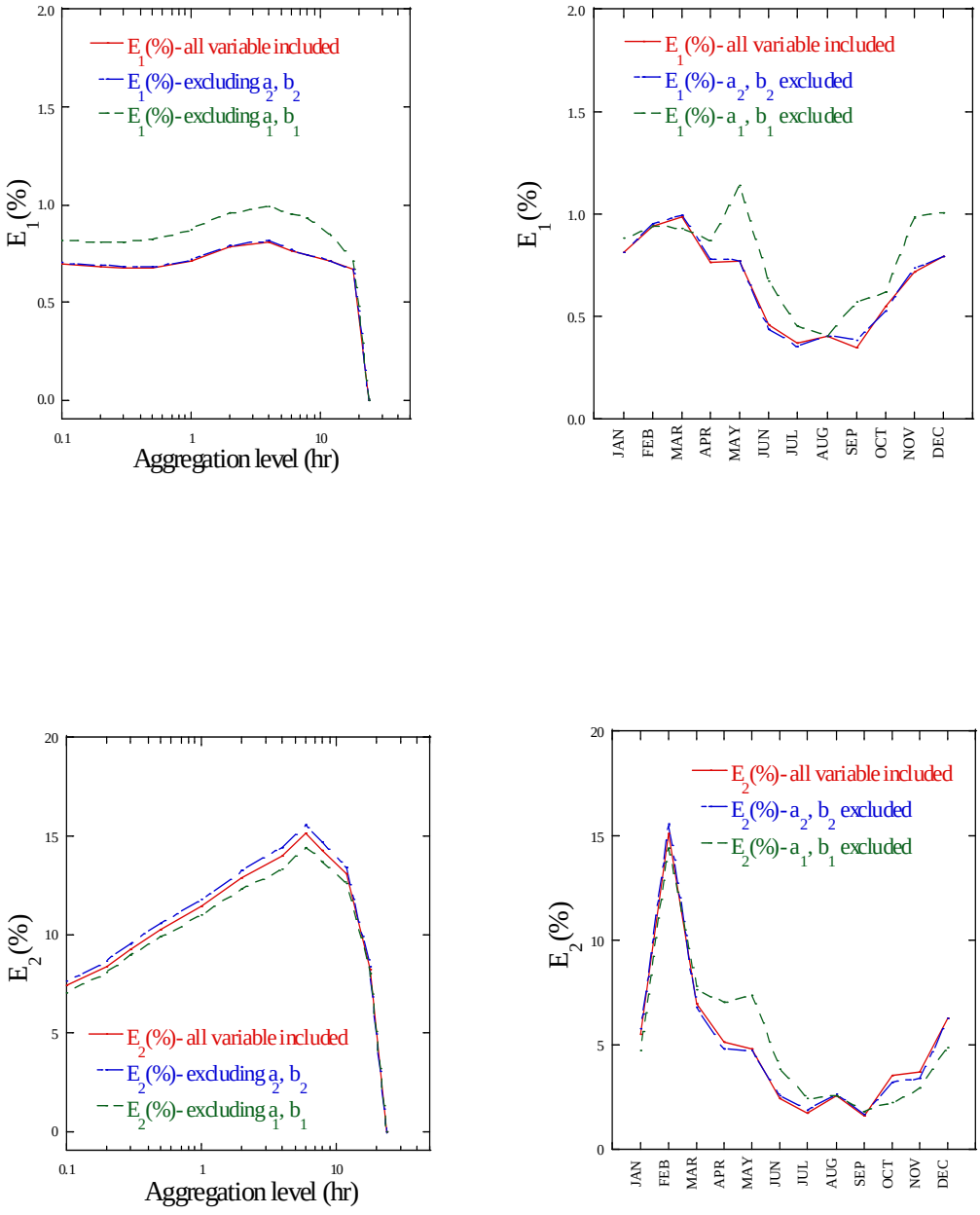

Figure 5.13 Comparison of E_1 and E_2 error statistics for all months and all aggregation levels under different sets of parameter arrangements

From the comparison it is seen that the exclusion of parameters a_2 and b_2 from the harmonic components of the parameter γ produces similar error statistics as shown earlier in Figure 5.10 for the joint calibration of all three regions. On the other hand, exclusion of

parameters a_1 and b_1 does not improve the error statistics a great deal as compared to Figure 5.10.

These findings open up the possibility of reducing of the number of parameters. Instead of carrying out the joint calibration with parameters β and η as fixed and a_0, a_1, b_1, a_2, b_2 as regional variables to calculate parameter γ, we can now rely on the joint calibration with parameters β, η, a_0, a_1 and b_1 as variables with similar results. In order to improve the error statistics and reduce the number of parameters further, the isolated station causing the maximum error of 15% in both Figure 5.10 and Figure 5.13 is taken out of the joint calibration. The error statistics are calculated again for the following arrangements:

- All the parameters β, η, a_0, a_1, b_1, a_2 and b_2 are taken as variables for the joint calibration of regions 35, 39 and 40 while the isolated station causing maximum error of 15% is omitted. This is a 7 parameter arrangement.
- Parameters a_2 and b_2 are taken out of the parameter space and the remaining five parameters β, η, a_0, a_1, b_1 are taken as variables for the joint calibration of regions 35, 39 and 40 while the isolated station causing maximum error of 15% is omitted.
- Parameters a_1, b_1, a_2 and b_2 are taken out of the parameter space and the remaining three parameters β, η *and* a_0 are taken as variables for the joint calibration of regions 35, 39 and 40 while the isolated station causing maximum error of 15% is omitted.

Figure 5.14 shows the comparison of the E_1, E_2 error statistics for all of the abovementioned parameter arrangements.

