Copyright 1992

by

George Michael Schmidt
This dissertation is dedicated to the men and women of the National Weather Service. Their loyalty and meticulous attention to the detail of collecting daily rainfall data during the last century made this research possible.
ACKNOWLEDGEMENTS

Dr. Allen G. Smajstrla (Department of Agricultural Engineering) was my dissertation advisor. His time, patience, encouragement and advice contributed greatly to my education and this dissertation. Dr. Fedro S. Zazueta (Department of Agricultural Engineering) was my coadvisor. His teaching skill and enthusiasm provided me with the tools to work, the confidence to make decisions and the energy to complete the research. Dr. M. K. Ochi (Department of Coastal and Oceanographic Engineering) was my outside examiner. He opened the door to probability theory and stochastic processes. It was an honor and a privilege to receive his opinions and guidance. Dr. James W. Jones (Department of Agricultural Engineering) provided great insight into the engineering application of stochastic rainfall models and the problems associated with their application.

Dr. G. W. Isaacs (professor emeritus, Department of Agricultural Engineering) supported my education and research activities for the better part of a decade. His encouragement and advice are gratefully acknowledged.

Mr. William Eshelman provided computer system support throughout the research. His support and encouragement to graduate student research in the Department of Agricultural Engineering contributed significantly to this dissertation.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>viii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xii</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>xxii</td>
</tr>
<tr>
<td>CHAPTERS</td>
<td></td>
</tr>
<tr>
<td>1  INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Statement of the Problem</td>
<td>5</td>
</tr>
<tr>
<td>1.2 Objectives of the Dissertation</td>
<td>5</td>
</tr>
<tr>
<td>1.3 Overview of the Dissertation</td>
<td>6</td>
</tr>
<tr>
<td>2  REVIEW OF THE LITERATURE</td>
<td>8</td>
</tr>
<tr>
<td>2.1 Rainfall Occurrence Models</td>
<td>9</td>
</tr>
<tr>
<td>2.1.1 Alternating Wet and Dry Interval Models</td>
<td>10</td>
</tr>
<tr>
<td>2.1.2 Wet and Dry Day Models</td>
<td>10</td>
</tr>
<tr>
<td>2.1.3 Point Process Models</td>
<td>15</td>
</tr>
<tr>
<td>2.2 Rainfall Amount Models</td>
<td>20</td>
</tr>
<tr>
<td>2.2.1 Chain-dependent Models</td>
<td>20</td>
</tr>
<tr>
<td>2.2.2 Chain-independent Models</td>
<td>21</td>
</tr>
<tr>
<td>2.3 Rainfall Model Parameter Estimation</td>
<td>24</td>
</tr>
<tr>
<td>2.3.1 Sample Selection</td>
<td>24</td>
</tr>
<tr>
<td>2.3.2 Parameter Estimation</td>
<td>25</td>
</tr>
<tr>
<td>2.3.3 Parameter Interpolation</td>
<td>27</td>
</tr>
<tr>
<td>2.4 Rainfall Model Applications</td>
<td>31</td>
</tr>
<tr>
<td>2.5 Rainfall Model Critique</td>
<td>33</td>
</tr>
<tr>
<td>3  DATA</td>
<td>36</td>
</tr>
<tr>
<td>3.1 Data Sources</td>
<td>36</td>
</tr>
<tr>
<td>3.2 Data Stations</td>
<td>37</td>
</tr>
<tr>
<td>3.3 Data Base Construction</td>
<td>40</td>
</tr>
<tr>
<td>3.4 Data Quality</td>
<td>41</td>
</tr>
</tbody>
</table>
4 PROBLEM JUSTIFICATION:
LONG-TERM VARIABILITY OF RAINFALL

4.1 Rainfall as a Stochastic Process
4.2 Amount of Rainfall
4.2.1 Averages
4.2.2 Standard Deviations
4.2.3 Quartile Ranges
4.2.4 Probability Distributions
4.2.5 Long-term Trends
4.3 Number of Wet Days
4.3.1 Averages
4.3.2 Standard Deviations
4.3.3 Quartile Ranges
4.3.4 Probability Distributions
4.3.5 Long-term Trends
4.4 Summary

5 PARAMETER ESTIMATION METHODOLOGY

5.1 Occurrence Model Parameter Estimation
5.1.1 Model Theory
5.1.2 Parameter Estimation Method
5.1.3 Results of Parameter Estimation Method
5.1.4 Summary and Conclusions
5.2 Amount Model Parameter Estimation
5.2.1 Model Theory
5.2.2 Parameter Estimation Method
5.2.3 Results of Parameter Estimation Method
5.2.4 Summary and Conclusions

6 PARAMETER SIMULATION METHODOLOGY

6.1 Occurrence Model Parameter Simulation
6.1.1 Parameter Simulation Method
6.1.2 Implementation of Parameter Simulation Method
6.1.3 Evaluation of Parameter Simulation Method
6.1.4 Results of Parameter Simulation Method
6.1.5 Summary and Conclusions
6.2 Amount Model Parameter Simulation
6.2.1 Parameter Simulation Method
6.2.2 Implementation of Parameter Simulation Method
6.2.3 Evaluation of Parameter Simulation Method
6.2.4 Results of Parameter Simulation Method
6.2.5 Summary and Conclusions

7 RESULTS AND DISCUSSION

7.1 Model Implementation
7.1.1 Rainfall Generation
7.1.2 Random Number Generation
7.2 Evaluation of Occurrence Model
7.2.1 Distribution of Number of Wet Days
7.2.2 Distribution of Dry Runs ........................................... 275
7.2.3 Long-term Trend of Number of Wet Days ................. 281
7.3 Evaluation of Amount Model ........................................ 282
  7.3.1 Distribution of Wet Day Amounts ......................... 288
  7.3.2 Distribution of Monthly Total Amounts ................. 296
  7.3.3 Long-term Trend of Monthly Total Amounts ........... 307

8 SUMMARY AND CONCLUSIONS ........................................ 313

APPENDICES
A GAMMA DISTRIBUTION PARAMETER ESTIMATION .......................... 318
B GOODNESS OF FIT TESTS ............................................. 320
C NUMERICAL EVALUATION OF THE STUDENT’S T DISTRIBUTION ................. 327
D SPECTRAL ANALYSIS AND FOURIER DOMAIN FILTERING METHODS .................. 330
E EVALUATION OF LARGE SAMPLE VARIANCES FOR THE MAXIMUM LIKELIHOOD PARAMETER ESTIMATES OF THE GAMMA DISTRIBUTION PARAMETERS ................. 334
F APPLICATION OF MULTIVARIATE AUTOREGRESSIVE TIME SERIES MODELS .................. 338
REFERENCES ................................................................... 343
BIOGRAPHICAL SKETCH .................................................. 354
<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Description of weather stations used in the data base</td>
<td>39</td>
</tr>
<tr>
<td>4.1</td>
<td>Kolmogorov-Smirnov test statistics for comparison of the distribution of monthly total amount of rainfall and the gamma distribution for Gainesville, Florida</td>
<td>57</td>
</tr>
<tr>
<td>4.2</td>
<td>Kolmogorov-Smirnov test statistics for comparison of the distribution of monthly total amount of rainfall and the gamma distribution for Orlando, Florida</td>
<td>58</td>
</tr>
<tr>
<td>4.3</td>
<td>Kolmogorov-Smirnov test statistics for comparison of the distribution of monthly total amount of rainfall and the gamma distribution for Tampa, Florida</td>
<td>59</td>
</tr>
<tr>
<td>4.4</td>
<td>Kolmogorov-Smirnov test statistics for comparison of the distribution of monthly total amount of rainfall and the gamma distribution for Ft. Myers, Florida</td>
<td>60</td>
</tr>
<tr>
<td>4.5</td>
<td>Parameter estimates and test statistics for linear regression of annual total amount of rainfall time series</td>
<td>74</td>
</tr>
<tr>
<td>4.6</td>
<td>Parameter estimates and test statistics for linear regression of winter total amount of rainfall time series</td>
<td>77</td>
</tr>
<tr>
<td>4.7</td>
<td>Parameter estimates and test statistics for linear regression of spring total amount of rainfall time series</td>
<td>80</td>
</tr>
<tr>
<td>4.8</td>
<td>Parameter estimates and test statistics for linear regression of summer total amount of rainfall time series</td>
<td>83</td>
</tr>
<tr>
<td>4.9</td>
<td>Parameter estimates and test statistics for linear regression of autumn total amount of rainfall time series</td>
<td>86</td>
</tr>
<tr>
<td>4.10</td>
<td>Chi-square test statistics for comparison of the distribution of monthly number of wet days and the binomial distribution for Gainesville, Florida</td>
<td>102</td>
</tr>
<tr>
<td>4.11</td>
<td>Chi-square test statistics for comparison of the distribution of monthly number of wet days and the binomial distribution for Orlando, Florida</td>
<td>103</td>
</tr>
<tr>
<td>4.12</td>
<td>Chi-square test statistics for comparison of the distribution of monthly number of wet days and the binomial distribution for Tampa, Florida</td>
<td>104</td>
</tr>
</tbody>
</table>
4.13 Chi-square test statistics for comparison of the distribution of monthly number of wet days and the binomial distribution for Ft. Myers, Florida

4.14 Parameter estimates and test statistics for linear regression of annual number of wet days time series

4.15 Parameter estimates and test statistics for linear regression of winter number of wet days time series

4.16 Parameter estimates and test statistics for linear regression of spring number of wet days time series

4.17 Parameter estimates and test statistics for linear regression of summer number of wet days time series

4.18 Parameter estimates and test statistics for linear regression of autumn number of wet days time series

4.19 Summary of estimated slopes from linear regression of seasonal and annual total amount of rainfall for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 to 1988)

4.20 Summary of estimated slopes from linear regression of seasonal and annual number of wet days for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 to 1988)

5.1 Estimated statistics for the time series \{\hat{p}_{DD}(vS + \tau)\}

5.2 Estimated statistics for the time series \{\hat{p}_{WD}(vS + \tau)\}

5.3 Estimated statistics for the time series \{\hat{p}_{DD}(vS + \tau)\} with \hat{p}_{WD} = 1.0 excluded

5.4 Kolmogorov-Smirnov test statistics, D, for the monthly samples, \{\hat{p}_{DD}(v, \tau)\}, for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988)

5.5 Kolmogorov-Smirnov test statistics, D, for the monthly samples, \{\hat{p}_{DD}(v, \tau)\}, for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988)

5.6 Maximum likelihood estimates of the gamma-distribution parameters \alpha(\tau) and \beta(\tau) with associated 99% confidence intervals by month for Gainesville, Florida

5.7 Maximum likelihood estimates of the gamma-distribution parameters \alpha(\tau) and \beta(\tau) with associated 99% confidence intervals by month for Orlando, Florida
5.8 Maximum likelihood estimates of the gamma distribution parameters \( \alpha(t) \) and \( \beta(t) \) with associated 99% confidence intervals by month for Tampa, Florida......................................................................................... 172

5.9 Maximum likelihood estimates of the gamma distribution parameters \( \alpha(t) \) and \( \beta(t) \) with associated 99% confidence intervals by month for Ft. Myers, Florida............................................... 173

5.10 Estimates of the wet day amount model parameter \( \delta(t) \) by month for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988)........................................................................................................ 174

5.11 Kolmogorov-Smirnov test statistics, \( D \), for monthly gamma distribution wet day amount model.................................................................................... 182

5.12 Kolmogorov-Smirnov p-values for monthly gamma distribution wet day amount model.................................................................................... 182

5.13 Kolmogorov-Smirnov sample sizes, \( n \), for monthly gamma distribution wet day amount model.................................................................................... 182

5.14 Maximum positive error, \( H^+ \) (mm/day), for monthly gamma distribution wet day amount model.................................................................................... 193

5.15 Location of error \( H^+ \), \( y_{\text{observed}} \) (mm), for monthly gamma distribution wet day amount model.................................................................................... 193

5.16 Maximum negative error, \( H^- \) (mm/day), for monthly gamma distribution wet day amount model.................................................................................... 194

5.17 Location of error \( H^- \), \( y_{\text{observed}} \) (mm), for monthly gamma distribution wet day amount model.................................................................................... 194

6.1 Kolmogorov-Smirnov test statistics, \( D \), for testing the hypothesis that the \( \{p_{wO}(v)\}_s \) time series are normally distributed for Gainesville, Orlando, Tampa, and Ft. Myers, Florida............................................................ 231

6.2 Kolmogorov-Smirnov test statistics, \( D \), for testing the hypothesis that the \( \{p_{wO}(v)\}_s \) time series are normally distributed for Gainesville, Orlando, Tampa, and Ft. Myers, Florida............................................................ 232

6.3 Multivariate AR(p) model orders for the \( \{P_{v, \text{Stochastic}}^e\} \) time series............................................... 233

6.4 Kolmogorov-Smirnov test statistics, \( D \), for testing the hypothesis that the residuals of the \( \{p_{wO}(v)\}_s \) time series are normally distributed for Gainesville, Orlando, Tampa, Ft. Myers, Florida............................................................ 235
6.5 Kolmogorov-Smirnov test statistics, $D$, for testing the hypothesis that the residuals of the $\{p_{WD}(v)\}_s$ time series are normally distributed for Gainesville, Orlando, Tampa, Ft. Myers, Florida.............................................. 236

6.6 Chi-square test statistics, $Q$, for the Porte Manteau goodness of fit test of the monthly $\{p_{WD}(v)\}_s$ series residuals.......................................................... 237

6.7 Chi-square degrees of freedom $(L-p)$ for the Porte Manteau goodness of fit test of the monthly $\{p_{WD}(v)\}_s$ series residuals......................................................... 237

6.8 Chi-square test statistics, $Q$, for the Porte Manteau goodness of fit test of the monthly $\{p_{WD}(v)\}_s$ series residuals.......................................................... 238

6.9 Chi-square degrees of freedom $(L-p)$ for the Porte Manteau goodness of fit test of the monthly $\{p_{WD}(v)\}_s$ series residuals......................................................... 238

7.1 Bin limits, midpoint of bin, and bin width for variable bin width histograms........................................................................................................... 290
**LIST OF FIGURES**

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Annually homogeneous Markov-chain transition probability estimates using data from Tallahassee, Florida</td>
<td>28</td>
</tr>
<tr>
<td>3.1</td>
<td>Station location map</td>
<td>38</td>
</tr>
<tr>
<td>4.1</td>
<td>Average monthly total amount of rainfall (mm) by month for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988)</td>
<td>46</td>
</tr>
<tr>
<td>4.2</td>
<td>Average monthly total amount of rainfall (mm) with plus and minus one standard deviation by month for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988)</td>
<td>48</td>
</tr>
<tr>
<td>4.3</td>
<td>Box-and-whiskers plots of monthly total amount of rainfall, showing the medians, inner quartile ranges, and extreme values by month for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988)</td>
<td>50</td>
</tr>
<tr>
<td>4.4</td>
<td>Monthly histograms and fitted gamma probability density function for monthly total amounts of rainfall (mm) for Gainesville, Florida (1900 through 1988)</td>
<td>52</td>
</tr>
<tr>
<td>4.5</td>
<td>Monthly histograms and fitted gamma probability density function for monthly total amounts of rainfall (mm) for Orlando, Florida (1900 through 1988)</td>
<td>53</td>
</tr>
<tr>
<td>4.6</td>
<td>Monthly histograms and fitted gamma probability density function for monthly total amounts of rainfall (mm) for Tampa, Florida (1900 through 1988)</td>
<td>54</td>
</tr>
<tr>
<td>4.7</td>
<td>Monthly histograms and fitted gamma probability density function for monthly total amounts of rainfall (mm) for Ft. Myers, Florida (1900 through 1988)</td>
<td>55</td>
</tr>
<tr>
<td>4.8</td>
<td>Gamma-distribution parameter $\alpha$ by month for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988)</td>
<td>61</td>
</tr>
<tr>
<td>4.9</td>
<td>Gamma-distribution parameter $\beta$ by month for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988)</td>
<td>62</td>
</tr>
<tr>
<td>4.10</td>
<td>Spatial and temporal variability in the gamma distribution fitted to monthly total amount of rainfall (mm) for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988)</td>
<td>64</td>
</tr>
</tbody>
</table>
4.11 Observed and filtered annual total amount of rainfall time series, with linear-regression line for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988) ................................................................. 72

4.12 Estimated spectral energy of annual total amount of rainfall time series for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988) ................................................................. 73

4.13 Observed and filtered winter total amount of rainfall time series, with linear-regression line for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988) ................................................................. 75

4.14 Estimated spectral energy of winter total amount of rainfall time series for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988) ................................................................. 76

4.15 Observed and filtered spring total amount of rainfall time series, with linear-regression line for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988) ................................................................. 78

4.16 Estimated spectral energy of spring total amount of rainfall time series for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988) ................................................................. 79

4.17 Observed and filtered summer total amount of rainfall time series, with linear-regression line for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988) ................................................................. 81

4.18 Estimated spectral energy of summer total amount of rainfall time series for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988) ................................................................. 82

4.19 Observed and filtered autumn total amount of rainfall time series, with linear-regression line for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988) ................................................................. 84

4.20 Estimated spectral energy of autumn total amount of rainfall time series for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988) ................................................................. 85

4.21 Average number of wet days per month for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988) ................................................................. 91

4.22 Average number of wet days per month with plus and minus one standard deviation for Gainesville, Orlando, Tampa, Ft. Myers, Florida (1900 through 1988) ................................................................. 93

4.23 Box-and-whiskers plots of the number of wet days per month, showing the medians, inner quartile ranges, and extreme values for Gainesville, Orlando, Tampa, Ft. Myers, Florida (1900 through 1988) ................................................................. 95

4.24 Monthly histograms and fitted binomial probability mass function for the number of wet days per month for Gainesville, Florida (1900 through 1988) ................................................................. 96
4.25 Monthly histograms and fitted binomial probability mass function for the number of wet days per month for Orlando, Florida (1900 through 1988)................................................................................................................. 97

4.26 Monthly histograms and fitted binomial probability mass function for the number of wet days per month for Tampa, Florida (1900 through 1988)................................................................................................................. 98

4.27 Monthly histograms and fitted binomial probability mass function for the number of wet days per month for Ft. Myers, Florida (1900 through 1988)................................................................................................................. 99

4.28 Binomial-distribution parameter $p$ by month for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988)....................................................................................................................... 106

4.29 Spatial and temporal variability in the binomial distribution fitted to the number of wet days per month for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988)................................................................................................................. 108

4.30 Observed and filtered annual number of wet days time series, with linear-regression line for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988)................................................................................................................. 110

4.31 Estimated spectral energy of annual number of wet days time series for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988)................................................................................................................. 111

4.32 Observed and filtered winter number of wet days time series, with linear-regression line for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988)................................................................................................................. 113

4.33 Estimated spectral energy of winter number of wet days time series for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988)................................................................................................................. 114

4.34 Observed and filtered spring number of wet days time series, with linear-regression line for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988)................................................................................................................. 116

4.35 Estimated spectral energy of spring number of wet days time series for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988)................................................................................................................. 117

4.36 Observed and filtered summer number of wet days time series, with linear-regression line for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988)................................................................................................................. 119

4.37 Estimated spectral energy of summer number of wet days time series for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988)................................................................................................................. 120
4.38 Observed and filtered autumn number of wet days time series, with linear-regression line for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988) .................................................................... 122

4.39 Estimated spectral energy of autumn number of wet days time series for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988) .................................................................... 123

5.1 Estimated time series of Markov-chain transition probabilities, \( \{ \hat{p}_{DD}(vS + \tau) \} \), for Gainesville, Florida from 1900 through 1988, with expanded views of 1900 through 1909, and 1900 through 1901. 137

5.2 Estimated time series of Markov-chain transition probabilities, \( \{ \hat{p}_{DD}(vS + \tau) \} \), for Orlando, Florida, from 1900 through 1988, with expanded views of 1900 through 1909, and 1900 through 1901. 138

5.3 Estimated time series of Markov-chain transition probabilities, \( \{ \hat{p}_{DD}(vS + \tau) \} \), for Tampa, Florida from 1900 through 1988, with expanded views of 1900 through 1909, and 1900 through 1901. 139

5.4 Estimated time series of Markov-chain transition probabilities, \( \{ \hat{p}_{DD}(vS + \tau) \} \), for Ft. Myers, Florida from 1900 through 1988, with expanded views of 1900 through 1909, and 1900 through 1901. 140

5.5 Estimated time series of Markov-chain transition probabilities, \( \{ \hat{p}_{wD}(vS + \tau) \} \), for Gainesville, Florida from 1900 through 1988, with expanded views of 1900 through 1909, and 1900 through 1901. 141

5.6 Estimated time series of Markov-chain transition probabilities, \( \{ \hat{p}_{wD}(vS + \tau) \} \), for Orlando, Florida from 1900 through 1988, with expanded views of 1900 through 1909, and 1900 through 1901. 142

5.7 Estimated time series of Markov-chain transition probabilities, \( \{ \hat{p}_{wD}(vS + \tau) \} \), for Tampa, Florida from 1900 through 1988, with expanded views of 1900 through 1909, and 1900 through 1901. 143

5.8 Estimated time series of Markov-chain transition probabilities, \( \{ \hat{p}_{wD}(vS + \tau) \} \), for Ft. Myers, Florida from 1900 through 1988, with expanded views of 1900 through 1909, and 1900 through 1901. 144

5.9 Comparisons between observed histograms of \( \{ \hat{p}_{DD}(vS + \tau) \} \), normal distributions, and Gram-Charlier distributions for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988) 147

5.10 Comparisons between observed histograms of \( \{ \hat{p}_{wD}(vS + \tau) \} \), normal distributions, and Gram-Charlier distributions for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988) 148
Comparisons between observed histograms of \( \{\hat{p}_{ws}(vS + \tau)\} \) with \( \hat{p}_{ws} = 1.0 \) excluded, normal distributions, and Gram-Charlier distributions for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988) .................................................. 152

Monthly observed histograms of \( \{\hat{p}_{ws}(vS + \tau)\} \) for Ft. Myers, Florida (1900 through 1988) .................................................. 156

Comparisons between December observed histograms of \( \{\hat{p}_{ws}(vS + \tau)\} \) with and without \( \hat{p}_{ws} = 1.0 \) excluded, normal distributions, and Gram-Charlier distributions for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988) .................................................. 157

Gamma-distribution parameter \( \alpha \) with plus and minus one standard deviation by month for Gainesville, Orlando, Tampa, and Ft. Myers, Florida .................................................................................. 175

Gamma-distribution parameter \( \beta \) with plus and minus one standard deviation by month for Gainesville, Orlando, Tampa, and Ft. Myers, Florida .................................................................................. 176

Gamma-distribution parameter \( \delta \) by month for Gainesville, Orlando, Tampa, and Ft. Myers, Florida .................................................................................. 177

Comparison of monthly gamma-distribution parameter \( \alpha \) for Gainesville, Orlando, Tampa, Ft. Myers, Florida .................................................................................. 178

Comparison of monthly gamma-distribution parameter \( \beta \) for Gainesville, Orlando, Tampa, Ft. Myers, Florida .................................................................................. 179

Comparison of monthly gamma-distribution parameter \( \delta \) for Gainesville, Orlando, Tampa, Ft. Myers, Florida .................................................................................. 180

Comparison of rainfall amount cumulative distributions \( F_{\text{Model}}, F_{\text{Observed}} \), \( F_{\text{UCL}}, \) and \( F_{\text{LCL}} \) for May at Gainesville, Florida .................................................................................. 184

Comparison of rainfall amount cumulative distributions \( F_{\text{Model}}, F_{\text{Observed}} \), \( F_{\text{UCL}}, \) and \( F_{\text{LCL}} \) by month at Gainesville, Florida .................................................................................. 185

Comparison of rainfall amount cumulative distributions \( F_{\text{Model}}, F_{\text{Observed}} \), \( F_{\text{UCL}}, \) and \( F_{\text{LCL}} \) by month at Orlando, Florida .................................................................................. 186

Comparison of rainfall amount cumulative distributions \( F_{\text{Model}}, F_{\text{Observed}} \), \( F_{\text{UCL}}, \) and \( F_{\text{LCL}} \) by month at Tampa, Florida .................................................................................. 187

Comparison of rainfall amount cumulative distributions \( F_{\text{Model}}, F_{\text{Observed}} \), \( F_{\text{UCL}}, \) and \( F_{\text{LCL}} \) by month at Ft. Myers, Florida .................................................................................. 188
5.25 Example of the simulation of wet day rainfall amounts using the gamma distribution for May at Gainesville, Florida ................................................................. 190

5.26 $H(Y_{\text{Observed}})$ error for the gamma distribution wet day amount model by month at Gainesville, Orlando, Tampa, and Ft. Myers, Florida ........................................................................................................ 195

5.27 Mapping of parametric confidence region into the associated distribution region of uncertainty for May at Gainesville, Florida ........................................................................................................ 199

5.28 Comparison of rainfall amount cumulative distributions $F_{\text{Model}}$, $F_{\text{Observed}}$, $F_{\text{UCL}}$, and $F_{\text{LCL}}$ for May at Gainesville, Florida .................................................................................................................. 200

6.1 Average monthly Markov-chain transition probability, $\overline{p}_{DD}^T$ ($T = 1, 2, \cdots, 12$), with plus and minus one standard deviation for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 to 1988).... 212

6.2 Average monthly Markov-chain transition probability, $\overline{p}_{WD}^T$ ($T = 1, 2, \cdots, 12$), with plus and minus one standard deviation for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 to 1988).... 213

6.3 Comparison of average monthly Markov-chain transition probabilities, $\overline{p}_{DD}^T$ and $\overline{p}_{WD}^T$ ($T = 1, 2, \cdots, 12$), for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 to 1988)............................................................................ 215

6.4 Comparison $\{p_{DD}(v)^T\}_{M+D}$, $\{p_{WD}(v)^T\}_{M+D}$, and monthly wet day count time series for February at Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988)................................................................. 218

6.5 Comparison of $\{p_{DD}(v)^T\}_{M+D}$, $\{p_{WD}(v)^T\}_{M+D}$, and monthly wet day count time series for April at Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988)................................................................. 219

6.6 Comparison of $\{p_{DD}(v)^T\}_{M+D}$, $\{p_{WD}(v)^T\}_{M+D}$, and monthly wet day count time series for July at Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988)................................................................. 220

6.7 Comparison of $\{p_{DD}(v)^T\}_{M+D}$, $\{p_{WD}(v)^T\}_{M+D}$, and monthly wet day count time series for November at Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988)................................................................. 221

6.8 Comparison of observed and low-pass filtered $\{p_{WD}(v)^T\}$ series, with observed and low-pass filtered monthly wet day count time series for April at Ft. Myers, Florida (1900 through 1988)................................................................. 223
6.9  Time series of Markov-chain transition probabilities, \( \{ p_{DD}(vS + \tau) \}_{M+D} \) and \( \{ p_{WD}(vS + \tau) \}_{M+D} \), for Gainesville, Florida from 1900 through 1988

6.10 Time series of Markov-chain transition probabilities, \( \{ p_{DD}(vS + \tau) \}_{M+D} \) and \( \{ p_{WD}(vS + \tau) \}_{M+D} \), for Orlando, Florida from 1900 through 1988

6.11 Time series of Markov-chain transition probabilities, \( \{ p_{DD}(vS + \tau) \}_{M+D} \) and \( \{ p_{WD}(vS + \tau) \}_{M+D} \), for Tampa, Florida from 1900 through 1988

6.12 Time series of Markov-chain transition probabilities, \( \{ p_{DD}(vS + \tau) \}_{M+D} \) and \( \{ p_{WD}(vS + \tau) \}_{M+D} \), for Ft. Myers, Florida from 1900 through 1988

6.13 Time series of Markov-chain transition probabilities, \( \{ p_{DD}(vS + \tau) \}_{M+D} \) and \( \{ p_{WD}(vS + \tau) \}_{M+D} \), for Gainesville, Florida from 1900 through 1929

6.14 Comparison of \( \alpha(\tau, v)_D, \beta(\tau, v)_D, \alpha(\tau, v)_D \), and low-pass filtered monthly total amount time series for February at Gainesville, Florida (1900 through 1988)

6.15 Comparison of \( \alpha(\beta(\tau, v)_D \) for \( W = 5, 7, 9, \) and 11 years with low-pass filtered monthly total amount time series for February at Gainesville, Florida (1900 through 1988)

6.16 Comparison of \( \alpha(\beta(\tau, v)_D \) and low-pass filtered monthly total amount time series for February at Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988)

6.17 Comparison of \( \alpha(\beta(\tau, v)_D \) and low-pass filtered monthly total amount time series for April at Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988)

6.18 Comparison of \( \alpha(\beta(\tau, v)_D \) and low-pass filtered monthly total amount time series for July at Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988)

6.19 Comparison of \( \alpha(\beta(\tau, v)_D \) and low-pass filtered monthly total amount time series for November at Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988)

6.20 Time series of gamma-distribution parameters, \( \alpha(vS + \tau)_D \) and \( \beta(vS + \tau)_D \), for Gainesville, Florida from 1903 through 1985
6.21 Time series of gamma-distribution parameters, $\alpha(vS + \tau)_D$ and $\beta(vS + \tau)_D$, for Orlando, Florida from 1903 through 1985 ........................................... 254

6.22 Time series of gamma-distribution parameters, $\alpha(vS + \tau)_D$ and $\beta(vS + \tau)_D$, for Tampa, Florida from 1903 through 1985 ........................................... 255

6.23 Time series of gamma-distribution parameters, $\alpha(vS + \tau)_D$ and $\beta(vS + \tau)_D$, for Ft. Myers, Florida from 1903 through 1985 ........................................... 256

6.24 Time series of gamma-distribution parameters, $\alpha(vS + \tau)_D$ and $\beta(vS + \tau)_D$, for Gainesville, Florida from 1903 through 1932 ........................................... 257

7.1 Comparison of generated uniform deviates, $R$, with uniform probability density function ....................................................................................... 265

7.2 Comparison of generated normal deviates, $\xi$, with normal probability density function ($\mu = 0, \sigma = 1$) ........................................................................... 267

7.3 Comparison of generated gamma deviates, $\Gamma$, with gamma probability density function ($\alpha = 0.792, \beta = 16.478$) ................................................. 268

7.4 Comparison of histograms for observed and generated number of wet days for February at Gainesville, Orlando, Tampa, and Ft. Myers, Florida ........................................................................................................ 271

7.5 Comparison of histograms for observed and generated number of wet days for April at Gainesville, Orlando, Tampa, and Ft. Myers, Florida ........................................................................................................ 272

7.6 Comparison of histograms for observed and generated number of wet days for July at Gainesville, Orlando, Tampa, and Ft. Myers, Florida ........................................................................................................ 273

7.7 Comparison of histograms for observed and generated number of wet days for November at Gainesville, Orlando, Tampa, and Ft. Myers, Florida ........................................................................................................ 274

7.8 Comparison of histograms for observed and generated dry run lengths for February at Gainesville, Orlando, Tampa, and Ft. Myers, Florida ........................................................................................................ 277

7.9 Comparison of histograms for observed and generated dry run lengths for April at Gainesville, Orlando, Tampa, and Ft. Myers, Florida ........................................................................................................ 278

7.10 Comparison of histograms for observed and generated dry run lengths for July at Gainesville, Orlando, Tampa, and Ft. Myers, Florida ........................................................................................................ 279
7.11 Comparison of histograms for observed and generated dry run lengths for November at Gainesville, Orlando, Tampa, and Ft. Myers, Florida.......................................................... 280
7.12 Comparison of observed and generated low-pass filtered time series of number of wet days for February at Gainesville, Orlando, Tampa, and Ft. Myers, Florida............................................................................. 283
7.13 Comparison of observed and generated low-pass filtered time series of number of wet days for April at Gainesville, Orlando, Tampa, and Ft. Myers, Florida.......................................................................... 284
7.14 Comparison of observed and generated low-pass filtered time series of number of wet days for July at Gainesville, Orlando, Tampa, and Ft. Myers, Florida............................................................................. 285
7.15 Comparison of observed and generated low-pass filtered time series of number of wet days for November at Gainesville, Orlando, Tampa, and Ft. Myers, Florida............................................................................. 286
7.16 Comparison of histograms and cumulative distributions for observed and generated wet day amounts of rainfall for February at Gainesville, Orlando, Tampa, and Ft. Myers, Florida .......................................................... 291
7.17 Comparison of histograms and cumulative distributions for observed and generated wet day amounts of rainfall for April at Gainesville, Orlando, Tampa, and Ft. Myers, Florida............................................................................. 292
7.18 Comparison of histograms and cumulative distributions for observed and generated wet day amounts of rainfall for July at Gainesville, Orlando, Tampa, and Ft. Myers, Florida............................................................................. 293
7.19 Comparison of histograms and cumulative distributions for observed and generated wet day amounts of rainfall for November at Gainesville, Orlando, Tampa, and Ft. Myers, Florida............................................................................. 294
7.20 Comparison of summary histograms for observed and generated wet day amounts of rainfall for February at Gainesville, Orlando, Tampa, and Ft. Myers, Florida.......................................................... 297
7.21 Comparison of summary histograms for observed and generated wet day amounts of rainfall for April at Gainesville, Orlando, Tampa, and Ft. Myers, Florida............................................................................. 298
7.22 Comparison of summary histograms for observed and generated wet day amounts of rainfall for July at Gainesville, Orlando, Tampa, and Ft. Myers, Florida............................................................................. 299
7.23 Comparison of summary histograms for observed and generated wet day amounts of rainfall for November at Gainesville, Orlando, Tampa, and Ft. Myers, Florida.......................................................... 300
7.24 Comparison of generated histograms and observed gamma density functions for monthly total amounts of rainfall at Gainesville, Florida. ................................................................. 302

7.25 Comparison of generated histograms and observed gamma density functions for monthly total amounts of rainfall at Orlando, Florida. ................................................................. 303

7.26 Comparison of generated histograms and observed gamma density functions for monthly total amounts of rainfall at Tampa, Florida. ................................................................. 304

7.27 Comparison of generated histograms and observed gamma density functions for monthly total amounts of rainfall at Ft. Myers, Florida. ................................................................. 305

7.28 Comparison of observed and generated low-pass filtered time series of monthly total amounts of rainfall for February at Gainesville, Orlando, Tampa, and Ft. Myers, Florida. ................................................................. 309

7.29 Comparison of observed and generated low-pass filtered time series of monthly total amounts of rainfall for April at Gainesville, Orlando, Tampa, and Ft. Myers, Florida. ................................................................. 310

7.30 Comparison of observed and generated low-pass filtered time series of monthly total amounts of rainfall for July at Gainesville, Orlando, Tampa, and Ft. Myers, Florida. ................................................................. 311

7.31 Comparison of observed and generated low-pass filtered time series of monthly total amounts of rainfall for November at Gainesville, Orlando, Tampa, and Ft. Myers, Florida. ................................................................. 312
Abstract of Dissertation Presented to the Graduate School of the University of Florida in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

STOCHASTIC RAINFALL MODELING AND LONG-TERM CLIMATIC VARIABILITY: MODEL PARAMETER ESTIMATION AND MODEL EVALUATION

By

George Michael Schmidt

December 1992

Chairman: Dr. Allen G. Smajstrla
Major Department: Agricultural Engineering

With increased interest in global climate change, stochastic rainfall models have been criticized for their inability to preserve the interannual variability and long-term trends found in long historical rainfall records. To address this problem, annually nonhomogeneous (time variant) parameter estimation methods and parameter simulation models were developed. The proposed methodology can be applied to a wide range of rainfall simulation models, however a first-order Markov-chain occurrence model and a gamma-distribution amount model were used for illustrative purposes.

The proposed methodology was implemented using 89 years of historical rainfall data from Gainesville, Orlando, Tampa, and Ft. Myers, Florida. Statistical and Fourier-domain analyses of the historical data indicated that considerable interannual variability and long-term trends were present. Implementation of the proposed parameter estimation method demonstrated that model parameters changed on a time scale of decades, and reflected long-term trends found in the historical data. Parameter simulation models were developed to allow for the generation of long time series of daily rainfall
Based on a limited sample of historical data. The parameter simulation models contained observed trend components that could be replaced with climate change scenarios or physically based models to study long-term effects of interdecade phenomena.

The proposed parameter simulation model was coupled to a rainfall simulation model and used to generate long time series of daily rainfall data. To evaluate for the preservation of long-term variability of rainfall, monthly histograms of wet day amounts, total amounts, number of wet days and dry run lengths were constructed from the observed and generated rainfall time series and compared. To evaluate for the presence of long-term trends in rainfall, observed and generated time series of monthly total amounts and monthly number of wet days were low-pass filtered and compared. During all seasons and at all locations considered, both the distributions and long-term trends found in the observed rainfall records were preserved in the generated rainfall time series.

The implementation and evaluation of the proposed annually nonhomogeneous parameter estimation methodology demonstrated that stochastic rainfall models can be designed which will preserve the interannual variability and long-term trends found in long historical rainfall records.
CHAPTER 1
INTRODUCTION

In the spring of 1900, Alexander J. Mitchell, Florida Section Director of the U. S. Weather Bureau, delivered the closing address at the Florida State Horticultural Society meeting. His report on this convention included the following observations concerning the ability of agriculturalists to adapt to highly uncertain climatological conditions.

In these meetings the weather bureau official can put his knowledge to practical use, and impart it to those for whom it is a pleasure to labor. The farmer of Florida is not properly mirrored by that oft much-abused-type-of-the-Swine-family facetiously denominated the "Florida razor-back." On the contrary, the Florida farmer and horticulturist is corpulent in girth, aggressive in action and fertile in mental resources. No misfortune, however great, has caused him to abandon his avocation for pursuits more alluring, though in the end less remunerative. (U.S. Department of Agriculture, 1900, May, p.4)

The accumulated agricultural knowledge and aggressive activity of farmers has increased dramatically since 1900. However, extended, irregularly occurring periods of drought and highly variable annual rainfall amounts have resulted in food shortages, economic hardship, and the abandonment of farming by many small farmers on a global scale. Faced with a high degree of climatic uncertainty, farmers, agricultural financiers and water resource managers require decision-making support that allows them to accurately assess the risk associated with proposed investment activities and management strategies.