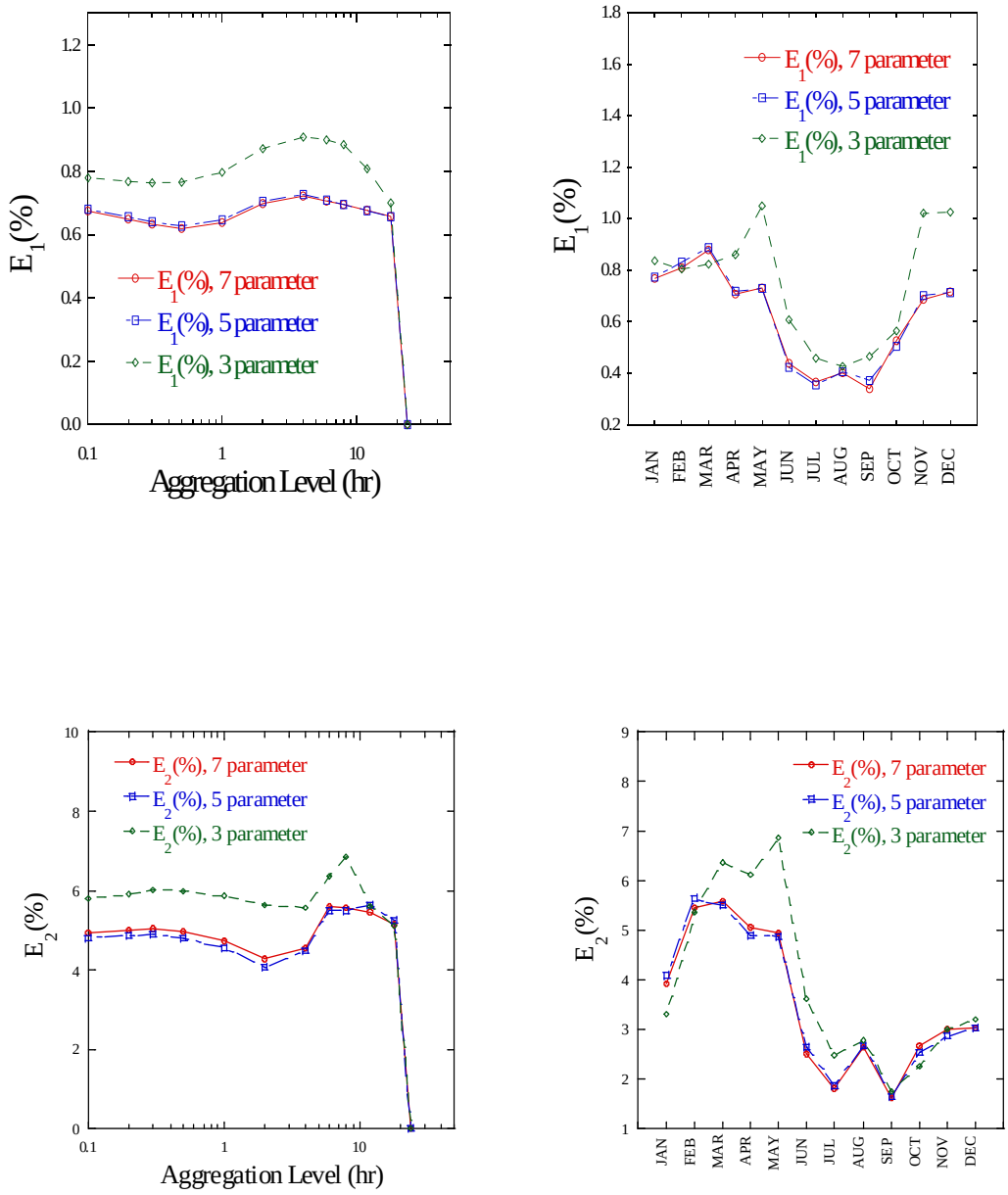

Figure 5.14 **Comparison of E_1 and E_2 error statistics for all months and all aggregation levels under different sets of parameter arrangements while the isolated station causing maximum error is omitted**

It is now evident from Figure 5.14 that the use of three parameters namely, β, η *and* a_0 gives a maximum error of 1.1% and 7% for E_1 and E_2 respectively. This is a significant improvement in the error statistics as well as a significant reduction of the number of parameters to calibrate.

The order of harmonics for parameter γ was investigated for first (7 parameter), second (5 parameter) and zero (3 parameter) orders for the Queensland regions by examining the absolute percentage error defined in equation 5.8. Figure 5.15 shows that there is no significant difference in the error statistics for these three arrangements with an absolute error being less than 10% for a non-exceedance probability of 99.9%.

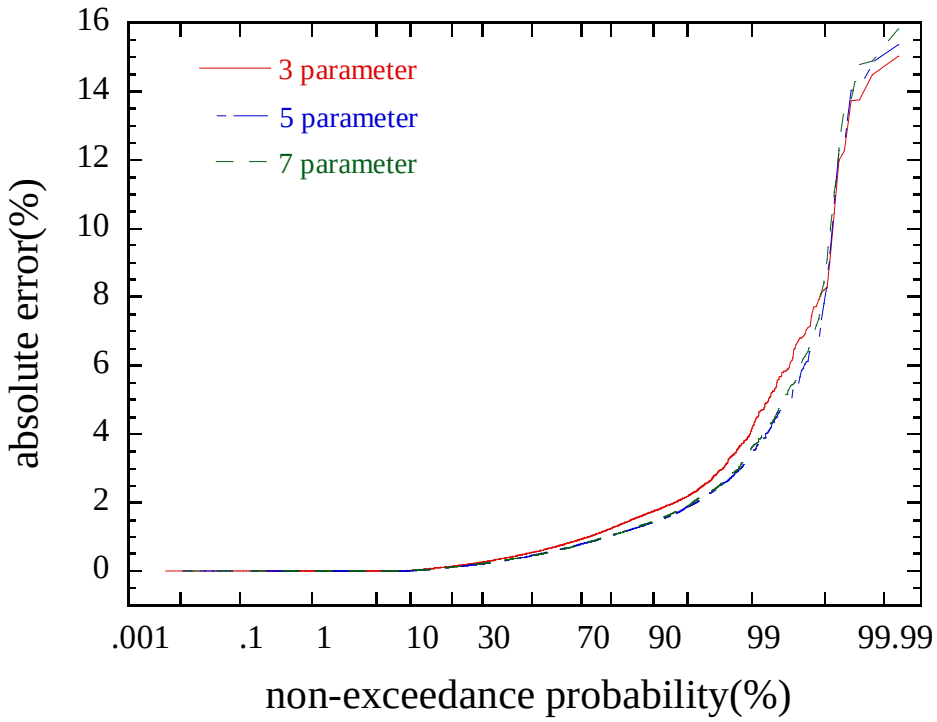

Figure 5.15 Absolute error plots for different order of harmonics of parameter γ

As it is evident from Figure 5.15 that using 3 parameters (i.e. zero order harmonics for parameter γ), the seasonal effect is absent in parameter γ. Hence γ will simply refer to a_0 hereafter. The distribution of the absolute error with the zero order harmonics is shown in Figure 5.16.

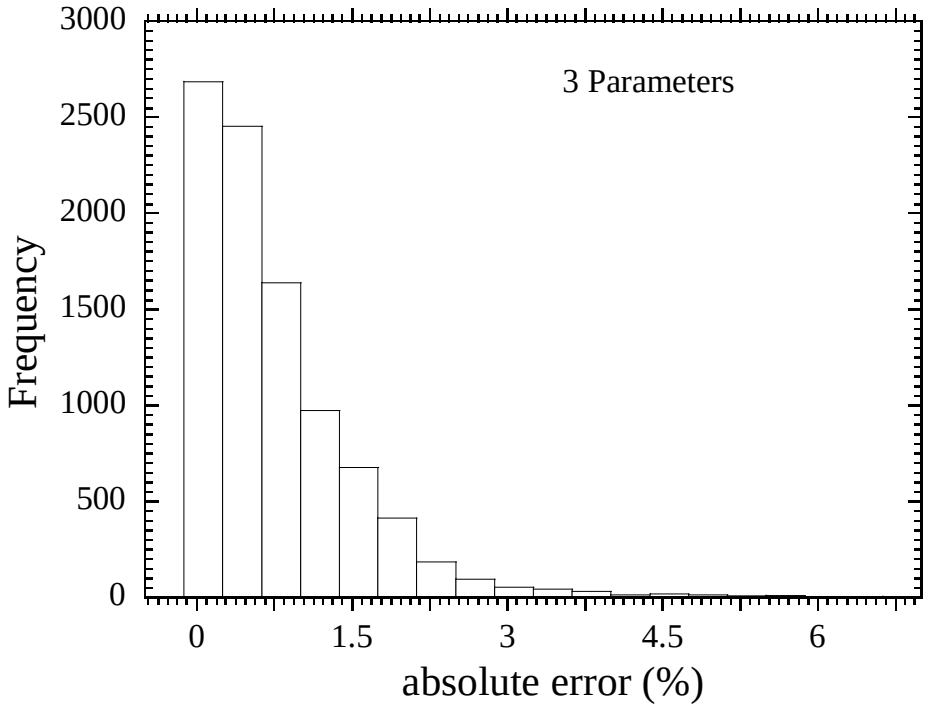

Figure 5.16 Absolute error distributions for zero order harmonics of parameter γ

The three parameter arrangement for the joint calibration of regions 35, 39 and 40 in the State of Queensland was put in comparison with an Australia wide joint parameter calibration performed by Gyasi-Agyei and Mahbub (2007). The traditional Multinormal Parameter Approximation and the Monte Carlo – based approximation algorithm known as Metropolis Algorithm (Kuczera and Parent, 1998) were used for calibration. The Duan et al. (1992) shuffled complex evolution of the global probabilistic search option of the NLFIT Bayesian non-linear regression software (Kuzcera, 1994) was employed to refine the posterior distribution of the multinormal approximation. The result of the calibrated parameters for QLD regions is given in Table 5.5.

Table 5.5 Statistics of calibrated parameters by multinormal approximation and metropolis algorithm for QLD regions

Parameter	Multinormal Approximation		Metropolis	
	Mean	Standard Deviation	Mean	Standard Deviation
β	0.200	0.011	0.166	0.010
η	0.510	0.013	0.473	0.012
γ	0.090	0.002	0.082	0.003

For the purpose of comparison, the Australia wide parameter values are reproduced from Gyasi-Agyei and Mahbub (2007) in Table 5.6.

Table 5.6 Statistics of calibrated parameters by multinormal approximation and metropolis algorithm for Australia wide regions

Parameter	Multinormal Approximation		Metropolis	
	Mean	Standard Deviation	Mean	Standard Deviation
β	0.273	0.012	0.278	0.013
η	0.746	0.019	0.756	0.022
γ	0.092	0.002	0.093	0.002

From Tables 5.5 and 5.6 it is evident that there is not much difference in the parameter calibration by the multinormal approximation and the metropolis algorithm. To see if there is any significant change in their posterior moment distributions, histogram analyses were carried out for Queensland regions and compared with those of Australia wide results from Gyasi-Agyei and Mahbub (2007). Figure 5.17 shows the histograms for both Queensland regions and the Australia wide stations.

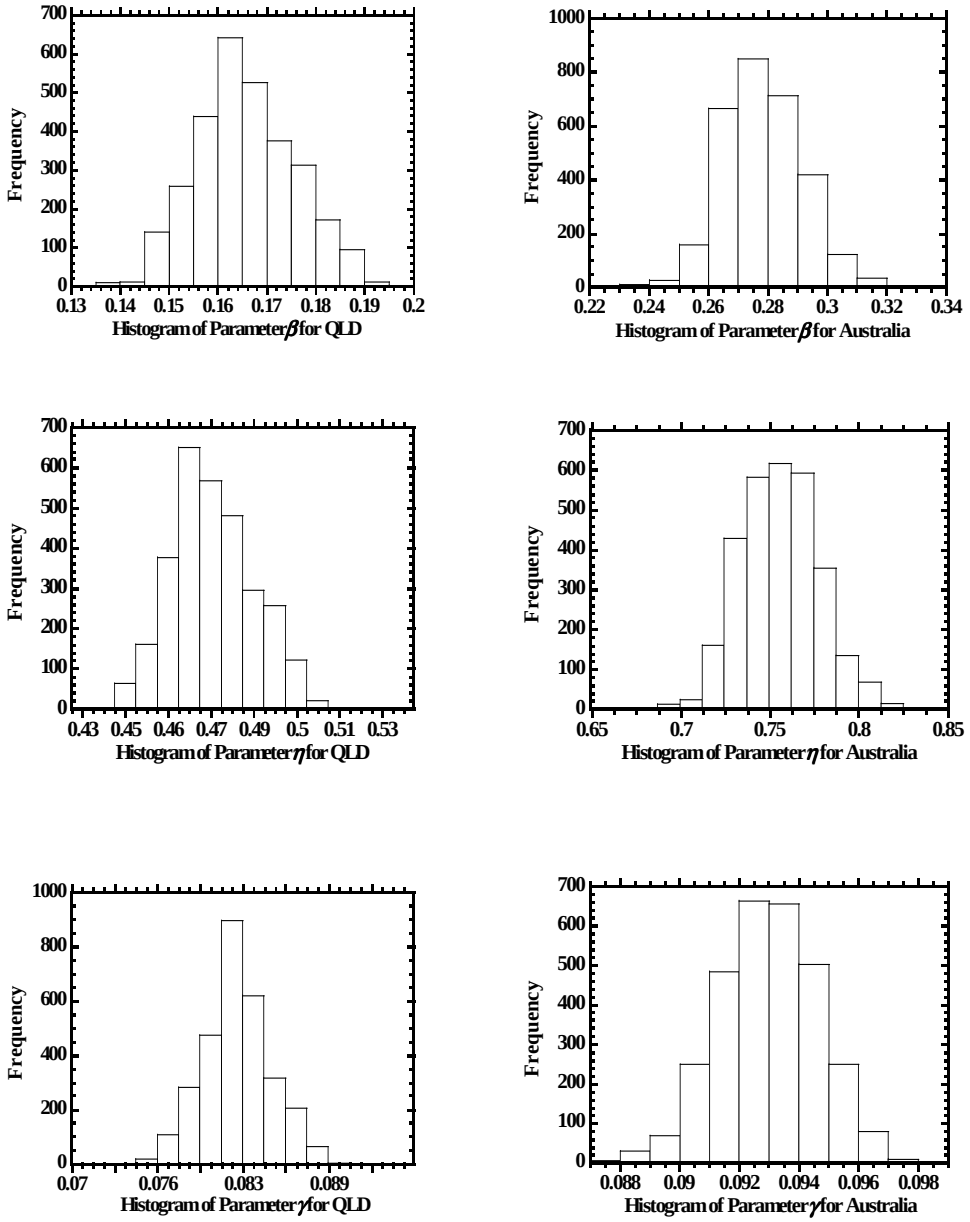

Figure 5.17 Histograms of the posterior mean values for the QLD regions

The posterior distribution of mean is nearly normal for both Queensland regions and the whole of Australia as shown in Figure 5.17. This is the reason why there were little differences between the means of calibrated parameters in Tables 5.5 and 5.6.

The covariance between parameters $\eta-\beta, \eta-\gamma$ and $\beta-\gamma$ were also analysed for Queensland regions and compared to that of Australia wide results from Gyasi-Agyei and Mahbub (2007). Figure 5.18 shows the correlations between the above parameters.

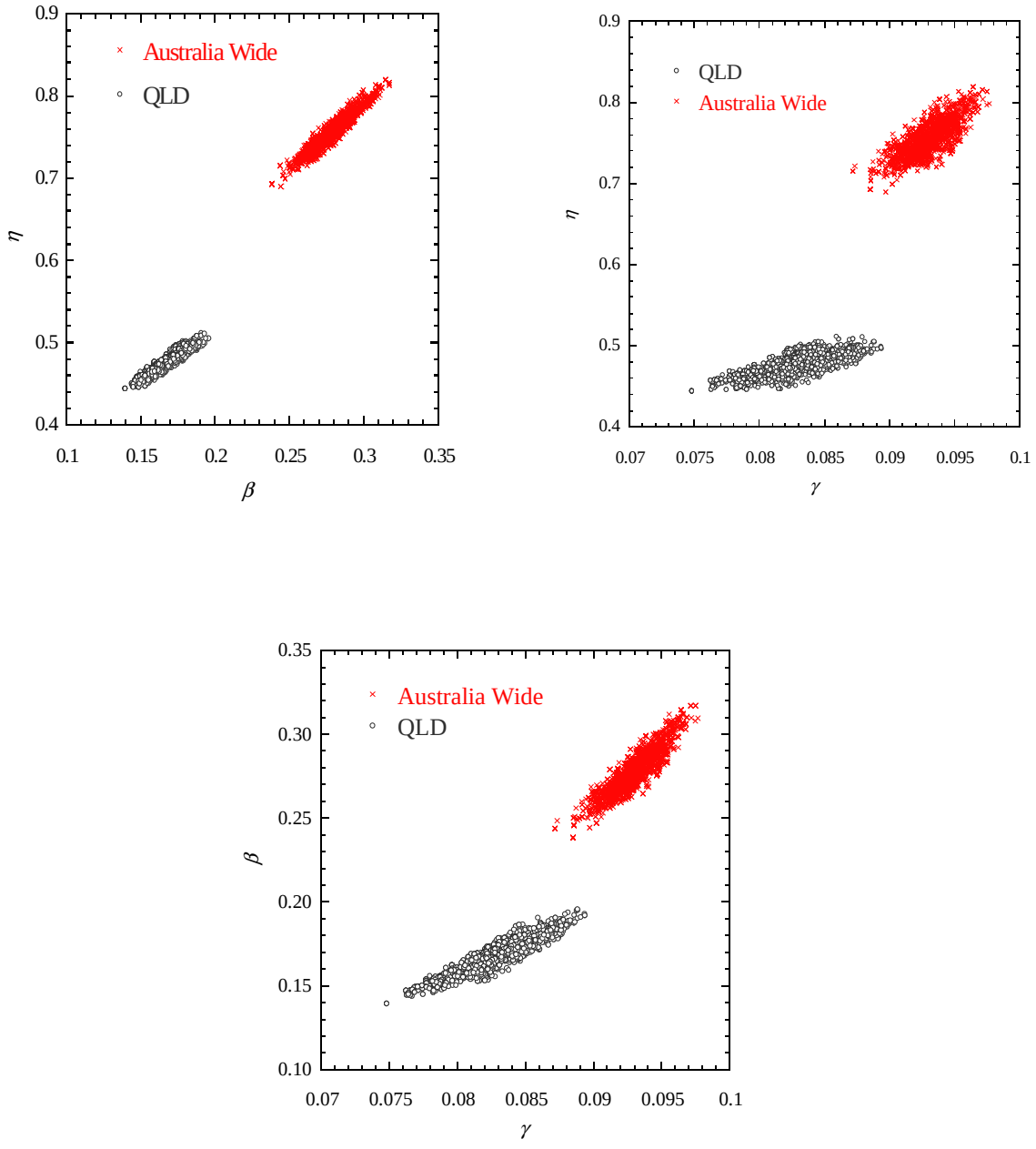

Figure 5.18 Covariance between parameters $\eta-\beta, \eta-\gamma$ and $\beta-\gamma$

Even though strong correlations are evident in Figure 5.18 individually for both Queensland regions and the Australia wide dataset parameters, the Queensland region fell outside the Australia wide dataset. This suggests that the model parameters have to be regionalised. Therefore, the parameters of Table 5.5 will be used as the binary chain parameters for the Queensland region of Australia.