This research was undertaken to support agricultural decision making in the presence of highly uncertain rainfall events over the long term (decades). The objective of this research was to develop a rainfall modeling approach that preserves the observed long-term variability of rainfall at a specific location.
The mathematical description or modeling of daily rainfall has been considered from probabilistic, stochastic and deterministic approaches. During the past two decades, considerable research and evaluation effort has focused on the stochastic modeling of daily rainfall, particularly in the agricultural and water resource management sectors. The stochastic approach was taken in this research.

The theory of stochastic processes is a large and growing field of study which is described in numerous classical texts including those of Parzen (1962), Cox and Miller (1965), Medhi (1982), and Ochi (1990). The engineering application of this body of theory to daily rainfall is also extensive and is reviewed in Chapter 2 of this work. Many rainfall models consider the rainfall status of a particular day as a combination of two stochastic processes in discrete time (days). The daily occurrence of rainfall events, \( \{X_1, \cdots, X_n, \cdots, X_t\} \), forms a discrete stochastic process where the value of \( X_t \) represents the wet or dry state of day \( t \). Daily rainfall, \( \{Y_1, \cdots, Y_n, \cdots, Y_t\} \), forms a continuous random process where \( Y_t \) represents the amount of rainfall on day \( t \). These two stochastic processes can be modeled as independent or dependent.

Historically, the modeling of daily rainfall has focused on the selection of a classical stochastic process to model the occurrence and amount of rainfall. Occurrence models used include alternating renewal models, Markov chains of varying order and continuous and discrete point process models. Daily rainfall amount models considered include a wide variety of continuous probability density functions with the exponential, gamma and mixed exponential functions receiving considerable attention. The engineering application of the theory of stochastic processes then becomes a question of selecting a set of models, fitting the selected models to an observed data sample, and evaluating the results. All of the stochastic rainfall models have some type of parameter that needs to be estimated from sampled data. Again the details of these models and parameter estimation methods are reviewed in Chapter 2.
Most climatological regions experience seasonality in rainfall events throughout the year. Summer thundershowers, occasional hurricanes and the passage of continental fronts during winter are the predominant rainfall mechanisms in Florida, while summer and winter monsoons dominant other regions. To account for this within year rainfall variability, seasonal data samples have been used to estimate seasonally varying rainfall model parameters. These seasonally varying parameters result in a stochastic rainfall model which is seasonally nonhomogeneous, but annually homogeneous. When this type of model is used to generate a long (decades) time series of daily rainfall, the model parameters vary from season to season within the year, but are constant or time invariant from year to year.

All stochastic rainfall models reviewed are of the annually homogeneous type. The most widely used annually homogeneous rainfall model appears to be the WGEN model (Richardson, 1981; Richardson and Wright, 1984) developed within the Agricultural Research Service (ARS) of the U.S. Department of Agriculture and distributed by the Climate Research Center of the Goddard Space Flight Center at Columbia University. The WGEN model is also included in the Decision Support System for Agrotechnology Transfer (DSSAT) distributed by the International Benchmark Sites Network for Agrotechnology Transfer (IBSNAT) (Jagtap et al., 1988). The WGEN model combines a first-order Markov chain occurrence model with a gamma-distribution amount model.

Stochastic rainfall models in the agricultural sector are frequently coupled with models of temperature, solar radiation, evapotranspiration, crop yield, irrigation water requirements, or agrochemical use. These coupled models are used to describe the behavior of agricultural or biological systems under highly variable rainfall regimes over long time periods (decades). The output of this type of agricultural system model is usually a time series of the form, \( \{Z_1, \cdots, Z_n, \cdots, Z_s\} \), where the random variable \( Z \) may be agricultural water use or crop yield at the end of growing season \( t \). These agricultural models often use a daily time step, with the output, \( Z_t \), accumulated or summed at
monthly, seasonal or annual intervals. To quantify the variability in time of the simulated variable, Z, a histogram or empirical cumulative probability distribution is often constructed. Understanding the nature of the random variable Z in distribution over time is an important objective of agricultural system modeling. This type of modeling approach allows planners and decision makers to make risk assessments based on the probability, \( Pr\{Z \leq z\} \) or \( Pr\{Z > z\} \), of the variable Z occurring under alternative management strategies and variable rainfall.

For example, consider the scenario where a particular agricultural crop such as citrus is modeled. The model may advance at a daily time step for thirty years, with an annual crop yield and irrigation water requirement determined at each annual harvest. Let the random variable \( Z_t \) be the irrigation water requirement for year \( t \). If the rainfall model component of the crop model does not accurately reproduce the relative frequency of dry intervals or rainfall amounts over the period of simulation, the crop model cannot generate a realistic distribution of annual irrigation water requirements, \( Z_t \). In particular, if the rainfall model underestimates the distribution of dry years, the crop model will underestimate the risk of years with large irrigation water requirements or the probability, \( Pr\{Z > z\} \).

The importance of long-term variability in rainfall to agricultural system modeling has led to some concern over the ability of stochastic rainfall models to reproduce the variability observed in historical records. Recently, Woolhiser and his colleagues have cautioned users of current rainfall models, saying:

We found that the MCME (Markov-chain/mixed-exponential) model preserved the important statistics within a year, but that caution should be taken when using it to study annual phenomena. (Hanson, Osborn, and Woolhiser, 1989, p. 873)
1.1 Statement of the Problem

The problem to be considered in this research is to quantitatively investigate the nature of long-term variability of daily rainfall and to determine the ability of stochastic rainfall models to preserve this variability at time scales of decades.

The basic hypothesis to be investigated in this research is that annually homogeneous parameter estimation methods do not allow stochastic rainfall models to simulate observed long-term (year-to-year) variability in rainfall. Alternatively, a stochastic rainfall model with annually nonhomogeneous model parameters that vary from year to year will be developed and evaluated.

1.2 Objectives of the Dissertation

The objective of rainfall simulation modeling is to generate a long time series of daily rainfall that preserves the amount, duration and frequency of rainfall events in distribution over time. Of particular interest to agricultural planners and engineering designers are irregularly occurring extreme events such as drought and flood-producing storms. It is critical that simulated rainfall time series accurately preserve the distribution of these events when simulated time series are used for design and risk analysis.

The estimation and simulation of nonhomogeneous rainfall model parameters will be investigated to determine the importance of these parameters in the generation of rainfall time series. An improvement in the ability of stochastic rainfall models to preserve the long-term distribution of rainfall events, not only the mean or standard deviation, is the major objective of the research. The specific objectives of the research are the following:

i. **Nature of Long-Term Variability of Rainfall**

   Investigate and quantify the nature of year-to-year variability in rainfall using long time series of observed rainfall records.
ii. **Nonhomogeneous Parameter Estimation Method**

Develop a method for rainfall model parameter estimation that reveals the year-to-year variability of the parameter as well as the annual seasonality.

iii. **Simulation Methodology for Rainfall Parameters**

Develop a method to simulate long time series of stochastic rainfall model parameters that preserves the variability observed using the above method.

iv. **Model Evaluation Method**

Develop a method to evaluate stochastic rainfall models and parameter estimation techniques.

### 1.3 Overview of the Dissertation

The primary goal of the dissertation is to develop an annually nonhomogeneous parameter estimation method for stochastic daily rainfall models at a specific point. A simple first-order Markov-chain occurrence model and a gamma-distribution amount model were used to illustrate the proposed method. The Markov-chain and gamma-distribution models were used to demonstrate the method due to their simplicity, familiarity in agricultural research, and widespread use in the WGEN model and the DSSAT system. Statistical methods to evaluate the parameter estimation method were developed.

The pertinent literature on stochastic rainfall modeling is reviewed in Chapter 2. The construction of an 89-year data base for daily rainfall is discussed in Chapter 3. The underlying premise that long-term variability in daily rainfall exists is investigated in Chapter 4. Nonhomogeneous parameter estimation methods are proposed in Chapter 5. The simulation of time series of nonhomogeneous rainfall model parameters is investigated in Chapter 6. The results of the investigation are presented in Chapter 7. A first-order Markov-chain occurrence model and a gamma-distribution amount model are
implemented to demonstrate the proposed parameter estimation and simulation methods. Statistical methods to evaluate and the proposed stochastic rainfall model are also presented. The conclusions of this research are summarized and discussed in Chapter 8.
CHAPTER 2
REVIEW OF THE LITERATURE

The mathematical description or modeling of daily rainfall at a point has used the theory of stochastic processes and probability theory extensively. The theory of stochastic processes is a well established and growing body of knowledge with introductions given in numerous texts, including Cox and Miller (1965), Medhi (1982), and Parzen (1962). The main emphasis in engineering research activities on stochastic rainfall modeling has been the selection and application of a limited number of simple stochastic processes to observed time series of daily rainfall. These research activities have resulted in a large body of competing literature, but limited information to aid a potential user in selecting a particular model for a specific application. Interest in the areas of agricultural system modeling and water resource management have created a growing demand for stochastic rainfall models.

Very often, rainfall models are selected for their simplicity, ease of use and availability, rather than their applicability or goodness of fit. Some rainfall models are becoming standards without the stringent review usually associated with engineering standards. To address this situation, the evolution of stochastic rainfall models is reviewed and an effort is made to point out the strengths, weaknesses and applicability of existing models.

First, some definitions and theoretical background on stochastic processes are given. According to Webster the word stochastic comes from the ancient Greek "stochastikos" which deals with aiming at a target. The word random is of ancient Germanic origin and comes to the English language through Old and Middle French, also having its origin in the uncertainty in aiming at a target. Webster considers the words,
stochastic and random, to be synonymous as do most texts in probability theory.

The concept of a random variable is central to the theory of stochastic processes and its definition is given following Hogg and Craig (1978). Assume we are given a random experiment with a sample space $S$. A function $X$, which assigns to each element $c$ in $S$ one and only one number $X(c) = x$, is called a random variable. Random variables may be either discrete or continuous. For example, the wet or dry state of a particular day is a discrete random variable, while the amount of rainfall on a wet day is a continuous random variable.

A random or stochastic process is then defined following Parzen (1962) as a collection of random variables $\{X(t), t \in T\}$. The set $T$ is called the index set of the process. If $T = \{0, 1, 2, \ldots\}$, the stochastic process is discrete in time, and if $T = \{t: \geq 0\}$, the stochastic process is continuous in time.

A combination of two stochastic processes is frequently used to model rainfall. The occurrence of wet or dry intervals forms a discrete stochastic process $\{X(t_1), X(t_2), \ldots, X(t_n)\}$ or equivalently $\{X_1, X_2, \ldots, X_n\}$, where the value of $X_t$ represents the wet or dry state of interval $t$. The occurrence process is modeled in discrete time where the interval is an integral multiple of days, and in continuous time where the interval is a real number. The wet day amount of rainfall, $\{Y_1, \ldots, Y_t, \ldots, Y_n\}$, forms a continuous stochastic process where $Y_t$ represents the amount of rainfall received on wet day $t$.

2.1 Rainfall Occurrence Models

Court (1979), Roldan and Woolhiser (1982), and Waymire and Gupta (1981) reviewed the extensive literature on the random nature of rainfall events. Foufoula-Georgiou (1985) classified the evolution of rainfall occurrence models by the nature of the stochastic process used in each model. Following this classification, the evolution and applicability of rainfall occurrence models was reviewed.
2.1.1 Alternating Wet and Dry Interval Models

An alternating-renewal model considers the rainfall status of intervals rather than the occurrence of wet and dry days. Consider two independent sequences of random variables \( \{X_{1}^{D}, X_{2}^{D}, \ldots \} \) and \( \{X_{1}^{W}, X_{2}^{W}, \ldots \} \), where the value of \( X_{i}^{D} \) represents the length of dry interval \( i \) and \( X_{i}^{W} \) the length of wet interval \( i \). Dry interval lengths are distributed according to the probability density \( f_{0}(x) \) and wet intervals are distributed according to \( f_{w}(x) \). Dry and wet intervals alternate, with the length of each interval selected randomly from \( f_{0}(x) \) and \( f_{w}(x) \). There is no correlation between consecutive wet and dry interval lengths as \( f_{0}(x) \) and \( f_{w}(x) \) are considered to be independent.

Various distributions were considered for \( f_{0}(x) \) and \( f_{w}(x) \). The exponential distribution was used for both wet and dry interval lengths by Thom (1958) and Green (1964). Grace and Eagleson (1966) used a Weibull distribution for wet interval lengths. The alternating renewal approach was also considered by Todorovic and Yevjevich (1969), Eagleson (1978), and Galloy et al. (1981). Roldan and Woolhiser (1982) compared an alternating-renewal process using a truncated geometric distribution for \( f_{w}(x) \) and a truncated negative binomial distribution for \( f_{0}(x) \) with a first-order Markov-chain model and found the Markov chain superior. A major problem with the alternating renewal type of model is the difficulty in partitioning the amount of rainfall received in an interval between the days in the interval.

2.1.2 Wet and Dry Day Models

In this class of models, the rainfall status of a particular day was described as a discrete random variable, \( X_{i} \), in discrete time (days). The daily sequencing of rainfall events, \( \{X_{1}, X_{2}, \ldots, X_{i}, \ldots, X_{n}\} \), then formed a random or stochastic process, where the value of \( X_{i} \) represented the rainfall state of day \( i \). Two states, wet and dry, were usually considered for the rainfall series \( \{X_{i}\} \). The rainfall state for day \( i \) was then Dry or Wet,
i.e., \( x_t = j \ (j = D, W) \), while for day \( t - 1 \), the possible states were \( x_{t-1} = i \ (i = D, W) \). The sequencing of wet and dry days was considered to be a dependent process using both Markov chains and autoregressive moving average models.

**Markov-chain models.** A first-order Markov chain is formed, if it is assumed that the probability of a rainfall event on any given day depends only on the state of the previous day. This assumption is the conditional probability

\[
Pr\{X_t = x_t \mid X_{t-1} = x_{t-1}, X_{t-2} = x_{t-2}, \ldots, X_1 = x_1\} = Pr\{X_t = x_t \mid X_{t-1} = x_{t-1}\} \tag{2.1}
\]

If \( X_{t-1} = i \) and \( X_t = j \), then there was a transition from state \( i \) to state \( j \) at day \( t \). The probability of the possible transition is called the transition probability and is written

\[
p_{ij} = Pr\{X_t = x_t \mid X_{t-1} = x_{t-1}\} \tag{2.2}
\]

For a first-order Markov chain with two states, there are four transition probabilities which can be represented by the matrix \( \overline{P} \) given by

\[
\overline{P} = [p_{ij}] = \begin{bmatrix}
p_{DD} & p_{DW} \\
p_{WD} & p_{WW}
\end{bmatrix} \tag{2.3}
\]

where \( p_{DW} = 1 - p_{DD} \) and \( p_{WW} = 1 - p_{WD} \).

Markov chains with multiple states can also be formed. Multiple state models are used to describe different classes of wet and dry days. For example, Schmidt et al. (1987) considered a three state Markov chain. The rainfall states for day \( t \) were DRY, TRACE or WET, i.e., \( x_t = j \ (j = 1, 2, 3) \), while for day \( t - 1 \), \( x_{t-1} = i \ (i = 1, 2, 3) \). In general, a Markov chain with \( s \) states has an \( s \times s \) transition probability matrix of the form
Higher order Markov chains can be formed, if it is assumed that the probability of a rainfall event on any given day depends on the states of $m$ previous days. This assumption is used to form a Markov chain of order $m$ and is the conditional probability

$$Pr\{X_i = x_i \mid X_{i-1} = x_{i-1}, X_{i-2} = x_{i-2}, \ldots, X_1 = x_1\}$$

$$= Pr\{X_i = x_i \mid X_{i-1} = x_{i-1}, X_{i-2} = x_{i-2}, \ldots, X_{i-m} = x_{i-m}\}$$

For an $m$th order Markov chain, the $m$ step transition probability matrix $\overline{P}(m)$ has the form

$$\overline{P}(m) = \overline{P}^m = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1s} \\ p_{21} & p_{22} & \cdots & p_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ p_{s1} & p_{s2} & \cdots & p_{ss} \end{bmatrix}$$

where $\overline{P}$ is the one step transition probability matrix given in equation 2.3 or 2.4.

The Markov-chain type of model is widely used due to its lack of complexity, ease of parameter estimation, and flexibility in order, seasonality, and number of states. Markov chains were implemented with varying degrees of success in the northern United States (Caskey, 1963; Weiss, 1964; Feyerherm and Bark, 1967), southeastern United States (Molten et al., 1985), Florida (Khanal and Hamrick, 1974; Schmidt et al., 1987), Canada (Hopkins and Robillard, 1964), Israel (Gabriel and Neumann, 1957, 1962), and India and Africa (Stern and Coe, 1984). The Markov-chain models appear to perform best in regions, such as the central United States, where rainfall events are strongly
influenced by the passage of fronts. The applicability of these models in the humid subtropics and tropics is under study by the International Benchmark Sites Network for Agrotechnology Transfer program (IBSNAT) (Jones, personal communication), but little is known about the goodness of fit in relation to different meteorological regimes. A pronounced decrease in the ability of a first-order Markov chain to preserve the distribution of wet and dry intervals was found (Schmidt et al., 1987) from north to south Florida. Florida is a transition region, where the influence of passing fronts and tropical disturbances changes from north to south. The need to identify spatial and temporal scales of meteorological regimes in rainfall model selection seems warranted.

Markov chains with multiple states were studied where each state represents a different range of rainfall amounts. With multiple state models, it is necessary to partition the observed rainfall into several classes and fit a distribution to each rainfall class. Khanal and Hamrick (1974) developed a fourteen-state model for Bithlo, Florida, while Haan et al. (1976) and Carey and Hann (1978) studied models with seven and three states, respectively. Molten et al. (1985) carefully evaluated several Markov chains with multiple states at Blacksburg, Virginia, and found that a five-state model performed well. Important drawbacks of multiple state models with multiple seasons are the need to estimate a large number of parameters, the limited sample size for each state, and the selection of class limits for each state. However, simple models with three states instead of two, where a trace state is used in addition to a wet and dry state, performed quite well in Jordan (Stern and Coe, 1984) and Florida (Schmidt et al., 1987).

Lowry and Guthrie (1968), Gates and Tong (1976) and Chin (1977) studied higher-order models. A second-order Markov chain is currently under study by the IBSNAT program (Jones, personal communication) for application to tropical regions. The selection of model order is a difficult problem which was addressed by using the Akaike Information Criterion (AIC) (Tong, 1975) and a likelihood ratio goodness of fit test (Hoel, 1977). In regions where the dominant meteorological regime changes
seasonally, the use of a variable model order seems reasonable. Chin (1977) studied this problem for the continental United States and showed that the optimal model order, using the AIC, varied seasonally and geographically.

Green (1965), Wiser (1965), and Foufoula-Georgiou (1985) criticized the Markov-chain class of models for their ability to predict long wet and dry spells, and the clustering of rainfall events. Alternatively, Markov chains were found to generate dry runs of a month or longer (Jones et al., 1972; Srikantan and McMahon, 1983; Schmidt et al., 1987); however, the distribution of run lengths was not accurately preserved for very short and very long runs. No evaluation of the ability of Markov-chain models to preserve clustering of rainfall events appears to exist.

**Discrete autoregressive moving average models.** The occurrence sequence \( \{X_n\} \) was described by a discrete mixed autoregressive moving average process (DARMA) by Jacobs and Lewis (1978a,b) in the following manner. Let \( \{Y_n\} \) be a sequence of discrete random variables. Also, let \( \{U_n\} \) and \( \{V_n\} \) be sequences of zeroes and ones such that \( \Pr\{U_n = 1\} = \beta \) and \( \Pr\{V_n = 1\} = \rho \). Finally, let \( \{S_n\} \) be the sequence \( \{0,1,...,N\} \). The DARMA(1,N+1) model has the form

\[
X_n = U_n Y_{n-S_n} + (1 - U_n) A_{n-S_n-1}, \quad n = 1, 2, \ldots
\]

where the first term on the right is the moving-average component and the second term is the autoregressive component where

\[
A_n = V_n A_{n-1} + (1 - V_n) Y_n, \quad n = -N, -N + 1, \ldots
\]

The DARMA model was first applied to the occurrence of wet days by Buishand (1978) in the Netherlands. Chang et al. (1982,1984) also applied the model to daily
rainfall in Indiana. Four seasons and three states were considered, for DARMA models of varying order. In three of the four seasons considered, the discrete autoregressive model, DAR(1), proved to be the best model. As Chang et al. (1984) pointed out, the DAR(1) model is equivalent to a Markov chain. For all seasons, the distribution of dry run lengths appears to be well maintained for the single station in Indiana that was considered. The DARMA models along with the Markov-chain models were criticized for their ability to preserve clustering in rainfall sequences associated with the passage of fronts by Foufoula-Georgiou (1985).

2.1.3 Point Process Models

A point process model considers a sequence of instantaneous events occurring at points in time. An event occurs when rainfall exceeds some predetermined threshold value. If events occur at times $t_1, t_2, t_3, \ldots$, then a random variable $X_r = t_r - t_{r-1}$ ($r = 1, 2, 3, \ldots$) can be defined as the time between events. The theory of point process is extensive and is considered in most introductory texts on stochastic processes including Cox and Miller (1965), Cox and Lewis (1978), Cox and Isham (1980), Cinlar (1975), Medhi (1982), Parzen (1962), Lewis (1972), and Vere-Jones (1970). For at least twenty years, point process theory was applied to rainfall using a continuous time index, and more recently using a discrete (daily) time index. In the last ten years, most theoretical studies of rainfall have used the point process approach, rather than the traditional wet/dry day Markov-chain approach.

Continuous point processes. If the time index is real valued, then the time between events is also a real number and the point process is continuous. A straightforward and widely used continuous point process is the Poisson process. In the Poisson process, the times between events are considered to be independent random variables with an exponential distribution, and the number of events in time interval $t$ are modeled as a Poisson distribution. The exponential distribution used to model the interarrival times, $X_i$, has the form
\[ f(x) = \lambda e^{-\lambda x} \quad \lambda > 0 \]

where \( \lambda \) is the rate of occurrence or the inverse of the mean interval length. The number of events in time \( t \), \( N_t \), has a Poisson distribution of the form

\[ Pr \{N_t = k\} = \frac{(\lambda t)^k e^{-\lambda t}}{k!} \quad k = 0, 1, 2, \ldots \]

where \( \lambda t \) is the mean number of events in time \( t \). The basic Poisson process was applied to rainfall occurrences by Todorovic and Yevjevich (1969) and Gupta and Duckstein (1975).

Daily rainfall events were observed to cluster (Kavvas and Delleur; 1975), where secondary rainfall events are distributed about a primary event. An important development in continuous point process models was the inclusion of clustering in the occurrence model. The theoretical basis for a clustered process was given by Neyman (1939) and is known as a Neyman-Scott process. This process was applied to rainfall occurrences by LeCam (1961) and Kavvas and Delleur (1975). Primary rainfall events or cluster centers are distributed in time according to the Poisson process described above. Secondary events are then clustered about a primary event with the number of events and their locations specified by a secondary point process. Cluster centers or primary events correspond to passing fronts, with a number of associated secondary storms. Kavvas and Delleur (1981) used a geometric distribution for the number of secondary events in a cluster and an exponential distribution for the location of these events.

Another important class of continuous point process models is the renewal Cox model with Markovian intensity introduced by Smith (1981) and Smith and Karr (1983).
This is a double stochastic Poisson process where wet and dry intervals alternate, with the interval lengths distributed exponentially. During a wet interval the secondary Poisson process is used to distribute rainfall events.

The development of continuous point process models for rainfall occurrences is important because they address the distribution of wet and dry interval lengths directly and include clustered rainfall events within a wet interval. However, continuous time models were criticized (Foufoula-Georgiou, 1985; Foufoula-Georgiou and Lettenmaier, 1986; and Foufoula-Georgiou and Gutterp, 1986) on the basis that most rainfall data are recorded on a discrete time (daily) basis. A discrete time alternative to the continuous time class of models is considered in the next section.

Discrete point processes. If the time index is integer valued (days) rather than real, the time between events is integer valued and the point process is discrete. An event is defined as a day during which a measurable amount of rainfall is observed. In discrete point process models the discrete analog of the Poisson process, the Bernoulli process, is used. In the Bernoulli process, the times between events are considered to be independent random variables with a geometric distribution, and the number of events in time interval \( r \) is modeled as a binomial distribution. The geometric distribution used to model the number of trials (days), \( X_n \), between the \( n \)th and \((n + 1)\)st successes (rainfall event) has the form

\[
Pr\{X_n = k\} = p(1 - p)^{k-1} \quad k = 1, 2, \ldots
\]

where \( p \) is the probability of success (rainfall event) at each trial. The number of successes (events) in \( r \) trials, \( N_r \), have a binomial distribution of the form

\[
Pr\{N_r = k\} = \binom{r}{k} p^k (1-p)^{r-k} \quad k = 0, 1, 2, \ldots
\]
The discrete-time Bernoulli process describes the occurrence of rainfall events or successes in a completely random manner, with the time between events, $X_n$, independent of previous events. This lack of time dependence in rainfall occurrences is not desirable, and it has led to the use of Markov-renewal models. Markov-renewal models use a Markov chain to provide time dependence to the sequencing of intervals given by a counting process such as the Bernoulli process.

Markov-renewal models were first studied by Smith (1955) with later investigations by Pyke (1961 a,b), Cox (1963), Teugels (1976), Cinlar (1975), and Cox and Lewis (1978). The application of a Markov renewal model to rainfall occurrences is based on the premise that several types of rainfall events exist resulting from different physical generating processes.

Foufoula-Georgiou and Lettenmaier (1987) studied a Markov-renewal model for rainfall occurrences using the discrete time point process framework. The justification for using a two-state Markov renewal model was stated by Foufoula-Georgiou and Lettenmaier:

Daily rainfall occurrences are the result of the interaction of several rainfall-generating mechanisms. For example, the first rainy day in a wet period may be the result of a frontal storm passing over a region, whereas subsequent rainy days in the same wet period may be considered secondary events. In that sense, times between events may come from different probability distributions, ... (1987, p. 875)

They assumed that there are two types of interarrival times or intervals, $\{X_i\}$, sampled from two different geometric distributions, $f_1(x)$ and $f_2(x)$. Historical rainfall data were analyzed (Foufoula-Georgiou, 1985) to justify the use of the geometric distributions.

In a 2-state Markov-renewal model, interval types do not strictly alternate, but Markov-chain transition probabilities are used to determine the sequencing of interval types. For two types of rainfall intervals the Markov-chain transition probability matrix, $\tilde{P}$, has the form
where $\alpha_{11} = a_1$ is the probability of making a transition from a Type 1 to a Type 1 interval and $\alpha_{22} = a_2$ is the probability of making a transition from a Type 2 to a Type 2 interval. The probability density function of the two types of intervals is given by the mixture

$$f(x) = e_1 f_1(x) + e_2 f_2(x)$$
threshold amount of daily rainfall to be an important factor in the performance of the models considered. Wet and dry days in historical records are defined by setting a threshold amount of rainfall, above which the day is considered wet. Smith found that for small (0.254 mm) thresholds the Markov-Bernoulli and Markov-chain models gave very similar results. For large (25.4 mm) thresholds the simple Bernoulli trial model performed as well as the more complex models. The Markov-Bernoulli model was found to be superior only for moderate thresholds (2.54 mm).

Markov-renewal and Markov-Bernoulli models are very general models, which contain Markov-chain models as a subclass. Smith's (1987) findings on the importance of wet day thresholds is an important factor to be considered by users of stochastic rainfall occurrence models. Extension of Smith's model comparison studies to Foufoula-Georgiou's Markov-renewal model would benefit users selecting a rainfall model for a specific application. The National Weather Service uses a 0.254 mm (0.01 in) threshold in reporting daily rainfall data, and this threshold is routinely used in Markov-chain rainfall models such as WGEN and CLIMATE. A small (0.254 mm) daily rainfall threshold is important in general weather simulation applications, because the occurrence of a wet day is used to condition daily temperature and solar radiation.

2.2 Rainfall Amount Models

The rainfall process is a combination of two stochastic processes, 
\{(X_t, Y_t); t = 1, 2, \ldots\}, where \(X_t\) is the wet or dry state of day \(t\) and \(Y_t\) is the amount of rainfall received on day \(t\). The rainfall amount model describes the distribution of rainfall on days that are given as wet by the occurrence model. The wet day amount of rainfall may be considered to be independent or dependent of the state of the previous day.

2.2.1 Chain-dependent Models

If the amount model is dependent on the state of the previous day, the cumulative distribution for the wet day amount has the form
where two different distributions are used, \( F_D(y) \) if the previous day was dry and \( F_W(y) \) if the previous day was wet. The importance of the chain-dependent nature of the wet day amount model was studied by Katz (1977 a,b), Haan et al. (1976), Buishand (1978), Woolhiser and Roldan (1982), and Guzman and Torrez (1985). The majority of the results suggest that there was no significant difference between the independent and chain-dependent models. However, both Chin and Miller (1980) and Foufoula-Georgiou (1985) reported some evidence of a dependent structure for the Pacific northwest, USA. The question of time dependence in rainfall amount models is important and deserves further study, since most of the locations considered in the above studies were from the continental USA.

2.2.2 Chain-independent Models

If the amount model is independent of the state of the previous day, the cumulative distribution of the wet day amount of rainfall has the form

\[
F(y) = Pr\{Y_i \leq y \mid X_i = \text{WET}\}
\]  

The wet day amount of rainfall generally follows an "exponential type" of distribution in which there is a high incidence of days with small amounts of rainfall and a low incidence of days with large amounts of rainfall. Several distributions for the wet day amount of rainfall were studied, including the exponential, gamma, Weibull, and mixed exponential distributions.

**Exponential distribution.** The random variable \( Y \) is said to have an exponential distribution if the probability density function is
where the model parameter, $\mu$, is the wet day mean amount of rainfall. Todorovic and Woolhiser (1971), Skees and Shenton (1974), and Mielke and Johnson (1974) studied the exponential distribution. Richardson (1981) used the exponential distribution for wet day amounts of rainfall in the original version of WGEN. Mielke and Johnson (1974) reported that the tail of the exponential distribution was thinner than observed distributions of daily rainfall amounts. This suggests that a rainfall model using an exponential amount model would underestimate the number of days with large amounts of rainfall.

**Gamma distribution.** The random variable $Y$ is said to have a gamma distribution if the probability density function is

$$f(y) = \begin{cases} \frac{1}{\mu} e^{-y/\mu}, & \mu > 0; 0 \leq y < \infty \\ 0, & \text{elsewhere} \end{cases} \quad 2.17$$

where the model parameters are $\alpha$, which controls the shape of the distribution by decreasing the skewness of $f(y)$ as $\alpha$ increases, and $\beta$, which scales the amount of rainfall $y$. The exponential distribution is a special case of the gamma distribution for $\alpha = 1$. The gamma distribution was used by Ison et al. (1971), Buishand (1978), and Carey and Haan (1978). The current version of the WGEN weather generation model uses the gamma distribution (Jagtap et al., 1988), however the shape parameter, $\alpha$, is constrained to be less than one at all times.

**Weibull distribution.** The random variable $Y$ is said to have a Weibull distribution if the probability density function is

$$f(y) = \begin{cases} \frac{y^{\alpha-1}e^{-y/\beta}}{\beta^\alpha \Gamma[\alpha]}, & \alpha, \beta > 0; 0 \leq y < \infty \\ 0, & \text{elsewhere} \end{cases} \quad 2.18$$
\[ f(y) = \begin{cases} \frac{a}{b} y^{b-1} e^{-y^{b}} & a, b > 0; 0 < y < \infty \\ 0 & \text{elsewhere} \end{cases} \]

where the model parameters are \( a \) and \( b \). The shape of the Weibull distribution is similar to that of the gamma distribution. The probability density function can be exponential, skewed or symmetrical as the parameter \( a \) increases. The Weibull distribution and the mixed exponential distribution were compared by Foufoula-Georgiou (1985) and the mixed exponential was found to be superior.

**Mixed exponential distribution.** The random variable \( Y \) is said to have a mixed exponential distribution (Everitt and Hand, 1981) if the probability density function is

\[ f_{s}(y) = \begin{cases} \frac{p}{\mu_{1}} e^{-y/\mu_{1}} + \frac{1-p}{\mu_{2}} e^{-y/\mu_{2}} & 0 \leq y < \infty \\ 0 & \text{elsewhere} \end{cases} \]

where the model parameters are \( p \), which is a weighting parameter for the two distributions varying between 0 and 1, \( \mu_{1} \) which is the mean of the smaller exponential distribution, and \( \mu_{2} \) which is the mean of the larger distribution. The overall wet day mean is related to the mixed exponential distribution parameters by

\[ \mu = p \mu_{1} + [1 - p] \mu_{2} \]

Smith and Schreiber (1973), Woolhisser and Pegram (1979), Woolhisser and Roland (1982) and Foufoula-Georgiou (1985) studied the mixed exponential distribution. Woolhisser and Roldan (1982) and Foufoula-Georgiou (1985) compared the mixed exponential with the exponential, gamma and Weibull rainfall amount distributions and found the mixed exponential superior in most but not all cases.
In addition to these distributions, the kappa or generalized beta distribution was studied by Mielke (1973) and Mielke and Johnson (1974), and the incomplete gamma distribution was used by Jones et al. (1972) with success.

2.3 Rainfall Model Parameter Estimation

In addition to the application of the theory of stochastic processes to daily rainfall at a point, the application of parameter estimation theory has also received attention. Estimation theory is considered in most texts on mathematical statistics, such as Mendenhall et al. (1981) and Mood et al. (1974). In estimation theory, it is first assumed that some characteristic of rainfall such as the wet day amount of rainfall or time between rainfall events can be described mathematically or modeled by a random variable $X$ with the probability density or mass $f_X(x; \theta_1, \theta_2, \ldots, \theta_r)$ where $\theta_i (i = 1, 2, \ldots, r)$ are the parameters of the model. Next, a limited sample $(x_1, x_2, \ldots, x_n)$ is selected that is used to estimate the model parameters. Finally, some method or rule is used to calculate an estimate of the parameters from the selected sample.

Parameter estimates are basically of two types, one a point estimate and the second an interval estimate (Mood et al., 1974). Point parameter estimates have been studied extensively for stochastic rainfall models and are reviewed in the following section of this chapter. Interval estimates assume that a model parameter is a random variable and that there is some interval of uncertainty associated with each parameter estimate. Interval estimates of model parameters have not been studied in conjunction with rainfall models, but will be considered in Chapter 5 of this work.

2.3.1 Sample Selection

Meteorological conditions in most regions change throughout the year, resulting in some type of intra-annual seasonal pattern of rainfall events. To account for this seasonal nature of rainfall, most implementations of stochastic rainfall models identify a number
of seasons within a year. The observed time series of daily rainfall is then sorted by season into subsamples, and model parameters are estimated using each seasonal subsample.

Seasons may be defined to coincide with observed temporal changes in rainfall patterns or they may be defined in an arbitrary manner. The number and length of the seasons will affect both the accuracy of the parameter estimate, if numerous short seasons are used, and the resolution of the annual parameter cycle, if a few long seasons are considered. Haan et al. (1976) considered season definition an important question and urged further study. Weekly seasons were used by Jones et al. (1972), biweekly seasons were used by Woolhiser in the model CLIMATE (Woolhiser et al., 1985, 1988), and Richardson used monthly seasons in the model WGEN (Richardson and Wright, 1984). Limited application of seasons with irregular lengths based on the analysis of rainfall data have been used by Gabriel and Neumann (1962), Foufoula-Georgiou (1985), and Schmidt et al. (1987).

2.3.2 Parameter Estimation

Once seasons are defined and subsamples selected from the observed data, parameter estimates are made using graphical procedures, method of moments, maximum likelihood methods or minimum chi-square methods. Moment and maximum likelihood parameter estimation methods have been used to calculate parameter estimates for stochastic rainfall models.

**Moment estimates.** Moment estimates of the parameters of density or mass function $f_X(x; \theta_1, \theta_2, \ldots, \theta_r)$ are obtained by equating the population and sample moments and solving for the parameters $(\theta_1, \theta_2, \ldots, \theta_r)$. The $r$th population moment about the origin for a discrete or continuous random variable is given by

$$
\mu_r = E(X^r) \equiv \sum_{j=1}^{n} x_j^r f(x)
$$

2.22
\[ \mu_r = E(X^r) = \int_{-\infty}^{\infty} x^r f(x) \]

and is given in most texts on probability and statistics. The corresponding \( r \)th sample moment is the average and is given by

\[ m_r = \frac{1}{n} \sum_{i=1}^{n} y_i^r \]

If \( r \) parameters are to be estimated, then \( r \) moment equations are used with the form

\[ m_r = \mu_r \]

This gives \( r \) equations in \( r \) unknowns which can be solved for the parameter set \{\( \theta_1, \theta_2, \ldots, \theta_r \}\).

The method of moments for parameter estimation is straightforward in application and works when other methods fail, but it does not always lead to estimates that have the smallest possible variance or uncertainty associated with the estimate. To obtain the "best" parameter estimate or the estimate with the smallest possible variance, maximum likelihood estimates have been used.

**Maximum likelihood estimates.** Consider the seasonal subsample \{\( x_1, x_2, \ldots, x_s \}\) which is assumed to have the joint probability distribution

\[ f_{x_1, x_2, \ldots, x_s}(x_1, x_2, \ldots, x_s; \theta_1, \theta_2, \ldots, \theta_r) \]

If it is assumed that the data points in the sample are independent, then the joint distribution can be written
which is the likelihood that the particular sample selected would be obtained by randomly sampling the population. The likelihood function is defined as

\[ L(\theta_1, \ldots, \theta_r) = \prod_{i=1}^{n} f(x; \theta_1, \ldots, \theta_r) \]

It is then desirable to choose parameters which maximize the likelihood of obtaining the observed sample. Since many distributions contain an exponential term, the log likelihood function, \( \ln L \), is usually maximized. The log likelihood function may be maximized by setting the derivatives with respect to each parameter equal to zero and solving a set of \( r \) equations in \( r \) unknowns. In practice, a numerical function optimization routine is usually employed.

2.3.3 Parameter Interpolation

Consider the collection of seasonal parameter sets \( \Theta^k = \{ \theta^k_1, \theta^k_2, \ldots, \theta^k_r \} \) for a model with \( r \) parameters and \( k \) seasons. The parameter set may be used directly in this discrete form, in which case the parameters are constant throughout each season. This results in discontinuous jumps in parameter values between each season (Fig. 2.1.a). Alternatively, an approximation function of the form \( \Theta(n) = \{ \theta_1(n), \theta_2(n), \ldots, \theta_r(n) \} \ (n = 1, 2, \ldots, 365) \) may be used to allow the parameter set to vary in a smooth and continuous manner from day to day (Fig. 2.1.b). Several approaches to fitting approximation functions to the parameter set \( \Theta^k \) have been used and are considered in the following.

Polynomial approximation. Given the set of discrete model parameters \( \{ \theta^1, \ldots, \theta^t, \theta^s \} \) where \( S \) is the number of seasons, a continuous function, \( \theta(n) \) \((n = 1, \ldots, 365)\) is required from which daily parameter values can be determined. An effective approach uses a polynomial with varying order \( M \) of the form
Figure 2.1  Annually homogeneous Markov-chain transition probability estimates using data from Tallahassee, FL (Woolhiser, 1979).
a. seasonally discrete; b. continuous.
Polynomials of varying order were fit to seasonal parameter estimates by Jones et al. (1972) and Stern (1980 a,b). Polynomial approximations provide a straightforward method in which the order of the polynomial can be selected in a statistically significant manner. However, higher order polynomials may be required if the discrete seasonal parameters change abruptly in time.

**Fourier series approximation.** The set of seasonal model parameters, \( \{\theta^1, ..., \theta^s, ..., \theta^S\} \) where \( S \) is the number of seasons, can also be approximated by a Fourier series, \( \theta(n) \) which has the form

\[
\theta(n) = A_0 + \sum_{j=1}^{m} \left( A_j \cos \left( \frac{n j}{T} \right) + B_j \sin \left( \frac{n j}{T} \right) \right)
\]

where \( A_0 \) is the annual mean of \( \theta \), \( A_j \) and \( B_j \) are amplitudes estimated by least squares, \( T = 365/2\pi \), \( m \) is the number of harmonics used to specify the parameter, and \( n \) is the day of the year. An equivalent and commonly used form is

\[
\theta(n) = C_0 + \sum_{j=1}^{m} \left( C_j \sin \left( \frac{n j}{T} + D_j \right) \right)
\]

where \( C_j \) are amplitudes and \( D_j \) are phase angles. The theory of Fourier series and least-squares parameter estimation is given in most engineering mathematics texts, such as Kreyszig (1979).
Fourier series approximations have been extensively used to characterize the periodic seasonal variations in rainfall statistics and model parameters. Horn and Bryson (1960), Sabbagh and Bryson (1962) and Fitzpatrick (1964) used Fourier series to model seasonal variations in monthly mean rainfall. Feyerherm and Bark (1965), Fitzpatrick and Krishnan (1967), and Woolhiser et al. (1973) used this approach to describe seasonal variations in first-order Markov chain transition probabilities, while Ison et al. (1971) and Woolhiser et al. (1973) used it in conjunction with wet day amount model parameters.

A major problem in the application of Fourier series approximation functions is the selection of significant harmonics. Woolhiser and Pegram (1979) make the point that there is no statistically significant method for testing the significance of individual harmonics. Frequency domain analysis has been widely used to estimate the spectral density associated with individual harmonics, but this approach has not been widely used in stochastic rainfall modeling.

**Combined Fourier series maximum likelihood approximation.** If it is assumed that a model parameter, θ, is a random variable, then there is some variance associated with each seasonal parameter estimate and the variance may change from season to season. Fitting an approximation function such as a polynomial or Fourier series to a set of discrete seasonal parameters does not account for the variance of the parameter estimate. To address this problem Woolhiser and Pegram (1979) proposed a method combining Fourier series approximation and maximum likelihood estimation.

Woolhiser's method consists of substituting Fourier series approximations for each model parameter into the likelihood function of the particular model under consideration. The observed data is then used in conjunction with a numerical optimization routine to maximize the likelihood function giving direct estimates of the Fourier coefficients. This procedure is repeated for an increasing number of harmonics, and a likelihood ratio test is used to select a significant number of harmonics. The combined log likelihood function, $U$ has the form
\[ U = \ln L[\bar{X} = \bar{x} | \Phi] \]

where \( \bar{x} \) is a vector containing the daily rainfall data, \( L \) is the likelihood function for the model being considered (equation 2.28), and \( \Phi \) is a vector containing the Fourier series coefficients.

Woolhiser and Roldan (1982, 1988) and Hanson et al. (1988) used the direct maximum likelihood method of estimating Fourier coefficients. This method considers the stochastic rainfall model parameters as random variables with an associated variance that may change seasonally. By using the maximum likelihood method, parameters are selected which minimize the variance. However, the significance or sensitivity of the rainfall model to seasonal changes in variance is not addressed.

All of the parameter estimation and interpolation methods studied assume the annual cycle (Fig. 2.1) of a model parameter is periodic with a period of one year. A stochastic process or model based on this assumption is classified as annually homogeneous (Parzen, 1962) or annually time invariant. The parameters of most probability distributions (Mood et al., 1963) and hydrologic models in general (Haan, 1989) are considered to be random variables with some associated uncertainty or variance. This basic principle of parameter estimation theory has not been considered in the application of stochastic processes to rainfall processes.

2.4 Rainfall Model Applications

The application of the theory of stochastic processes to the phenomena of daily rainfall has resulted in an extensive but inconclusive body of literature on stochastic rainfall models. As in many areas of engineering applications, the immediate need for a working stochastic rainfall model has forced the user to select a particular model based on incomplete information. The model selected may not be the most mathematically
elegant or provide the best possible fit to the historical data, but practical decisions were necessary in a field where there is considerable theoretical diversity and minimal comparative evaluation.

Increased concern over global-scale climate changes and accelerated research in the area of agricultural system modeling have increased the need for the best possible stochastic rainfall model. Jones et al. stated this need twenty years ago saying,

Weather is a primary forcing function for decisions concerning crop production. Plant models have been developed to simulate plant production based upon the weather that is imposed upon the "computer plant". For studying a complete crop production system, it is desirable to have an environmental model to provide simulated field conditions for a particular location in which the "computer plant" is to be grown. (1972, p. 372)

Agricultural system models are complex mathematical descriptions of the atmosphere-plant-soil continuum. They contain components which describe plant production, nutrient utilization, water flux, and weather. The weather component is often composed of models for temperature and solar radiation, as well as rainfall.

Very little attention has been given to the development of daily weather generation models which recognize the interaction between rainfall, solar radiation, temperature, and evapotranspiration. One of the first models of this type was developed by Jones et al. (1972) in which daily rainfall, temperature and evaporation were generated simultaneously. A first-order Markov-chain occurrence model and incomplete gamma distribution amount model with weekly seasons were used in this model. Jones et al. emphasized the importance of the rainfall component of this type of model saying,

Based on the literature review, it was hypothesized that rainfall is the most basic weather variable, independent of evaporation and temperature, and dependent only on the time of year and the tendency for the persistency of rainfall described by the Markov chain. (1972, p. 367)

Following Jones et al. (1972), Richardson (1981) proposed a weather generation model (WGEN) for rainfall, temperature and solar radiation. This model used a two-state first-order Markov-chain occurrence model and an exponential amount model with monthly parameter estimates approximated by a Fourier series. More recently,
Woolhiser et al., 1985, 1988) proposed a similar model in which a two-state first-order Markov-chain occurrence model was combined with a mixed exponential amount model. Fourteen day seasons were used with combined Fourier series, maximum likelihood parameter estimates.

The most widely used agricultural or crop system model is the Decision Support System for Agrotechnology Transfer (DSSAT) developed by the International Benchmark Sites Network for Agrotechnology Transfer (IBSNAT) (Jagtap et al., 1988). This model contains a weather generation component based upon WGEN, which was originally developed by Richardson (1981) and Richardson and Wright (1984). The rainfall component of WGEN is composed of a two-state, first-order Markov-chain occurrence model and a gamma distribution amount model. Monthly seasons are used in WGEN with discrete, seasonally homogeneous parameter estimates for the rainfall component. In all, cases the two-state, first-order Markov-chain type of model is used with little attention to order selection, state classification, or season definition.

2.5 Rainfall Model Critique

Research on stochastic modeling of daily rainfall at a point has considered basically two problems. First, a limited number of stochastic processes have been applied to the occurrence of rainfall events and the wet day amount of rainfall. Second, statistical estimation theory has been used to study seasonal point parameter estimation methods using observed rainfall time series from a limited number of locations. Although there are several excellent review articles, the extensive stochastic rainfall modeling literature provides limited guidance to the potential user in selecting the best possible model for a particular location or application.

This literature review suggests that there are basically two types of models, the Markov chain and discrete point process, that show promise for practitioners. The Markov chain was studied extensively for several decades, while the discrete point process was considered only in the last five years, primarily by Foufoula-Georgiou and
Smith. The main difference in these two types of models appears to be the accurate reproduction of rainfall event clustering. There has been no systematic comparison made to demonstrate this point. Each of these models was evaluated at a limited number of locations with positive results.

Model fitting or parameter estimation research has focused exclusively on point parameter estimates. There has been no consideration of interval parameter estimates that recognize that model parameters are in fact random variables with an associated variance or uncertainty that may change throughout the year. Arbitrary, equal-length seasons were studied extensively, resulting in models with a large number of parameters. To attain more parsimonious models, preliminary data analysis and consideration of regional meteorological cycles to define appropriate seasons is suggested. Model parameters were also assumed to vary seasonally in a smooth, continuous manner with Fourier series used to approximate an annual parameter cycle. If model parameters vary abruptly from season to season, or if the seasonal parameters are not normally distributed, seasonally discrete parameter estimates may be superior to continuous Fourier series approximations as shown in Figure 2.1 for Tallahassee, Florida. Finally, rainfall model parameters were assumed to be annually homogeneous or interannually time invariant. This assumption ignores the possibility of long-term trends or naturally occurring multi-year cycles in rainfall events.

Very little attention has focused on a systematic approach to model selection and parameter estimation. For example, most of the extensive research on Markov-chain rainfall models was conducted by selecting the model order, number of states and seasons in a somewhat arbitrary manner. In contrast to this approach, Chin (1977) has shown, using the AIC, that the optimal model order does vary seasonally and geographically for the continental United States. It is therefore suggested that the following steps be considered in stochastic rainfall modeling building:

1. define a wet day threshold for the particular application;
2. select the number and length of seasons based on preliminary analysis of historical data and meteorological conditions;

3. select model order using criteria such as AIC;

4. select the number of states for each season;

5. select rainfall classes for each state;

6. estimate model parameters once model form is selected;

7. generate time series of daily rainfall and evaluate model performance;

8. update form of model if necessary and repeat the procedure.

This approach to rainfall model building suggests that there is not a particular model that is suitable for all locations, seasons or applications.
CHAPTER 3
DATA

To investigate and model the long-term (decades) variability of rainfall, the longest possible time series of daily rainfall was sought as a statistical sample. Through careful review of historical weather records, 89 year time series of daily rainfall (1900 to 1988) were compiled for four locations in Florida (Gainesville, Orlando, Tampa, Fort Myers). The sources used to compile these samples, the completeness of the records, the location of the study stations and the methods used to compile and verify the historical time series are discussed below.

3.1 Data Sources

The daily historical rainfall data used in this study were obtained from several primary sources. Because the administration and reporting practices of the U. S. Weather Bureau/National Weather Service have changed considerably during the twentieth century, the publications used to construct the data base are discussed in detail. From January, 1900 through January, 1906, data were obtained from the Florida Section of the Climate and Crop Service of the Weather Bureau, published by the U. S. Department of Agricultural. Beginning in February, 1906 and continuing through June, 1909, the title of the above publication was changed to the Florida Section of the Climatological Service of the Weather Bureau, again published by the U. S. Department of Agriculture. The Monthly Weather Review, published by the Weather Bureau under the authority of the U. S. Department of Agriculture, was used as a source of daily rainfall from July, 1909 to December, 1913. Incomplete or missing documents in the University of Florida Library required this shift to the Monthly Weather Review. From January, 1914 through 1988 the primary source for daily rainfall data was Climatological Data: Florida Section,
published by the Weather Bureau under the authority of the U. S. Department of Agriculture until 1940 and administered by the U. S. Department of Commerce since then. In 1970, the U. S. Weather Bureau became known as the National Weather Service, administered by National Oceanic and Atmospheric Administration.

3.2 Data Stations

To gain some understanding of spatial variability of long time series of daily rainfall, four historical weather stations were selected throughout the Florida peninsula. The National Weather Service has grouped weather stations into divisions which are areas within a state which have similar climatological characteristics. One data station was selected from each of four divisions. The location and descriptive information on each station are given in Figure 3.1 and Table 3.1. Historical information on the location of weather stations was taken from the U. S. Department of Commerce publication Substation History, Florida.

Throughout the twentieth century, some of the weather stations used in this study were relocated. In all cases, the relocation distance (length scale of 1 km) was small compared to the length scales of weather systems influencing rainfall patterns in Florida. The relocation distances were ignored and the daily rainfall time series were concatenated to form a single series for each of the four locations. The details of the relocations are discussed below.

During the study period, there were minor changes in the location of the Gainesville station (Table 3.1). Station number 3316, operated by the Agronomy Department of the University, was used for the period from January, 1900 until May, 1957. From June, 1957 through 1988, station number 3322 was used. In February of 1970, station number 3322 was relocated from 2SW to 3WSW. The daily rainfall amount at this station was recorded in the morning which differs from the standard National Weather Service practice of recording the daily amount at midnight.
Figure 3.1 Station location map (modified from U. S. Department of Commerce, 1988).
Table 3.1 Description of weather stations used in the data base.

<table>
<thead>
<tr>
<th>Division Number</th>
<th>Location and Name</th>
<th>Index Number</th>
<th>Latitude (°, ′ N)</th>
<th>Longitude (°, ′ W)</th>
<th>Elevation (ft)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Gainesville</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>University</td>
<td>3316</td>
<td>29 39</td>
<td>82 21</td>
<td>165 (175?)</td>
</tr>
<tr>
<td></td>
<td>2 SW</td>
<td>3321</td>
<td>29 38</td>
<td>82 22</td>
<td>135</td>
</tr>
<tr>
<td></td>
<td>3 WSW</td>
<td>3321</td>
<td>29 38</td>
<td>82 22</td>
<td>86</td>
</tr>
<tr>
<td>3</td>
<td>Orlando</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Water Plant</td>
<td>6633</td>
<td>28 33</td>
<td>81 21</td>
<td>111</td>
</tr>
<tr>
<td></td>
<td>WSO AP</td>
<td>6638</td>
<td>28 33</td>
<td>81 20</td>
<td>110</td>
</tr>
<tr>
<td></td>
<td>WSO McCoy</td>
<td>6628</td>
<td>28 27</td>
<td>81 19</td>
<td>100</td>
</tr>
<tr>
<td>4</td>
<td>Tampa</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Post Office</td>
<td>27 57</td>
<td>82 27</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Davis Island</td>
<td>27 55</td>
<td>82 27</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>WSO AP</td>
<td>8788</td>
<td>27 58</td>
<td>82 32</td>
<td>19</td>
</tr>
<tr>
<td>5</td>
<td>Fort Myers</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Water Plant</td>
<td>3181</td>
<td>26 37</td>
<td>81 51</td>
<td>?</td>
</tr>
<tr>
<td></td>
<td>FAA AP</td>
<td>3186</td>
<td>26 35</td>
<td>81 52</td>
<td>15</td>
</tr>
</tbody>
</table>


NOTE: The symbol ? was used in the source reference to indicate missing or uncertain information.
There were also minor changes in the location of the Orlando station (Table 3.1). Station number 6633, located at the Orlando water plant, was used from 1900 to 1942. From January, 1943 through January, 1974, station number 6638 was used. In February, 1974, station number 6638 was replaced with station number 6628. The daily rainfall amount at this station was also recorded in the morning which differs from the standard National Weather Service practice of recording the daily amount at midnight.

The weather station at Tampa experienced minor relocations (Table 3.1) during the study period (C. Eagleton, NWS, Tampa, personal communication). From 1900 to 1940 the weather station was located at the post office in downtown Tampa. During the Second World War, the station was relocated to Peter O’Night Airport on Davis Island (2.5 SSW). From 1947 until the present time, the weather station has been located at the Tampa International Airport (4.4 WNW). Tampa is a standard National Weather Service station at which daily rainfall is recorded for the 24-hour period, midnight to midnight.

The Ft. Myers station, index number 3181, was originally located at the city water plant (Table 3.1). On December 7, 1959, the Fort Myers station was moved 3500 feet to the SW. The daily amount of rainfall was recorded in the morning, similar to the procedure used at Gainesville and Orlando.

### 3.3 Data Base Construction

The primary historical data sources discussed above were not available in computer compatible format for analysis. It was therefore necessary to construct secondary data sources or computer compatible files to conduct this study. The methods used to construct these files, possible sources of errors and methods used to check the data base for errors are discussed below.

Daily rainfall data from 1900 to the 1930’s, and from 1987 to 1988 were available only in bound atlases published by the National Weather Service and were not available in computer compatible format. Daily data from this period were entered by hand using a spread sheet. This spread sheet was then translated to a formatted ASCII file which was
printed for data base validation. After data entry was completed, a two person "reader-checker" system was used to compare every daily entry in the data base with the primary data source. This approach was assumed to produce a highly accurate duplication of the primary data source. Based on the results of the validation procedure, corrections were made to the spread sheet file and translated to a corrected formatted ASCII file. Daily rainfall amounts reported as a trace or less than 0.254 mm (0.01 in) were entered into the data base as 0.127 mm (0.005 in).

Daily rainfall data were available for the later part of the twentieth century in computer compatible format from Climatedata (Earth Info Inc.). This data base contained daily rainfall data for Gainesville, Tampa, and Fort Myers from 1933 to 1986, and data for Orlando from 1949 to 1987. Climatedata was verified against the primary sources described in Section 3.1 by checking monthly total amounts of rainfall. A few missing data points were identified in Climatedata, however all missing data were recovered from the primary sources.

3.4 Data Quality

The primary data sources from the National Weather Service were found to be very complete. From 1900 to 1910 there were several instances where the reported daily amount of rainfall was included in the next day’s observation. In all cases, the day in question was taken as dry and the succeeding day was recorded as given without any effort made to apportion the rainfall amount between the two days in question. The number of days with combined amounts of this type by station were: 25 days for Gainesville; 2 days for Orlando; no days for Tampa; and no days for Fort Myers.

For the four stations considered, Gainesville, Orlando, and Tampa reported no periods when observations were not made, or when data were missing. The primary source did not list the Fort Myers station for May and June of 1912. It was not possible to determine if the Fort Myers station was operating during this time, or if these observations were accidently excluded from the primary source (National Weather
Service, personal communication). During this time, there were very few weather stations operating in the southwest region of Florida. For this reason, classical approaches for interpolating missing data from surrounding stations were not considered appropriate. Daily rainfall data for May and June were generated using the WGEN model (Jagtap et al., 1987). Observed data from 1913 to 1988 for Ft. Myers were used for parameter estimation in the WGEN model. This substitution represented less than 0.2% of the total rainfall time series used in the data base.
CHAPTER 4
PROBLEM JUSTIFICATION:
LONG-TERM VARIABILITY OF RAINFALL

The underlying premise of this research was the existence of significant long-term variability in rainfall. Before attempting to formulate a stochastic rainfall model which preserves long-term variability in rainfall, the validity of this basic premise needs justification. The objective of this chapter was to quantitatively investigate the degree of long-term variability in rainfall.

4.1 Rainfall as a Stochastic Process

Rainfall at a point may be considered to be continuous-time intermittent process with intensity $\xi(t)$ (Foufoula-Georgiou, 1985). Rainfall amounts are recorded and analyzed as cumulative amounts over discrete time such as minutes, hours, days, months or years. By integrating the continuous process $\xi(t)$ over some time interval $T$, a discrete stochastic process, $\{Y_1, Y_2, \ldots \}$, of rainfall amounts is obtained from

$$Y_{i(T)} = \int_{t_{i-1}}^{t_i} \xi(\tau)d\tau$$  \hspace{1cm} 4.1

where $t_i - t_{i-1} = T$ is the time of accumulation.

For a century or longer, rain gauges with a period of accumulation of one day have been used throughout the world to integrate the continuous rainfall intensity, $\xi(t)$. Periods of accumulation less than one day have been used in recent years, but the length of these records was not sufficient to study long-term variability at a temporal scale of decades. Historical rain gauge data were selected as the data base for these analyses.
(Chapter 3) resulting in observed discrete time series of daily rainfall amounts, 
\[ \{Y_1, Y_2, \ldots\}_{T=1\text{ day}} \]

The daily time series \( \{Y_1, Y_2, \ldots\}_{T=1\text{ day}} \) contains numerous days with no accumulated rainfall which makes statistical or time series analysis difficult. To overcome this difficulty, rainfall time series with periods of accumulation longer than one day were formed. For a period of accumulation of one month, the series
\[ \{Y_1, Y_2, \ldots\}_{T=1\text{ month}} \] results from

\[
Y_i(T = 1 \text{ month}) = \sum_{j=1}^{n_i} Y_j(T = 1 \text{ day}) \tag{4.2}
\]

\[ Y_j(T = 1 \text{ day}) \geq 0.254 \text{ mm (0.01 in)} \]

where \( Y_j(T = 1 \text{ day}) \) is the accumulated daily amount of rainfall for day \( j \) of month \( i \),
\( Y_i(T = 1 \text{ month}) \) is the accumulated rainfall for month \( i \), and \( n_i \) is the number of days in month \( i \). Trace amounts of rainfall (<0.254 mm) (Chapter 3) were not included in the rainfall totals, \( Y_i(T = 1 \text{ month}) \). Associated with this monthly series of rainfall amounts is the series,
\[ \{X_1, X_2, \ldots\}_{T=1\text{ month}} \] where \( X_i \) is the number of wet days in month \( i \).

To investigate the nature of long-term variability of rainfall, a rainfall amount series, \( \{Y_1, Y_2, \ldots\}_{T=1\text{ month}} \) and a wet day count time series, \( \{X_1, X_2, \ldots\}_{T=1\text{ month}} \) were constructed using equation 4.2 with the time of accumulation, \( T \), set equal to one month. These monthly time series were constructed from 89 years (1900 through 1988) of daily rainfall data for Gainesville, Orlando, Tampa and Ft. Myers, Florida (Chapter 3).

To analyze these time series, twelve monthly seasons were defined and the data were sorted by month. This resulted in twelve statistical samples for each location of the form

\[ \{Y_{i1}^k, Y_{i2}^k, \ldots, Y_{in_i}^k, Y_{89}^k\} \quad k = 1, 2, \ldots, 12 \tag{4.3} \]
where $y^k_i$ is the monthly total amount of rainfall for month $k$ in year $i$, and

$$\{X^k_1, X^k_2, \ldots, X^k_i, \ldots, X^k_{89}\} \quad k = 1, 2, \ldots, 12$$

where $X^k_i$ is the integer count of the number of wet days with accumulated rainfall greater than 0.254 mm (0.01 in) for month $k$ in year $i$.

### 4.2 Amount of Rainfall

A methodology consisting of descriptive statistics, histograms, and Fourier domain analysis was developed to quantitatively assess and describe long-term variability in the monthly total amount of rainfall sample shown in equation 4.3.

#### 4.2.1 Averages

An average monthly total amount of rainfall was estimated for each monthly season. The average monthly total was calculated by

$$\bar{y}_k = \frac{1}{n} \sum_{i=1}^{n} y^k_i \quad k = 1, 2, \ldots, 12$$

where $k$ is the month index, $i$ is the year index, $n$ is the number of years in the sample, $y^k_i$ is the total amount of rainfall received in month $k$ during year $i$, and $\bar{y}_k$ is the average monthly amount of rainfall. Average monthly rainfall totals were calculated at the four locations considered and were presented graphically (Fig. 4.1).

A pronounced annual cycle in $\bar{y}_k$ was observed at all locations. The annual cycle was composed of a small wet winter (February-March) season, a short spring (April-May) dry season, a long summer (June-September) wet season and a long autumn (October-January) dry season.

Spatially, average monthly rainfall totals varied significantly from Gainesville in the north to Ft. Myers in the south (Fig. 4.1). During the autumn, winter and spring
Figure 4.1  Average monthly total amount of rainfall (mm) by month for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988).
seasons, average rainfall totals decreased from north to south, which is probably related to the decreased influence of passing continental fronts in south Florida. Average monthly totals increased southward during the summer wet season which may be related to the influence of tropical marine weather patterns, specifically a higher incidence of hurricane and tropical storm related rainfall.

4.2.2 Standard Deviations

To investigate the degree of variability associated with the monthly total amount of rainfall, the standard deviation of the 89 year sample (equation 4.3) of monthly total amounts was estimated. The standard deviation was calculated by

\[
\hat{\sigma}_k = \left\{ \frac{1}{n-1} \sum_{i=1}^{n} (Y_i^k - \overline{Y}_k)^2 \right\}^{\frac{1}{2}} \quad k = 1, 2, \ldots, 12
\]

where \(\hat{\sigma}_k\) is the standard deviation for month \(k\) of the sample of monthly total amounts (equation 4.3). The monthly mean amount of rainfall, \(\overline{Y}_k\), plus and minus one standard deviation, \(\hat{\sigma}_k\) were presented graphically (Fig. 4.2) for the four locations considered.

The estimated standard deviations were found to be large (40 to 75 mm) in all months of the year suggesting considerable long-term variability in the monthly total amount of rainfall received in any particular month. The monthly standard deviations were largest during the summer wet season and smallest during the autumn dry season at all locations.

The standard deviations did not vary spatially to the extent that the average totals varied. There was very little difference from north to south for any particular month, with the exception of Ft. Myers, where \(\hat{\sigma}\) was slightly larger than for the rest of the state during the summer wet season.
Figure 4.2  Average monthly total amount of rainfall (mm) with plus and minus one standard deviation by month for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988).
4.2.3 Quartile Ranges

To obtain a more detailed representation of the long-term distribution of the monthly total amount of rainfall, the median, quartile ranges, minima and maxima values were determined for each monthly sample. These descriptors were presented in monthly box-and-whiskers plots (Fig. 4.3) for the four locations considered. For each month, the height of the box indicates the extent of the interquartile range, the tails on each box the monthly maximum and minimum total amount observed, and the cross mark, the median value.

The median monthly amount of rainfall followed an annual cycle similar to the mean (Section 4.2.1) with summer and winter wet season highs and spring and autumn lows. In addition, there was an annual cycle in both the inner (Q₂ to Q₃) and inter (maximum to minimum) quartile ranges. Both the inner and inter quartile ranges were smallest during November, increased slightly during the winter, decreases slightly in the spring and increased during the summer wet season. In a similar manner, the upper quartiles were large during both wet seasons and small during the dry seasons. The lower quartiles were usually smaller than the upper quartiles throughout the year. The inner two quartiles, Q₂ and Q₃, were unequal throughout the year, with no clearly defined pattern either spatially or temporally. This distribution of monthly total amount of rainfall by quartiles suggests that there is significant year-to-year variability in monthly total rainfall and that this distribution is not symmetrically distributed about the median. The quartile ranges varied spatially in an irregular manner and it was not possible to clearly discern regularities in spatial variability from these descriptors.

4.2.4 Probability Distributions

To further investigate the lack of symmetry in the distribution of monthly total rainfall suggested in Figure 4.3, histograms were constructed for each monthly sample given in equation 4.3. Histograms with 20 bins and a bin width of 25 mm were constructed with monthly rainfall totals greater than 500 mm included in the last bin.
Figure 4.3  Box-and-whiskers plots of monthly total amount of rainfall, showing the medians, inner quartile ranges, and extreme values by month for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988).
The relative frequency divided by the bin width, or probability density, was plotted so that the area under each histogram was equal to unity. The histograms for all months at the four locations considered are shown in Figures 4.4 to 4.7.

The histograms of monthly total amounts of rainfall showed a definite annual cycle that was similar at all locations. During the driest month of November, a very high frequency of small monthly totals and a low frequency of moderate to large totals was observed. Throughout the year, the shape of the monthly histograms gradually changed, showing a decreased frequency of small monthly totals and an increased frequency of moderate to large monthly totals. From December to July or August, the monthly histogram gradually changed from an extreme exponential type distribution to a broad, flat distribution slightly shifted to the left. Progressing from summer into winter, the histograms gradually returned to the extreme exponential shape.

To concisely summarize and quantitatively describe the observed (Figs. 4.4 to 4.7) annual cycle of histogram shape, a general probability density function was sought which would retain the observed histogram shape from month to month. The gamma probability density function was selected for its flexibility in shape. The gamma probability density has the form

\[
 f_k(y) = \frac{(y/\beta_k)^{\alpha_k-1} e^{-y/\beta_k}}{\beta_k \Gamma(\alpha_k)}
\]

where \( \alpha_k \) is the shape parameter for month \( k \) and \( \beta_k \) is the scale parameter for month \( k \).

For \( \alpha = 1.0 \), the gamma density function is the exponential density, while for \( \alpha > 1.0 \), the function is a bell shaped distribution. As \( \alpha \) increases, the gamma density changes from a highly non-symmetrical function with large positive skewness to a broad, flat bell shaped curve.
Figure 4.4  Monthly histograms and fitted gamma probability density function for monthly total amounts of rainfall (mm) for Gainesville, Florida (1900 through 1988).
Figure 4.5 Monthly histograms and fitted gamma probability density function for monthly total amounts of rainfall (mm) for Orlando, Florida (1900 through 1988).
Figure 4.6  Monthly histograms and fitted gamma probability density function for monthly total amounts of rainfall (mm) for Tampa, Florida (1900 through 1988).
Figure 4.7 Monthly histograms and fitted gamma probability density function for monthly total amounts of rainfall (mm) for Ft. Myers, Florida (1900 through 1988).
The gamma distribution was fit to each monthly sample at the four locations considered. Maximum likelihood estimates of the gamma-distribution parameters $\alpha$ and $\beta$ were made (Appendix A) for each monthly sample. The gamma-distribution curve was then calculated and overlaid on each monthly histogram (Figs. 4.4 to 4.7).

The gamma probability density function was observed to fit the monthly shape of the observed histogram quite well. The annual cycle in histogram shape was also preserved by the gamma-distribution model with the density function gradually changing from an extreme exponential to a broad flat bell shape and back during the year. The null hypothesis that monthly total amounts of rainfall are distributed according to a gamma distribution was tested (Appendix B). Since monthly total rainfall are continuous data, the Kolmogorov-Smirnov test was selected. Test statistics and attained significance levels or p-values were calculated (Tables 4.1 to 4.4). The critical value of the Kolmogorov-Smirnov D statistic for a significance level of 0.01 and sample size of 89 is 0.173 (Table E.9, Haan, 1977). High D statistics and associated low p-values offered evidence to reject the null hypothesis during midsummer at all four locations. During the rest of the year, low D statistics and high p-values did not offer significant evidence to reject the null hypothesis. However, visual inspection of the histograms and density functions (Figures 4.4 to 4.7) offers additional evidence that the gamma distribution is a useful model for the monthly total amount of rainfall.

To concisely summarize the annual cycle of the distribution of monthly total rainfall modeled by the gamma distribution, the gamma-distribution parameters, $\alpha$ and $\beta$, were plotted against time, in months (Figs. 4.8 and 4.9). The shape parameter, $\alpha$ (Fig. 4.8), varied in a very regular manner throughout the year. From November to April, $\alpha$ changed little, attaining an annual minimum in the dry month of November and a small relative maximum in the wet month of February. Beginning in May, $\alpha$ increased considerably in a regular manner reaching an annual maximum in mid-summer, and then decreased to the November minimum. The scale parameter, $\beta$, showed a more irregular
Table 4.1  Kolmogorov-Smirnov test statistics for comparison of the distribution of monthly total amount of rainfall and the gamma distribution for Gainesville, Florida.

<table>
<thead>
<tr>
<th>Month</th>
<th>Kolmogorov-Smirnov D</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>January</td>
<td>0.106</td>
<td>0.2701</td>
</tr>
<tr>
<td>February</td>
<td>0.083</td>
<td>0.5765</td>
</tr>
<tr>
<td>March</td>
<td>0.058</td>
<td>0.9193</td>
</tr>
<tr>
<td>April</td>
<td>0.060</td>
<td>0.9013</td>
</tr>
<tr>
<td>May</td>
<td>0.092</td>
<td>0.4353</td>
</tr>
<tr>
<td>June</td>
<td>0.188</td>
<td>0.0037</td>
</tr>
<tr>
<td>July</td>
<td>0.175</td>
<td>0.0086</td>
</tr>
<tr>
<td>August</td>
<td>0.118</td>
<td>0.1662</td>
</tr>
<tr>
<td>September</td>
<td>0.085</td>
<td>0.5384</td>
</tr>
<tr>
<td>October</td>
<td>0.047</td>
<td>0.9358</td>
</tr>
<tr>
<td>November</td>
<td>0.081</td>
<td>0.6193</td>
</tr>
<tr>
<td>December</td>
<td>0.055</td>
<td>0.9350</td>
</tr>
</tbody>
</table>

NOTE: The critical values of the Kolmogorov-Smirnov test statistic (Haan, 1977) for a specified level of significance using a sample size of \( n = 89 \) are: \( D = 0.129 \) for \( \alpha = 0.10 \); \( D = 0.144 \) for \( \alpha = 0.05 \); and \( D = 0.173 \) for \( \alpha = 0.01 \).
Table 4.2  Kolmogorov-Smirnov test statistics for comparison of the distribution of monthly total amount of rainfall and the gamma distribution for Orlando, Florida.

<table>
<thead>
<tr>
<th>Month</th>
<th>Kolmogorov-Smirnov D</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>January</td>
<td>0.069</td>
<td>0.7942</td>
</tr>
<tr>
<td>February</td>
<td>0.104</td>
<td>0.2950</td>
</tr>
<tr>
<td>March</td>
<td>0.058</td>
<td>0.9149</td>
</tr>
<tr>
<td>April</td>
<td>0.072</td>
<td>0.7454</td>
</tr>
<tr>
<td>May</td>
<td>0.084</td>
<td>0.5582</td>
</tr>
<tr>
<td>June</td>
<td>0.153</td>
<td>0.0331</td>
</tr>
<tr>
<td>July</td>
<td>0.177</td>
<td>0.0088</td>
</tr>
<tr>
<td>August</td>
<td>0.112</td>
<td>0.2188</td>
</tr>
<tr>
<td>September</td>
<td>0.132</td>
<td>0.0896</td>
</tr>
<tr>
<td>October</td>
<td>0.064</td>
<td>0.8539</td>
</tr>
<tr>
<td>November</td>
<td>0.050</td>
<td>0.9444</td>
</tr>
<tr>
<td>December</td>
<td>0.051</td>
<td>0.9448</td>
</tr>
</tbody>
</table>

NOTE: The critical values of the Kolmogorov-Smirnov test statistic (Haan, 1977) for a specified level of significance using a sample size of $n = 89$ are: $D = 0.129$ for $\alpha = 0.10$; $D = 0.144$ for $\alpha = 0.05$; and $D = 0.173$ for $\alpha = 0.01$. 
Table 4.3  Kolmogorov-Smirnov statistics for comparison of the distribution of monthly total amount of rainfall and the gamma distribution for Tampa, Florida.

<table>
<thead>
<tr>
<th>Month</th>
<th>Kolmogorov-Smirnov D</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>January</td>
<td>0.045</td>
<td>0.9202</td>
</tr>
<tr>
<td>February</td>
<td>0.085</td>
<td>0.5457</td>
</tr>
<tr>
<td>March</td>
<td>0.066</td>
<td>0.8425</td>
</tr>
<tr>
<td>April</td>
<td>0.075</td>
<td>0.7157</td>
</tr>
<tr>
<td>May</td>
<td>0.064</td>
<td>0.8619</td>
</tr>
<tr>
<td>June</td>
<td>0.097</td>
<td>0.3696</td>
</tr>
<tr>
<td>July</td>
<td>0.161</td>
<td>0.0208</td>
</tr>
<tr>
<td>August</td>
<td>0.164</td>
<td>0.0165</td>
</tr>
<tr>
<td>September</td>
<td>0.116</td>
<td>0.1838</td>
</tr>
<tr>
<td>October</td>
<td>0.073</td>
<td>0.7365</td>
</tr>
<tr>
<td>November</td>
<td>0.056</td>
<td>0.9322</td>
</tr>
<tr>
<td>December</td>
<td>0.046</td>
<td>0.9278</td>
</tr>
</tbody>
</table>

NOTE: The critical values of the Kolmogorov-Smirnov test statistic (Haan, 1977) for a specified level of significance using a sample size of $n = 89$ are: $D = 0.129$ for $\alpha = 0.10$; $D = 0.144$ for $\alpha = 0.05$; and $D = 0.173$ for $\alpha = 0.01$. 
Table 4.4  Kolmogorov-Smirnov test statistics for comparison of the distribution of monthly total amount of rainfall and the gamma distribution for Ft. Myers, Florida.