5.2 Variance Relationship Parameter Estimation for Queensland

This research has adopted the empirical scaling relationship between the daily variance, $\sigma^2_{\bar{Y}}(24)$ and the fine timescale variance $\sigma^2_{\bar{Y}_k}(h)$ suggested by Gyasi-Agyei (1999). Gyasi-Agyei and Mahbub (2007) established the empirical scaling relationship between daily variance and the fine scale variance (6-minute) for selected sites throughout Australia as:

$$\sigma^2_{\bar{Y}}(0.1) = 0.000272\sigma^2_{\bar{Y}}(24)^{0.913} \qquad (5.9)$$

No apparent seasonality was observed in this data set. To find out whether the data-lengths have effects on similar variance relationship for Queensland, the pluviograph stations were ranked according to their monthly data lengths. The variance relationships were then separately established between the top and bottom 50% of the data to see if there is any change in the spread of the station-months from the fitted power curves. The following two figures show the separately established variance relationship for the ranked data in Queensland.

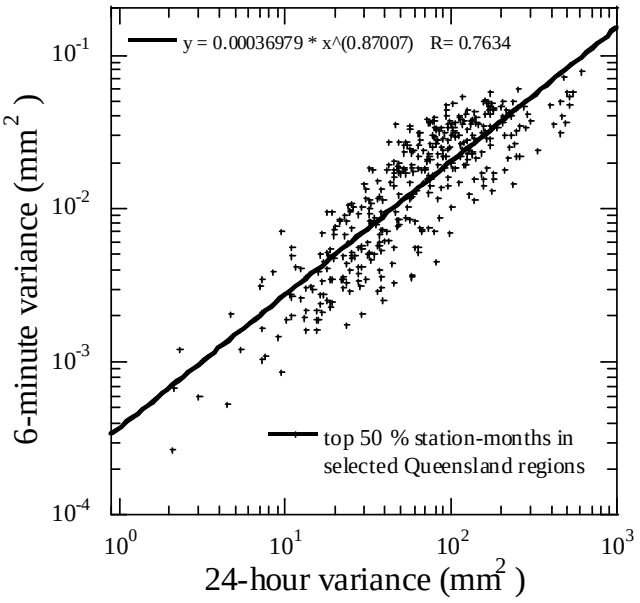

Figure 5.19 Variance relationship for the top 50% ranked station-months in Queensland

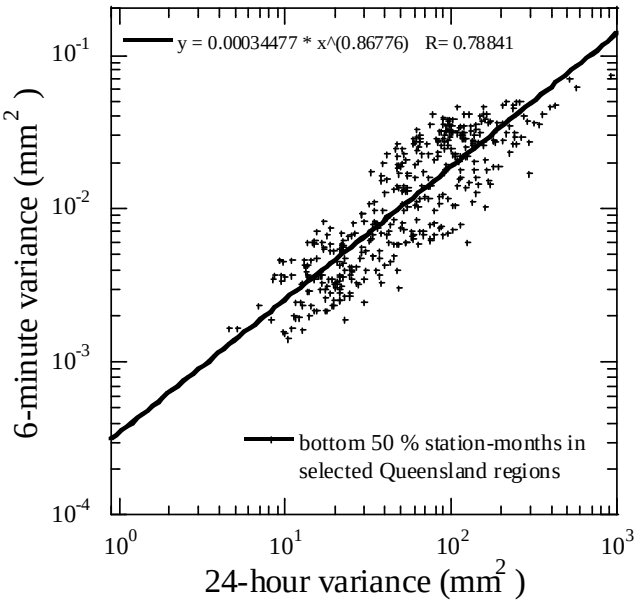

Figure 5.20 Variance relationship for the bottom 50% ranked station-months in Queensland

It is evident from Figures 5.19 and 5.20 that spread of the station-months from the fitted power curves does not really vary with the data lengths. Hence it was decided that the data lengths of different months did not influence this variance relationship.

The scaling relationship for Queensland regions was compared to that of Australia wide. Figure 5.21 shows that there is a subtle difference between the regionalised fit and the Australia wide fit power curves. Both display significant scatter of many station months from the fitted power curves. To minimise the scatters, seasonality was investigated for the relationship between 6-minute and daily variances in the selected three regions (region 35, 39 and 40) of Queensland.

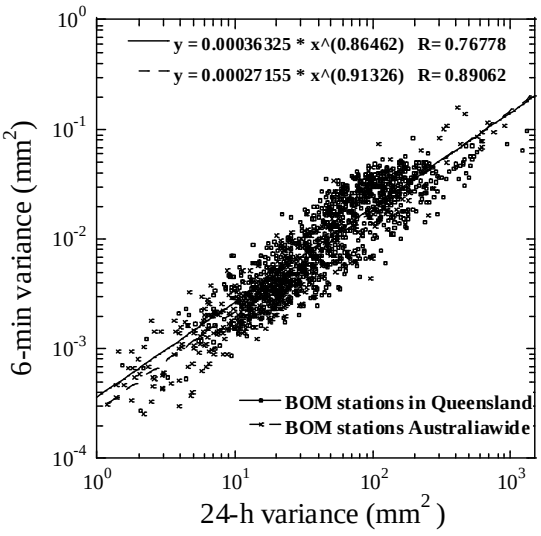

Figure 5.21 Fitted power curves between six-minute and 24-hour variances in Queensland and Australia wide

A distinctive pattern of scaled 6-minute variance to 24-hour variance with the 24-hour variance was observed for the regions 35, 39 and 40 in Queensland. Figure 5.22 shows this pattern for both summer and winter seasons.

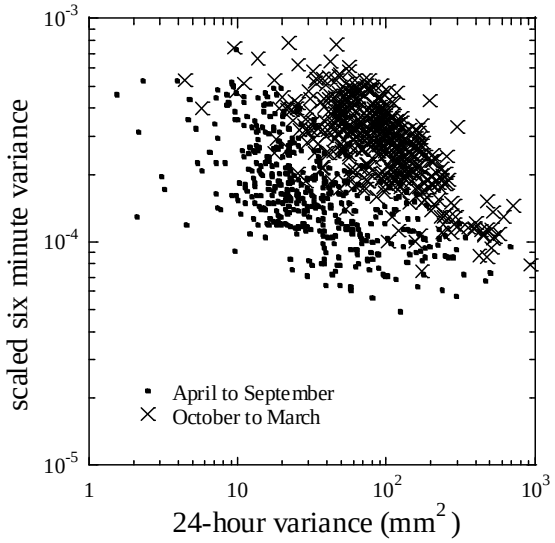

Figure 5.22 Comparison of seasonal patterns of scaled 6-minute variance with 24-hour variance in the selected Queensland regions

The distinctive pattern of Fig. 5.22 does not fully reveal how the individual months are affecting the variance relationship of Figure 5.21. Hence an investigation of 6-minute variance of each month in the selected Queensland regions is performed. The following figures show the monthly variance patterns for Queensland.