<table>
<thead>
<tr>
<th>Month</th>
<th>Kolmogorov-Smirnov D</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>January</td>
<td>0.087</td>
<td>0.5150</td>
</tr>
<tr>
<td>February</td>
<td>0.064</td>
<td>0.8600</td>
</tr>
<tr>
<td>March</td>
<td>0.056</td>
<td>0.9293</td>
</tr>
<tr>
<td>April</td>
<td>0.050</td>
<td>0.9435</td>
</tr>
<tr>
<td>May</td>
<td>0.101</td>
<td>0.3273</td>
</tr>
<tr>
<td>June</td>
<td>0.130</td>
<td>0.1027</td>
</tr>
<tr>
<td>July</td>
<td>0.218</td>
<td>0.0004</td>
</tr>
<tr>
<td>August</td>
<td>0.162</td>
<td>0.0201</td>
</tr>
<tr>
<td>September</td>
<td>0.137</td>
<td>0.0710</td>
</tr>
<tr>
<td>October</td>
<td>0.065</td>
<td>0.8454</td>
</tr>
<tr>
<td>November</td>
<td>0.046</td>
<td>0.9286</td>
</tr>
<tr>
<td>December</td>
<td>0.041</td>
<td>0.8746</td>
</tr>
</tbody>
</table>

NOTE: The critical values of the Kolmogorov-Smirnov test statistic (Haan, 1977) for a specified level of significance using a sample size of $n = 89$ are: $D = 0.129$ for $\alpha = 0.10$; $D = 0.144$ for $\alpha = 0.05$; and $D = 0.173$ for $\alpha = 0.01$. 
Figure 4.8 Gamma-distribution parameter, $\alpha$, by month for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988).
Figure 4.9  Gamma-distribution parameter, $\beta$, by month for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988).
annual cycle. In general, however, $\beta$ was small during wet seasons, and large during the dry seasons.

Spatial variability in the distribution of monthly total amounts of rainfall was investigated by comparing the spatial and temporal variability of the estimated gamma shape parameter $\alpha$. For the four locations considered, $\alpha$ vs time curves were overlaid in Figure 4.10.a. For Gainesville, Orlando and Ft. Myers, a regular north to south increase in $\alpha$ was observed during the summer wet season, and a north to south decrease in $\alpha$ was observed during winter and spring. The $\alpha$ values for Tampa did not follow this regular longitudinal spatial trend.

The observed spatial variability in the gamma-distribution shape parameter, $\alpha$, suggested that the distribution of monthly total amounts of rainfall changes from north to south. The southward increase in $\alpha$ during the summer suggested that the distribution of monthly totals broadens and flattens to the south. From north to south, there was a slightly higher frequency of months with moderate to large monthly totals and a lower frequency of months with small monthly totals. Similarly, the southward decrease in $\alpha$ during the winter and spring suggested that the distribution of monthly total rainfall becomes more exponential in shape to the south. From north to south there was a slightly higher frequency of months with small monthly totals and a slightly lower frequency of months with moderate to large totals. To illustrate the spatial change in the distribution of monthly rainfall totals, the fitted gamma distributions for March were presented in Figure 4.10.b. The observed spatial change in the distribution is based on three stations, Gainesville, Orlando and Ft. Myers, with Tampa showing some irregularities. To clarify this question, data from additional stations should be analyzed and a distinction made between coastal and inland locations.
Figure 4.10 Spatial and temporal variability in the gamma distribution fitted to monthly total amount of rainfall (mm) for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988).

a. monthly values of the gamma-distribution parameter $\alpha$; b. fitted gamma distribution for March.
4.2.5 Long-term Trends

In sections 4.2.1 to 4.2.4 above, it was shown that there was significant year-to-year variability in the monthly total amount of rainfall time series given by equation 4.3. The nature of this long-term variability was quantified temporally and spatially by applying the gamma distribution. This description of the distribution of accumulated rainfall amounts did not address the time sequencing of events or the possibility of long-term trends. The possible presence and nature of long-term trends was investigated in this section.

Background. The literature on long-term trends in rainfall and their relation to geophysical phenomena is extensive. Tyson (1986) and Currie and O'Brien (1988) cited over 70 papers on this topic. Long time series of annual total rainfall were analyzed for the possible presence of trends for Brazil (Kane and Trivedi, 1988), China (Hameed et al., 1983), East Africa (Rhode and Virji, 1976), Greece (Katsoulsi and Kambetzids, 1989), India (Mooley and Parthasarathy, 1984), Italy (Camuffo, 1984), North America (Currie and O'Brien, 1988, 1989, 1990; Huff et al., 1987; Mitchell, 1979; Vines, 1982, 1984), and South Africa (Tyson, 1980). Analyses of long rainfall records in Florida are limited. Foufoula-Georgiou (1985) analyzed 30 years of data for Miami using linear regression and found no evidence of a significant linear trend. Huber et al. (1982) analyzed 48 years of data from two stations in the Kissimmee River basin using spectral analysis and found periodicities of 4 to 6 years. Shih (1984) analyzed time series of varying lengths from 15 South Florida stations using spectral analysis and found periodicities of 6 and 12 years, but cause-effect relations were not considered.

Methods of trend analysis considered have included statistical analyses of monthly and annual totals, running means, linear regression, time and Fourier domain low-pass filtering, empirical-orthogonal function analysis (McGuirk, 1982), autocorrelation analysis, periodogram spectral analysis, and most recently, maximum-entropy spectral analysis (Currie and O'Brien, 1990). Long-period, cyclic trends were identified at all of
the locations considered, however the nature of these trends appears to exhibit some spatial variability. Periodicities of 2.0-2.9 years (quasi-biennial oscillation) (Wright, 1977), 3.0-3.9 years (no explanation), 5.5 years (half sunspot cycle), 10.0-12.0 years (sunspot cycle), 18.5 years (lunisolar cycle and atmospheric tidal waves), 21-23 years (possibly double sunspot cycle), and 84 years (interaction of planetary waves with the Himalayas) were all found at locations throughout the world.

Considerable discussion and speculation exists in the literature on the relation between specific geophysical processes and observed periodic trends in rainfall, particularly drought. Global scale phenomena such as 700 mb circulation (Namias, 1972), quasi-biennial oscillation (QBO) of winds and temperatures in the lower stratosphere (Wright, 1977), sea surface temperature (Markham and McLain, 1977; Hastenrath and Heller, 1977), sea level pressure (Hastenrath, 1984), perturbations in the Intertropical Convergence Zone (Hastenrath and Heller, 1977; Moura and Shukla, 1981), tropical disturbances (Ramos, 1975; Yamazaki and Rao, 1977), passage of cold fronts (Kousky and Chu, 1978; Kousky, 1979), and teleconnection between rainfall patterns in different regions (Nicholls, 1980; Moura and Kagano (1983) were all found to correlate with observed periodicities in rainfall. All of these studies suggested that numerous, complex physical mechanisms influence the observed long-term, cyclic nature of rainfall, however observed periods of drought do not appear to be physically explainable at this time.

The prediction of drought has been approached both physically and statistically. Without a detailed understanding of the physical processes responsible for the occurrence of drought, it is difficult to construct a physically based model. Several efforts to predict drought statistically were reviewed by Kane and Trivedi (1986). Basically two methods were considered, multiple regression and the extension of observed trends using significant periodicities identified from spectral analysis. Attempts at using predictors such as sunspots, sea level pressure, sea surface temperature, and zonal winds were made
(King et al., 1974; Hastenrath, 1984) with limited success. Recently, maximum-entropy spectral analysis to identify periodicities, coupled with least-squares regression to estimate Fourier series amplitudes were used to extend observed trend lines. Kane and Trivedi (1988) reported success with this approach in northeast Brazil. However, they qualify this success by noting that the method used is applicable only where the observed rainfall time series exhibit periodicities of 10 years or longer. Also, methods of this type do not account for short period oscillations about the prediction, which can be considerable. All of these methods assume that the rainfall time series is stationary, which may not be the case, particularly if physical processes are being influenced by relatively recent industrial or urban factors.

Sample selection. To investigate the possible presence of long-term trends in rainfall, it was first necessary to define an appropriate time of accumulation (equation 4.1) and a time series of rainfall amounts. Both annual and seasonal accumulation times were considered in an attempt to relate seasonal shifts in meteorological phenomena to seasonal rainfall patterns. Annual and seasonal time series were constructed as described below.

A discrete time series of annual total rainfall amounts was defined as

\[
\{Y_{y1}, Y_{y2}, \ldots, Y_{y88}\}_{T = 1 \text{ year}}
\]

resulting from

\[
y_{i(T = 1 \text{ year})} = \sum_{j=1}^{n_i} y_{j(T = 1 \text{ day})}
\]
where $Y_{j(T=1\ day)}$ is the accumulated daily amount of rainfall for day $j$ of year $i$, $Y_{i(T=1\ year)}$ is the annual total amount of rainfall for year $i$, and $n_i$ is the total number of days in year $i$.

In addition to the annual time series of rainfall amounts, four seasonal time series were also defined. This approach was selected to allow the investigation of possible trends with different periodicities during different seasons. The use of seasonal time series also allowed the different meteorological regimes influencing Florida to be compared with the results of the trend analyses. Four seasonal time series were defined as

$$\{Y^k_1, Y^k_2, \ldots, Y^k_m, \ldots, Y^k_{88}\}_{T=1\ season} \quad k = 1, 2, 3, 4$$

resulting from

$$Y^k_{i(T=1\ season)} = \sum_{j=1}^{n_i} Y^k_{j(T=1\ day)}$$

where $Y^k_i$ is the accumulated rainfall for season $k$ in year $i$. In comparison to the rest of the continental United States, Florida is unique in that it is located in a transition region where continental, tropical and oceanic meteorological regimes dominate during different seasons of the year. The location and strength of the North Atlantic Subtropical (Bermuda) High is the major meteorological feature controlling precipitation in Florida (Dohrenwend, 1978). Based on the seasonal dominance of the various meteorological regimes and the analysis of monthly rainfall totals given in section 4.2.1, four unequal length seasons were defined as follows:
1. A winter season was defined as February and March during which the Florida peninsula is dominated by the irregular passage of continental fronts and associated rainfall. During the winter there is high degree of atmospheric stability resulting from differential cooling of the land and ocean, a high rate of nocturnal cooling, frequent high pressure cells, and trade wind inversions. This stability limits the formation of convective rainfall events.

2. A spring season was defined as April and May during which agricultural activity is intense and rainfall is low. Spring is a transition period, during which the incidence of passing continental fronts decreases into summer.

3. A summer season was defined as June through September during which Florida receives most of its rainfall. Convective thunderstorms fed by sea breeze convergences from both the Gulf of Mexico and the Atlantic Ocean predominate during the summer season. Rainfall from tropical storms and hurricanes is also received during this season.

4. An autumn season was defined as October through December. October is the end of the hurricane season in Florida during which large rainfall is occasionally received over short intervals. November and December are the driest months of the year. Again, autumn is a transition period between the convective storms of summer and frontal rains of winter.

Due to the season definitions, a specialized year-indexing method was used. A year was considered to begin on the first day of February, and end on the last day of January. Year 1 began with February 1900 and ended with January 1901, while year 88 started with February 1987 and ended with January 1988. Thus, January 1900 and February through December 1988 were not included in the sample. This time indexing convention resulted in 88-year seasonal time series.

Methods of analysis. Least-squares linear regression and Fourier-domain spectral analysis and filtering were the methods selected to analyze the selected rainfall time series for the presence of long-term trends. Classical least-squares regression was selected for its ease of application, and wide use. This approach can only be used to detect the presence of a single linear trend in the time series to be analyzed. To investigate the possibility of short-term cyclic trends, frequency domain analysis of the time series described above was also used.
Least-squares linear regression (Ott, 1984) was applied to both the annual and seasonal total amount of rainfall time series. The linear-regression model used had the form

\[ y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \]  \hspace{1cm} (4.13)

where the \( y_i \)'s are the total rainfall amounts (mm) for year \( i \), \( x_i \) are times (yrs), \( \beta_0 \) is the intercept, \( \beta_1 \) is the slope of the regression line (mm/yr), and \( \varepsilon_i \) is the random error associated with \( y_i \). In addition to the intercept and slope, the mean and a 99% confidence interval for the slope were determined. The confidence interval for the slope has the form

\[ \hat{\beta}_1 - t_{\alpha/2} \sigma_{\beta_1} \leq \beta_1 \leq \hat{\beta}_1 + t_{\alpha/2} \sigma_{\beta_1} \]  \hspace{1cm} (4.14)

where the standard error for the slope is

\[ \sigma_{\beta_1} = \frac{\sigma_e}{\sqrt{S_{xx}}} = \frac{\sqrt{\sum_{i=1}^{n} (y_i - \hat{y})^2/(n - 2)}}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2}} \]  \hspace{1cm} (4.15)

with \( \hat{y} \) the least squares prediction equation.

To determine whether a long-term linear trend was present in the total amount time series, it was necessary to decide whether the estimated slopes from the regression analyses were significantly different from zero. The research hypothesis tested was

\[ H_0: \beta_1 = 0 \]  \hspace{1cm} (4.16)

\[ H_a: \beta_1 \neq 0 \]
If \( |t| \geq t_{\alpha/2} \), the null hypothesis, \( H_0 \), was rejected. In addition, the attained significance level, or p-value, was calculated (Appendix C) for each test.

In addition to linear regression, Fourier-domain analyses were conducted on the annual and seasonal total rainfall amount time series. The main objective of these analyses was to investigate the possible presence and nature of long-term cycles in rainfall amounts. First, the observed time series was decomposed into component cosine terms using a fast Fourier transform (FFT). The output of the FFT was used to make periodogram estimates of the spectral energy. In addition, the time series of rainfall totals were low-pass filtered using a cut off period of 10 years \( (f_c = 1/10 \text{ yrs}) \). The details of the spectral estimation techniques and filtering methods used are presented in Appendix D.

**Results of trend analysis.** The least-squares linear regression, spectral analysis, and frequency domain filtering methods described above were applied to the annual and seasonal total amount of rainfall time series given in expressions 4.9 and 4.11. These time series represented data from Gainesville, Orlando, Tampa, and Ft. Myers, Florida from 1900 through 1988. The observed time series, linear regression lines, and low-pass filter output are given in Figures 4.11, 4.13, 4.15, 4.17 and 4.19. The linear-regression parameter estimates, associated confidence intervals, and test statistics are given in Tables 4.5 to 4.9. The research hypothesis that the slopes of the regression lines were different from zero was tested using a significance level of \( \alpha/2 = 0.1 \) in which case \( t_{\alpha/2} = 1.282 \) (Ott, 1984). Also, the spectral energy estimates are given in Figures 4.12, 4.14, 4.16, 4.18 and 4.20. Using the seasons defined above, the results of the regression and spectral analysis were interpreted in terms of the dominant physical factors influencing rainfall in Florida.

For the winter season, the slopes of the regression lines (Table 4.6 and 4.19) were all found to be significantly different from zero and positive. The maximum slope occurred at Gainesville, decreased to a minimum at Orlando and then increased.
Figure 4.11 Observed and filtered annual total amount of rainfall time series, with linear-regression line for Gainesville, Orlando, Tampa and Ft. Myers, Florida (1900 through 1988).
Figure 4.12 Estimated spectral energy of annual total amount of rainfall time series for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988).
Table 4.5  Parameter estimates and test statistics for linear regression of annual total amount of rainfall time series.

<table>
<thead>
<tr>
<th>Station</th>
<th>Mean (mm)</th>
<th>Intercept $\beta_0$ (mm)</th>
<th>Lower CI (mm/yr)</th>
<th>Slope $\beta_1$ (mm/yr)</th>
<th>Upper CI (mm/yr)</th>
<th>$r^2$</th>
<th>Test Statistic $t$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gainesville</td>
<td>1308</td>
<td>1253</td>
<td>-1.2</td>
<td>1.2</td>
<td>3.7</td>
<td>0.01992</td>
<td>1.330</td>
<td>0.1871</td>
</tr>
<tr>
<td>Orlando</td>
<td>1292</td>
<td>1345</td>
<td>-3.6</td>
<td>-1.2</td>
<td>1.2</td>
<td>0.01990</td>
<td>-1.329</td>
<td>0.1874</td>
</tr>
<tr>
<td>Tampa</td>
<td>1211</td>
<td>1231</td>
<td>-3.1</td>
<td>-0.46</td>
<td>2.2</td>
<td>0.00242</td>
<td>-0.460</td>
<td>0.6469</td>
</tr>
<tr>
<td>Ft. Myers</td>
<td>1345</td>
<td>1296</td>
<td>-1.6</td>
<td>1.1</td>
<td>3.9</td>
<td>0.01293</td>
<td>1.068</td>
<td>0.2887</td>
</tr>
</tbody>
</table>
Figure 4.13  Observed and filtered winter total amount of rainfall time series, with linear-regression line for Gainesville, Orlando, Tampa and Ft. Myers, Florida (1900 through 1988).
Figure 4.14 Estimated spectral energy of winter total amount of rainfall time series for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988).
Table 4.6  Parameter estimates and test statistics for linear regression of winter total amount of rainfall time series.

<table>
<thead>
<tr>
<th>Station</th>
<th>Mean (mm)</th>
<th>Intercept $\hat{\beta}_0$ (mm)</th>
<th>Lower CI on $\hat{\beta}_1$ (mm/yr)</th>
<th>Slope $\hat{\beta}_1$ (mm/yr)</th>
<th>Upper CI on $\hat{\beta}_1$ (mm/yr)</th>
<th>$r^2$</th>
<th>Test Statistic $t$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gainesville</td>
<td>183</td>
<td>150</td>
<td>-0.23</td>
<td>0.74</td>
<td>1.71</td>
<td>0.04553</td>
<td>2.025</td>
<td>0.0459</td>
</tr>
<tr>
<td>Orlando</td>
<td>150</td>
<td>127</td>
<td>-0.48</td>
<td>0.50</td>
<td>1.48</td>
<td>0.02087</td>
<td>1.354</td>
<td>0.1793</td>
</tr>
<tr>
<td>Tampa</td>
<td>150</td>
<td>121</td>
<td>-0.35</td>
<td>0.66</td>
<td>1.66</td>
<td>0.03346</td>
<td>1.725</td>
<td>0.0880</td>
</tr>
<tr>
<td>Ft. Myers</td>
<td>117</td>
<td>87</td>
<td>-0.37</td>
<td>0.68</td>
<td>1.73</td>
<td>0.03306</td>
<td>1.715</td>
<td>0.0900</td>
</tr>
</tbody>
</table>
Figure 4.15  Observed and filtered spring total amount of rainfall time series, with linear-regression line for Gainesville, Orlando, Tampa and Ft. Myers, Florida (1900 through 1988).
Figure 4.16  Estimated spectral energy of spring total amount of rainfall time series for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988).
Table 4.7 Parameter estimates and test statistics for linear regression of spring total amount of rainfall time series.

<table>
<thead>
<tr>
<th>Station</th>
<th>Mean (mm)</th>
<th>Intercept $\hat{\beta}_0$ (mm)</th>
<th>Lower CI on $\hat{\beta}_1$ (mm/yr)</th>
<th>Slope $\hat{\beta}_1$ (mm/yr)</th>
<th>Upper CI on $\hat{\beta}_1$ (mm/yr)</th>
<th>$r^2$</th>
<th>Test Statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gainesville</td>
<td>167</td>
<td>140</td>
<td>-0.31</td>
<td>0.61</td>
<td>1.52</td>
<td>0.03438</td>
<td>1.750</td>
<td>0.0837</td>
</tr>
<tr>
<td>Orlando</td>
<td>166</td>
<td>191</td>
<td>-1.42</td>
<td>-0.56</td>
<td>0.30</td>
<td>0.03272</td>
<td>-1.706</td>
<td>0.0917</td>
</tr>
<tr>
<td>Tampa</td>
<td>132</td>
<td>139</td>
<td>-1.00</td>
<td>-0.15</td>
<td>0.70</td>
<td>0.00258</td>
<td>-0.471</td>
<td>0.6386</td>
</tr>
<tr>
<td>Ft. Myers</td>
<td>149</td>
<td>165</td>
<td>-1.15</td>
<td>-0.36</td>
<td>0.43</td>
<td>0.01620</td>
<td>-1.190</td>
<td>0.2373</td>
</tr>
</tbody>
</table>
Figure 4.17  Observed and filtered summer total amount of rainfall time series, with linear-regression line for Gainesville, Orlando, Tampa and Ft. Myers, Florida (1900 through 1988).
Figure 4.18  Estimated spectral energy of summer total amount of rainfall time series for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988).
Table 4.8  Parameter estimates and test statistics for linear regression of summer total amount of rainfall time series.

<table>
<thead>
<tr>
<th>Station</th>
<th>Mean (mm)</th>
<th>Intercept $\hat{\beta}_0$ (mm)</th>
<th>Lower CI $\hat{\beta}_1$ (mm/yr)</th>
<th>Slope $\hat{\beta}_1$ (mm/yr)</th>
<th>Upper CI $\hat{\beta}_1$ (mm/yr)</th>
<th>$r^2$</th>
<th>Test Statistic $t$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gainesville</td>
<td>673</td>
<td>685</td>
<td>-1.97</td>
<td>-0.27</td>
<td>1.43</td>
<td>0.00209</td>
<td>-0.424</td>
<td>0.6726</td>
</tr>
<tr>
<td>Orlando</td>
<td>726</td>
<td>747</td>
<td>-2.35</td>
<td>-0.48</td>
<td>1.39</td>
<td>0.00525</td>
<td>-0.674</td>
<td>0.5024</td>
</tr>
<tr>
<td>Tampa</td>
<td>714</td>
<td>749</td>
<td>-2.79</td>
<td>-0.79</td>
<td>1.21</td>
<td>0.01242</td>
<td>-1.040</td>
<td>0.3012</td>
</tr>
<tr>
<td>Ft. Myers</td>
<td>867</td>
<td>805</td>
<td>-0.57</td>
<td>1.40</td>
<td>3.36</td>
<td>0.03951</td>
<td>1.881</td>
<td>0.0634</td>
</tr>
</tbody>
</table>
Figure 4.19 Observed and filtered autumn total amount of rainfall time series, with linear-regression line for Gainesville, Orlando, Tampa and Ft. Myers, Florida (1900 through 1988).
Figure 4.20  Estimated spectral energy of autumn total amount of rainfall time series for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988).
Table 4.9 Parameter estimates and test statistics for linear regression of autumn total amount of rainfall time series.

<table>
<thead>
<tr>
<th>Station</th>
<th>Mean (mm)</th>
<th>Intercept $\beta_0$ (mm)</th>
<th>Lower CI $\beta_1$ (mm/yr)</th>
<th>Slope $\beta_1$ (mm/yr)</th>
<th>Upper CI $\beta_1$ (mm/yr)</th>
<th>$r^2$</th>
<th>Test Statistic $t$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gainesville</td>
<td>282</td>
<td>280</td>
<td>-1.13</td>
<td>0.06</td>
<td>1.25</td>
<td>0.00022</td>
<td>0.137</td>
<td>0.8911</td>
</tr>
<tr>
<td>Orlando</td>
<td>251</td>
<td>281</td>
<td>-1.97</td>
<td>-0.70</td>
<td>0.57</td>
<td>0.02376</td>
<td>-1.447</td>
<td>0.1516</td>
</tr>
<tr>
<td>Tampa</td>
<td>213</td>
<td>223</td>
<td>-1.31</td>
<td>-0.23</td>
<td>0.85</td>
<td>0.00356</td>
<td>-0.555</td>
<td>0.5806</td>
</tr>
<tr>
<td>Ft. Myers</td>
<td>217</td>
<td>225</td>
<td>-1.40</td>
<td>-0.17</td>
<td>1.06</td>
<td>0.00160</td>
<td>-0.371</td>
<td>0.7118</td>
</tr>
</tbody>
</table>
southward to Ft. Myers. Evidence suggesting long-term trends in rainfall was also found in the spectral energies for winter (Fig. 4.14) which showed the presence of periodicities greater than 100 years. The largest spectral peaks occurred at periods of 5.5 years with associated secondary peaks at approximately 11.6 and 21.3 years. Periodicities of this type are usually associated with the solar cycle. Schmidt (1986), of the Max-Plank Institute in Germany, has presented evidence that sun spots and the solar cycle influence circulation in the troposphere and stratosphere.

It is suggested that the observed spectral peaks in winter are related to cycles in rainfall associated with the passage of frontal systems. In addition, there was a pronounced spectral peak at 32 years. This periodicity has been observed by other researchers (Tyson, 1986; Currie and O’Brien, 1988) but there does not seem to be a physical explanation for it at this time. The low-pass filtered time series consisted of long period (10 to 15 yr) cycles with irregular period and amplitude.

In contrast to the winter season, the spectral estimates for the summer season (Fig. 4.18) were quite different. The winter spectral peaks ranging from 5.5 to 32 years had diminished, and a single pronounced peak at approximately 3.5 years was present. Periodicities of 3.5 years are usually related to the quasi-biennial oscillation (QBO) (Wright, 1977) of temperature and zonal winds over the Equator. Rainfall in Florida during the summer is predominantly due to convective thunderstorms feed by sea breezes and occasional tropical disturbances. It seems reasonable to suggests that the QBO may influence this rainfall mechanism. The slopes of the regression lines (Tables 4.8 and 4.19) were not significantly different from zero at the 3 northern locations but positive at Ft. Myers. The low-pass filtered time series was quite flat during most of the century, with considerable short period oscillations. Relative maxima occurred at the 3 northern locations during the late 1940s and relative minima around 1920.
Spring, a typically dry season, appears to be a time of transition between the frontally dominated winter rainfall and the convective thunderstorms of summer. Relative to the winter season, spectral peaks (Fig. 4.16) at periods of 10 years or greater had decreased, while peaks at shorter periods of 2 to 5 years had increased. During this season, spectral peaks occur at 18.3 years, but these peaks are not observed at any other season. This periodicity was attributed to the lunisolar cycle by Currie and O’Brien (1990) who suggested that atmospheric tidal waves are the responsible physical cause. A peak at 64 years also appears during the spring. This peak may be related to the 32 year peak observed during the winter season, however there does not seem to be a physical explanation for this periodicity. Linear-regression slopes (Table 4.7 and 4.19) were irregular with zero slopes at Tampa and Ft. Myers, a positive slope at Gainesville and negative slope at Orlando. The low-pass filtered data showed very irregular long-period oscillations with highly irregular amplitudes.

From summer (Fig 4.18) to autumn (Fig. 4.20), the spectra also changed significantly. The 3.6 year solitary peak of summer was replaced by numerous peaks with periods of 2 to 4 years. The spectral peaks at 5.5, 11.6 and 21.3 years associated with the solar cycle and observed during the winter season were present, along with the 32 year peak. The 18.3 year lunisolar peak was still present, but diminished in magnitude. The slopes of the regression lines (Table 4.9 and 4.19) were all zero, except for Orlando, which was negative. The low-pass filtered data showed a very flat curve. As in the spring, the autumn season appeared to be a transition season, where several physical factors were influencing the rainfall process in Florida, including the occasional occurrence of hurricanes.

In addition to the seasonal time series of rainfall totals, the annual totals were also analyzed. The spectra for the annual totals (Fig. 4.12) showed most of the features discussed above. The largest spectral peaks occurred at the solar cycles of approximately 5.5 and 11.5 years. There were a large number of secondary peaks at periods of 2 to 4
years and only minor peaks at periods greater than 10 years. Obviously, all the seasonal physical factors discussed above are active in the annual total data. Linear regression did not show any clear spatial variability in long-term trends in Florida rainfall. The regression slopes (Tables 4.5 to 4.9) were positive for Gainesville, negative for Orlando and zero for Tampa and Ft. Myers. The low-pass filtered time series showed irregular, long-term oscillations with periods of approximately 10 to 12 years with very irregular amplitudes.

**Conclusions.** The main objective of this research was to investigate the possible presence of long-term trends in Florida rainfall and identify physical mechanisms responsible for these trends. Most research of this type has focused on the analysis of time series of annual total rainfall. Rainfall in the humid subtropics is fundamentally different from continental rainfall, in that there are several different physical factors affecting the rainfall process during different times of the year. By defining seasons according to the dominant physical factors producing rainfall, it was possible to obtain an improved understanding of the nature of long-term trends in rainfall.

Rainfall in Florida is basically controlled by seasonal shifts in the Bermuda High which results in rainfall associated with the passage of continental fronts during winter and convective thunderstorms during summer. Long-term trends were observed with a periodicity of approximately 11 years during winter, and a periodicity of 3.5 years during summer. During both seasons, the period and amplitude of the low-pass filtered data were not constant, but varied from cycle to cycle. Numerous investigations of rainfall trends, discussed above, have related the solar cycle to observed 11 year periodicities, and the QBO to 3.5 year periodicities. However, imposed upon the observed trends, there is considerable year-to-year variability in rainfall which probably results from the interaction of complex, global-scale astronomical-atmosphere-ocean processes. Finally, the observed cyclic nature of rainfall time series suggests that the exclusive use of linear regression is not an appropriate analysis technique. The $r^2$ values for all regression
analyses were small, indicating that the proportion of the total variability of rainfall amounts (dependent variable) that was accounted for by time (independent variable) was small. Ott (1984) gives examples of regression analyses performed on data that are randomly dispersed or curvilinear with the resulting $r^2 = 0$.

### 4.3 Number of Wet Days

A methodology consisting of descriptive statistics, histograms, and Fourier-domain analysis was developed to quantitatively assess and describe long-term variability in the number of wet days per month sample given in equation 4.4. This approach was similar to the approach used in Section 4.2 to study long-term variability in amount of rainfall.

#### 4.3.1 Averages

An average number of wet days was estimated for each monthly season. The average count was calculated by

$$
\bar{X}_k = \frac{1}{n} \sum_{i=1}^{n} X_i^k \quad k = 1, 2, \cdots, 12
$$

where $k$ is the month index, $i$ is the year index, $n$ is the number of years in the sample, $X_i^k$ is the number of wet days in month $k$ during year $i$, and $\bar{X}_k$ is the average number of wet days per month. The average number of wet days per month were calculated for the four locations considered and presented graphically (Fig. 4.21).

An annual cycle in the average number of wet days per month was observed, similar to the cycle observed for the monthly total amount of rainfall. The annual cycle was composed of a long winter season (December to March), a short transitional spring season (April to May), a long summer season (June to September), and a short transitional autumn season (October to November). The number of wet days per month was quite uniform throughout the winter season, and did not reflect the winter wet season indicated in Figure 4.1. Beginning in April with relatively few wet days, the monthly
Figure 4.21  Average number of wet days per month for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988).
count increased to an annual maximum by mid-summer, and then decreased to a minimum in November. The number of wet days in April and November was quite similar and did not reflect the observation that November was drier than April (Fig. 4.1).

The monthly number of wet days varied spatially in a manner similar to the monthly total amounts of rainfall. The number of wet days during winter decreased from north to south, while the wet day count during summer increased toward the south. Large, regular changes in the longitudinal wet day count were not observed during the spring and autumn.

4.3.2 Standard Deviations

To investigate the degree of variability associated with the wet day counts, the standard deviation of the 89 year sample of monthly number of wet days was estimated. The standard deviation was calculated by

\[
\hat{\sigma}_k = \left\{ \frac{1}{n-1} \sum_{i=1}^{n} (X_i^k - \overline{X}_k)^2 \right\}^{\frac{1}{2}} \quad k = 1, 2, \ldots, 12
\]

where \(\hat{\sigma}_k\) is the standard deviation for month \(k\) of the sample of monthly counts (equation 4.4). The monthly mean number of wet days, \(\overline{X}_k\), plus and minus one standard deviation, \(\hat{\sigma}_k\) were presented graphically (Fig. 4.22) for the four locations considered.

The standard deviations of the wet day counts were not as large or as variable as those observed for the monthly rainfall totals (Fig. 4.2). The standard deviations were small and similar during the spring and autumn (2 to 3 days) and largest during the summer (3 to 4 days). During the winter, the standard deviations were seasonally uniform. Spatially, there was very little variability in the observed standard deviations.
Figure 4.22  Average number of wet days per month with plus and minus one standard deviation month for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988).
4.3.3 Quartile Ranges

To further clarify the long-term distribution of the monthly number of wet days the median, quartile ranges, minima and maxima values were determined for each monthly sample. These descriptors were presented in monthly box-and-whiskers plots (Fig. 4.23) for the four locations considered.

As with the monthly total rainfall, the median number of wet days followed an annual cycle similar to the mean. The inner quartile ranges were rather small (2.5 to 5 days) and varied in an irregular manner with no apparent seasonal cycle in their magnitudes. Spatially, there was no clear regular change in the quartile ranges.

4.3.4 Probability Distributions

The average, standard deviation, and quartile ranges discussed above did not give a clear picture of the distribution of the number of wet days per month. To clarify the distribution of the number of wet days per month, histograms were constructed for each monthly sample shown in equation 4.4. The monthly number of wet days is an integer count and therefore a discrete random variable as opposed to the monthly total amount of rainfall which is a continuous random variable. For this reason, the histograms for the number of wet days show the discrete probability mass, rather than the continuous probability density. Discrete histograms with 31 bins corresponding to the maximum number of wet days in a month were constructed for each monthly sample of wet day counts. The histograms for all months at the four locations considered are shown in Figures 4.24 to 4.27.

The histograms of the monthly number of wet days showed a definite annual cycle that was similar at all locations. The histogram shifted to the right during the winter and summer wet seasons and shifted to the left during the spring and autumn dry seasons. Throughout the year, the histograms had a bell-type shape, with only slight changes in the degree of symmetry during the year. The observed bell-shaped of all of the histograms
Figure 4.23  Box-and-whiskers plots of the number of wet days per month, showing the medians, inner quartile ranges, and extreme values for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988).
Figure 4.24 Monthly histograms and fitted binomial probability mass function for the number of wet days per month for Gainesville, Florida (1900 through 1988).
Figure 4.25  Monthly histograms and fitted binomial probability mass function for the number of wet days per month for Orlando, Florida (1900 through 1988).
Figure 4.26 Monthly histograms and fitted binomial probability mass function for the number of wet days per month for Tampa, Florida (1900 through 1988).
Figure 4.27 Monthly histograms and fitted binomial probability density function for the number of wet days per month for Ft. Myers, Florida (1900 through 1988).
suggested that most years had a moderate number of wet days per month centered about
the monthly mean $\bar{X}_k$, but a few years had a larger or smaller number of wet days per
month.

To quantitatively describe the observed annual cycle in the histograms of monthly
number of wet days, a discrete bell-shaped probability mass function was needed. The
binomial mass function was selected for its flexibility in shape and its importance in
Markov-renewal stochastic rainfall models (Foufoula-Georgiou et al., 1987) as discussed
in Chapter 2. The binomial distribution has the form

$$g_k(x) = Pr\{X = x\}$$

$$= \binom{n_k}{x} p_k^x q_k^{n_k-x}$$

$$= \frac{n_k!}{x!(n_k-x)!} p_k^x q_k^{n_k-x}$$

$$x = 0, 1, 2, \ldots, n_k$$

where $g(x)$ is the probability that $x$ wet days will occur during the $n_k$ days in month $k$, $p_k$
is the probability of a wet day occurring on a particular day in month $k$, and $q_k = 1 - p_k$.
The binomial distribution has a bell shape centered about the mean $(n_k p_k)$. As $n_k$
becomes large, and if neither $p$ or $q$ are near zero, the binomial distribution approaches
the normal distribution.

The binomial distribution was fit to each monthly sample (equation 4.4) at the four
locations considered. Moment estimates of the parameters $p_k$ and $q_k$ were made for each
monthly sample where

$$\hat{p}_k = \frac{\bar{X}_k}{n_k}$$

4.20
Using these parameter estimates, the discrete binomial distribution was calculated and overlaid on the observed monthly histograms (Figs. 4.24 to 4.27).

The binomial distribution was observed to fit the shape of the observed histogram moderately well. The binomial distribution retained the annual cycle of the wet day histograms, with the distribution gradually shifting to the right from spring to summer and then shifting to the left from summer to autumn. The null hypothesis that the monthly number of wet days are distributed according to the binomial distribution was tested (Appendix B). Because the monthly number of wet days are discrete data, the chi-square goodness of fit test was selected. The critical value of the chi-square statistic for a significance level of 0.01 with 28 degrees of freedom is 48.3 (Table E.6, Haan, 1977). High values of $\chi^2$ and low p-values for most months at all four locations offered evidence to reject the null hypothesis in most instances (Tables 4.10 to 4.13). Only for April and November at Gainesville, and February and November at Tampa, were the $\chi^2$ values low enough and p-values large enough that the null hypothesis could not be rejected. Inspection of Figures 4.24 to 4.27 suggests that the binomial distribution under predicts the probability of months with very few or very many wet days. In addition, the number of months with moderate numbers of wet days is also under predicted. This type of disagreement was observed during most seasons at all locations considered. The importance of the binomial distribution to the theoretical formulation of Markov-renewal rainfall models (Chapter 2), suggests that future investigations should be considered for modeling the number of wet days per month, particularly in the humid subtropics and tropics.

The binomial-distribution parameter, $p$, was plotted against time (Fig. 4.28) to summarize the annual cycle in the distribution of the number of wet days per month. The parameter, $p$, varied in a very regular manner throughout the year. From a minimum in April, $p$ increased to a maximum in July and then decreased throughout the late summer.
Table 4.10  Chi-square test statistics for comparison of the distribution of monthly number of wet days and the binomial distribution for Gainesville, Florida.

<table>
<thead>
<tr>
<th>Month</th>
<th>$\chi^2$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>January</td>
<td>69.56</td>
<td>2.19E-5</td>
</tr>
<tr>
<td>February</td>
<td>44.42</td>
<td>0.897E-3</td>
</tr>
<tr>
<td>March</td>
<td>163.49</td>
<td>0.000</td>
</tr>
<tr>
<td>April</td>
<td>12.92</td>
<td>0.993</td>
</tr>
<tr>
<td>May</td>
<td>1.02E5</td>
<td>0.000</td>
</tr>
<tr>
<td>June</td>
<td>192.70</td>
<td>0.000</td>
</tr>
<tr>
<td>July</td>
<td>129.93</td>
<td>0.000</td>
</tr>
<tr>
<td>August</td>
<td>95.51</td>
<td>0.000</td>
</tr>
<tr>
<td>September</td>
<td>146.93</td>
<td>0.000</td>
</tr>
<tr>
<td>October</td>
<td>135.47</td>
<td>0.000</td>
</tr>
<tr>
<td>November</td>
<td>21.46</td>
<td>0.806</td>
</tr>
<tr>
<td>December</td>
<td>146.44</td>
<td>0.000</td>
</tr>
</tbody>
</table>
Table 4.11  Chi-square test statistics for comparison of the distribution of monthly number of wet days and the binomial distribution for Orlando, Florida.

<table>
<thead>
<tr>
<th>Month</th>
<th>$\chi^2$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>January</td>
<td>34.08</td>
<td>7.26E-2</td>
</tr>
<tr>
<td>February</td>
<td>53.94</td>
<td>8.71E-4</td>
</tr>
<tr>
<td>March</td>
<td>79.04</td>
<td>1.07E-6</td>
</tr>
<tr>
<td>April</td>
<td>131.84</td>
<td>0.000</td>
</tr>
<tr>
<td>May</td>
<td>1215.16</td>
<td>0.000</td>
</tr>
<tr>
<td>June</td>
<td>2.96E6</td>
<td>0.000</td>
</tr>
<tr>
<td>July</td>
<td>3.95E9</td>
<td>0.000</td>
</tr>
<tr>
<td>August</td>
<td>6.24E7</td>
<td>0.000</td>
</tr>
<tr>
<td>September</td>
<td>486.58</td>
<td>0.000</td>
</tr>
<tr>
<td>October</td>
<td>154.40</td>
<td>0.000</td>
</tr>
<tr>
<td>November</td>
<td>1130.46</td>
<td>0.000</td>
</tr>
<tr>
<td>December</td>
<td>5451.35</td>
<td>0.000</td>
</tr>
</tbody>
</table>
Table 4.12  Chi-square test statistics for comparison of the distribution of monthly number of wet days and the binomial distribution for Tampa, Florida.

<table>
<thead>
<tr>
<th>Month</th>
<th>$\chi^2$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>January</td>
<td>56.82</td>
<td>8.33E-4</td>
</tr>
<tr>
<td>February</td>
<td>26.19</td>
<td>0.508</td>
</tr>
<tr>
<td>March</td>
<td>104.37</td>
<td>0.000</td>
</tr>
<tr>
<td>April</td>
<td>25.84</td>
<td>0.582</td>
</tr>
<tr>
<td>May</td>
<td>33.76</td>
<td>0.075</td>
</tr>
<tr>
<td>June</td>
<td>63.82</td>
<td>7.92E-5</td>
</tr>
<tr>
<td>July</td>
<td>1.56E8</td>
<td>0.000</td>
</tr>
<tr>
<td>August</td>
<td>91.41</td>
<td>2.00E-8</td>
</tr>
<tr>
<td>September</td>
<td>82.40</td>
<td>2.00E-7</td>
</tr>
<tr>
<td>October</td>
<td>33.91</td>
<td>0.074</td>
</tr>
<tr>
<td>November</td>
<td>24.75</td>
<td>0.642</td>
</tr>
<tr>
<td>December</td>
<td>107.31</td>
<td>0.000</td>
</tr>
</tbody>
</table>
Table 4.13  Chi-square test statistics for comparison of the distribution of monthly number of wet days and the binomial distribution for Ft. Myers, Florida.

<table>
<thead>
<tr>
<th>Month</th>
<th>$\chi^2$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>January</td>
<td>68.12</td>
<td>3.39E-5</td>
</tr>
<tr>
<td>February</td>
<td>53.62</td>
<td>9.49E-4</td>
</tr>
<tr>
<td>March</td>
<td>121.32</td>
<td>0.000</td>
</tr>
<tr>
<td>April</td>
<td>169.76</td>
<td>0.000</td>
</tr>
<tr>
<td>May</td>
<td>292.72</td>
<td>0.000</td>
</tr>
<tr>
<td>June</td>
<td>2.14E7</td>
<td>0.000</td>
</tr>
<tr>
<td>July</td>
<td>480.82</td>
<td>0.000</td>
</tr>
<tr>
<td>August</td>
<td>2.01E9</td>
<td>0.000</td>
</tr>
<tr>
<td>September</td>
<td>264.10</td>
<td>0.000</td>
</tr>
<tr>
<td>October</td>
<td>292.14</td>
<td>0.000</td>
</tr>
<tr>
<td>November</td>
<td>55.12</td>
<td>9.12E-4</td>
</tr>
<tr>
<td>December</td>
<td>45.46</td>
<td>1.20E-2</td>
</tr>
</tbody>
</table>
Figure 4.28  Binomial-distribution parameter, $p$, by month for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988).
until November. From November to April, \( p \) changed only slightly, but attained a small relative winter maximum in February.

The spatial and temporal variability of the binomial parameter, \( p \), were investigated to study the spatial variability of the distribution of the number of wet days per month. The \( p \) versus time curves shown in Figure 4.28 were overlaid in Figure 4.29.a. With the exception of Tampa, there was a regular north to south increase in \( p \) during the summer wet season, and a north to south decrease during the remainder of the year.

As the binomial parameter \( p_k \) increases for a fixed sample size, the mean increases and the distribution is shifted to the right. To illustrate the spatial variability of the distribution of the number of wet days suggested by latitudinal changes in \( p \), the fitted binomial distributions for March were plotted (Fig. 4.29.b.) From north to south during March, the frequency of a small number of wet days per month increased and the frequency of a large number of wet days decreased.

4.3.5 Long-term Trends

In Chapter 4.3.1 to 4.3.4 above, it was shown that there was significant year-to-year variability in the monthly number of wet day time series given in equation 4.4. The nature of this long-term variability was quantified temporally and spatially by applying the binomial distribution. This description of the distribution of accumulated rainfall amounts does not address the time sequencing of events or the possibility of long-term trends. The possible presence and nature of long-term trends was investigated in this section.

Background and sample selection. The background information, sample selection procedure, and methods of trend analysis described in Chapter 4.2.5 were also used to analyze the number of wet day time series. Most of the trend analysis research reported in the literature deals with the annual total amount of rainfall received at a point. Less
Figure 4.29 Spatial and temporal variability in the binomial distribution fitted to the number of wet days per month for Gainesville, Orlando, Tampa, Ft. Myers, Florida (1900 through 1988).

a. monthly values of the binomial-distribution parameter $p$; b. fitted binomial distribution for March.
consideration was given to trends in the number of wet days, or the duration and frequency of drought.

Methods of analysis. The least-squares linear regression, spectral analysis, and Fourier-domain filtering methods described in Chapter 4.2.5 were applied to the annual and seasonal number of wet day time series. The observed time series, linear-regression lines, and low-pass filter output are given in Figures 4.30, 4.32, 4.34, 4.36 and 4.38. The linear-regression parameter estimates, associated confidence intervals, and test statistics are given in Tables 4.14 to 4.18. The research hypothesis tested was that the slopes of the regression lines were significantly different from zero. The spectral energy estimates for the wet day time series are given in Figures 4.31, 4.33, 4.35, 4.37 and 4.39.

Results of trend analysis. For the winter season, the largest spectral peak (Fig. 4.33) occurred at a period of approximately 5.5 years, with secondary peaks at 11.6 and 21.3 years. This finding coincides with the prominent spectral peaks in the winter total amount time series (Fig. 4.14) which are usually associated with the solar cycle and the passage of continental frontal systems. In addition, there was a minor peak at 32 years in both the amount and wet day count time series. The slopes of the regression lines (Tables 4.15 and 4.20) varied over the state. The slopes for Gainesville and Tampa were zero, while the slopes for Orlando and Ft. Myers were positive suggesting an increase in the number of wet days. The low-pass filtered time series was flatter than that for the amount time series, showing irregular, long-period, low-amplitude cycles during the century. Both the spectra (Fig. 4.33) and the observed data (Fig. 4.32) indicated considerable short period noise about the mean for the period of record.

As with the rainfall amount time series, the spectra for the summer season (Fig. 4.37) were quite different from those for the winter season. The winter season spectral peaks for periods from 5 to 32 years had disappeared or diminished from the summer season data. Most of the spectral energy had shifted to periods from 2 to 4 years, which may be associated with short period oscillations in wind and temperature over the
Figure 4.30  Observed and filtered annual number of wet days time series, with linear-regression line for Gainesville, Orlando, Tampa and Ft. Myers, Florida (1900 through 1988).
Figure 4.31 Estimated spectral energy of annual number of wet days time series for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988).
Table 4.14  Parameter estimates and test statistics for linear regression of annual number of wet days time series.

<table>
<thead>
<tr>
<th>Station</th>
<th>Mean (days)</th>
<th>Intercept $\hat{\beta}_0$ (days)</th>
<th>Lower CI on $\hat{\beta}_1$ (days/yr)</th>
<th>Slope $\hat{\beta}_1$ (days/yr)</th>
<th>Upper CI on $\hat{\beta}_1$ (days/yr)</th>
<th>$r^2$</th>
<th>Test Statistic $t$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gainesville</td>
<td>115.9</td>
<td>117.1</td>
<td>-0.164</td>
<td>-0.026</td>
<td>0.111</td>
<td>0.00289</td>
<td>-0.502</td>
<td>0.6169</td>
</tr>
<tr>
<td>Orlando</td>
<td>116.4</td>
<td>116.2</td>
<td>-0.174</td>
<td>0.003</td>
<td>0.180</td>
<td>0.00003</td>
<td>0.046</td>
<td>0.9632</td>
</tr>
<tr>
<td>Tampa</td>
<td>108.9</td>
<td>114.0</td>
<td>-0.243</td>
<td>-0.114</td>
<td>0.016</td>
<td>0.05836</td>
<td>-2.322</td>
<td>0.0226</td>
</tr>
<tr>
<td>Ft. Myers</td>
<td>111.3</td>
<td>108.4</td>
<td>-0.112</td>
<td>0.067</td>
<td>0.246</td>
<td>0.01107</td>
<td>0.987</td>
<td>0.3266</td>
</tr>
</tbody>
</table>
Figure 4.32  Observed and filtered winter number of wet days time series, with linear-regression line for Gainesville, Orlando, Tampa and Ft. Myers, Florida (1900 through 1988).
Figure 4.33 Estimated spectral energy of winter number of wet days time series for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988).
Table 4.15 Parameter estimates and test statistics for linear regression of winter number of wet days time series.

<table>
<thead>
<tr>
<th>Station</th>
<th>Mean (days)</th>
<th>Intercept $\hat{\beta}_0$ (days)</th>
<th>Lower CI on $\hat{\beta}_1$ (days/yr)</th>
<th>Slope $\hat{\beta}_1$ (days/yr)</th>
<th>Upper CI on $\hat{\beta}_1$ (days/yr)</th>
<th>$r^2$</th>
<th>Test Statistic $t$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gainesville</td>
<td>15.4</td>
<td>14.4</td>
<td>-0.029</td>
<td>0.024</td>
<td>0.076</td>
<td>0.01632</td>
<td>1.195</td>
<td>0.2355</td>
</tr>
<tr>
<td>Orlando</td>
<td>13.7</td>
<td>12.0</td>
<td>-0.012</td>
<td>0.039</td>
<td>0.089</td>
<td>0.04557</td>
<td>2.026</td>
<td>0.0458</td>
</tr>
<tr>
<td>Tampa</td>
<td>13.3</td>
<td>12.6</td>
<td>-0.036</td>
<td>0.016</td>
<td>0.068</td>
<td>0.00745</td>
<td>0.803</td>
<td>0.4241</td>
</tr>
<tr>
<td>Ft. Myers</td>
<td>10.7</td>
<td>9.0</td>
<td>-0.013</td>
<td>0.037</td>
<td>0.088</td>
<td>0.04335</td>
<td>1.974</td>
<td>0.0516</td>
</tr>
</tbody>
</table>
Figure 4.34  Observed and filtered spring number of wet days time series, with linear-regression line for Gainesville, Orlando, Tampa and Ft. Myers, Florida (1900 through 1988).
Figure 4.35  Estimated spectral energy of spring number of wet days time series for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988).
Table 4.16 Parameter estimates and test statistics for linear regression of spring number of wet days time series.

<table>
<thead>
<tr>
<th>Station</th>
<th>Mean (days)</th>
<th>Intercept $\hat{\beta}_0$ (days)</th>
<th>Lower CI on $\hat{\beta}_1$ (days/yr)</th>
<th>Slope $\hat{\beta}_1$ (days/yr)</th>
<th>Upper CI on $\hat{\beta}_1$ (days/yr)</th>
<th>$r^2$</th>
<th>Test Statistic $t$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gainesville</td>
<td>14.1</td>
<td>14.1</td>
<td>-0.050</td>
<td>0.001</td>
<td>0.051</td>
<td>0.00002</td>
<td>0.037</td>
<td>0.9706</td>
</tr>
<tr>
<td>Orlando</td>
<td>14.6</td>
<td>15.3</td>
<td>-0.067</td>
<td>-0.016</td>
<td>0.035</td>
<td>0.00792</td>
<td>-0.829</td>
<td>0.4096</td>
</tr>
<tr>
<td>Tampa</td>
<td>11.7</td>
<td>13.2</td>
<td>-0.076</td>
<td>-0.033</td>
<td>0.010</td>
<td>0.04452</td>
<td>-2.002</td>
<td>0.0485</td>
</tr>
<tr>
<td>Ft. Myers</td>
<td>13.1</td>
<td>14.1</td>
<td>-0.078</td>
<td>-0.022</td>
<td>0.033</td>
<td>0.01311</td>
<td>-1.069</td>
<td>0.2881</td>
</tr>
</tbody>
</table>
Figure 4.36  Observed and filtered summer number of wet days time series, with linear-regression line for Gainesville, Orlando, Tampa and Ft. Myers, Florida (1900 through 1988).
Figure 4.37  Estimated spectral energy of summer number of wet days time series for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988).
Table 4.17  Parameter estimates and test statistics for linear regression of summer number of wet days time series.

<table>
<thead>
<tr>
<th>Station</th>
<th>Mean (days)</th>
<th>Intercept $\hat{\beta}_0$ (days)</th>
<th>Lower CI on $\hat{\beta}_1$ (days/yr)</th>
<th>Slope $\hat{\beta}_1$ (days/yr)</th>
<th>Upper CI on $\hat{\beta}_1$ (days/yr)</th>
<th>$r^2$</th>
<th>Test Statistic $t$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gainesville</td>
<td>58.1</td>
<td>60.6</td>
<td>-0.145</td>
<td>-0.055</td>
<td>0.034</td>
<td>0.03018</td>
<td>-1.636</td>
<td>0.1055</td>
</tr>
<tr>
<td>Orlando</td>
<td>60.7</td>
<td>60.5</td>
<td>-0.101</td>
<td>0.007</td>
<td>0.114</td>
<td>0.00029</td>
<td>0.159</td>
<td>0.8742</td>
</tr>
<tr>
<td>Tampa</td>
<td>58.9</td>
<td>63.2</td>
<td>-0.173</td>
<td>-0.095</td>
<td>-0.017</td>
<td>0.10716</td>
<td>-3.213</td>
<td>0.0019</td>
</tr>
<tr>
<td>Ft. Myers</td>
<td>65.2</td>
<td>62.2</td>
<td>-0.035</td>
<td>0.064</td>
<td>0.164</td>
<td>0.03259</td>
<td>1.702</td>
<td>0.0924</td>
</tr>
</tbody>
</table>
Figure 4.38  Observed and filtered autumn number of wet days time series, with linear-regression line for Gainesville, Orlando, Tampa and Ft. Myers, Florida (1900 through 1988).
Figure 4.39 Estimated spectral energy of autumn number of wet days time series for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988).
Table 4.18  Parameter estimates and test statistics for linear regression of autumn number of wet days time series.

<table>
<thead>
<tr>
<th>Station</th>
<th>Mean (days)</th>
<th>Intercept $\hat{\beta}_0$ (days)</th>
<th>Lower CI on $\hat{\beta}_1$ (days/yr)</th>
<th>Slope $\hat{\beta}_1$ (days/yr)</th>
<th>Upper CI on $\hat{\beta}_1$ (days/yr)</th>
<th>$r^2$</th>
<th>Test Statistic $t$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gainesville</td>
<td>28.3</td>
<td>27.8</td>
<td>-0.069</td>
<td>0.011</td>
<td>0.091</td>
<td>0.00150</td>
<td>0.360</td>
<td>0.7201</td>
</tr>
<tr>
<td>Orlando</td>
<td>27.4</td>
<td>27.9</td>
<td>-0.094</td>
<td>-0.013</td>
<td>0.069</td>
<td>0.00191</td>
<td>-0.406</td>
<td>0.6856</td>
</tr>
<tr>
<td>Tampa</td>
<td>25.0</td>
<td>24.9</td>
<td>-0.063</td>
<td>0.002</td>
<td>0.067</td>
<td>0.00006</td>
<td>0.069</td>
<td>0.9454</td>
</tr>
<tr>
<td>Ft. Myers</td>
<td>22.6</td>
<td>22.3</td>
<td>-0.062</td>
<td>0.008</td>
<td>0.078</td>
<td>0.00105</td>
<td>0.300</td>
<td>0.7648</td>
</tr>
</tbody>
</table>
Table 4.19  Summary of estimated slopes from linear regression of seasonal and annual total amount of rainfall for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988).

<table>
<thead>
<tr>
<th>Station</th>
<th>Winter</th>
<th>Spring</th>
<th>Summer</th>
<th>Autumn</th>
<th>Annual</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gainesville</td>
<td>+ slope</td>
<td>+ slope</td>
<td>0 slope</td>
<td>0 slope</td>
<td>+ slope</td>
</tr>
<tr>
<td>Orlando</td>
<td>+ slope</td>
<td>- slope</td>
<td>0 slope</td>
<td>- slope</td>
<td>- slope</td>
</tr>
<tr>
<td>Tampa</td>
<td>+ slope</td>
<td>0 slope</td>
<td>0 slope</td>
<td>0 slope</td>
<td>0 slope</td>
</tr>
<tr>
<td>Ft. Myers</td>
<td>+ slope</td>
<td>0 slope</td>
<td>+ slope</td>
<td>0 slope</td>
<td>0 slope</td>
</tr>
</tbody>
</table>

Table 4.20  Summary of estimated slopes from linear regression of seasonal and annual number of wet days for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988).

<table>
<thead>
<tr>
<th>Station</th>
<th>Winter</th>
<th>Spring</th>
<th>Summer</th>
<th>Autumn</th>
<th>Annual</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gainesville</td>
<td>0 slope</td>
<td>0 slope</td>
<td>- slope</td>
<td>0 slope</td>
<td>0 slope</td>
</tr>
<tr>
<td>Orlando</td>
<td>+ slope</td>
<td>0 slope</td>
<td>0 slope</td>
<td>0 slope</td>
<td>0 slope</td>
</tr>
<tr>
<td>Tampa</td>
<td>0 slope</td>
<td>- slope</td>
<td>- slope</td>
<td>0 slope</td>
<td>- slope</td>
</tr>
<tr>
<td>Ft. Myers</td>
<td>+ slope</td>
<td>0 slope</td>
<td>+ slope</td>
<td>0 slope</td>
<td>0 slope</td>
</tr>
</tbody>
</table>
Equator (QBO). However, the largest peak occurred at approximately 64 years. The physical reason for this is not clear at this time. As in winter, the slopes (Tables 4.17 and 4.20) of the regression lines for the summer season did not show a regular pattern. The slope was zero for Orlando, negative for Gainesville and Tampa, and positive for Ft. Myers. The slopes for Ft. Myers, located on the Gulf of Mexico coast in south Florida, were positive during the winter and summer, for both the total amount and wet day time series. The low-pass filtered time series were again quite flat for Gainesville and Tampa, while those for Orlando and Ft. Myers showed a long-period oscillation from 1900 to about 1940. From 1900 to 1920 the number of wet days was below average and from 1920 to about 1940 there was an increase to an above average number of wet days. This occurrence in the early part of the century certainly influences the linear regression and spectral analyses, and demonstrates the importance of using long time series in the analysis of rainfall phenomena.

During the dry spring season, the predominant winter spectral peak at 5.5 years (Fig. 4.33) disappeared, and a peak at 18.3 years (Fig. 4.35) appeared. The rest of the spectra for this season was quite flat. This shift to the 18.3 year periodicity was also observed in the total amount time series, and was related to the lunisolar cycle by Currie and O’Brien (1990). In addition, a spectral peak at 64 years increased in magnitude and continued to increase in the summer season (Fig. 4.37). The slopes of the regression lines (Tables 4.16 and 4.20) were all zero except for Tampa, which had a negative slope. The low-pass filtered time series were not as flat as those for winter and summer, and showed a long-period oscillation during the early part of the century, particularly for Ft. Myers.

From summer (Fig. 4.37) to autumn (Fig. 4.39) the spectra for the wet day count changed in a manner similar to the total amount spectra for these seasons. During autumn, the spectral peak at about 5.5 years was again present, and the large peak at 64 years diminished. The spectral peak at 18.3 years observed during the spring season and absent during the winter and summer seasons in southern and central Florida also
increased in prominence. The slopes of the regression lines were zero at all locations (Table 4.18 and Table 4.20). The low-pass filtered time series were again quite flat, except for the early part of the century at Orlando and Ft. Myers as observed during other seasons.

The annual wet day count time series was also analyzed. The most prominent peak in the annual spectra occurred at 64 years. In addition, peaks occurred at 18.3 years, 5.5 years, and in the range of 2 to 4 years. The low-pass filtered wet day count series (Fig. 4.30) showed a period of about 32 years (1900 to 1932) during which the occurrence of rainfall events cycled from a minimum in 1908 to a maximum in 1922. This feature was not present in the annual total amount time series. This cycle during the early part of the century was probably responsible for the spectral peak at 64 years. The annual total spectra and the wet day count spectra were different in that the annual total spectra did not show evidence of periodicities of 18.3 and 64 years. This suggests that the occurrence of rainfall events may be influenced by some long-period physical factors which were not evident in the rainfall amount process. Linear regression of the annual wet day count time series gave slopes of zero for Gainesville, Orlando and Ft. Myers, but a negative slope at Tampa (Tables 4.14 and 4.20).

**Conclusions.** The analyses of the seasonal and annual wet day count time series were similar to the seasonal and annual total time series in that periodicities of approximately 18.5, 11, 5.5, and 2 to 4 years were present. Periodicities of 18.5 years were present in the spring and autumn seasons and may be related to the lunisolar cycle. The 5.5 and 11 year periods were strongest during the winter season and were related to the solar cycle. The 5.5 and 11 year periodicities were weak or absent during the summer when short period (2 to 4 years) cycles dominated the spectra. These short periodicities may be related to quasi-biennial oscillations in temperature and wind at the Equator. In addition, periodicities of 32 and 64 years were observed in the annual and seasonal number of wet day time series, but were absent in the total amount data. It was suggested
that some long period physical factor may influence the occurrence of rainfall but not the rainfall amount processes. For all regression analyses, the $r^2$ values were approximately equal to zero, which was similar to the findings in regression analyses for the total rainfall amount time series discussed above.

4.4 Summary

The problem considered was an investigation and mathematical description of long-term (decades) variability in rainfall. To investigate the problem, the total amount of rainfall and the number of wet days accumulated over various times (months, seasons and years) were considered. These two descriptors of rainfall were selected because of their widespread use in stochastic rainfall model evaluation. The monthly mean, standard deviation, and quartile ranges of the sample sets were calculated. These summary statistics did not give a straightforward understanding of the long-term distribution of the monthly rainfall totals or wet day counts. To clarify the nature of the distribution of the descriptors considered, monthly histograms of total rainfall and number of wet days were constructed.

The histograms were observed to vary in a very regular manner throughout the year. A continuous gamma distribution was successfully fit to the monthly total amount histograms. A discrete binomial distribution was fit to the monthly number of wet days with limited success. The binomial distribution is the theoretical basis for discrete point process stochastic rainfall models (Foufoula-Georgiou, 1985). The observed lack of fit between the binomial distribution and Florida rainfall warrants further consideration. Using the parameters of the gamma and binomial distributions, it was possible to study the spatial and temporal nature of long-term rainfall variability in rainfall.

The possible presence of long-term trends in rainfall was investigated using classical linear regression and Fourier-domain methods. Periodicities of 32, 18.5, 11, 5.5, and 3.5 years were found in the observed time series. The relative importance of each of these periodicities varies seasonally for the locations considered. The seasonal shifts in
dominant periodicities were related to geophysical phenomena discussed in the literature. These analyses suggested that: the solar cycle (11 years) was related to cycles in winter rainfall associated with passing frontal systems; the lunisolar cycle (18.5 years) was related to spring rainfall; and quasi-biennial oscillations (3.5 years) in wind and temperature were related to summer rainfall associated with convective thunderstorms fed by sea breeze convergence from the Gulf and the Atlantic. A physical explanation for the observed 32 and 64 year periodicity of rainfall occurrence or wet day counts, particularly in south Florida, is not apparent at this time.

The inability of simple descriptive statistics such as the mean and standard deviation to quantify the nature of long-term variability was demonstrated. The implications of these findings to stochastic rainfall model parameter estimation, and to model evaluation will be considered in the following chapters.
CHAPTER 5
PARAMETER ESTIMATION METHODOLOGY

An annually nonhomogeneous parameter estimation (ANPE) methodology for stochastic rainfall models with a daily time step was developed. A time dependent parameter estimation method was used for rainfall occurrence models, and interval parameter estimates were used for rainfall amount models. To illustrate the methodology, a first-order Markov chain was used for the occurrence model, and the gamma distribution was used for the amount model. The Markov-chain / gamma-distribution stochastic rainfall model (MCG) was selected for illustrative purposes because of its widespread use in the WGEN weather generation model (Richardson and Wright, 1984). The parameter estimation methodology was applied to 89 years of daily rainfall data from Gainesville, Orlando, Tampa, and Ft. Myers, Florida (Chapter 3).

Hydrologic models in general and rainfall models in particular are constructed following several basic steps. First, a hydrologic variable is selected. Variables such as the wet or dry state of a particular day, the time between rainfall events, or the wet day amount of rainfall are frequently used in rainfall models. Debate over the choice of hydrologic variable is extensive with numerous unresolved questions remaining (Chapter 2). Once a hydrologic variable is selected, a model is constructed. The model may be a simplified conceptual picture, a scaled physical replica in a laboratory, or a collection of mathematical expressions describing the hydrologic variable under consideration. If physical principles are used in constructing the mathematical model, the model is often labeled deterministic, while if probability theory is used, the model is labeled stochastic. Both deterministic and stochastic rainfall models have been proposed, however stochastic
models were the focus of this research. Extensive research on the construction of stochastic rainfall models has been conducted (Chapter 2), however there is little consensus of opinion to assist users in selecting the “best possible” model for application. Once a model is constructed, it must be fit to observed data by estimating parameters. All research studies found on stochastic rainfall model parameter estimation during the literature review for this work (Chapter 2) have assumed that model parameters are annually homogeneous or annually time invariant.

Consider the model, \( f(z; \theta_1, \theta_2, \ldots, \theta_r) \), where \( Z \) is the variable under consideration, and \( \theta_i \) (\( i = 1, 2, \ldots, r \)) are the \( r \) parameters of the model. To fit the model to observed data, it is necessary to estimate the model parameters. Model parameters often vary seasonally to account for intraannual cycles in meteorological conditions, in which case, observed data are sorted seasonally prior to parameter estimation. This results in a set of \( r \) model parameters for each season or \( \{\theta_{1\tau}, \theta_{2\tau}, \ldots, \theta_{r\tau}\} \), where \( \tau \) is the season index (\( \tau = 1, 2, \ldots, S \)), and \( S \) is the number of seasons. To simulate a long (decades) time series of the variable \( Z \) using the model \( f(z; \theta_{1\tau}, \theta_{2\tau}, \ldots, \theta_{r\tau}) \), the parameter set \( \{\theta_{1\tau}, \theta_{2\tau}, \ldots, \theta_{r\tau}\} \) is used repeatedly for each year generated. This annually homogeneous parameter estimation (AHPE) method was used in all stochastic rainfall models reviewed in Chapter 2.

The main objective of this research was to evaluate the hypothesis that stochastic rainfall model parameters may be annually nonhomogeneous or annually time variant. The reasons for proposing this hypothesis included:

i. Physical factors affecting global atmospheric circulation patterns, such as solar radiation and lunar-solar gravitational forces, vary from year to year.

ii. Urbanization, industrialization, and deforestation may influence regional or global precipitation patterns.

iii. Historical rainfall data (Chapter 4) suggested that there is considerable year-to-year variability in rainfall, including irregularly occurring periods of drought and aperiodic long-term trends, that are not retained in existing stochastic rainfall models (Schmidt et al., 1989).
iv. Any observed data base represents only a limited sample of the total population of information.

A parameter estimation methodology was developed based on this hypothesis.

The stochastic rainfall model parameter estimation methodology developed was based on the premise that model parameters are random variables with some degree of uncertainty. The uncertainty in a parameter may be completely random, or there may be some underlying order to the uncertainty. Two different approaches were used to develop the proposed ANPE estimation methodology. An interval parameter estimation method was developed for rainfall amount models and a time dependent parameter estimation method was developed for rainfall occurrence models.

Statistical estimation theory was used to construct confidence intervals about wet day amount model parameters. This approach assumed that the parameters were random variables which are distributed according to some density function, such as the normal distribution. Rather than attempting to estimate the model parameters exactly, this approach placed an interval of uncertainty about each parameter. All stochastic rainfall model parameter estimation methods studied (Chapter 2) and used in applications have focused on point estimates. These methods all attempted to obtain a "best possible fit" of a model using a limited amount of historical data to estimate a parameter as exactly as possible. The proposed ANPE method was essentially opposite to earlier approaches. A "relaxed" fit of the model was sought which allows the model to "vibrate" within the limits of the estimated parameter interval. To illustrate the proposed interval estimation method, a gamma distribution was used for the wet day amount model.

A time dependent annually nonhomogeneous parameter estimation method which produced a time series of parameter estimates was developed for a Markov-chain rainfall occurrence model. This approach allowed the time sequencing of model parameters to be investigated (Chapter 6). All occurrence model parameter estimation methods considered (Chapter 2) used annually homogeneous parameter sets that were repeated exactly from
year to year. The proposed time dependent ANPE approach allowed the model parameters to be treated as stochastic processes in which there may be long-term trends and time dependent structure which can be used (Chapter 6) to develop a parameter model for the Markov chain. This approach allowed the occurrence model to "vibrate" in a time dependent manner preserving the correlation structure of the historical data. To illustrate the proposed parameter estimation method, a first-order Markov chain was used for the occurrence model.

5.1 Occurrence Model Parameter Estimation

5.1.1 Model Theory

The rainfall status of a particular day can be modeled as a discrete random variable, \( X_t \), in discrete time (days) (Parzen, 1962). The daily sequence of rainfall events \( \{X_1, X_2, \ldots, X_r, \ldots, X_n\} \) or \( \{X_i\} \) then forms a random or stochastic process called a Markov chain, where the value of \( X_t \) represents the rainfall state of day \( t \). The states considered for the random variable \( X_t \) were:

i. \( X_t = \text{Dry}, \) if rainfall \( < 0.254 \text{ mm (0.01 in)} \) for day \( t \);

ii. \( X_t = \text{Wet}, \) if rainfall \( \geq 0.254 \text{ mm (0.01 in)} \) for day \( t \).

The rainfall state for day \( t \) can then be Dry or Wet, i.e. \( x_t = j \) \( (j = D, W) \), while for day \( t - 1 \), \( x_{t-1} = i \) \( (i = D, W) \).

If it is assumed that the probability of a rainfall event on any given day depends only on the state of the previous day, the Markov chain is of first order. This assumption is the conditional probability

\[
Pr\{X_t = x_t \mid X_{t-1} = x_{t-1}, X_{t-2} = x_{t-2}, \ldots, X_2 = x_2, \ldots, X_1 = x_1\} = \frac{Pr\{X_t = x_t, X_{t-1} = x_{t-1}\}}{Pr\{X_{t-1} = x_{t-1}\}}
\]
If $X_{t-1} = i$ and $X_t = j$, then there has been a transition from state $i$ to state $j$ at day $t$. The probability of the possible transition is called the transition probability and is written

$$p_{ij} = Pr\{X_t = x_t | X_{t-1} = x_{t-1}\} \quad \text{(5.3)}$$

For a first-order Markov chain with two states, there are four transition probabilities which can be represented by the matrix $\overline{P}$ given by

$$\overline{P} = [p_{ij}] = \begin{bmatrix} p_{DD} & p_{DW} \\ p_{WD} & p_{WW} \end{bmatrix} \quad \text{(5.4)}$$

where $p_{DW} = 1 - p_{DD}$ and $p_{WW} = 1 - p_{WD}$.

5.1.2 Parameter Estimation Method

The elements of the Markov chain transition probability matrix, $\overline{P}$, are estimated by counting the number of transitions from state $i$ to state $j$ in the observed time series $\{X_t\}$. Estimates for $p_{ij}$ can be obtained from

$$\hat{p}_{ij} = \frac{f_{ij}}{f_{iD} + f_{iW}} \quad i, j = D, W \quad \text{(5.5)}$$

where $f_{ij}$ is an integer count of the number of days in state $j$ preceded by state $i$.

Temporal changes in climatological conditions throughout the year have been accounted for by defining appropriate seasons and counting the transitions in each season. Monthly (WGEN) (Richardson and Wright, 1984) and biweekly (CLIMATE) (Woolhiser et al., 1985, 1988) seasons have been used extensively. This approach results in seasonally varying, annually homogeneous estimates of the transition probabilities which have the form
\[ \bar{P}_\tau = [p_{ij}(\tau)] \]

\[ = \begin{bmatrix} p_{DD}(\tau) & p_{DW}(\tau) \\ p_{WD}(\tau) & p_{WW}(\tau) \end{bmatrix} \]

where \( \tau \) is the season index. The AHPE method used to estimate \( \bar{P}_\tau \) has the form

\[ \hat{p}_{ij}(\tau) = \frac{f_{ij}(\tau)}{f_{ii}(\tau) + f_{ij}(\tau)} \quad i, j = D, W \]

where \( f_{ij}(\tau) \) is an integer count of the number of days in state \( j \) preceded by state \( i \) contained in a sample for season \( \tau \). To obtain accurate estimates of \( \bar{P}_\tau \), large seasonal samples from long time histories of daily rainfall are sought. When long time histories of daily rainfall are used to make annually homogeneous parameter estimates, long-term uncertainty in the parameter may be lost.

To investigate the importance of parametric uncertainty in Markov-chain models, an annually nonhomogeneous estimation method was investigated. Consider a Markov-chain transition probability matrix that varies annually as well as seasonally. This annually nonhomogeneous matrix has the form

\[ \bar{P}_{\nu, S + \tau} = [p_{ij}(\nu S + \tau)] \]

\[ = \begin{bmatrix} p_{DD}(\nu S + \tau) & p_{DW}(\nu S + \tau) \\ p_{WD}(\nu S + \tau) & p_{WW}(\nu S + \tau) \end{bmatrix} \]

where \( \nu \) is the year index \( (\nu = 0, 1, 2, \ldots, N) \), \( \tau \) is the season index \( (\tau = 1, 2, \ldots, S) \), and \( S \) is the number of seasons. The transition probabilities can be estimated from the observed daily series of rainfall occurrences, \( \{X_i\} \), by

\[ \hat{p}_{ij}(\nu S + \tau) = \frac{f_{ij}(\nu, \tau)}{f_{ii}(\nu, \tau) + f_{ij}(\nu, \tau)} \quad i, j = D, W \]
where $f_g(v, \tau)$ is an integer count of the number of days in state $j$ preceded by state $i$ contained in a sample for year $v$ and season $\tau$. Transitions are counted beginning with the last day of the previous season and ending on the last day of the current season.

5.1.3 Results of Parameter Estimation Method

Annually nonhomogeneous Markov-chain transition probabilities were estimated using daily rainfall data (Chapter 3) for Gainesville, Orlando, Tampa, and Ft. Myers, Florida from 1900 through 1988 ($v = 0, 1, \ldots, 88$). Monthly seasons ($S = 12$) were selected following the approach used by Richardson and Wright (1985) in the WGEN stochastic rainfall model (Chapter 2). Equation 5.9 was used to estimate the time series of Markov-chain transition probabilities, \($P_{vs + \tau}$\). The parameter time series \(\{\hat{p}_{DD}(vS + \tau)\}\) and \(\{\hat{p}_{WD}(vS + \tau)\}\) were estimated directly using equation 5.9 while the time series \(\{\hat{p}_{DW}(vS + \tau)\}\) and \(\{\hat{p}_{WW}(vS + \tau)\}\) were calculated using the properties that

\[ p_{DW} = 1 - p_{DD} \]
\[ p_{WW} = 1 - p_{WD} \]

for each data point in the time series. The estimated time series of parameters \(\{\hat{p}_{DD}(vS + \tau)\}\) are given in Figures 5.1 to 5.4 and the time series \(\{\hat{p}_{WD}(vS + \tau)\}\) are given in Figures 5.5 to 5.8 for the four locations considered. Each figure shows a complete 89 year time series from 1900 through 1988, with 10-year expanded views for 1900 through 1909, and 2-year expanded views for 1900 through 1901.