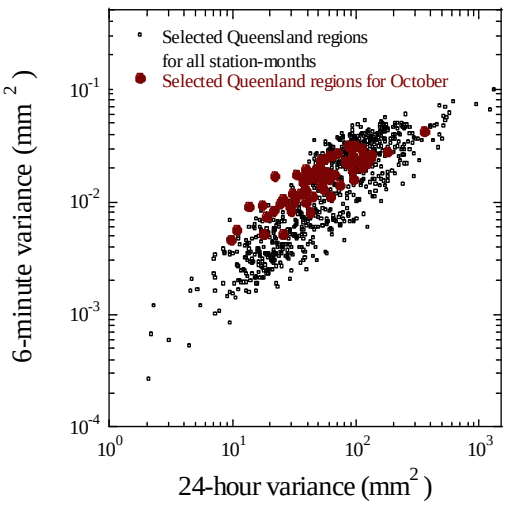

Figure 5.23 Comparison of QLD October patterns with whole QLD station months

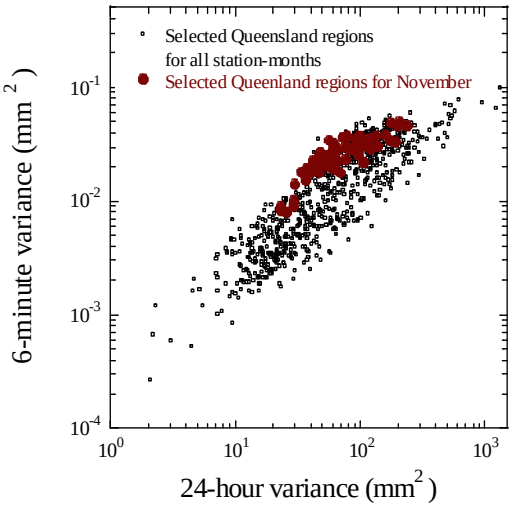

Figure 5.24 Comparison of QLD November patterns with whole QLD station months

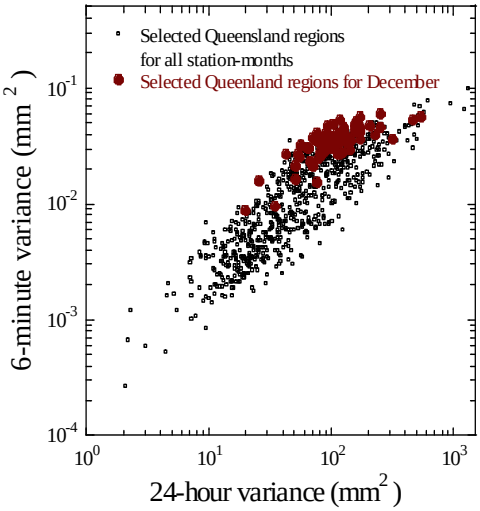

Figure 5.25 Comparison of QLD December patterns with whole QLD station months

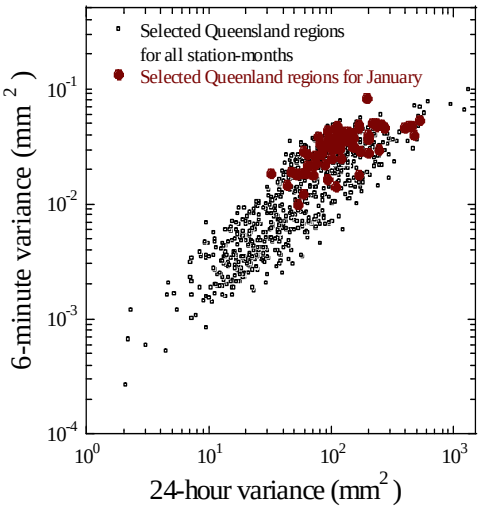

Figure 5.26 Comparison of QLD January patterns with whole QLD station months

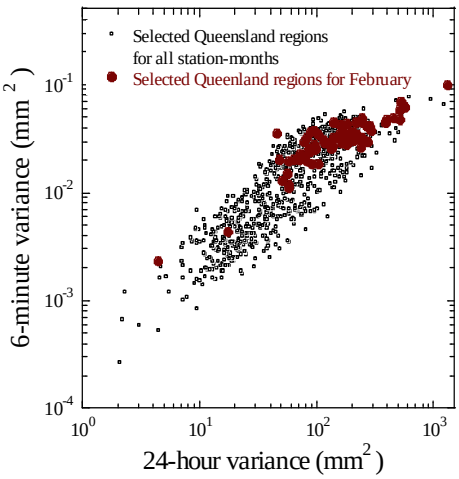

Figure 5.27 Comparison of QLD February patterns with whole QLD station months

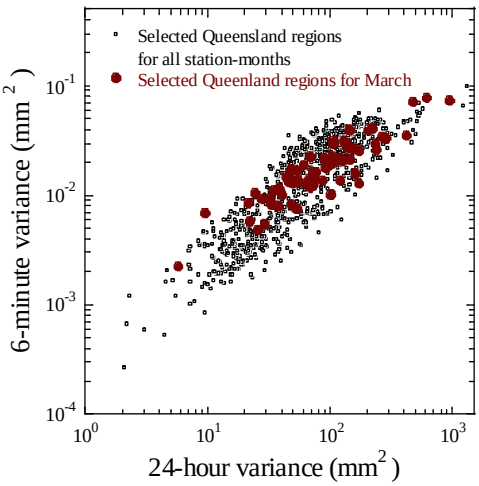

Figure 5.28 Comparison of QLD March patterns with whole QLD station months

Figures 5.23 to 5.28 show that there exists a specific downward shift pattern in the variance relationship between 6-minute and 24-hour for months from October to March

in Queensland regions. The following figures show the variance patterns for months from April to September in the Queensland regions.