The ANPE time series, \(\{\hat{p}_{DD}(vS + \tau)\}\) and \(\{\hat{p}_{WD}(vS + \tau)\}\), shown in Figures 5.1 to 5.8 all exhibited a high degree of interannual variability compared to the AHPE series represented in Figure 2.1. An annual cycle was discernable in both the \(\{\hat{p}_{DD}(vS + \tau)\}\) and \(\{\hat{p}_{WD}(vS + \tau)\}\) time series, with low values observed during the summer and high values observed during winter. However, the shape of this annual cycle varied considerably from year to year. Relative annual maxima and minima did not occur during the same month each year, and the annual maxima to minima distance was highly variable. The annual cycle of the \(\{\hat{p}_{DD}(vS + \tau)\}\) time series exhibited broad, flat peaks during winter and sharp, plunging troughs during summer. The \(\{\hat{p}_{WD}(vS + \tau)\}\) time series was
Figure 5.1  Estimated time series of Markov-chain transition probabilities, \( \{\hat{p}_{D}(vS + \tau)\} \), for Gainesville, Florida from 1900 through 1988 with expanded views of 1900 through 1909 and 1900 through 1901.
Figure 5.2  Estimated time series of Markov-chain transition probabilities, \( \{ \hat{p}_{ij}(\nu S + \tau) \} \), for Orlando, Florida from 1900 through 1988 with expanded views of 1900 through 1909 and 1900 through 1901.
Figure 5.3  Estimated time series of Markov-chain transition probabilities, $\{ \hat{\rho}_{B0}(vS + t) \}$, for Tampa, Florida from 1900 through 1988 with expanded views of 1900 through 1909 and 1900 through 1901.
Figure 5.4 Estimated time series of Markov-chain transition probabilities, \( \hat{p}_{00}(tS + t) \), for Ft. Myers, Florida from 1900 through 1988 with expanded views of 1900 through 1909 and 1900 through 1901.
Figure 5.5  Estimated time series of Markov-chain transition probabilities, \( \{ \hat{P}_{wo}(vS + \tau) \} \), for Gainesville, Florida from 1900 through 1988 with expanded views of 1900 through 1909 and 1900 through 1901.
Figure 5.6 Estimated time series of Markov-chain transition probabilities, \( \{p_{\omega \omega}(\nu S + \tau)\} \), for Orlando, Florida from 1900 through 1988 with expanded views of 1900 through 1909 and 1900 through 1901.
Figure 5.7  Estimated time series of Markov-chain transition probabilities, \( \{ p_{wd}(vS + \tau) \} \), for Tampa, Florida from 1900 through 1988 with expanded views of 1900 through 1909 and 1900 through 1901.
Figure 5.8  Estimated time series of Markov-chain transition probabilities, \( \{ \beta_{wd}(vS + \tau) \} \), for Ft. Myers, Florida from 1900 through 1988 with expanded views of 1900 through 1909 and 1900 through 1901.
fundamentally different from the $\{\dot{\rho}_{DD}(vS+\tau)\}$ series with winter peaks similar in shape to summer troughs. In addition, the $\{\dot{\rho}_{WD}(vS+\tau)\}$ time series contained a large number of values equal to 1.0 during the winter months.

To quantitatively describe the ANPE time series shown in Figures 5.1 to 5.8, the mean ($\mu$), variance ($\sigma^2$), skewness ($\lambda_3$), and kurtosis - 3 ($\lambda_4$) were estimated for the $\{\dot{\rho}_{DD}(vS+\tau)\}$ series (Table 5.1) and the $\{\dot{\rho}_{WD}(vS+\tau)\}$ series (Table 5.2) at the four locations considered. For the $\{\dot{\rho}_{DD}(vS+\tau)\}$ series, there was little variability in the means from north to south, while the variance increased from Gainesville in the north to Ft. Myers in the south. At all locations, the skewness was negative, and decreased slightly from north to south. The kurtosis - 3 was positive throughout Florida, with a relative minimum at Ft. Myers. For the $\{\dot{\rho}_{WD}(vS+\tau)\}$ series, both the mean and variance increased slightly from north to south. The skewness was positive for all locations with a pronounced decrease from north to south. The kurtosis - 3 was negative, and became more negative from north to south. Both $\lambda_3$ and $\lambda_4$ for the $\{\dot{\rho}_{WD}(vS+\tau)\}$ were roughly an order of magnitude less and opposite in sign than those for the $\{\dot{\rho}_{DD}(vS+\tau)\}$ series.

The nonzero values of $\lambda_3$ and $\lambda_4$ for the $\{\dot{\rho}_{DD}(vS+\tau)\}$ and $\{\dot{\rho}_{WD}(vS+\tau)\}$ time series suggested that they were not normally distributed. To investigate the distribution of these data, histograms were constructed (Figs. 5.9 and 5.10) for all locations considered. The $\{\dot{\rho}_{DD}(vS+\tau)\}$ series were found to have highly non-Gaussian histograms, reflecting the negative skewness and positive kurtosis described previously, while the $\{\dot{\rho}_{WD}(vS+\tau)\}$ series did not exhibit a pronounced non-Gaussian distribution. The $\{\dot{\rho}_{WD}(vS+\tau)\}$ series appeared to be bimodal in distribution with the $Pr\{\rho_{WD} > 0.9\}$ increasing from north to south. The high incidence of observed values with $\dot{\rho}_{WD} = 1.0$ discussed above was the reason for the bimodal shape of the histogram. This may be related to the latitudinal decrease in rainfall associated with passing continental fronts during winter.
Table 5.1  Estimated statistics for the time series \( \{ \hat{p}_{DD}(vS + \tau) \} \).

<table>
<thead>
<tr>
<th>Station</th>
<th>Mean ( \mu )</th>
<th>Variance ( \sigma^2 )</th>
<th>Skewness ( \lambda_3 )</th>
<th>Kurtosis - 3 ( \lambda_4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gainesville</td>
<td>0.7433</td>
<td>0.0216</td>
<td>-1.1071</td>
<td>1.2722</td>
</tr>
<tr>
<td>Orlando</td>
<td>0.7424</td>
<td>0.0245</td>
<td>-1.0601</td>
<td>1.1986</td>
</tr>
<tr>
<td>Tampa</td>
<td>0.7536</td>
<td>0.0224</td>
<td>-1.0877</td>
<td>1.2833</td>
</tr>
<tr>
<td>Ft. Myers</td>
<td>0.7452</td>
<td>0.0328</td>
<td>-1.0408</td>
<td>0.7238</td>
</tr>
</tbody>
</table>

Table 5.2  Estimated statistics for the time series \( \{ \hat{p}_{WD}(vS + \tau) \} \).

<table>
<thead>
<tr>
<th>Station</th>
<th>Mean ( \mu )</th>
<th>Variance ( \sigma^2 )</th>
<th>Skewness ( \lambda_3 )</th>
<th>Kurtosis - 3 ( \lambda_4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gainesville</td>
<td>0.5702</td>
<td>0.0480</td>
<td>0.3421</td>
<td>-0.4920</td>
</tr>
<tr>
<td>Orlando</td>
<td>0.5632</td>
<td>0.0497</td>
<td>0.3006</td>
<td>-0.4356</td>
</tr>
<tr>
<td>Tampa</td>
<td>0.5999</td>
<td>0.0518</td>
<td>0.2134</td>
<td>-0.6443</td>
</tr>
<tr>
<td>Ft. Myers</td>
<td>0.5986</td>
<td>0.0670</td>
<td>0.1286</td>
<td>-0.9457</td>
</tr>
</tbody>
</table>

Table 5.3  Estimated statistics for the time series \( \{ \hat{p}_{WD}(vS + \tau) \} \) with \( \hat{p}_{WD} = 1.0 \) excluded.

<table>
<thead>
<tr>
<th>Station (sample size)</th>
<th>Mean ( \mu )</th>
<th>Variance ( \sigma^2 )</th>
<th>Skewness ( \lambda_3 )</th>
<th>Kurtosis - 3 ( \lambda_4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gainesville (964)</td>
<td>0.5239</td>
<td>0.0311</td>
<td>-0.0303</td>
<td>-0.6371</td>
</tr>
<tr>
<td>Orlando (962)</td>
<td>0.5151</td>
<td>0.0319</td>
<td>-0.1364</td>
<td>-0.6016</td>
</tr>
<tr>
<td>Tampa (926)</td>
<td>0.5385</td>
<td>0.0314</td>
<td>-0.1649</td>
<td>-0.5555</td>
</tr>
<tr>
<td>Ft. Myers (873)</td>
<td>0.5089</td>
<td>0.0379</td>
<td>-0.1084</td>
<td>-0.6983</td>
</tr>
</tbody>
</table>
Figure 5.9  Comparisons between observed histograms of $\hat{\rho}_{D}(V \Delta + \tau)$, normal distributions, and Gram-Charlier distributions for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988).
Figure 5.10  Comparisons between observed histograms of \( \{ p_{wd}(vS + \tau) \} \), normal distributions, and Gram-Charlier distributions for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988).
To further describe the observed distribution of the ANPE time series, Gaussian and non-Gaussian probability density functions were investigated. If either of the Markov-chain transition probabilities, \( \hat{p}_{DD} \), or \( \hat{p}_{WB} \) are denoted by the non-standardized random variable \( X \), the normal or Gaussian probability density function with mean \( \mu \) and variance \( \sigma^2 \) has the form

\[
f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}
\]

The engineering application of non-Gaussian stochastic processes is a relatively recent and active area of research (Longuet-Higgins, 1963; Ochi, 1990; Shinozuka, 1977). Ochi (1986) compared several distributions and found the Gram-Charlier probability density function to be useful in characterizing natural phenomena. The Gram-Charlier density with mean \( \mu \) and variance \( \sigma^2 \) has the form

\[
f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \left[ 1 + \frac{\lambda_3}{3!} H_3\left( \frac{x-\mu}{\sigma} \right) + \frac{\lambda_4}{4!} H_4\left( \frac{x-\mu}{\sigma} \right) \right. \\
\left. + \frac{\lambda_5}{5!} H_5\left( \frac{x-\mu}{\sigma} \right) + \frac{\lambda_6}{6!} + \frac{\lambda_3^2}{72} \right] H_6\left( \frac{x-\mu}{\sigma} \right) + \ldots \]

\[
\lambda_3 \text{ is the skewness, } \lambda_4 \text{ is the kurtosis - 3, and } H_n\left( \frac{x-\mu}{\sigma} \right) \text{ are Hermite polynomials of degree } n \text{ given by}
\]

\[
H_3(z) = z^3 - 3z \\
H_4(z) = z^4 - 6z^2 + 3 \\
H_5(z) = z^5 - 10z^3 + 15z \\
H_6(z) = z^6 - 15z^4 + 45z^2 - 15
\]
Ochi (1986) has found that an approximate form of the Gram-Charlier distribution given by

\[ f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \left[ 1 + \frac{\lambda_3}{3!} H_3\left(\frac{x-\mu}{\sigma}\right) + \frac{\lambda_4}{4!} H_4\left(\frac{x-\mu}{\sigma}\right) + \frac{\lambda_3^2}{72} H_6\left(\frac{x-\mu}{\sigma}\right) \right] \cdot L \]

where \( L \) is a normalization factor, agrees well with observed histograms.

The estimated statistics of the series \( \{ \rho_{DD}(vS + \tau) \} \) (Table 5.1) and \( \{ \rho_{WD}(vS + \tau) \} \) (Table 5.2) were used to calculate Gaussian (equation 5.10) and non-Gaussian (equation 5.13) density functions. These density functions were normalized by using

\[ f_n(x) = \frac{f(x)}{\int_0^1 f(x) dx} \]

over the sample space \( \{ x \mid 0 \leq x \leq 1 \} \). The normalized density functions, \( f_n(x) \), were truncated outside the sample space and compared with the observed histograms given in Figures 5.9 and 5.10, respectively, for the four Florida locations considered. For the \( \{ \rho_{DD}(vS + \tau) \} \) series, the Gaussian density functions deviated significantly from the observed histograms, while the non-Gaussian density functions agreed quite well. The shape of the Gram-Charlier distribution reflected the large observed values of \( \lambda_3 \) and \( \lambda_4 \) not accounted for by the normal distribution. For the \( \{ \rho_{WD}(vS + \tau) \} \) series, the Gaussian and non-Gaussian density functions were similar and fit the observed histograms moderately well, with the exception of the last bin \( (0.9 < \rho_{WD} \leq 1.0) \). The probability density of the last bin increased significantly from north to south for the four locations considered.
When $\hat{p}_{wd} = 1.0$ (i.e. $\hat{p}_{ww} = 0.0$) for a particular month, there are no wet day to wet day transitions. Wet spells of more than one day are often associated with frontally induced rainfall, while convective rainfall events are usually less than one day in duration. To further investigate the distribution of the $\{\hat{p}_{wd}(vS + \tau)\}$ series, values of $\hat{p}_{wd} = 1.0$ were excluded from the monthly samples, and the histograms, statistics (Table 5.3), and density functions (Fig. 5.11) were recalculated. After excluding $\hat{p}_{wd} = 1.0$ values, the shapes of the Gaussian and non-Gaussian distributions were very similar and agreed with the observed histograms quite well. The essentially Gaussian distribution of the $\{\hat{p}_{wd}(vS + \tau)\}$ series with $\hat{p}_{wd} = 1.0$ excluded (Fig. 5.11) and the north to south increase in the relative frequency of months where $\hat{p}_{wd} = 1.0$ (Fig. 5.10 and Table 5.3) suggested that it may be possible to separate convective and frontally induced rainfall generating process by using the distributions of their observed parameter sets.

As discussed above, the entire parameter time series $\{\hat{p}_{dd}(vS + \tau)\}$ was found to be non-Gaussian and was successfully modeled using the Gram-Charlier distribution, while the $\{\hat{p}_{wd}(vS + \tau)\}$ series (with $\hat{p}_{wd} = 1.0$ excluded) was found to be Gaussian and was modeled using the normal distribution. To further describe the nature of the observed parameter time series, the distribution of individual monthly samples was considered. The entire observed parameter time series was sorted into monthly subsamples, $\{\hat{p}_{dd}(v, \tau)\}$ and $\{\hat{p}_{wd}(v, \tau)\}$, where $v$ is the year index and $\tau$ is the monthly season index. The null hypothesis that each monthly sample was distributed according to a normal distribution was tested using the Kolmogorov-Smirnov test (Appendix B). The Kolmogorov-Smirnov test statistics are given in Table 5.4 for the $\{\hat{p}_{dd}(v, \tau)\}$ samples, and in Table 5.5 for the $\{\hat{p}_{wd}(v, \tau)\}$ samples. For the $\{\hat{p}_{dd}(v, \tau)\}$ samples, low D statistics for all months and locations considered did not offer significant evidence to reject the null hypothesis of normality at a significance level of $\alpha = 0.01$. For the $\{\hat{p}_{wd}(v, \tau)\}$ samples, high D statistics offered evidence to reject the null hypothesis for
Figure 5.11 Comparisons between observed histograms of \( p_{WD}(vS + \tau) \) with \( p_{WD} = 1.0 \) excluded, normal distributions, and Gram-Charlier distributions for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988).
Table 5.4  Kolmogorov-Smirnov test statistics, D, for the monthly samples \( \{ \hat{p}_{DD}(v, \tau) \} \) for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988).

<table>
<thead>
<tr>
<th>Month</th>
<th>Gainesville</th>
<th>Orlando</th>
<th>Tampa</th>
<th>Ft. Myers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.067</td>
<td>0.111</td>
<td>0.084</td>
<td>0.113</td>
</tr>
<tr>
<td>2</td>
<td>0.111</td>
<td>0.072</td>
<td>0.089</td>
<td>0.129</td>
</tr>
<tr>
<td>3</td>
<td>0.087</td>
<td>0.093</td>
<td>0.074</td>
<td>0.141</td>
</tr>
<tr>
<td>4</td>
<td>0.108</td>
<td>0.151</td>
<td>0.103</td>
<td>0.134</td>
</tr>
<tr>
<td>5</td>
<td>0.113</td>
<td>0.066</td>
<td>0.122</td>
<td>0.133</td>
</tr>
<tr>
<td>6</td>
<td>0.078</td>
<td>0.106</td>
<td>0.109</td>
<td>0.062</td>
</tr>
<tr>
<td>7</td>
<td>0.121</td>
<td>0.123</td>
<td>0.111</td>
<td>0.075</td>
</tr>
<tr>
<td>8</td>
<td>0.106</td>
<td>0.107</td>
<td>0.107</td>
<td>0.067</td>
</tr>
<tr>
<td>9</td>
<td>0.110</td>
<td>0.078</td>
<td>0.114</td>
<td>0.152</td>
</tr>
<tr>
<td>10</td>
<td>0.105</td>
<td>0.096</td>
<td>0.103</td>
<td>0.092</td>
</tr>
<tr>
<td>11</td>
<td>0.097</td>
<td>0.083</td>
<td>0.115</td>
<td>0.141</td>
</tr>
<tr>
<td>12</td>
<td>0.106</td>
<td>0.132</td>
<td>0.109</td>
<td>0.137</td>
</tr>
</tbody>
</table>

NOTE: The critical values of the Kolmogorov-Smirnov test statistic (Haan, 1977) for a specified level of significance using a sample size of \( n = 89 \) are: \( D = 0.129 \) for \( \alpha = 0.10 \); \( D = 0.144 \) for \( \alpha = 0.05 \); and \( D = 0.173 \) for \( \alpha = 0.01 \).
November through April at Ft. Myers, while low D values for all other months and locations did not offer evidence to reject the hypothesis of normality.

The Kolmogorov-Smirnov D statistic (Table 5.5) for $\{\hat{p}_{wd}(v, \tau)\}$ at Ft. Myers cycled annually, with a minimum in June/July and a maximum in December, suggesting an annual cycle in the degree of normality of the samples. To explore this annual cycle, histograms were constructed for the $\{\hat{p}_{wd}(v, \tau)\}$ samples at Ft. Myers (Fig. 5.12). During summer, the observed histograms were unimodal and bell-shaped, with low D values suggesting a normal distribution. From autumn into winter, the histograms became increasingly bimodal, with the probability density shifting from the primary bell-shaped portion of the histogram to a pronounced secondary peak ($0.9 < \hat{p}_{wd} \leq 1.0$). From winter into summer, the secondary peak diminished, and the histogram returned to a Gaussian shape by July. This observed annual cycle in histogram shape of the $\{\hat{p}_{wd}(v, \tau)\}$ samples coincides with the annual cycle of Florida rainfall processes where frontally induced rainfall dominates during winter and convective thunderstorms dominate during summer. These observations offered further evidence that seasonal shifts in physical processes generating rainfall may be reflected in stochastic rainfall model parameter sets.

Table 5.5 suggested that monthly $\{\hat{p}_{wd}(v, \tau)\}$ samples may not be normally distributed in all cases, and Figure 5.12 suggested that the distribution of these samples cycle between bell-shaped and bimodal on an annual basis. To expand this analysis spatially, an illustrative example for the month of December was considered. Histograms of $\{\hat{p}_{wd}(v, 12)\}$ showed (Fig. 5.13) a pronounced increase in $Pr\{p_{wd} > 0.9\}$ from north to south. After excluding values of $\hat{p}_{wd} = 1.0$ from $\{\hat{p}_{wd}(v, 12)\}$, observed histograms were compared with the normal and Gram-Charlier density functions (Fig. 5.13). The observed histograms were not Gaussian in all cases, however the Gram-Charlier function agreed well with observations. This example suggested that the inherent flexibility of the non-Gaussian Gram-Charlier density function may provide a concise and accurate model for monthly distributions of $\{\hat{p}_{wd}(v, \tau)\}$ after excluding $\hat{p}_{wd} = 1.0$. A model for the
Table 5.5  Kolmogorov-Smirnov test statistics, D, for the monthly samples \( \{ \hat{p}_{wD}(v, \tau) \} \) for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 to 1988).

<table>
<thead>
<tr>
<th>Month</th>
<th>Gainesville</th>
<th>Orlando</th>
<th>Tampa</th>
<th>Ft. Myers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.132</td>
<td>0.114</td>
<td>0.104</td>
<td>0.175</td>
</tr>
<tr>
<td>2</td>
<td>0.129</td>
<td>0.146</td>
<td>0.124</td>
<td>0.190</td>
</tr>
<tr>
<td>3</td>
<td>0.078</td>
<td>0.126</td>
<td>0.153</td>
<td>0.183</td>
</tr>
<tr>
<td>4</td>
<td>0.108</td>
<td>0.137</td>
<td>0.168</td>
<td>0.187</td>
</tr>
<tr>
<td>5</td>
<td>0.154</td>
<td>0.084</td>
<td>0.148</td>
<td>0.145</td>
</tr>
<tr>
<td>6</td>
<td>0.101</td>
<td>0.114</td>
<td>0.105</td>
<td>0.096</td>
</tr>
<tr>
<td>7</td>
<td>0.096</td>
<td>0.093</td>
<td>0.120</td>
<td>0.096</td>
</tr>
<tr>
<td>8</td>
<td>0.090</td>
<td>0.060</td>
<td>0.076</td>
<td>0.121</td>
</tr>
<tr>
<td>9</td>
<td>0.118</td>
<td>0.081</td>
<td>0.091</td>
<td>0.120</td>
</tr>
<tr>
<td>10</td>
<td>0.082</td>
<td>0.108</td>
<td>0.143</td>
<td>0.129</td>
</tr>
<tr>
<td>11</td>
<td>0.154</td>
<td>0.096</td>
<td>0.149</td>
<td>0.173</td>
</tr>
<tr>
<td>12</td>
<td>0.078</td>
<td>0.137</td>
<td>0.153</td>
<td>0.238</td>
</tr>
</tbody>
</table>

NOTE: The critical values of the Kolmogorov-Smirnov test statistic (Haan, 1977) for a specified level of significance using a sample size of \( n = 89 \) are: \( D = 0.129 \) for \( \alpha = 0.10 \); \( D = 0.144 \) for \( \alpha = 0.05 \); and \( D = 0.173 \) for \( \alpha = 0.01 \).
Figure 5.12  Monthly observed histograms of \( \{p_{wd}(v,\tau)\} \) for Ft. Myers, Florida (1900 through 1988).
Figure 5.13 Comparisons between December observed histograms of $\{\hat{p}_{WD}(v, \tau)\}$ with and without $\hat{p}_{WD} = 1.0$ excluded, normal distributions, and Gram-Charlier distributions for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988).
distribution of monthly parameter sets is critical to the development of time series simulation models developed in Chapter 6 for the \( \{ \hat{p}_{DD}(vS + \tau) \} \) and \( \{ \hat{p}_{WB}(vS + \tau) \} \) series.

5.1.4. Summary and Conclusions

An annually nonhomogeneous parameter estimation (ANPE) methodology was proposed for a stochastic rainfall occurrence model. The methodology was implemented using a first-order Markov chain applied at four Florida locations. The resulting ANPE time series, \( \{ \bar{P}_{vs + \tau} \} \), suggested several features of the rainfall occurrence process that may not be accounted for by AHPE rainfall models. The shape of the annual cycle of the ANPE time series varied considerably from cycle to cycle, suggesting that there is a high degree of interannual variability in the rainfall occurrence process that is not accounted for by an AHPE model. The time of occurrence of relative annual maxima and minima in the ANPE time series shifted from year to year, suggesting that pronounced dry and wet spells do not occur at exactly the same season each year. Also, seasons that are usually wet or dry were observed to experience occasional, anomalous droughts or wet spells not accounted for in an AHPE model. The high degree of observed interannual variability in the peak to trough distance, particularly in the \( \{ \hat{p}_{DD}(vS + \tau) \} \) time series suggested that the relative frequency of spring to summer and summer to autumn drought varied considerably from year to year. Finally, the proposed ANPE method produced an observed time series of parameters, \( \{ P_{vs + \tau} \} \), which can be analyzed for possible long-term trends, and time dependence correlation structure. Using the results of these analyses, parameter time series models can be constructed and evaluated. The analysis and modeling of the ANPE time series will be investigated in detail in Chapter 6.

At first inspection, the parameter time series, \( \{ \bar{P}_{vs + \tau} \} \), obtained using the proposed ANPE method appeared complex (Figs. 5.1 to 5.8), compared to the parameter time series obtained from the AHPE method (Fig. 2.1). However, considerable order was observed when annual (Figs. 5.9 to 5.11) and seasonal (Figs. 5.12 and 5.13) histograms
were constructed for the \( \{\hat{p}_{DD}(vS + t)\} \) and \( \{\hat{p}_{WD}(vS + t)\} \) time series. The shape of the observed histograms suggested that it may be possible to relate spatial and temporal variations in convective and frontally induced rainfall processes to stochastic rainfall model parameters. The Gram-Charlier probability density function was found to be a flexible and useful model for the observed annual and seasonal histograms.

5.2 Amount Model Parameter Estimation

In Section 5.1, a time dependent annually nonhomogeneous parameter estimation method which produced a time series of parameter estimates was developed for a Markov-chain rainfall occurrence model. Each point in the parameter time series was determined using a single month of data, which provided a data sample with approximately 30 day-to-day transitions. Admittedly, this sample is small, however it is argued that this approach does give some indication of the year-to-year variability in occurrence model parameters. This time dependent approach, using single monthly samples, was not applicable to the wet day amount model due to sample size limitations. Many months contain no or very few (<5) wet days, making parameter estimation unreliable. To obtain an adequate data sample for parameter estimation, the entire observed rainfall record was sorted into monthly sub-samples. These monthly samples were then used to make point parameter estimates and to construct confidence intervals about these parameters. The details of this approach to wet day amount model parameter estimation are given below.

5.2.1 Model Theory

The rainfall process is a combination of two stochastic processes, \( \{(X_t, Y_t); t = 1, 2, \ldots\} \) where \( X_t \) is the wet or dry state of day \( t \) and \( Y_t \) is the amount of rainfall received on day \( t \) if \( X_t = WET \). If it is assumed that the wet day amount of rainfall is independent of the previous day following Richardson and Wright (1985) and Woolhiser et al. (1985, 1988), then the cumulative distribution of the wet day amount has the form
\[ F(y) = Pr\{Y \leq y \mid X_i = WET\} \quad 5.15 \]

To illustrate the proposed parameter estimation methodology, the gamma probability density function was selected to model the wet day amount of rainfall due to its wide spread use in the WGEN model (Richardson and Wright, 1985). The random variable \(U\) is said to have a gamma distribution if the probability density function\(^1\) is

\[
f_U(u; \alpha, \beta) = \begin{cases} 
0 & u \leq 0 \\
\frac{u^{\alpha-1}e^{-u/\beta}}{\beta^\alpha \Gamma(\alpha)} & u > 0 
\end{cases} \quad 5.16
\]

where the model parameters are \(\alpha\) which controls the shape of the distribution by decreasing the skewness of \(f_U(u; \alpha, \beta)\) as \(\alpha\) increases, and \(\beta\) which scales \(u\).

The U. S. National Weather Service measures and reports daily rainfall amounts in inches with a lower threshold amount of 0.01 inches (0.254 mm). Let the random variable \(Y\) be the wet day amount of rainfall, where \(y \geq 0.254\) mm. A gamma probability density model, \(f_Y(y; \alpha, \beta)\), is required for the wet day amount of rainfall. To account for the threshold amount of rainfall, the variable transformation \(Y = U + 0.254\) mm is used. To obtain the required density, \(f_Y(y; \alpha, \beta)\), the method of transformation of variables (Mendenhall et al., 1981) is used where

\[
f_Y(y) = f_U(u) \left| \frac{du}{dy} \right| \quad \text{where} \quad u = y - 0.254 \quad \text{mm} \quad 5.17
\]

Applying the transformation method of equation 5.17 gives the required wet day amount model

---

1. The notation \(f_X(x; \theta_1, \theta_2)\) is used to denote a probability density function for random variable \(X\) with parameters \(\theta_1\) and \(\theta_2\).
\[
f_r(y; \alpha, \beta) = \begin{cases} 
0 & y < 0.254 \text{ mm} \\
\delta & y = 0.254 \text{ mm} \\
\frac{(y - 0.254)^{\alpha - 1} e^{-(y - 0.254)\beta}}{\beta^\alpha \Gamma(\alpha)} & y > 0.254 \text{ mm}
\end{cases} \quad 5.18
\]

where the discrete spike, \( \delta \), is included to account for the threshold rainfall amounts.

When the year is divided into seasons, a separate gamma density function is used for each season, given by

\[
f_r(y; \alpha(\tau), \beta(\tau)) = \begin{cases} 
0 & y < 0.254 \text{ mm} \\
\delta(\tau) & y = 0.254 \text{ mm} \\
\frac{(y - 0.254)^{\alpha(\tau) - 1} e^{-(y - 0.254)\beta(\tau)}}{\beta(\tau)^\alpha \Gamma(\alpha(\tau))} & y > 0.254 \text{ mm}
\end{cases} \quad 5.19
\]

where \( \tau \) is the season index. An equivalent form of expression for equation 5.19 is the mixture (Everitt and Hand, 1981) given by

\[
f_r(y; \alpha(\tau), \beta(\tau)) = \delta(\tau) + [1 - \delta(\tau)] \left[ \frac{(y - 0.254)^{\alpha(\tau) - 1} e^{-(y - 0.254)\beta(\tau)}}{\beta(\tau)^\alpha \Gamma(\alpha(\tau))} \right] \quad 5.20
\]

To implement the wet day amount model, the cumulative probability distribution is required. The cumulative distribution for the gamma density function given in equation 5.16 is by definition

\[
F_U(u) \equiv Pr\{U \leq u\} \equiv \int_0^u f_U(u; \alpha, \beta) du \quad 5.21
\]
After variable transformation, the cumulative distribution for the random variable $Y$, the wet day amount of rainfall is

$$F_Y(y; \alpha(\tau), \beta(\tau)) \equiv Pr \{ Y \leq y \} = \int_0^y f_Y(y; \alpha(\tau), \beta(\tau))dy$$

$$= \left[\begin{array}{c}
0 \mid_{y=0}^{y<0.254} + \delta(\tau) \mid_{y=0.254} \\
+[1 - \delta(\tau)] \int_{0.254}^y \frac{(y - 0.254)^{\alpha(\tau)-1}e^{-(y-0.254)/\beta(\tau)}}{\beta(\tau)^{\alpha(\tau)}\Gamma[\alpha(\tau)]}dy
\end{array}\right]$$

$$= \left[\begin{array}{c}
0 \mid_{y=0}^{y<0.254} + \delta(\tau) \mid_{y=0.254} + [1 - \delta(\tau)]P(\alpha, y - 0.254)
\end{array}\right]$$

where $P(\alpha, \cdot)$ is the incomplete gamma function, which can be evaluated from tables (Abramowitz and Stegun, 1965) or numerically (Press et al., 1986). Equation 5.22 is then the final form of the wet day rainfall amount model based upon the gamma distribution.

### 5.2.2 Parameter Estimation Method

The proposed parameter estimation methodology for the wet day amount model is composed of two parts. First, point estimates of the gamma distribution parameters $\alpha(\tau)$ and $\beta(\tau)$ are made. Second, confidence intervals are constructed for each of these estimated parameters.

**Point parameter estimation.** The method of maximum likelihood is used to estimate the parameters $\alpha(\tau)$ and $\beta(\tau)$ of the gamma density function. The details of this point parameter estimation method are given in Appendix A. The parameter $\delta(\tau)$ is estimated by $n(\tau)_{0.254}/n(\tau)$ where $n(\tau)_{0.254}$ is the number of wet days that received 0.254 mm (0.01 in) of rainfall and $n(\tau)$ is the number of wet days in season $\tau$. 
Interval parameter estimation. To quantify the uncertainty associated with the estimates of the gamma density function parameters, $\hat{\alpha}$ and $\hat{\beta}$, a confidence region is constructed about these jointly estimated parameters. It is assumed that the joint distribution of the maximum likelihood parameter estimates $(\hat{\alpha}, \hat{\beta})$ is asymptotically $(n \to \infty)$ normal. Using the pivotal quantity method, confidence intervals are constructed for $\hat{\alpha}$ and $\hat{\beta}$ (Mood et al., 1963).

The point parameter estimates obtained above are determined from a single sample of the population of all wet day rainfall amounts. If sub-samples are selected or samples from other time periods are available, then it is reasonable to expect that the parameter estimates will vary from sample to sample. Let \( \{\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_m\} \) and \( \{\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_m\} \) be parameter samples obtained in repeated sampling with replacement of observed rainfall data. In order to construct a confidence interval about the single set of estimated parameters, \( \{\hat{\alpha}, \hat{\beta}\} \), it is first necessary to know the sampling distribution of these parameters.

Associated with maximum likelihood parameter estimates are certain large sample properties (Mood et al., 1963) which provide useful information on the distribution of these estimates. A relevant property is given in the following theorem following Larson (1974)

**Theorem 5.1.** If \( Y \) is a random variable with probability density function \( f(y; \theta) \) which depends on an unknown parameter \( \theta \), and \( \hat{\theta} \) is the maximum likelihood parameter estimate based on a random sample of size \( n \), then \( \hat{\theta} \) is approximately normally distributed for large \( n \) with mean \( \theta \) and variance

\[
\sigma^2_\theta = \left\{ nE \left[ \left( \frac{d}{d\theta} \log f(y; \theta) \right)^2 \right] \right\}^{-1}
\]

2. Population parameters are denoted by \( \theta \), while parameters estimated from observed samples are denoted by \( \hat{\theta} \).
3. A large sample is one for which the sample size tends to infinity.
In equation 5.23 and all following expressions, \( \log f(y, \theta) \) indicates the natural logarithm of the probability density, and \( E[\cdot] \) indicates the expected value.\(^4\) This large sample property also applies to maximum likelihood parameter estimates for probability density functions with more than one parameter. For a density function with two parameters, \( \theta_1 \) and \( \theta_2 \), the following theorem applies (Mood et al., 1963)

**Theorem 5.2.** If \( Y \) is a random variable with probability density function \( f(y; \theta_1, \theta_2) \) which depends on the unknown parameters \( (\theta_1, \theta_2) \) and \( (\hat{\theta}_1, \hat{\theta}_2) \) are the maximum likelihood estimators based on a random sample of size \( n \) of \( Y \), then the joint distribution of \( (\hat{\theta}_1, \hat{\theta}_2) \) is asymptotically distributed as a bivariate normal distribution with the means \( \mu_1 = \theta_1, \mu_2 = \theta_1 \), and variances

\[
\sigma^2_{\hat{\theta}_1} = \frac{-E \left[ \frac{\partial^2}{\partial \theta_1^2} \log f(y; \theta_1, \theta_2) \right]}{n \Delta} \quad \text{(5.24)}
\]

\[
\sigma^2_{\hat{\theta}_2} = \frac{-E \left[ \frac{\partial^2}{\partial \theta_2^2} \log f(y; \theta_1, \theta_2) \right]}{n \Delta} \quad \text{(5.25)}
\]

where

\[
\Delta = E \left[ \frac{\partial^2}{\partial \theta_1^2} \log f(y; \theta_1, \theta_2) \right] E \left[ \frac{\partial^2}{\partial \theta_2^2} \log f(y; \theta_1, \theta_2) \right] - \left( E \left[ \frac{\partial^2}{\partial \theta_2 \partial \theta_1} \log f(y; \theta_1, \theta_2) \right] \right)^2
\]

A formal proof of Theorem 5.1 was given by Cramér (1946) following the work of Dugué (1937). In this proof, the log likelihood function is approximated by a polynomial using a Taylor series expansion. The root of the polynomial is the estimated parameter \( \hat{\theta} \). By allowing the sample size, \( n \), to approach infinity, and applying the Central Limit Theorem, it is shown that the maximum likelihood estimate, \( \hat{\theta} \), is normally distributed.

---

4. If \( X \) is a continuous random variable and \( g(X) \) is a function of \( X \), then the expected value of \( g(X) \) is defined as (Mendenhall, 1981)

\[
E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) dx
\]

where \( f(x) \) is the probability density of the variable \( X \).
with mean equal to the population parameter, \( \theta \), and variance given by equation 5.23.

The main justification for this approach comes from the Central Limit Theorem, by which we know that the sum and consequently the mean of a set of random variables is asymptotically \( (n \to \infty) \) normal, regardless of the distribution of the individual random variables. Following the approach of Cramér, Schmetter (1956) extended Theorem 5.1 to Theorem 5.2 for distributions with more than one parameter.

For maximum likelihood parameter estimates made from large samples \( (n \to \infty) \), it is possible to obtain confidence intervals using the method of pivotal quantities. From Theorem 5.2 we know that the gamma density (equation 5.18) parameters \( \hat{\alpha} \) and \( \hat{\beta} \) are asymptotically normally distributed with means \( \alpha \) and \( \beta \), and variances

\[
\sigma^2_\alpha = \frac{-E\left[ \frac{\partial^2}{\partial \alpha^2} \log f(y; \alpha, \beta) \right]}{n\Delta}
\]

\[
\sigma^2_\beta = \frac{-E\left[ \frac{\partial^2}{\partial \beta^2} \log f(y; \alpha, \beta) \right]}{n\Delta}
\]

where

\[
\Delta = E\left[ \frac{\partial^2}{\partial \alpha^2} \log f(y; \alpha, \beta) \right] E\left[ \frac{\partial^2}{\partial \beta^2} \log f(y; \alpha, \beta) \right] - \left( E\left[ \frac{\partial^2}{\partial \beta \partial \alpha} \log f(y; \alpha, \beta) \right] \right)^2
\]

For the gamma probability density pivotal quantities are

\[
Z = \frac{\hat{\alpha} - \alpha}{\sigma_\alpha}
\]

\[
W = \frac{\hat{\beta} - \beta}{\sigma_\beta}
\]

---

5. The method of pivotal quantities is based on finding a random variable \( Z \) that is a function of the parameter \( \theta \) with a sampling distribution that does not contain any unknown parameters (Mood et al., 1963).
which are normally distributed for large samples. Using these pivotal quantities the following inequalities can be written

\[ Pr\left\{-z < \frac{\hat{\alpha} - \alpha}{\sigma_\alpha} < z \right\} = \gamma_\alpha \]

\[ Pr\left\{-w < \frac{\hat{\beta} - \beta}{\sigma_\beta} < w \right\} = \gamma_\beta \]

where \( z = z_{(1 + \gamma/2)} \) and \( w = w_{(1 + \gamma/2)} \) are defined by

\[ \gamma_\alpha = \lim_{n \to \infty} [Pr\{Z \leq z\} - Pr\{Z \leq -z\}] \]

\[ = \frac{1}{\sqrt{2\pi}} \int_{-z}^{z} e^{-t^2/2} dt - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-z} e^{-t^2/2} dt \]

\[ \gamma_\beta = \lim_{n \to \infty} [Pr\{W \leq w\} - Pr\{W \leq -w\}] \]

\[ = \frac{1}{\sqrt{2\pi}} \int_{-w}^{w} e^{-t^2/2} dt - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-w} e^{-t^2/2} dt \]

ie. \( \gamma_\alpha \) and \( \gamma_\beta \) are the areas under the standard normal curve between \( \pm z \) and \( \pm w \) respectively. Rearranging or "pivoting" 5.30 and 5.31 gives 100\( \gamma_\alpha \) and 100\( \gamma_\beta \)% confidence intervals for \( \alpha \) and \( \beta \)

\[ Pr\{\hat{\alpha} - z \sigma_\alpha < \alpha < \hat{\alpha} + z \sigma_\alpha\} = \gamma_\alpha \]

\[ Pr\{\hat{\beta} - w \sigma_\beta < \beta < \hat{\beta} + w \sigma_\beta\} = \gamma_\beta \]

Recall, that the joint distribution of \( \hat{\alpha} \) and \( \hat{\beta} \) is a bivariate normal distribution which is a nonsymmetrical bell shaped surface, \( f(\hat{\alpha}, \hat{\beta}) \), with an elliptical base (Mood et al., 1974).

A two dimensional confidence region, \( \mathcal{R} \), for the dependent random variables \( \hat{\alpha} \) and \( \hat{\beta} \) is
an ellipse with minor and major axis given by the intervals in equations 5.34 and 5.35. The parameters \( \alpha \) and \( \beta \) will be contained in this confidence region \( 100\gamma \% \) of the time where

\[
Pr\{(\alpha, \beta) \in \mathcal{R}\} = \int_{\mathcal{R}} f(\alpha, \beta) d\alpha d\beta = \gamma
\]

5.36

Note that the rectangle denoted \( \mathcal{R} \) formed by equations 5.34 and 5.35 is not a true confidence region, however this rectangle bounds the elliptical region, \( \mathcal{R} \), and contains the parameters \( 100\gamma \% \) of the time.

To evaluate the confidence intervals (equations 5.34 and 5.35) or confidence region (equation 5.36) it is necessary to determine the standard deviation of \( \alpha, \sigma_\alpha, \) and \( \beta, \sigma_\beta. \) This requires the evaluation of equations 5.26 and 5.27 for the gamma probability density (Appendix E). The resulting standard deviations are

\[
\hat{\sigma}_\alpha = \sqrt{\frac{\alpha}{n(\hat{\alpha}\Sigma_\alpha - 1)}}
\]

5.37

\[
\hat{\sigma}_\beta = \sqrt{\frac{\Sigma_\beta^2}{n(\hat{\alpha}\Sigma_\alpha - 1)}}
\]

5.38

Using these estimates of the standard deviation of \( \alpha \) and \( \beta \) it is possible to construct confidence intervals for \( \alpha \) and \( \beta \) by substituting (5.37) into (5.34) and (5.38) into (5.35). The resulting confidence intervals are

\[
Pr\{\hat{\alpha} - z\sqrt{\frac{\alpha}{n(\hat{\alpha}\Sigma_\alpha - 1)}} < \alpha < \hat{\alpha} + z\sqrt{\frac{\alpha}{n(\hat{\alpha}\Sigma_\alpha - 1)}}\} = \gamma_\alpha
\]

5.39

\[
Pr\{\hat{\beta} - w\sqrt{\frac{\Sigma_\beta^2}{n(\hat{\alpha}\Sigma_\alpha - 1)}} < \beta < \hat{\beta} + w\sqrt{\frac{\Sigma_\beta^2}{n(\hat{\alpha}\Sigma_\alpha - 1)}}\} = \gamma_\beta
\]

5.40
The unknown population parameters, $\alpha$ and $\beta$, of the gamma distribution rainfall amount model will lie within these intervals $100\gamma_a$ and $100\gamma_b$ % of the time, respectively. To implement equations 5.39 and 5.40, the probabilities $\gamma_a$ and $\gamma_b$ are specified by the user. Using $\gamma_a$ and $\gamma_b$, $z$ and $w$ are then determined from a table for the standard normal density function where $\gamma_a = \Phi(z) - \Phi(-z)$ and $\gamma_b = \Phi(w) - \Phi(-w)$. The confidence intervals are then completely specified with all needed parameters known.

5.2.3 Results of Parameter Estimation Method

Point and interval parameter estimates were made for the gamma distribution wet day amount model using daily rainfall data (Chapter 3) for Gainesville, Orlando, Tampa, and Ft. Myers, Florida from 1900 through 1988. Monthly seasons ($\tau = 1, 2, \ldots, 12$) were selected following the approach used by Richardson and Wright (1985) in the WGEN stochastic rainfall model (Chapter 2). The method of maximum likelihood (Appendix A) was used to estimate the gamma distribution parameters $\hat{\alpha}(\tau)$ and $\hat{\beta}(\tau)$ of equation 5.22. To apply this method, the wet day rainfall amounts were sorted by month, and only amounts with $y > 0.254$ mm (0.01 in) were used for estimation. The wet day threshold amount parameter, $\delta(\tau)$, was estimated by $n(\tau)_{0.254}/n(\tau)$ where $n(\tau)_{0.254}$ is the number of wet days in month $\tau$ that received 0.254 mm (0.01 in) of rainfall and $n(\tau)$ is the total number of wet days ($y \geq 0.254$ mm) in month $\tau$. Confidence intervals were then estimated for the gamma distribution parameters, $\alpha(\tau)$ and $\beta(\tau)$. First, the standard deviations, $\sigma_\alpha(\tau)$ and $\sigma_\beta(\tau)$ were estimated using equations 5.37 and 5.38 respectively. Second, 99% confidence intervals ($\gamma_a = 0.99, z = 2.575; \gamma_b = 0.99, w = 2.575$) were constructed using equations 5.39 and 5.40. Maximum likelihood parameter estimates, $\hat{\alpha}(\tau)$ and $\hat{\beta}(\tau)$, for the gamma distribution and associated 99% confidence intervals are

6. The notation $\Phi(x)$ indicates the area under the standard normal density function given in most statistical texts and defined by

$$
\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt
$$
given in Tables 5.6 through 5.9, while estimates for the threshold amount parameter, \( \delta(\tau) \), are given in Table 5.10, for the four locations considered.

**Temporal variability of parameters.** To investigate possible intraannual seasonality of parameters, \( \alpha(\tau) \pm \tilde{\alpha}(\tau) \), \( \beta(\tau) \pm \tilde{\beta}(\tau) \), and \( \delta(\tau) \) were plotted (Figs. 5.14 through 5.16). The shape parameter, \( \alpha(\tau) \), varied in a regular manner throughout the year, with relative maxima observed during the winter and summer wet seasons, and relative minima observed during the spring and autumn dry seasons. The standard deviations of the shape parameter, \( \tilde{\alpha}(\tau) \), also varied throughout the year, with relatively low values observed during the summer and higher values observed during the rest of the year. The scale parameter, \( \tilde{\beta}(\tau) \), varied in a pronounced annual cycle, with relative maxima coinciding with scale parameter minima and relative minima coinciding with scale parameter maxima. The standard deviations of the scale parameter, \( \tilde{\beta}(\tau) \), varied in a manner similar to those of the shape parameter, with low values observed during summer and higher values observed during the rest of the year. The wet day threshold amount parameter, \( \delta(\tau) \), also exhibited an annual cycle, with low values observed from spring through summer, and higher values observed from autumn through winter.

**Spatial variability of parameters.** The possibility of regular latitudinal increases or decreases in the wet day amount model parameters was investigated. For the four locations considered, \( \alpha(\tau) \) vs \( \tau \) (Fig. 5.17), \( \tilde{\beta}(\tau) \) vs \( \tau \) (Fig. 5.18), and \( \delta(\tau) \) vs \( \tau \) (Fig. 5.19) were plotted. There was evidence of regular latitudinal gradients in parameter values for Gainesville, Orlando and Tampa during some seasons, however parameters from Ft. Myers did not follow this pattern. For these three stations, the shape parameter, \( \tilde{\alpha}(\tau) \), decreased from north to south during midsummer and increased during midwinter. The scale parameter, \( \tilde{\beta}(\tau) \), exhibited similar, but opposite seasonal gradients with the parameter increasing from north to south during summer and decreasing during winter. The strongest evidence for latitudinal gradients was observed for the threshold parameter, \( \delta(\tau) \), which generally increased from north to south throughout the year. For all
Table 5.6  Maximum likelihood estimates of the gamma-distribution parameters $\hat{\alpha}(\tau)$ and $\hat{\beta}(\tau)$ with associated 99% confidence intervals by month for Gainesville, Florida.

<table>
<thead>
<tr>
<th>Month $\tau$</th>
<th>$\hat{\alpha}(\tau) - z\sigma_{\alpha}(\tau)$</th>
<th>$\hat{\alpha}(\tau)$</th>
<th>$\hat{\alpha}(\tau) + z\sigma_{\alpha}(\tau)$</th>
<th>$\hat{\beta}(\tau) - w\sigma_{\beta}(\tau)$</th>
<th>$\hat{\beta}(\tau)$</th>
<th>$\hat{\beta}(\tau) + w\sigma_{\beta}(\tau)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.595</td>
<td>0.677</td>
<td>0.760</td>
<td>13.103</td>
<td>15.841</td>
<td>18.579</td>
</tr>
<tr>
<td>2</td>
<td>0.599</td>
<td>0.683</td>
<td>0.767</td>
<td>14.239</td>
<td>17.251</td>
<td>20.263</td>
</tr>
<tr>
<td>3</td>
<td>0.661</td>
<td>0.753</td>
<td>0.846</td>
<td>13.971</td>
<td>16.811</td>
<td>19.652</td>
</tr>
<tr>
<td>4</td>
<td>0.569</td>
<td>0.661</td>
<td>0.752</td>
<td>15.968</td>
<td>19.917</td>
<td>23.866</td>
</tr>
<tr>
<td>5</td>
<td>0.624</td>
<td>0.707</td>
<td>0.790</td>
<td>13.419</td>
<td>16.060</td>
<td>18.701</td>
</tr>
<tr>
<td>6</td>
<td>0.699</td>
<td>0.773</td>
<td>0.846</td>
<td>15.221</td>
<td>17.496</td>
<td>19.771</td>
</tr>
<tr>
<td>7</td>
<td>0.746</td>
<td>0.814</td>
<td>0.882</td>
<td>11.853</td>
<td>13.353</td>
<td>14.854</td>
</tr>
<tr>
<td>8</td>
<td>0.637</td>
<td>0.695</td>
<td>0.753</td>
<td>14.647</td>
<td>16.616</td>
<td>18.585</td>
</tr>
<tr>
<td>9</td>
<td>0.600</td>
<td>0.664</td>
<td>0.728</td>
<td>14.472</td>
<td>16.770</td>
<td>19.068</td>
</tr>
<tr>
<td>10</td>
<td>0.428</td>
<td>0.489</td>
<td>0.549</td>
<td>18.323</td>
<td>22.786</td>
<td>27.249</td>
</tr>
<tr>
<td>11</td>
<td>0.558</td>
<td>0.653</td>
<td>0.749</td>
<td>11.798</td>
<td>14.944</td>
<td>18.089</td>
</tr>
<tr>
<td>12</td>
<td>0.549</td>
<td>0.625</td>
<td>0.702</td>
<td>13.842</td>
<td>16.830</td>
<td>19.818</td>
</tr>
</tbody>
</table>
Table 5.7  Maximum likelihood estimates of the gamma-distribution parameters $\hat{\alpha}(\tau)$ and $\hat{\beta}(\tau)$ with associated 99% confidence intervals by month for Orlando, Florida.

<table>
<thead>
<tr>
<th>Month $\tau$</th>
<th>$\hat{\alpha}(\tau) - z\sigma_\alpha(\tau)$</th>
<th>$\hat{\alpha}(\tau)$</th>
<th>$\hat{\alpha}(\tau) + z\sigma_\alpha(\tau)$</th>
<th>$\hat{\beta}(\tau) - w\sigma_\beta(\tau)$</th>
<th>$\hat{\beta}(\tau)$</th>
<th>$\hat{\beta}(\tau) + w\sigma_\beta(\tau)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.591</td>
<td>0.683</td>
<td>0.774</td>
<td>11.453</td>
<td>14.149</td>
<td>16.846</td>
</tr>
<tr>
<td>2</td>
<td>0.690</td>
<td>0.797</td>
<td>0.905</td>
<td>10.954</td>
<td>13.401</td>
<td>15.848</td>
</tr>
<tr>
<td>3</td>
<td>0.640</td>
<td>0.734</td>
<td>0.828</td>
<td>13.666</td>
<td>16.617</td>
<td>19.567</td>
</tr>
<tr>
<td>4</td>
<td>0.655</td>
<td>0.765</td>
<td>0.874</td>
<td>12.849</td>
<td>15.989</td>
<td>19.128</td>
</tr>
<tr>
<td>5</td>
<td>0.638</td>
<td>0.720</td>
<td>0.802</td>
<td>13.725</td>
<td>16.320</td>
<td>18.915</td>
</tr>
<tr>
<td>6</td>
<td>0.656</td>
<td>0.721</td>
<td>0.787</td>
<td>16.024</td>
<td>18.345</td>
<td>20.666</td>
</tr>
<tr>
<td>7</td>
<td>0.701</td>
<td>0.765</td>
<td>0.828</td>
<td>14.464</td>
<td>16.328</td>
<td>18.192</td>
</tr>
<tr>
<td>8</td>
<td>0.708</td>
<td>0.775</td>
<td>0.842</td>
<td>12.970</td>
<td>14.714</td>
<td>16.457</td>
</tr>
<tr>
<td>9</td>
<td>0.641</td>
<td>0.705</td>
<td>0.769</td>
<td>15.415</td>
<td>17.662</td>
<td>19.910</td>
</tr>
<tr>
<td>10</td>
<td>0.562</td>
<td>0.635</td>
<td>0.708</td>
<td>14.824</td>
<td>17.779</td>
<td>20.734</td>
</tr>
<tr>
<td>11</td>
<td>0.516</td>
<td>0.602</td>
<td>0.689</td>
<td>11.155</td>
<td>14.165</td>
<td>17.176</td>
</tr>
<tr>
<td>12</td>
<td>0.598</td>
<td>0.692</td>
<td>0.786</td>
<td>10.200</td>
<td>12.622</td>
<td>15.044</td>
</tr>
</tbody>
</table>
Table 5.8  Maximum likelihood estimates of the gamma-distribution parameters $\hat{\alpha}(\tau)$ and $\hat{\beta}(\tau)$ with associated 99% confidence intervals by month for Tampa, Florida.

<table>
<thead>
<tr>
<th>Month $\tau$</th>
<th>$\hat{\alpha}(\tau) - z\sigma_{\alpha}(\tau)$</th>
<th>$\hat{\alpha}(\tau)$</th>
<th>$\hat{\alpha}(\tau) + z\sigma_{\alpha}(\tau)$</th>
<th>$\hat{\beta}(\tau) - w\sigma_{\beta}(\tau)$</th>
<th>$\hat{\beta}(\tau)$</th>
<th>$\hat{\beta}(\tau) + w\sigma_{\beta}(\tau)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.612</td>
<td>0.710</td>
<td>0.807</td>
<td>10.425</td>
<td>12.904</td>
<td>15.382</td>
</tr>
<tr>
<td>2</td>
<td>0.645</td>
<td>0.746</td>
<td>0.847</td>
<td>12.788</td>
<td>15.727</td>
<td>18.665</td>
</tr>
<tr>
<td>3</td>
<td>0.570</td>
<td>0.658</td>
<td>0.746</td>
<td>15.958</td>
<td>19.716</td>
<td>23.474</td>
</tr>
<tr>
<td>4</td>
<td>0.628</td>
<td>0.746</td>
<td>0.864</td>
<td>12.145</td>
<td>15.529</td>
<td>18.913</td>
</tr>
<tr>
<td>5</td>
<td>0.558</td>
<td>0.642</td>
<td>0.726</td>
<td>15.903</td>
<td>19.608</td>
<td>23.314</td>
</tr>
<tr>
<td>6</td>
<td>0.609</td>
<td>0.673</td>
<td>0.738</td>
<td>17.852</td>
<td>20.660</td>
<td>23.467</td>
</tr>
<tr>
<td>7</td>
<td>0.630</td>
<td>0.689</td>
<td>0.748</td>
<td>15.593</td>
<td>17.743</td>
<td>19.893</td>
</tr>
<tr>
<td>8</td>
<td>0.675</td>
<td>0.738</td>
<td>0.800</td>
<td>15.364</td>
<td>17.401</td>
<td>19.438</td>
</tr>
<tr>
<td>9</td>
<td>0.598</td>
<td>0.659</td>
<td>0.720</td>
<td>16.535</td>
<td>19.073</td>
<td>21.611</td>
</tr>
<tr>
<td>10</td>
<td>0.552</td>
<td>0.637</td>
<td>0.721</td>
<td>13.014</td>
<td>16.112</td>
<td>19.210</td>
</tr>
<tr>
<td>11</td>
<td>0.525</td>
<td>0.621</td>
<td>0.716</td>
<td>11.138</td>
<td>14.355</td>
<td>17.573</td>
</tr>
<tr>
<td>12</td>
<td>0.600</td>
<td>0.695</td>
<td>0.791</td>
<td>10.114</td>
<td>12.544</td>
<td>14.975</td>
</tr>
</tbody>
</table>
Table 5.9  Maximum likelihood estimates of the gamma-distribution parameters $\hat{\alpha}(\tau)$ and $\hat{\beta}(\tau)$ with associated 99% confidence intervals by month for Ft. Myers, Florida.

<table>
<thead>
<tr>
<th>Month $\tau$</th>
<th>$\hat{\alpha}(\tau) - z \sigma_{\alpha}(\tau)$</th>
<th>$\hat{\alpha}(\tau)$</th>
<th>$\hat{\alpha}(\tau) + z \sigma_{\alpha}(\tau)$</th>
<th>$\hat{\beta}(\tau) - w \sigma_{\beta}(\tau)$</th>
<th>$\hat{\beta}(\tau)$</th>
<th>$\hat{\beta}(\tau) + w \sigma_{\beta}(\tau)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.589</td>
<td>0.693</td>
<td>0.796</td>
<td>9.581</td>
<td>12.145</td>
<td>14.709</td>
</tr>
<tr>
<td>2</td>
<td>0.579</td>
<td>0.680</td>
<td>0.782</td>
<td>12.504</td>
<td>15.856</td>
<td>19.207</td>
</tr>
<tr>
<td>3</td>
<td>0.569</td>
<td>0.668</td>
<td>0.767</td>
<td>14.638</td>
<td>18.551</td>
<td>22.464</td>
</tr>
<tr>
<td>4</td>
<td>0.620</td>
<td>0.737</td>
<td>0.854</td>
<td>11.778</td>
<td>15.096</td>
<td>18.414</td>
</tr>
<tr>
<td>5</td>
<td>0.643</td>
<td>0.728</td>
<td>0.813</td>
<td>13.763</td>
<td>16.429</td>
<td>19.095</td>
</tr>
<tr>
<td>6</td>
<td>0.625</td>
<td>0.684</td>
<td>0.743</td>
<td>20.451</td>
<td>23.317</td>
<td>26.184</td>
</tr>
<tr>
<td>7</td>
<td>0.710</td>
<td>0.772</td>
<td>0.834</td>
<td>14.650</td>
<td>16.462</td>
<td>18.275</td>
</tr>
<tr>
<td>8</td>
<td>0.727</td>
<td>0.792</td>
<td>0.857</td>
<td>14.633</td>
<td>16.478</td>
<td>18.322</td>
</tr>
<tr>
<td>9</td>
<td>0.603</td>
<td>0.660</td>
<td>0.717</td>
<td>18.649</td>
<td>21.295</td>
<td>23.941</td>
</tr>
<tr>
<td>10</td>
<td>0.563</td>
<td>0.640</td>
<td>0.717</td>
<td>17.270</td>
<td>20.881</td>
<td>24.491</td>
</tr>
<tr>
<td>11</td>
<td>0.585</td>
<td>0.702</td>
<td>0.820</td>
<td>9.882</td>
<td>12.915</td>
<td>15.949</td>
</tr>
<tr>
<td>12</td>
<td>0.617</td>
<td>0.733</td>
<td>0.849</td>
<td>8.941</td>
<td>11.452</td>
<td>13.962</td>
</tr>
</tbody>
</table>
Table 5.10  Estimates of the wet day amount model parameter $\delta(\tau)$ by month for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988).

<table>
<thead>
<tr>
<th>Month</th>
<th>Gainesville</th>
<th>Orlando</th>
<th>Tampa</th>
<th>Ft. Myers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.073</td>
<td>0.086</td>
<td>0.130</td>
<td>0.115</td>
</tr>
<tr>
<td>2</td>
<td>0.063</td>
<td>0.076</td>
<td>0.111</td>
<td>0.079</td>
</tr>
<tr>
<td>3</td>
<td>0.060</td>
<td>0.048</td>
<td>0.075</td>
<td>0.091</td>
</tr>
<tr>
<td>4</td>
<td>0.033</td>
<td>0.049</td>
<td>0.079</td>
<td>0.047</td>
</tr>
<tr>
<td>5</td>
<td>0.041</td>
<td>0.051</td>
<td>0.090</td>
<td>0.047</td>
</tr>
<tr>
<td>6</td>
<td>0.058</td>
<td>0.047</td>
<td>0.072</td>
<td>0.043</td>
</tr>
<tr>
<td>7</td>
<td>0.037</td>
<td>0.053</td>
<td>0.081</td>
<td>0.041</td>
</tr>
<tr>
<td>8</td>
<td>0.041</td>
<td>0.055</td>
<td>0.067</td>
<td>0.046</td>
</tr>
<tr>
<td>9</td>
<td>0.048</td>
<td>0.067</td>
<td>0.083</td>
<td>0.050</td>
</tr>
<tr>
<td>10</td>
<td>0.083</td>
<td>0.062</td>
<td>0.113</td>
<td>0.059</td>
</tr>
<tr>
<td>11</td>
<td>0.079</td>
<td>0.126</td>
<td>0.123</td>
<td>0.089</td>
</tr>
<tr>
<td>12</td>
<td>0.074</td>
<td>0.079</td>
<td>0.092</td>
<td>0.088</td>
</tr>
</tbody>
</table>
Figure 5.14 Gamma-distribution parameter, \( \alpha \), with plus and minus one standard deviation by month for Gainesville, Orlando, Tampa, and Ft. Myers, Florida.
Figure 5.15 Gamma-distribution parameter, $\beta$, with plus and minus one standard deviation by month for Gainesville, Orlando, Tampa, and Ft. Myers, Florida.
Figure 5.16  Gamma-distribution parameter, $\delta$, by month for Gainesville, Orlando, Tampa, and Ft. Myers, Florida.
Figure 5.17 Comparison of monthly gamma-distribution parameter, $\alpha$, for Gainesville, Orlando, Tampa, Ft. Myers, Florida.
Figure 5.18 Comparison of monthly gamma-distribution parameter, $\beta$, for Gainesville, Orlando, Tampa, Ft. Myers, Florida.
Figure 5.19 Comparison of monthly gamma distribution-parameter, $\delta$, for Gainesville, Orlando, Tampa, Ft. Myers, Florida.
parameters, there were months in which no clearly defined latitudinal gradients were observed. This may be related to periods of transition between seasonal meteorological regimes. Geographically, the four stations considered varied considerably, with the sample composed of two inland stations and two Gulf of Mexico stations. More stations, with a broader distribution of geographic and climatological conditions would be required to quantify the parameter gradients suggested by these results.

**Evaluation of point parameter estimates.** To implement the wet day amount portion of a stochastic rainfall model, the cumulative probability distribution is used, rather than the probability density function. For this reason, the gamma distribution (equation 5.22) was implemented and evaluated, rather than the gamma density (equation 5.21). Maximum likelihood point parameter estimates, $\hat{\alpha}(\tau)$ and $\hat{\beta}(\tau)$, were used to implement the gamma distribution model, $F_{\text{Model}} = F_y(y; \hat{\alpha}, \hat{\beta})$. The threshold parameter, $\delta(\tau)$, was used in all of the models. The models were implemented with monthly seasons ($\tau = 1, 2, \ldots, 12$) at all locations considered.

The classical Kolmogorov-Smirnov goodness of fit test (Appendix B) was used to test the null hypothesis that each monthly sample of wet day rainfall amounts was drawn from a population having a gamma distribution. The Kolmogorov-Smirnov test statistics, $D$, associated p-values, and sample sizes are given in Tables 5.11, 5.12 and 5.13 respectively. The observed $D$ values were small and the associated p-values were large (> 0.20) for all months and locations, except for October at Gainesville. These small $D$ values and large p-values do not offer evidence to reject the null hypothesis. For the October sample from Gainesville, the null hypothesis can be rejected for significance.

---

7. The critical values of the Kolmogorov-Smirnov test statistic (Haan, 1977) for a specified level of significance using a large sample of size $n$ are: $D = 1.07/\sqrt{n}$ for $\alpha = 0.20$; $D = 1.14/\sqrt{n}$ for $\alpha = 0.15$; $D = 1.22/\sqrt{n}$ for $\alpha = 0.10$; $D = 1.36/\sqrt{n}$ for $\alpha = 0.05$; $D = 1.63/\sqrt{n}$ for $\alpha = 0.01$. If a significance level, $\alpha$, is selected which is greater than or equal to the p-value, the null hypothesis is rejected; if $\alpha$ is less than the p-value, the null hypothesis can not be rejected (Mendenhall, 1981).
Table 5.11  Kolmogorov-Smirnov test statistics, D, for monthly gamma distribution wet day amount model.

<table>
<thead>
<tr>
<th>City</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>GVL</td>
<td>0.062</td>
<td>0.058</td>
<td>0.055</td>
<td>0.065</td>
<td>0.062</td>
<td>0.051</td>
<td>0.041</td>
<td>0.057</td>
<td>0.065</td>
<td>0.115</td>
<td>0.073</td>
<td>0.085</td>
</tr>
<tr>
<td>ORL</td>
<td>0.089</td>
<td>0.062</td>
<td>0.079</td>
<td>0.068</td>
<td>0.070</td>
<td>0.047</td>
<td>0.064</td>
<td>0.058</td>
<td>0.051</td>
<td>0.084</td>
<td>0.087</td>
<td>0.068</td>
</tr>
<tr>
<td>TAM</td>
<td>0.060</td>
<td>0.056</td>
<td>0.058</td>
<td>0.060</td>
<td>0.069</td>
<td>0.058</td>
<td>0.054</td>
<td>0.045</td>
<td>0.064</td>
<td>0.070</td>
<td>0.101</td>
<td>0.074</td>
</tr>
<tr>
<td>FTM</td>
<td>0.067</td>
<td>0.063</td>
<td>0.057</td>
<td>0.080</td>
<td>0.057</td>
<td>0.048</td>
<td>0.062</td>
<td>0.049</td>
<td>0.061</td>
<td>0.077</td>
<td>0.079</td>
<td>0.077</td>
</tr>
</tbody>
</table>

Table 5.12  Kolmogorov-Smirnov p-values for monthly gamma distribution wet day amount model.

<table>
<thead>
<tr>
<th>City</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>GVL</td>
<td>0.583</td>
<td>0.657</td>
<td>0.714</td>
<td>0.538</td>
<td>0.579</td>
<td>0.645</td>
<td>0.890</td>
<td>0.501</td>
<td>0.390</td>
<td>0.041</td>
<td>0.527</td>
<td>0.210</td>
</tr>
<tr>
<td>ORL</td>
<td>0.274</td>
<td>0.670</td>
<td>0.278</td>
<td>0.509</td>
<td>0.393</td>
<td>0.727</td>
<td>0.364</td>
<td>0.542</td>
<td>0.661</td>
<td>0.205</td>
<td>0.372</td>
<td>0.664</td>
</tr>
<tr>
<td>TAM</td>
<td>0.755</td>
<td>0.751</td>
<td>0.645</td>
<td>0.754</td>
<td>0.485</td>
<td>0.498</td>
<td>0.563</td>
<td>0.788</td>
<td>0.362</td>
<td>0.521</td>
<td>0.237</td>
<td>0.502</td>
</tr>
<tr>
<td>FTM</td>
<td>0.694</td>
<td>0.692</td>
<td>0.779</td>
<td>0.426</td>
<td>0.601</td>
<td>0.611</td>
<td>0.374</td>
<td>0.706</td>
<td>0.380</td>
<td>0.259</td>
<td>0.526</td>
<td>0.573</td>
</tr>
</tbody>
</table>

Table 5.13  Kolmogorov-Smirnov sample sizes, n, for monthly gamma distribution wet day amount model.

<table>
<thead>
<tr>
<th>City</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>GVL</td>
<td>157</td>
<td>161</td>
<td>162</td>
<td>153</td>
<td>156</td>
<td>211</td>
<td>194</td>
<td>207</td>
<td>193</td>
<td>148</td>
<td>123</td>
<td>156</td>
</tr>
<tr>
<td>ORL</td>
<td>125</td>
<td>136</td>
<td>157</td>
<td>145</td>
<td>166</td>
<td>212</td>
<td>210</td>
<td>188</td>
<td>202</td>
<td>160</td>
<td>112</td>
<td>116</td>
</tr>
<tr>
<td>TAM</td>
<td>126</td>
<td>143</td>
<td>161</td>
<td>124</td>
<td>149</td>
<td>207</td>
<td>213</td>
<td>212</td>
<td>210</td>
<td>136</td>
<td>105</td>
<td>123</td>
</tr>
<tr>
<td>FTM</td>
<td>111</td>
<td>127</td>
<td>133</td>
<td>120</td>
<td>182</td>
<td>253</td>
<td>215</td>
<td>210</td>
<td>225</td>
<td>171</td>
<td>105</td>
<td>103</td>
</tr>
</tbody>
</table>
levels with $\alpha \geq 0.05$, but can not be rejected for a significance level with $\alpha = 0.01$. These results offer strong evidence to support the choice of the gamma distribution for the wet day amount model. However, the results of the Kolmogorov-Smirnov test can be misleading as Haan suggested,

Many hydrologists discourage the use of the chi-square and Kolmogorov-Smirnov tests when testing hydrologic frequency distributions. The reason for this is the importance of the tails of the hydrologic frequency distributions and the insensitivity of these tests in the tails of these distributions. (1977, p. 178)

To further investigate the goodness of fit of the gamma distribution to the wet day amount of rainfall data, $F_{\text{Model}}$ (equation 5.22) and $F_{\text{Observed}}$ were presented as semilog plots for each month at the four locations considered (Figs. 5.20 to 5.24). $F_{\text{Observed}}$ (Appendix B) was shown as a step function, which is the fraction of observed data points $(y_1, y_2, ..., y_n)$ to the left of a particular rainfall amount, $y_{\text{Observed}}$. The distributions for the month of May at Gainesville were given at an enlarged scale in Figure 5.20 as an illustrative example. May is a relatively dry month, with agricultural production normally requiring irrigation to bring crops to harvest. For all seasons considered, $F_{\text{Model}}$ agreed very closely with $F_{\text{Observed}}$ for small daily rainfall amounts ($y < 1.0 \text{ mm}$). For moderate daily rainfall amounts ($1.0 \text{ mm} < y < 30.0 \text{ mm}$), $F_{\text{Observed}}$ exceeded $F_{\text{Model}}$, while for large rainfall amounts ($y > 30.0 \text{ mm}$), $F_{\text{Model}}$ exceeded $F_{\text{Observed}}$. This general pattern was clearly evident for May/Gainesville (Fig. 5.20), and was present in most other months with only slight variations (Fig. 5.21 through Fig. 5.24). These figures showed that the shape of $F_{\text{Observed}}$ is slightly, but consistently different from $F_{\text{Model}}$ throughout the year at all four locations.

The results of the Kolmogorov-Smirnov tests (Tables 5.11 to 5.13) suggested that gamma distribution agreed well with the observed distributions, while the graphical comparisons of $F_{\text{Observed}}$ and $F_{\text{Model}}$ suggested that there were regular deviations during all seasons and at all locations. Neither of these tests provided a clear, quantitative measure
Figure 5.20 Comparison of rainfall amount cumulative distributions, $F_{\text{Model}}$, $F_{\text{Observed}}$, $F_{\text{UCL}}$, and $F_{\text{LCL}}$, for May at Gainesville, Florida.
Figure 5.21 Comparison of rainfall amount cumulative distributions, $F_{\text{Model}}$, $F_{\text{Observed}}$, $F_{\text{UCL}}$, and $F_{\text{LCL}}$, by month at Gainesville, Florida.
Figure 5.22  Comparison of rainfall amount cumulative distributions, $F_{\text{Model}}$, $F_{\text{Observed}}$, $F_{\text{UCL}}$, and $F_{\text{LCL}}$, by month at Orlando, Florida.
Figure 5.23  Comparison of rainfall amount cumulative distributions, $F_{\text{Model}}$, $F_{\text{Observed}}$, $F_{\text{UCL}}$, and $F_{\text{LCL}}$, by month at Tampa, Florida.
Figure 5.24  Comparison of rainfall amount cumulative distributions, $F_{\text{Model}}$, $F_{\text{Observed}}$, $F_{\text{UCL}}$, and $F_{\text{LCL}}$, by month at Ft. Myers, Florida.
of the goodness of fit of the gamma distribution model as applied to the simulation of wet
day amounts of rainfall. To further evaluate the goodness of fit in application, an
illustrative example of rainfall simulation was considered.

Given that a particular day is wet, the wet day amount of rainfall, $y_{\text{Model}}$, is obtained
by generating a uniformly distributed random number, $R$, between 0 and 1, and solving
the equation $R = F_{\text{Model}}(y_{\text{Model}})$ for $y_{\text{Model}}$. This equation can be solved (Pritsker, 1986)
analytically or by numerical iteration, however in this example, the solution was obtained
graphically (Fig. 5.25) to clarify the process. First, generate a random number, $R$, enter
the graph at point $R$ along the cumulative probability axis, and move horizontally to
$F_{\text{Model}} = R$ and $F_{\text{Observed}} = R$. Second, move vertically to the amount axis obtaining the
rainfall amounts $y_{\text{Observed}} = F_{\text{Observed}}^{-1}(R)$ and $y_{\text{Model}} = F_{\text{Model}}^{-1}(R)$. For example, with $R = 0.482$,
$y_{\text{Observed}} = 4.8$ mm and $y_{\text{Model}} = 6.0$ mm, resulting in an over prediction of
$y_{\text{Model}} - y_{\text{Observed}} = 1.2$ mm. While for $R = 0.982$, $y_{\text{Observed}} = 66$ mm and $y_{\text{Model}} = 55$ mm,
resulting in an under prediction of $y_{\text{Model}} - y_{\text{Observed}} = -11$ mm. This example suggested
that the gamma distribution model, when used to simulate wet day rainfall amounts,
slightly over predicted moderate amounts of daily rainfall ($1 \text{ mm} < y < 30 \text{ mm}$), and
significantly under predicted large daily amounts ($y > 30 \text{ mm}$). It is useful to note, that
the largest error, $y_{\text{Model}} - y_{\text{Observed}}$, occurred in the upper tail of the distribution where the
vertical separation or Kolmogorov-Smirnov $D = F_{\text{Model}} - F_{\text{Observed}}$ was small, and that
smaller errors occurred in the midrange of the distribution where the
Kolmogorov-Smirnov $D$ was large. This example confirmed Haan’s (1977) caution and
suggested that the standard Kolmogorov-Smirnov test may not adequately quantify the
goodness of fit of a distribution used to simulate wet day rainfall amounts. As
demonstrated, small $D = F_{\text{Model}} - F_{\text{Observed}}$ values can result in significant errors
($y_{\text{Model}} - y_{\text{Observed}}$), if they occur in the upper tail of $F_{\text{Model}}$. 
Figure 5.25  Example of the simulation of wet day rainfall amounts using the gamma distribution with associated errors for May at Gainesville, Florida.
To investigate the significance of the reported deviations (Table 5.11 and Figs. 5.21 to 5.24) between the gamma distribution and the observed distribution of daily rainfall amounts, an alternative to the Kolmogorov-Smirnov D was investigated. The alternative evaluation method used was:

i. Let $F_{\text{Model}}(y)$ be the theoretical cumulative distribution under consideration.

ii. Let $F_{\text{Observed}}(y)$ be the observed cumulative distribution of wet day rainfall amounts determined by sorting (Appendix B).

iii. Determine the maximum positive and negative horizontal deviations, $H^+$ and $H^-$ defined by:

$$H^+(y_{\text{Observed}}) = \max[y_{\text{Model}} - y_{\text{Observed}}]$$

$$= \max[F^{-1}_{\text{Model}} - F^{-1}_{\text{Observed}}]$$

$$H^-(y_{\text{Observed}}) = \max[y_{\text{Observed}} - y_{\text{Model}}]$$

$$= \max[F^{-1}_{\text{Observed}} - F^{-1}_{\text{Model}}]$$

The method is conducted by numerically iterating on $y$ beginning at $y_{\text{Observed}}$ until a $y_{\text{Model}}$ value is found which satisfies the equation, $F_{\text{Observed}}(y_{\text{Observed}}) = F_{\text{Model}}(y_{\text{Model}})$. This approach is used at each step of the $F_{\text{Observed}}$ function. The maximum horizontal deviations, $H^+(y_{\text{Observed}})$ and $H^-(y_{\text{Observed}})$, are then located. These horizontal deviations are simply the maximum errors in mm between observed and simulated wet day amounts. $H^+(y_{\text{Observed}})$ is the maximum over prediction of the model located at $y_{\text{Observed}}$ on the amount axis, while $H^-(y_{\text{Observed}})$ is the maximum under prediction.