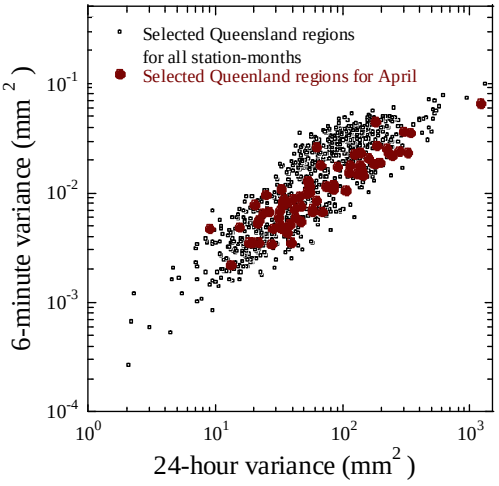

Figure 5.29 Comparison of QLD April patterns with whole QLD station months

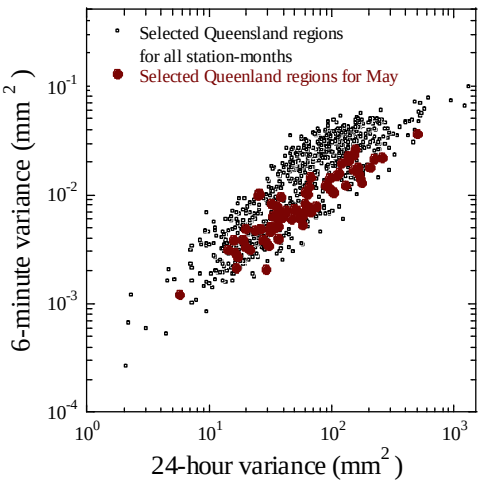

Figure 5.30 Comparison of QLD May patterns with whole QLD station months

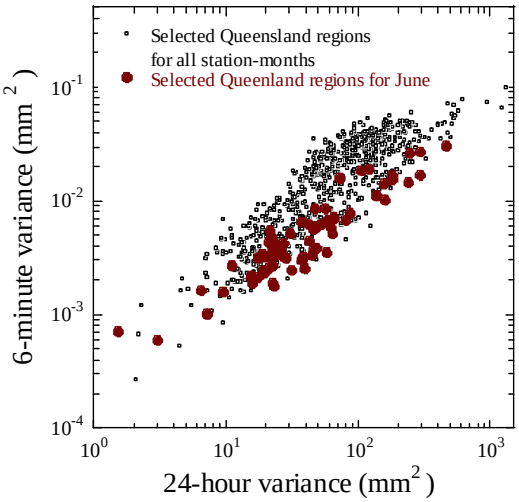

Figure 5.31 Comparison of QLD June patterns with whole QLD station months

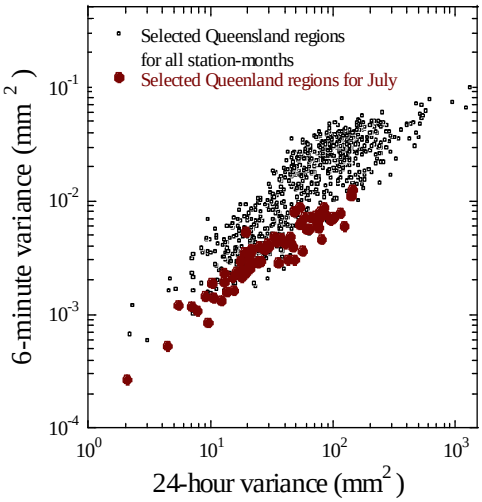

Figure 5.32 Comparison of QLD July patterns with whole QLD station months

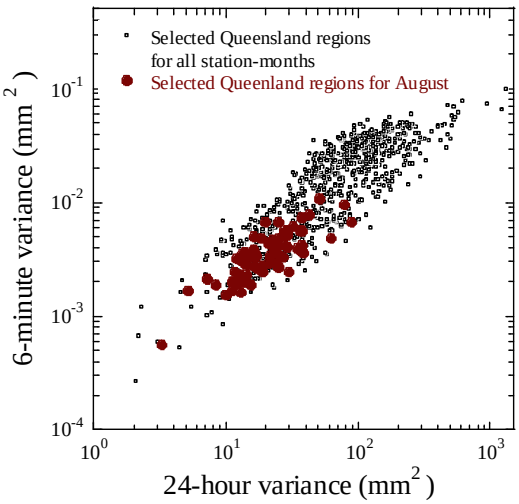

Figure 5.33 **Comparison of QLD August patterns with whole QLD station months**

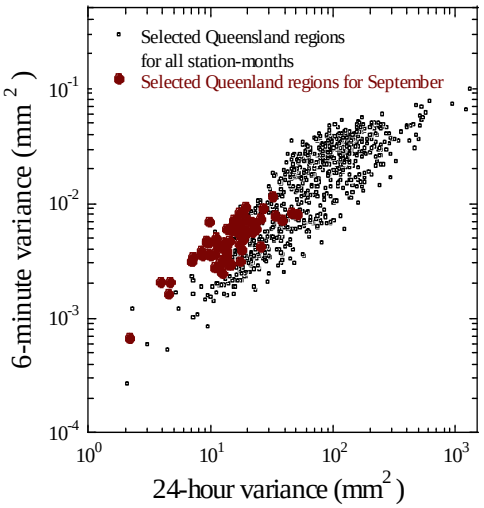

Figure 5.34 **Comparison of QLD September patterns with whole QLD station months**

Figures 5.29 to 5.34 show that the downward shift in the variance relationship continues in the winter months and at the end of the season it starts reversing back to the top (Figures 5.34 and 5.23). These findings suggested that the use of a single variance relationship power curve between the 6-minute and 24-hour variance for the whole of Queensland might cause problems in simulating the fine scale variances in Queensland. Hence, the seasonal approach associated with the variance relationship proposed by Mahbub et al. (2007) is adopted in this research. This seasonal approach has been described in section 4.1.6 of the previous chapter. As per equations 4.21 to 4.23, this approach incorporates parameter a comprising of coefficients a_0, a_1 and a_2 and parameter b comprising of coefficients b_0, b_1 and b_2 into the disaggregation model. Parameter a is defined as the scaling constant and parameter b is defined as the scaling exponent in equation 4.21. The global optimisation search strategy (Duan et al., 1992) of the Bayesian Non-Linear regression software NLFIT (Kuczera, 1994) is used again to

determine these seasonal coefficient values for the selected Queensland regions. Table 5.7 shows the optimised coefficients for Queensland regions determined by NLFIT.

Table 5.7 Seasonal parameter set for Queensland

Seasonal coefficients	Optimised values for Queensland
a_0	0.001329
a_1	0.0007735
a_2	0.0004887
b_0	0.5943
b_1	-0.007314
b_2	-0.06866

Figures 5.35 and 5.36 show the seasonal variations of the scaling constant a and the exponent b for the Queensland regions.

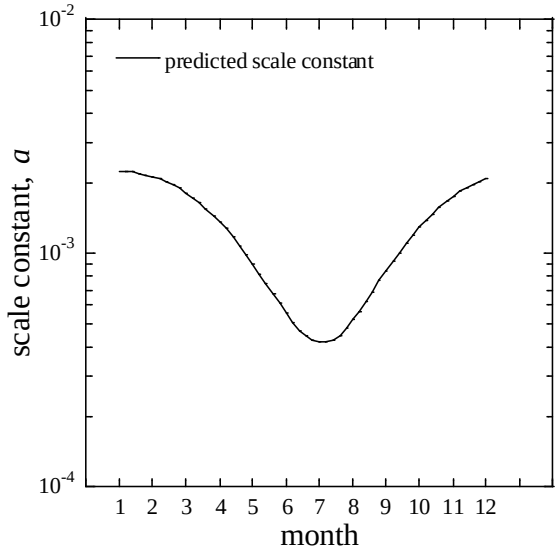

Figure 5.35 Seasonal variation of the scaling constant a for Queensland

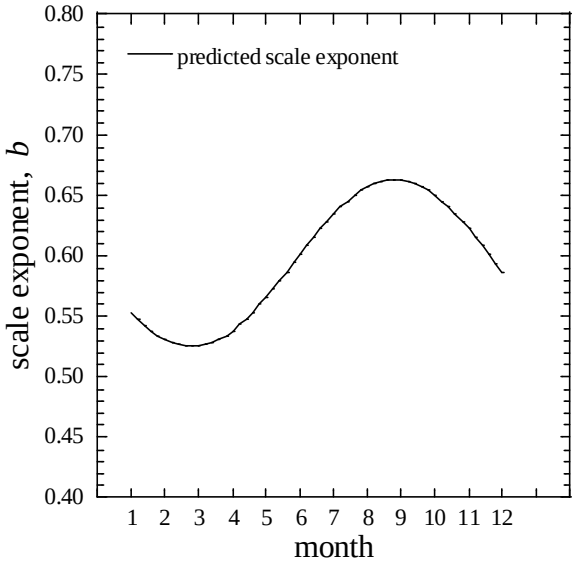

Figure 5.36 Seasonal variation of the scaling exponent b for Queensland

The scaling constant a and the exponent b are determined by six seasonal coefficients: a_0, a_1, a_2, b_0, b_1 and b_2 according to equations 4.22 and 4.23.

The following figure shows the monthly variation of the downscaling relationship of equation 4.21 for four equally spaced months in Queensland after the inclusion of seasonality.

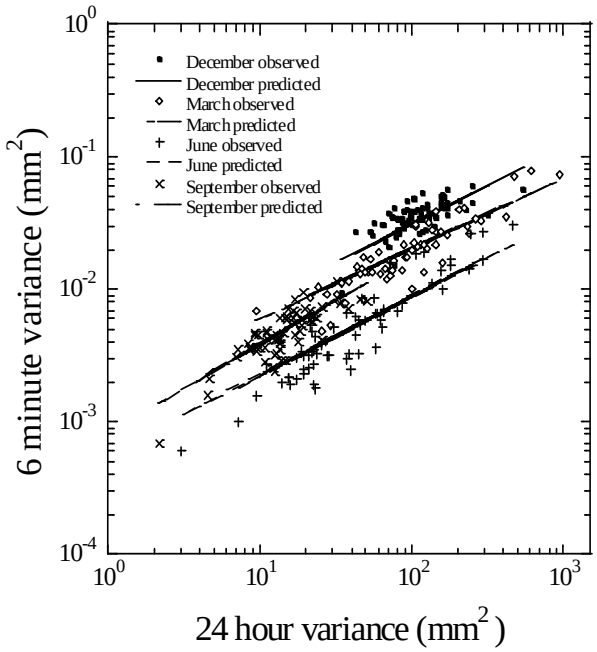

Figure 5.37 Seasonal variation of the variance relationship for Queensland

From Figure 5.37 it is more evident that the variance relationship for the 6 minute and 24 hour variances for the Queensland regions follows a clear seasonal pattern. This research will investigate the simulation results of variance by the disaggregation model using both seasonal and non-seasonal approach in its variance relationship. For the non-seasonal approach, either the Australia wide parameter values or the Queensland parameter values of Figure 5.21 can be used for Queensland. For the seasonal approach, values from Table 5.7 will be used for Queensland.

5.3 Capping Parameter Estimation for Queensland

The capping technique, i.e. placing an upper limit on the simulated timescale rainfall depths, is required to tackle the possible overestimation of the simulated timescale rainfall depth during the disaggregation process. The capping value is allowed to vary depending on the daily depth between the rainfall stations. According to equation 4.20 of

the previous chapter, the capping parameters *A* and *B* are to be estimated for each stations in Queensland regions. Figures 5.38 to 5.40 represent this capping relationship for 6-minute timescale at Dingo, Rockhampton and Brisbane which are situated in regions 35, 39 and 40 in Queensland, respectively. The parameter values are calculated from the fitted capping equations of the corresponding stations.

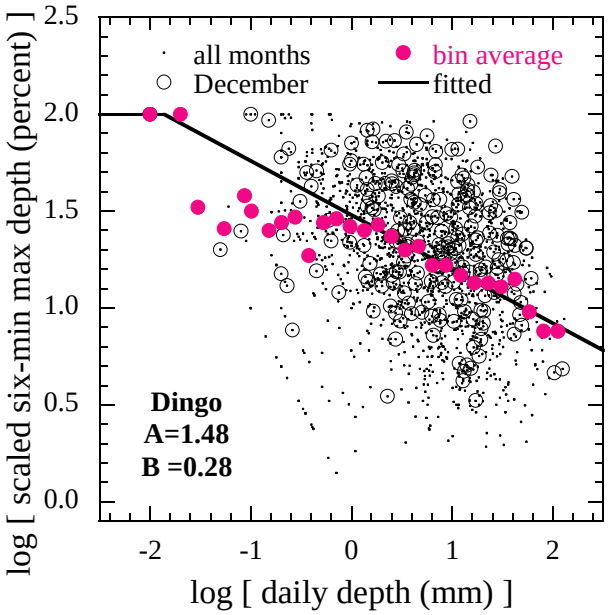

Figure 5.38 Fitted Capping relationship at Dingo, Queensland Australia

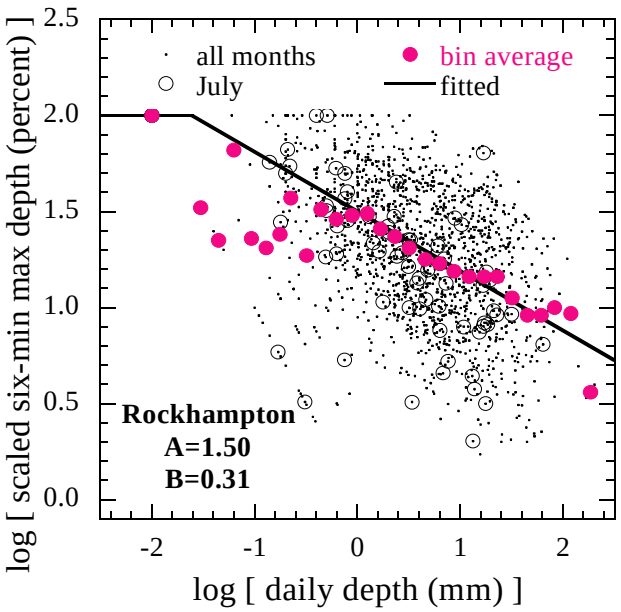

Figure 5.39 Fitted Capping relationship at Rockhampton, Queensland Australia

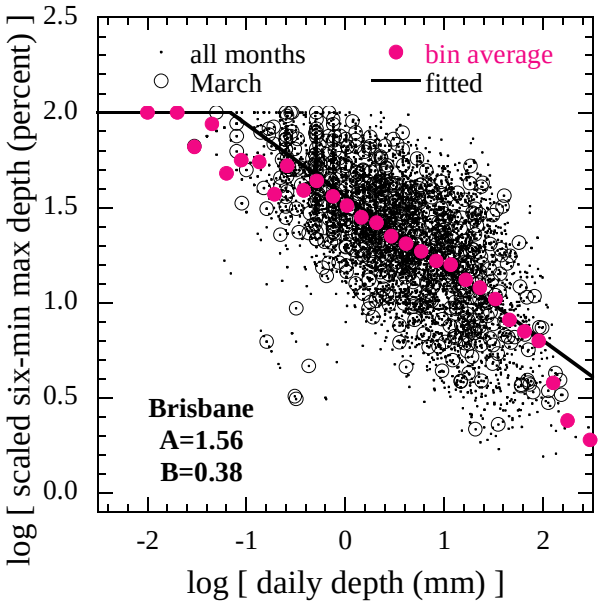

Figure 5.40 Fitted Capping relationship at Brisbane, Queensland Australia

No seasonality is observed in this scaled 6-minute depth versus daily depth relationship as the months (for example, December in Fig. 5.38) are randomly distributed. The logarithmic transformed daily depths were ranked and divided into 30 groups (bins) of

equal intervals and the mean and the standard deviation of the logarithmic transformed scaled 6-minute maximum depths within each bin were estimated. During the simulation for Queensland regions, each wet day's capping value is estimated by randomly drawing a number from a normal distribution of the binned log[scaled 6-minute maximum depth (%)] points with the mean estimated by equation 4.20 and the standard deviation set to the maximum of the bins' standard deviation. The following figures show that normal distribution can be approximately estimated for the bins in all three regions:

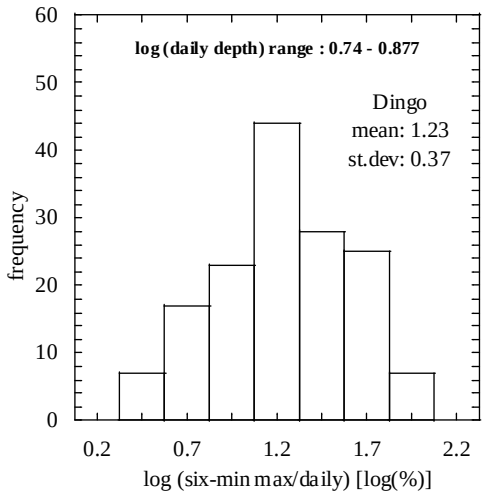

Figure 5.41 Histogram of log[scaled 6-min (%)] at Dingo; bin range: 0.74 - 0.877

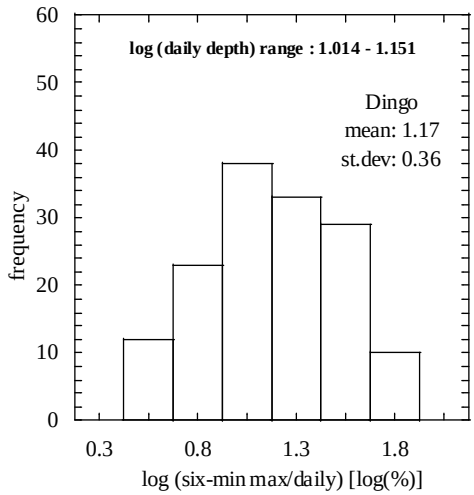

Figure 5.42 Histogram of log[scaled 6-min(%)]at Dingo; bin range: 1.014 - 1.151

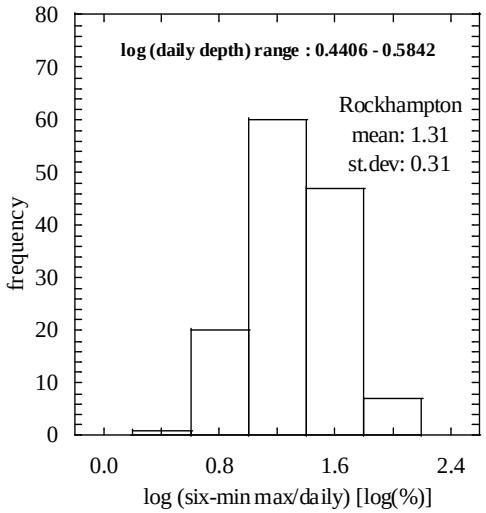

Figure 5.43 Histogram of log[scaled 6-min (%)] at Rockhampton; bin range: 0.44 - 0.58

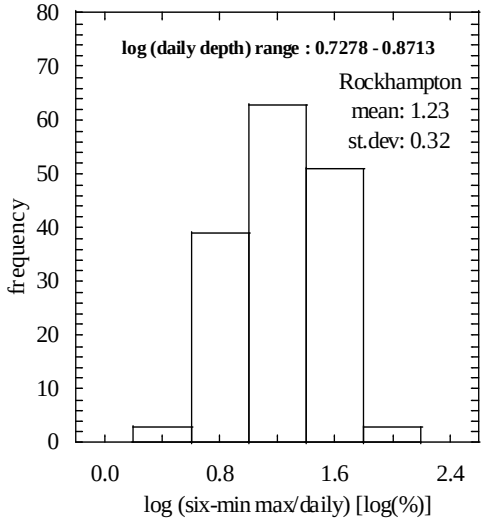

Figure 5.44 Histogram of log[scaled 6- min(%)] at Rockhampton; bin range: 0.73 – 0.87

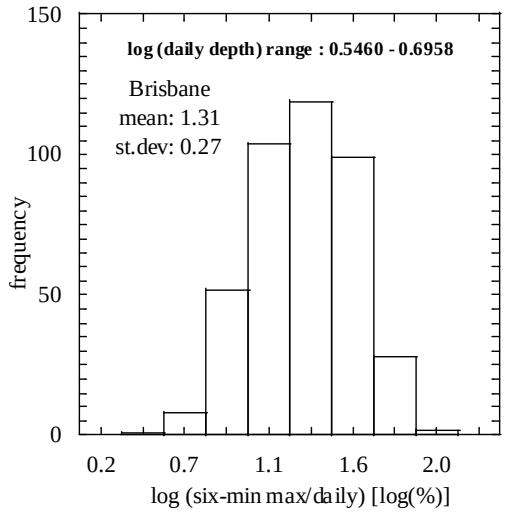

Figure 5.45 Histogram of log[scaled 6-min (%)] at Brisbane; bin range: 0.55 - 0.70

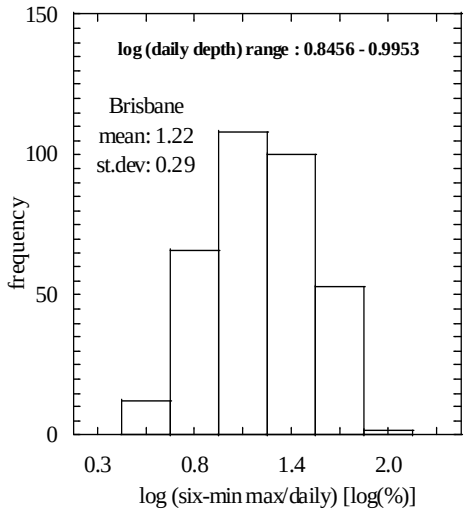

Figure 5.46 Histogram of log[scaled 6- min(%)] at Brisbane; bin range: 0.85 – 0.99

Figures 5.41 to 5.46 validate the approximation of normal distribution for the binned log[scaled 6-min(%)] points in all three sites at the selected regions of Queensland. From this site-specific normal distribution, the maximum capping value for the simulated rainfall depth will be randomly chosen by the model. Table 5.8 shows the site- specific

capping parameter values for all 65 selected stations in the three regions (35, 39 and 40) in Queensland.

Table 5.8 Capping parameter values for selected Queensland stations

BOM Pluviometer-station ID	Parameter A	Parameter B	Standard Deviation	BOM Pluviometer-station ID	Parameter A	Parameter B	Standard Deviation
35000	1.52	0.33	0.37	40112	1.40	0.23	0.40
35025	1.48	0.28	0.37	40126	1.64	0.42	0.32
35029	1.39	0.23	0.42	40135	1.46	0.30	0.49
35059	1.43	0.27	0.39	40152	1.48	0.29	0.40
35065	1.46	0.26	0.38	40160	1.59	0.42	0.35
35069	1.58	0.36	0.36	40178	1.43	0.29	0.39
35070	1.44	0.29	0.40	40180	1.39	0.27	0.45
35090	1.60	0.45	0.45	40189	1.26	0.37	0.37
35104	1.50	0.29	0.34	40192	1.50	0.46	0.31
35147	1.55	0.33	0.38	40197	1.54	0.40	0.38
35267	1.50	0.32	0.36	40214	1.56	0.38	0.33
39006	1.40	0.25	0.36	40222	1.53	0.39	0.38
39069	1.38	0.26	0.40	40223	1.59	0.44	0.36
39070	1.44	0.30	0.37	40241	1.56	0.44	0.35
39083	1.42	0.23	0.37	40265	1.51	0.33	0.33
39090	1.57	0.42	0.35	40282	1.56	0.40	0.31
39123	1.55	0.32	0.33	40308	1.48	0.38	0.36
39128	1.61	0.38	0.32	40312	1.50	0.37	0.37
39140	1.46	0.33	0.38	40406	1.54	0.38	0.35
39297	1.50	0.26	0.41	40458	1.47	0.32	0.35
39303	1.61	0.36	0.37	40459	1.43	0.32	0.34
39314	1.71	0.43	0.31	40460	1.50	0.34	0.35
40004	1.54	0.38	0.50	40461	1.41	0.32	0.40
40014	1.51	0.38	0.41	40469	1.42	0.30	0.38

BOM Pluviometer-station ID (contd.)	Parameter A (contd.)	Parameter B (contd.)	Standard Deviation (contd.)	BOM Pluviometer-station ID (contd.)	Parameter A (contd.)	Parameter B (contd.)	Standard Deviation (contd.)
40019	1.31	0.20	0.47	40496	1.52	0.34	0.32
40059	1.54	0.37	0.31	40537	1.56	0.30	0.29
40062	1.44	0.39	0.35	40584	1.53	0.37	0.31
40063	1.49	0.38	0.37	40606	1.57	0.40	0.37
40082	1.46	0.29	0.36	40609	1.70	0.46	0.33
40093	1.56	0.41	0.37	40659	1.49	0.37	0.36
40094	1.48	0.39	0.40	40677	1.46	0.26	0.38
40106	1.57	0.43	0.37	40715	1.56	0.38	0.33
40111	1.50	0.34	0.39				

Table 5.8 enables the use of the three parameter stochastic disaggregation model at all the selected stations around Queensland for the capping purposes.

5.4 Estimation of Autocorrelation for Queensland

For fine timescale simulation purposes, Gyasi-Agyei (1999) and Gyasi-Agyei and Mahbub (2007) suggested the use of a fixed autocorrelation, $\rho_{\bar{Y}}(h)$ for Australia wide due to the presence of strong correlation between higher lag autocorrelation and lag-1 autocorrelation. The Australia wide 6-minute lag-1 autocorrelation is suggested as 0.71 by Gyasi-Agyei and Mahbub (2007). For Queensland, the monthly mean values of the 6 minute lag-1 autocorrelation did not show any seasonality (Figure 5.47). A constant value of 0.73 was observed for Queensland for almost all the year round. This research will investigate the model performance using both the Australia wide autocorrelation value of 0.71 and the regionalised Queensland value of 0.73 for 6-minute timescale simulation.

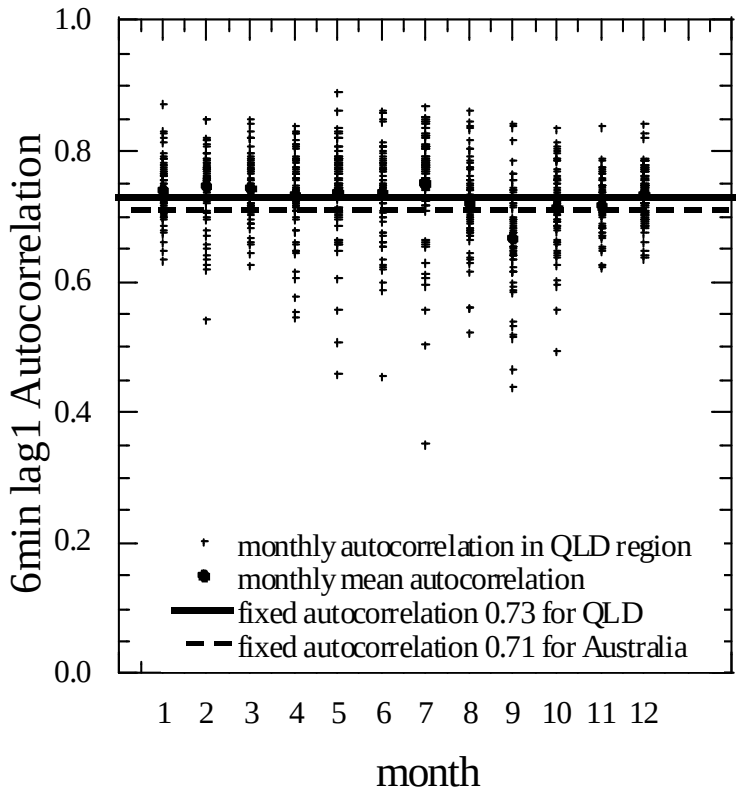

Figure 5.47 Monthly autocorrelation in Queensland

The next chapter will discuss on the application of the stochastic model and its performance.