---

8. Given a functional relation $y = f(x)$, the value of $y$ can be determined for any particular $x$ by substitution into $f(x)$. However, the value of $x$ associated with a particular $y$ is determined by making the inverse transformation $x = f^{-1}(y)$. This transformation can be made analytically or numerically.
The $H$ error method described above was used to evaluate the monthly gamma-distribution wet day amount model for Gainesville, Orlando, Tampa and Ft. Myers, Florida. The maximum errors, $H^+$ and $H^-$, are given in Tables 5.14 and 5.16, while the location of these errors, $y_{\text{observed}}$, on the amount axis of the gamma distribution are given in Tables 5.15 and 5.17. For example, consider the month of April which is a dry month, during which considerable irrigated agricultural production occurs in Florida. The maximum errors for this month at Gainesville were $H^+(32 \text{ mm}) = 4 \text{ mm/day}$ and $H^-(116 \text{ mm}) = -36 \text{ mm/day}$. This example suggested that moderate wet day amounts of rainfall on the order of 30 mm will be over predicted by at most 4 mm/day, while large wet day amounts of rainfall (110 to 120 mm) will be under predicted by at most 36 mm/day. The $H(y_{\text{observed}})$ value gives the maximum error at a specific location, $y_{\text{observed}}$, in the gamma distribution, however, the gradual and progressive deviations of $F_{\text{observed}}$ from $F_{\text{Model}}$ (Figs. 5.21 to 5.24) suggests that $H(y_{\text{observed}})$ is a reasonable indicator of the error order of magnitude in a region of the distribution centered about $y_{\text{observed}}$.

To concisely summarize these errors, $H(y_{\text{observed}})$ vs $y_{\text{observed}}$ values were plotted in Figure 5.26, where each error value was labeled by month. The largest errors (-30 to -90 mm/day) occurred in the upper tail of the gamma distribution (> 120 mm) during the wettest time of the year (summer through early autumn). Moderate positive (5 to 20 mm/day) and negative (-10 to -30 mm/day) errors occurred in the middle of the gamma distribution (60 mm to 120 mm) primarily from late autumn through late spring. The smallest errors (1 to 5 mm/day), which were largely positive, occurred in the lower tail of the gamma distribution during all seasons of the year. Spatially, the errors were observed to be of a similar order of magnitude with no regular latitudinal gradients.

**Evaluation of interval parameter estimates.** To determine interval parameter estimates for the wet day amount model (equation 5.22), the standard deviations, $\sigma_{\alpha}(\tau)$ and $\sigma_{\beta}(\tau)$, were first estimated using equations 5.37 and 5.38. Then, using equations 5.39
Table 5.14  Maximum positive error, $H^+$ (mm/day), for monthly gamma distribution wet day amount model.

<table>
<thead>
<tr>
<th>City</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>GVL</td>
<td>3</td>
<td>4</td>
<td>19</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>8</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>ORL</td>
<td>6</td>
<td>3</td>
<td>2</td>
<td>13</td>
<td>17</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>TAM</td>
<td>2</td>
<td>3</td>
<td>11</td>
<td>5</td>
<td>7</td>
<td>4</td>
<td>2</td>
<td>6</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>FTM</td>
<td>2</td>
<td>18</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>14</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>6</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 5.15  Location of error $H^+$, $y_{\text{observed}}$ (mm), for monthly gamma distribution wet day amount model.

<table>
<thead>
<tr>
<th>City</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>GVL</td>
<td>26</td>
<td>50</td>
<td>80</td>
<td>32</td>
<td>21</td>
<td>17</td>
<td>10</td>
<td>31</td>
<td>11</td>
<td>36</td>
<td>61</td>
<td>32</td>
</tr>
<tr>
<td>ORL</td>
<td>42</td>
<td>19</td>
<td>10</td>
<td>77</td>
<td>80</td>
<td>6</td>
<td>74</td>
<td>10</td>
<td>29</td>
<td>12</td>
<td>7</td>
<td>14</td>
</tr>
<tr>
<td>TAM</td>
<td>17</td>
<td>14</td>
<td>84</td>
<td>30</td>
<td>100</td>
<td>57</td>
<td>12</td>
<td>90</td>
<td>12</td>
<td>16</td>
<td>26</td>
<td>11</td>
</tr>
<tr>
<td>FTM</td>
<td>10</td>
<td>66</td>
<td>16</td>
<td>9</td>
<td>9</td>
<td>38</td>
<td>98</td>
<td>30</td>
<td>21</td>
<td>38</td>
<td>53</td>
<td>11</td>
</tr>
</tbody>
</table>
Table 5.16  Maximum negative error, $H^-$ (mm/day), for monthly gamma distribution wet day amount model.

<table>
<thead>
<tr>
<th>City</th>
<th>Month</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ORL</td>
<td>2</td>
<td>-19</td>
<td>-15</td>
<td>-26</td>
<td>-4</td>
<td>-5</td>
<td>-34</td>
<td>-13</td>
<td>-25</td>
<td>-86</td>
<td>-84</td>
<td>-28</td>
</tr>
<tr>
<td>TAM</td>
<td>3</td>
<td>-18</td>
<td>-11</td>
<td>-4</td>
<td>-7</td>
<td>-10</td>
<td>-41</td>
<td>-30</td>
<td>-14</td>
<td>-30</td>
<td>-10</td>
<td>-23</td>
</tr>
<tr>
<td>FLM</td>
<td>4</td>
<td>-9</td>
<td>-5</td>
<td>-22</td>
<td>-26</td>
<td>-23</td>
<td>-33</td>
<td>-3</td>
<td>-64</td>
<td>-64</td>
<td>-41</td>
<td>-15</td>
</tr>
</tbody>
</table>

Table 5.17  Location of error $H^-$, $y_{\text{Observed}}$ (mm), for monthly gamma distribution wet day amount model.

<table>
<thead>
<tr>
<th>City</th>
<th>Month</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>GVL</td>
<td>1</td>
<td>101</td>
<td>112</td>
<td>23</td>
<td>116</td>
<td>87</td>
<td>94</td>
<td>111</td>
<td>144</td>
<td>131</td>
<td>188</td>
<td>82</td>
</tr>
<tr>
<td>ORL</td>
<td>2</td>
<td>87</td>
<td>75</td>
<td>103</td>
<td>45</td>
<td>52</td>
<td>150</td>
<td>103</td>
<td>122</td>
<td>196</td>
<td>183</td>
<td>91</td>
</tr>
<tr>
<td>TAM</td>
<td>3</td>
<td>80</td>
<td>99</td>
<td>49</td>
<td>57</td>
<td>77</td>
<td>166</td>
<td>118</td>
<td>115</td>
<td>146</td>
<td>56</td>
<td>77</td>
</tr>
<tr>
<td>FLM</td>
<td>4</td>
<td>50</td>
<td>40</td>
<td>119</td>
<td>96</td>
<td>98</td>
<td>164</td>
<td>60</td>
<td>173</td>
<td>198</td>
<td>156</td>
<td>81</td>
</tr>
</tbody>
</table>
Figure 5.26 $H(y_{\text{observed}})$ error for the gamma-distribution wet day amount model by month at Gainesville, Orlando, Tampa, and Ft. Myers, Florida.
Figure 5.26--continued
and 5.40, 99% confidence intervals \((\gamma_a = 0.99, z = 2.575; \gamma_b = 0.99, w = 2.575)\) were constructed (Tables 5.6 through 5.9). The values for \(\gamma_a\) and \(\gamma_b\) were selected to illustrate the method.

For a model with two parameters, such as the gamma distribution, a two dimensional parameter space exists, in which all possible parameter values can occur. Using the one dimensional confidence intervals for the individual parameters, a two dimensional confidence region in parameter space can be determined. The confidence region, \(\mathcal{R}\), is an ellipse with major and minor axes given by the confidence intervals for \(\beta\) and \(\alpha\) respectively (Chapter 5.2.1). The parameter pairs \((\alpha, \beta)\) have a joint normal distribution with a bell shaped density surface centered over the ellipse. Thus, the distribution of possible parameter pairs \((\alpha, \beta)\) is dense about the center of the ellipse and decreases toward the edges.

To evaluate the effect of parametric uncertainty on the gamma distribution amount model, the parametric confidence region, \(\mathcal{R}\), was mapped into a distribution region using equation 5.22. This mapping was intended to provide a graphical representation of the degree of model uncertainty resulting from parametric uncertainty. To map \(\mathcal{R}\) into a distribution region, two pairs of parameter coordinates were selected which bracket all points in \(\mathcal{R}\). The coordinates selected were the upper (UCL) and lower (LCL) confidence limits on \(\alpha\) and \(\beta\) (equations 5.34 and 5.35) which are the extreme corners of \(\mathcal{R}'\), \((\alpha + z \sigma_{\alpha}, \beta + w \sigma_{\beta})\) and \((\alpha - z \sigma_{\alpha}, \beta - w \sigma_{\beta})\). The corners of \(\mathcal{R}'\) were selected, as no two coordinate pairs on the ellipse bounding \(\mathcal{R}\) will bracket all points within \(\mathcal{R}\). Thus, the region of distribution uncertainty obtained from the mapping

\[
F_{\text{UCL}} = F_y(y; \hat{\alpha} + z \hat{\sigma}_{\alpha}, \hat{\beta} + w \hat{\sigma}_{\beta}) \quad 5.42
\]

\[
F_{\text{LCL}} = F_y(y; \hat{\alpha} - z \hat{\sigma}_{\alpha}, \hat{\beta} - w \hat{\sigma}_{\beta}) \quad 5.43
\]
will be an outer bound for the actual distribution region associated with the elliptical parameter region $\mathcal{R}$. To illustrate this method, data for the month of May at Gainesville were used to map the parametric confidence region (Fig. 5.27a) into the associated distribution region (Fig. 5.27b).

The 99% parametric confidence intervals given in Tables 5.6 to 5.9 were mapped into regions of distribution uncertainty using equations 5.42 and 5.43, for each month at the four locations considered (Figs. 5.21 to 5.24) with an illustrative example at expanded scale given for May at Gainesville (Fig. 5.20). For all months and locations considered, the region of uncertainty for the gamma distribution model contained by $F_{\text{LCL}}$ and $F_{\text{UCL}}$ was relatively broad, and bracketed both $F_{\text{Model}}$ and $F_{\text{Observed}}$ with $F_{\text{LCL}}$ forming the upper bound and $F_{\text{UCL}}$ the lower bound. In the lower and upper tails of the gamma distribution, the region was very narrow, while in the middle of the distribution it was broad. In the lower portion (< 50 mm) of the gamma distribution, $F_{\text{Observed}}$ was tightly bracketed between $F_{\text{LCL}}$ and $F_{\text{Model}}$, while in the upper portion (> 50 mm) of the distribution $F_{\text{Observed}}$ was tightly bracketed between $F_{\text{Model}}$ and $F_{\text{UCL}}$. This was illustrated by replotting Figure 5.20 in Figure 5.28 using an arithmetic amount axis rather than a logarithmic axis. Reducing the values of $\gamma_a$ and $\gamma_\beta$ would reduce the area of region $\mathcal{R}$, thus reducing the region $[F_{\text{UCL}}, F_{\text{LCL}}]$. This could result in $F_{\text{Observed}}$ protruding from $[F_{\text{UCL}}, F_{\text{LCL}}]$, however this is not unexpected, as there is always some probability that the parameters $\alpha$ and $\beta$ will occur outside of $\mathcal{R}$.

The mapping of parametric confidence limits described above is in effect a type of apriori sensitivity analysis. Rather than generating numerous rainfall time series using stochastic models with slightly different parameters, the limits of $F_{\text{Model}}$'s sensitivity to parametric uncertainty can be roughly assessed from the nature of the region bounded by $[F_{\text{LCL}}, F_{\text{UCL}}]$. If $F_{\text{Model}}$ was insensitive to small changes in parameters, then the region of distribution uncertainty, $[F_{\text{LCL}}, F_{\text{UCL}}]$, would be relatively small or barely distinguishable from $F_{\text{Model}}$. The graphical results presented above suggested that $F_{\text{Model}}$ was sensitive to
Figure 5.27 Mapping of parametric confidence region to distribution region of uncertainty for May at Gainesville, Florida.
a. parametric region; b. distribution region.
Figure 5.28 Comparison of rainfall amount cumulative distributions, \( F_{\text{Model}} \), \( F_{\text{Observed}} \), \( F_{\text{UCL}} \), and \( F_{\text{LCL}} \), for May at Gainesville, Florida.
small uncertainties in the model parameters $\alpha$ and $\beta$. The region of distribution uncertainty was relatively broad and generally contained $F_{\text{observed}}$ throughout the possible range of daily rainfall amounts for all locations.

A quantitative measure of fit, similar to those used for point parameter estimates, was not suggested for interval parameter estimates, or for the region of distribution uncertainty contained by $[F_{\text{LCL}}, F_{\text{UCL}}]$. Improvements in fit between distributions of observed and simulated rainfall amounts are best assessed after the incorporation of parametric uncertainty in a stochastic rainfall model. The wet day amount model is implemented (Chapter 6) by randomly selecting monthly parameters $[\alpha(t), \beta(t)]$ for $F_{\text{Model}}$ from a bivariate normal distribution. This results in a slightly different $F(t)_{\text{Model}}$ for each year of daily rainfall generated, none of which will perfectly "fit" $F_{\text{observed}}$. The basic hypothesis in this approach is that $F_{\text{Model}}$ will over predict and under predict daily rainfall amounts, in response to the induced parametric variability. It is anticipated that this approach will result in a distribution of seasonal and annual rainfall totals that range from extremely wet through average to extremely dry. After implementation of the stochastic rainfall model, the daily rainfall time series generated using $F_{\text{Model}}$ can be used (Chapter 7) to obtain monthly distributions of wet day amounts, $F_{\text{Simulated}}$. $F_{\text{Simulated}}$ is obtained by sorting the simulated data, just as the observed data were sorted to obtain $F_{\text{observed}}$. The goodness of fit techniques discussed above and in Appendix B can then be used to evaluate the effects of including parametric uncertainty in a stochastic rainfall model.

5.2.4 Summary and Conclusions

An annually nonhomogeneous parameter estimation methodology was developed for the wet day amount component of a stochastic rainfall model. Traditional point parameter estimates were combined with interval parameter estimates to include parametric uncertainty in the rainfall simulation models. The methodology was demonstrated using the gamma distribution at four locations in Florida. Due to its
wide-spread use in the WGEN model (Richardson and Wright, 1985), the gamma distribution was selected for illustrative purposes. The form of the gamma distribution used was slightly different from that used in WGEN with a variable transformation made, maximum likelihood parameter estimates used instead of moment estimates, and a rainfall threshold parameter included. The Kolmogorov-Smirnov goodness of fit test provided evidence suggesting that the observed wet day rainfall amounts were drawn from a gamma distribution during all seasons at all locations. However, detailed evaluation of the gamma distribution suggested that the Kolmogrov-Smirnov D statistic may not accurately indicate the errors associated with the application of a particular distribution to rainfall simulation. An alternative evaluation measure, H error, was proposed and applied to the gamma distribution. This evaluation method suggested that over predictions of 1 to 10 mm/day are possible for moderate wet day amounts, and under predictions of 30 mm/day or greater are possible for large wet day amounts generated in the upper tail of the gamma distribution. Research on wet day amount models (Chapter 2) has focused on obtaining the best possible fit to observed data through improved parameter estimation methods and distribution selection. The extreme sensitivity of wet day amount models to small deviations \( (F_{\text{Model}} - F_{\text{Observed}}) \) in the upper tails of distributions (demonstrated above) suggested the possibility that no one distribution function may exist which "fits", to an acceptable degree, during all seasons at all locations.

Rather than continue the search for the best possible distribution function or point parameter estimation technique, an alternative approach was considered in which the model parameters were assumed to be normally distributed random variables. Using large sample properties, estimates of the variance associated with parameter estimates were made and confidence intervals constructed. Using the parameter confidence intervals, regions of distribution uncertainty were constructed for the gamma distribution. The derived regions, \( \left[ F_{\text{LCL}}, F_{\text{UCL}} \right] \), were shown to be relatively broad, suggesting that the wet day amount model is sensitive to small variations in parameters. To implement the
proposed method, the gamma distribution is perturbed on a monthly basis by allowing the parameters $\alpha$ and $\beta$ to vary as normally distributed random variables (Chapter 6).

Implementation of a daily stochastic rainfall model is required to fully evaluate (Chapter 7) the effect of incorporating parametric uncertainty in the wet day amount model.
CHAPTER 6
PARAMETER SIMULATION METHODOLOGY

In Chapter 4, the long-term variability of rainfall was investigated. This investigation showed that considerable long-term variability existed, and that irregular long-term trends were present in the historical rainfall records considered. In Chapter 5, a parameter estimation methodology for stochastic rainfall models was developed. The implementation of this parameter estimation method resulted in parameter time series with significant long-term variability. The objective of this chapter was to develop parameter simulation methods which preserved the observed long-term variability, time dependence, and trends observed in the investigations presented in Chapters 4 and 5. In Chapter 7, the proposed parameter simulation methodology was coupled to a stochastic rainfall model with a daily time step, and the effects of parametric uncertainty on stochastically generated rainfall were evaluated in detail.

An annually nonhomogeneous parameter simulation methodology for stochastic rainfall models with a daily time step was developed. To illustrate this methodology, a first-order Markov chain was used for the rainfall occurrence model, and the gamma distribution was used for the wet day amount model. The parameter simulation methodology was applied to parameter estimates made using the methodology developed in Chapter 5, at Gainesville, Orlando, Tampa, and Ft. Myers, Florida.

6.1 Occurrence Model Parameter Simulation

6.1.1 Parameter Simulation Method

In Chapter 5, an annually nonhomogeneous parameter estimation method was proposed and implemented for a first-order Markov-chain rainfall occurrence model. This parameter estimation method resulted in multivariate or matrix time series of
Markov-chain transition probabilities (equation 5.8) denoted \( \{ P_{vS + \tau} \} = \{ [p_{ij}(vS + \tau)] \} \) where \((i, j = D, W)\), \(v\) is the annual time index, \(S\) is the number of seasons, and \(\tau\) is the seasonal time index. Thus, four transition probability time series resulted, where \(\{ \hat{p}_{DD}(vS + \tau) \}\) and \(\{ \hat{p}_{WD}(vS + \tau) \}\) were estimated directly from rainfall data, while \(\{ \hat{p}_{DW}(vS + \tau) \}\) and \(\{ \hat{p}_{WW}(vS + \tau) \}\) were calculated using the properties \(\hat{p}_{DW} = 1 - \hat{p}_{DD}\) and \(\hat{p}_{WW} = 1 - \hat{p}_{WD}.\)

In Chapter 4, considerable differences in the degree of seasonal variability in rainfall amounts (Figs. 4.4 to 4.7) and the number of wet days (Figs. 4.24 to 4.27) were observed. Long-term trends and associated spectral properties in rainfall amounts (Figs. 4.13 to 4.20) and wet day counts (Figs. 4.32 to 4.39) were also observed to vary considerably from season to season. The results of the trend analyses (Chapter 4) interpreted with respect to the climatological literature reviewed (Chapter 4.2.5 and 4.3.5) suggested that different geophysical factors may influence the long-term nature of rainfall at different seasons of the year. For example, long-term trends during the winter wet season showed dominant periods of 11.5 years associated with solar radiation cycles, while trends during the summer wet season showed dominant periodicities of approximately 2 years associated with fluctuations of winds and temperatures in the lower stratosphere over the Equator. These findings suggested that long-term seasonal time dependence or correlation in rainfall processes was significant and should be considered in the formulation of parameter time series models.

---

1. During a particular year, it is possible to observe a season during which only dry days occur. In this case, the two state Markov-chain model is not appropriate. This possibility is accounted for by setting \( \hat{p}_{DD} = 1 \) and \( \hat{p}_{DW} = \hat{p}_{WD} = \hat{p}_{WW} = 0. \)
To account for the different degrees of variability and types of long-term trends observed at different seasons of the year, S seasonal parameter time series of the form

\[ \{P^t_v\} = \{P^1_v, P^2_v, \ldots, P^N_v\} \]

were considered rather than the single continuous time series, \( \{P_{vs+\tau}\} \). For each season, \( \tau \), the time series \( \{P^t_v\} \) is a series of 2x2 square matrices given by

\[
P^t_v = \begin{bmatrix}
p^t_{DD}(v) & p^t_{DW}(v) \\
p^t_{WD}(v) & p^t_{WW}(v)
\end{bmatrix}
\]

where \( v \) is the annual time index. Noting that the elements of the second column of \( P^t_v \) are functions of the elements of the first column, it is possible to reformulate the 2x2 matrix time series of equation 6.2 in terms of the 2x1 vector time series given by

\[
P^t_v = \begin{bmatrix}
p^\tau_{DD}(v) \\
p^\tau_{WD}(v)
\end{bmatrix}
\]

where \( p^\tau_{DD}(v) \) and \( p^\tau_{WD}(v) \) are the parameter time series estimated in Chapter 5, sorted by season.
Based on the findings reported in Chapter 4, a model for the parameter time series (equation 6.3) estimated in Chapter 5 was developed. The objectives of this model development were to:

i. retain the statistical properties of the estimated parameter time series;

ii. retain the time dependence or auto-correlation structure within each individual estimated parameter time series;

iii. retain the time dependence or cross-correlation structure between each estimated parameter time series;

iv. retain the observed long-term trends of the rainfall occurrence process.

Based on these objectives, a composite model was developed having the form

$$\mathbf{P}_v^r = \mathbf{P}_{v,\text{Mean}}^r + \mathbf{P}_{v,\text{Deterministic}}^r + \mathbf{P}_{v,\text{Stochastic}}^r$$  \hspace{1cm} (6.4)

where $\mathbf{P}_{v,\text{Mean}}^r$ is the seasonal mean which is constant over time. The deterministic or physically based component, $\mathbf{P}_{v,\text{Deterministic}}^r$, was included to describe any long-term trends contained in the estimated parameter time series. For example, this deterministic component would contain the long-term effects of variations in solar radiation on the parameter time series in particular, or the rainfall occurrence process in general. The remaining component, $\mathbf{P}_{v,\text{Stochastic}}^r$ is the stochastic component, which contains any short-term auto-correlation or cross-correlation structure remaining after the mean and long-term trend were removed from the parameter series. A bivariate autoregressive time series model of order $p$, AR($p$), was developed for this stochastic component. The stochastic component has the form

$$\mathbf{P}_{v,\text{Stochastic}}^r = \mathbf{A}_1^r \mathbf{P}_{v-1}^r + \mathbf{A}_2^r \mathbf{P}_{v-2}^r + \cdots + \mathbf{A}_p^r \mathbf{P}_{v-p}^r + \mathbf{B}^r \mathbf{e}_v^r$$  \hspace{1cm} (6.5)
where $A_i$ and $B_i$ are 2x2 parameter matrices, and $\vec{e}_v$ is a 2x1 normally distributed white noise vector with zero mean, $E[\vec{e}_v, \vec{e}_v] = 0$, and unit variance $E[\vec{e}_v, \vec{e}_v] = I$. The details of implementing or fitting the proposed model (equation 6.4) to the estimated parameter time series of equation 6.3 are given below.

6.1.2 Implementation of Parameter Simulation Method

Each component of the proposed parameter simulation model (equation 6.4) was fit to the estimated parameter time series (equation 6.3) on a seasonal basis.

**Mean.** For each season $(\tau)$, $P^\tau_{\text{Mean}}$ is an annually time invariant (homogeneous) 2x1 vector given by

$$P^\tau_{\text{Mean}} = \begin{bmatrix} \bar{p}^\tau_{DD} \\ \bar{p}^\tau_{WD} \end{bmatrix} \quad (\tau = 1, 2, \cdots, S) \quad 6.6$$

The vector elements, $\bar{p}^\tau_{DD}$ and $\bar{p}^\tau_{WD}$, are estimated by calculating the overall mean of the estimated time series, $\{p^\tau_{DD}(v)\}$ and $\{p^\tau_{WD}(v)\}$ or

$$\bar{p}^\tau_{DD} = \frac{1}{N} \sum_{v=1}^{N} p^\tau_{DD}(v) \quad 6.7$$

$$\bar{p}^\tau_{WD} = \frac{1}{N} \sum_{v=1}^{N} p^\tau_{WD}(v)$$

where $N$ is the number of years in the sample.

**Deterministic.** Subtracting $P^\tau_{\text{Mean}}$ from $\{P^\tau_v\}$, point by point, results in zero mean time series vectors, $\{P^\tau_v - P^\tau_{\text{Mean}}\}$ which are used to estimate the remaining components of equation 6.4. At this point, the deterministic and stochastic components are separated by Fourier-domain low-pass filtering (Appendix D). Detrending or removal of the deterministic component of a time series before developing a stochastic model is a standard step in time series analysis (Box and Jenkins, 1976; Salas et al., 1980). The
daily rainfall occurrence process is a complex process, influenced by many physical factors of varying time and spatial scales. Separating a time series into deterministic and stochastic components is not straight-forward, but can be accomplished by defining limits on the physical processes relevant to the application of the rainfall model. For example, long-period astronomical processes, such as sunspots, are studied and described on a physical basis. Defining these long-period processes as deterministic, and separating them by filtering would enable their effects on the rainfall process to be evaluated.

**Stochastic.** The stochastic component \((P_{\text{Stochastic}} - P_{\text{Mean}} - P_{\text{Deterministic}})\) of equation 6.4 is determined after the mean and deterministic components are separated. A bivariate autoregressive time series model (equation 6.6) is used for the stochastic component of the parameter simulation model. Fitting a time series model of this type is a complex process considered in detail in Appendix F.

6.1.3 Evaluation of Parameter Simulation Method

Evaluating a stochastic rainfall model coupled to the proposed parameter simulation model is the primary means of method evaluation. The effects of parametric uncertainty on stochastically generated rainfall time series are considered in detail in Chapter 7. However, a preliminary evaluation of the proposed parameter simulation model can be made, based on regional climatological conditions and the results of historical data analyses discussed in Chapter 4.

Florida experiences wet seasons during the winter and summer, and relatively dry seasons during spring and autumn. The meteorological conditions (Dohrenwend, 1978) producing the winter and summer wet seasons are fundamentally different, with the passage of continental fronts prevailing during winter, and convective thundershowers dominating during the summer. During winter, rainfall events are frequently greater than one day in duration, and dry events normally last several days. During summer, rainfall events are normally more frequent and shorter in duration, while dry events are also shorter. Spring and autumn are transition seasons with an irregular pattern of
meteorological factors affecting the rainfall process. These transition seasons are typically temperate and dry.

The number of wet days per season is a criteria frequently used for evaluating rainfall occurrence models (Chapter 2). Additional evaluation criteria, such as the distribution of wet interval lengths, are considered in Chapter 7, however seasonal wet day counts can provide useful information on the rainfall occurrence process, and are of interest in water resource management problems. In Chapter 4, descriptive statistics, distributions and long-term trend analyses were presented for monthly and seasonal wet day counts. As discussed in Chapter 4, annual cycles in wet day counts were in close agreement with the annual meteorological cycle (wet and dry seasonality) discussed above.

Based on these meteorological cycles and historical observations, inferences can be made on the expected nature of occurrence model parameters estimated from historical data. During the winter wet seasons, the probabilities of dry day to dry day transitions \( p_{DD} \) are expected to be higher, while the probabilities of wet day to wet day transitions \( p_{ww} \) are expected to be lower than during the summer wet seasons. Winter rains occur over several days interspersed between longer dry intervals, while summer thundershowers occur frequently with infrequent prolonged dry spells. During the spring and autumn dry seasons, the probabilities of dry day to dry day transitions are expected to be high based on the dry nature of the seasons and low wet day counts (Figs. 4.2 and 4.22). The probabilities of wet day to wet day transitions are expected to increase into summer and decrease into winter in response to the summer thunderstorm season.

Spatially, a north to south increase in \( p_{DD} \) and decrease in \( p_{ww} \) during winter are expected, based on the latitudinal decrease in observed wet day counts and decreasing influence of continental fronts toward the tropics. During summer, a north to south decrease in \( p_{DD} \) and increase in \( p_{ww} \) are expected based on the increased incidence of thunderstorms form north to south (Butson and Pyne, 1968; Davis and Sakamoto, 1976).
Temporally, long periods (years) with below average wet day counts are expected to be associated with lower $p_{ww}$ and higher $p_{DD}$ values.

This proposed evaluation method is suggested only to provide a preliminary evaluation of the proposed parameter estimation and simulation methodologies. Discrepancies between wet day count analyses and parameter analyses do not necessarily indicate inadequacy in the proposed methodologies, but may indicate weaknesses in the evaluation method. As in any statistical analysis, the sample size will affect the results. For example during a dry season, a given low wet day count could be associated with a single wet event of several days (high $p_{ww}$) or several isolated wet days (zero $p_{ww}$) while the $p_{DD}$ parameter would be high in either case. Also, a long-term climatological shift, such as a latitudinal increase in the influence of frontally induced rainfall with an associated displacement of tropically influenced rainfall, could increase the $p_{ww}$ parameter without significantly altering the wet day count.

6.1.4 Results of Parameter Simulation Method

The proposed parameter simulation model (equation 6.4) was fit to the estimated Markov-chain transition probability time series, $\{P^v\}$, (Figs. 5.1 to 5.8) using the methods proposed above. Each Markov-chain parameter time series, $\{P^v\}$, was estimated from daily rainfall data (Chapter 3) from 1900 through 1988, ($v = 0, 1, 2, \cdots, 88$) for Gainesville, Orlando, Tampa, and Ft. Myers, Florida. Monthly seasons ($\tau = 1, 2, \cdots, 12$) were selected following the approach used by Richardson and Wright (1985).

Mean. Using equation 6.7, the monthly means, $\bar{p}^{\tau}_{DD}$ and $\bar{p}^{\tau}_{WD}$, were estimated using the 89 year samples, $\{p^{\tau}_{DD}(v)\}$ and $\{p^{\tau}_{WD}(v)\}$ ($v = 0, 1, 2, \cdots, 88$). The estimated means, plus and minus one standard deviation, are given in Figures 6.1 and 6.2 for the four locations considered.

Both $\bar{p}^{\tau}_{DD}$ and $\bar{p}^{\tau}_{WD}$ exhibited pronounced annual cycles very similar to the annual cycles in monthly wet day counts discussed in Chapter 4 and shown in Figure 4.22. Temporally, $\bar{p}^{\tau}_{DD}$ varied in a regular manner (Fig. 6.1) throughout the year, with primary
Figure 6.1 Average monthly Markov-chain transition probability, $\bar{P}_{DD}^\tau$ ($\tau = 1, 2, \cdots, 12$), with plus and minus one standard deviation for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 to 1988).
Figure 6.2  Average monthly Markov-chain transition probability, $P_{wd}$ ($\tau = 1, 2, \ldots, 12$), with plus and minus one standard deviation for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 to 1988).
minima observed during the July-summer wet season, and secondary minima observed during the February-winter wet season. Large values of $\bar{p}_{dd}$ occurred during both the April-spring and November-autumn dry seasons. The standard deviations associated with $\bar{p}_{dd}$ were larger during the summer than during the rest of the year. This suggests a high degree of dry run length variability during summer, and is reflected in the relatively large standard deviations of summer wet day counts (Fig. 4.22). Spatially, there was a definite north to south gradient (Fig. 6.3) between $\bar{p}_{dd}$ values for Gainesville (north) and Ft. Myers (south), however, the intermediate stations of Orlando and Tampa did not exhibit the same latitudinal gradients. From autumn through spring, the $\bar{p}_{dd}$ values increased from north to south, suggesting fewer rain events and longer dry spells from north to south. This is in accord with the expected decreasing influence of continental fronts toward the tropics. During summer, the $\bar{p}_{dd}$ values decreased from north to south, suggesting fewer dry runs and more frequent rain events. The summer $\bar{p}_{dd}$ gradient agrees with the observed increase in summer thunderstorm frequency toward the topics (Davis and Sakamoto, 1976).

The $\bar{p}_{wd} \left(1 - \bar{p}_{ww}^*\right)^2$ parameter also exhibited an annual cycle (Fig. 6.2) with pronounced summer wet season minima and higher values throughout the rest of the year. Pronounced seasonal maxima occurring during a specific month were not present as in the $\bar{p}_{dd}$ cycle, and the well defined secondary minima and maxima of the $\bar{p}_{dd}$ parameter were also absent. The standard deviations associated with $\bar{p}_{wd}$ were smaller during the summer than during the rest of the year, and generally increased from north to south. The large autumn through spring standard deviations and lack of pronounced $\bar{p}_{wd}^*$ maxima ($\bar{p}_{ww}^*$ minima) suggested a high degree of variability in the duration of rainstorms during this part of the year, compared with relatively regular afternoon thunderstorms of summer. Spatially, regular north to south gradients in $\bar{p}_{wd}^*$, similar to

2. Large values of $\bar{p}_{ww}^*$ are equivalent to small values of $\bar{p}_{wd}^*$. 

Figure 6.3  Comparison of average monthly Markov-chain transition probabilities, $\overline{p}_{DD}$ and $\overline{p}_{WD}$ ($\tau = 1, 2, \cdots, 12$), for Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988).
the $\bar{p}_{DD}$ gradients, were only observed between Gainesville and Ft. Myers. The $\bar{p}_{WW}$

$(1 - \bar{p}_{WD})$ values decreased from Gainesville to Ft. Myers during the winter wet season, and increased during the summer wet season. The annual cycle in these gradients agrees with the annual cycle in meteorological conditions discussed above.

In general, the observed annual cycles of $\bar{p}_{DD}$ and $\bar{p}_{WD}$ agreed with the annual cycle in meteorological conditions, wet day counts, and standard deviations discussed in Chapter 4. The annual cycles of $\bar{p}_{DD}$ and $\bar{p}_{WD}$ are time invariant (homogeneous), and are repeated every year in the proposed parameter simulation method. The deterministic and stochastic components of the method will perturb this homogeneous component resulting in a composite time series of parameters that is nonhomogeneous or time variant from year to year.

**Deterministic.** To calculate the deterministic component, $\{P_{\nu,\text{Deterministic}}\}$, of equation 6.4, the monthly means, $P_{\nu,\text{Mean}}$, were first subtracted from the estimated parameter time series, $\{P_{\nu}\}$, point by point for the 89 years of record at the four locations considered. Each zero mean monthly parameter series, $\{P_{\nu} - P_{\nu,\text{Mean}}\}$, was then low-pass filtered (Appendix D) using a cutoff frequency of $f_c = 1/10$ years. After filtering, the deterministic component, $\{P_{\nu,\text{Deterministic}}\}$, contained only those components of the original estimates series, $\{P_{\nu}\}$, with periods of 10 years or greater. The 10 year cutoff frequency was selected to retain solar cycles with periodicities of approximately 11 to 12 years (Chapter 4).

A preliminary evaluation of the deterministic component of equation 6.4 was made using the wet day count comparison method discussed above. Monthly time series consisting of integer counts of the number of wet days per month (equation 4.4) were determined. Each monthly wet day count time series was then low-pass filtered in the same manner as the monthly parameter time series using a cutoff frequency of $f_c = 1/10$ years. To investigate the possible presence of long-term trends in the estimated
parameter time series, and evaluate the fit of the deterministic component, \( \{\mathbf{P}_{v,\text{Deterministic}}^r\} \), the filtered time series of parameters and wet day counts were plotted on a monthly basis. For comparison purposes, a representative month from each season (Chapter 4.2.5) was selected for the four locations considered. Comparison plots of the filtered series \( \{p_{DD}(v)\}_D \) and \( \{p_{ww}(v)\}_D \), along with the filtered monthly wet day counts are given in Figures 6.4 to 6.7. Agreement between the parameter series and the observed wet day count series is considered to be preliminary evidence supporting the hypothesis that long-term trends in the observed rainfall occurrence process are retained by the proposed parameter estimation and simulation methodologies.

Very similar long-term trends were found in both the parameter series and wet day count series at all seasons and locations. These trends were not linear, but cyclic with each cycle having a slightly different period. Pronounced periods in the 10 to 12-year range were present, suggesting the importance of solar cycles to the rainfall process. As discussed in Chapter 4, the period and amplitude of solar cycles vary from cycle to cycle and this is reflected in the irregularities of the parameter and count series. In addition to irregularities in the solar cycle, a complex combination of long-period global to local scale geophysical factors are expected to combine and influence rainfall processes at a particular point. Long-period global-scale factors such as lunar cycles, atmospheric tides, and perturbed planetary waves, combined with regional and local scale atmosphere-ocean-land interactions\(^3\) are expected to contribute to irregularities observed in the parameter and count trends.

The monthly \( \{p_{DD}(v)\}_D \) series were consistently opposite in phase to the wet day count series, while the monthly \( \{p_{ww}(v)\}_D \) series were generally in phase. In general, minima in the \( \{p_{DD}(v)\}_D \) series coincided with maxima in the count series, and maxima in

---

\(^3\) The four Florida locations considered are located on a peninsula where oceanographic and meteorological regimes range from continental to tropical.
Figure 6.4 Comparison of $\{p_{DD}(v)\}_{M+D}$, $\{p_{WW}(v)\}_{M+D}$, and monthly wet day count
Figure 6.5 Comparison of $\{p_{DD}^*(v)\}_{M+D}$, $\{p_{WW}^*(v)\}_{M+D}$, and monthly wet day count time series for April at Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988).
Figure 6.6  Comparison of \( \{p_{DD}(v)\}_{M+D} \), \( \{p_{WW}(v)\}_{M+D} \), and monthly wet day count
time series for July at Gainesville, Orlando, Tampa, and Ft. Myers, Florida
(1900 through 1988).
Figure 6.7 Comparison of \( \{p_{DD}(v)\}^\ast_{M+D} \), \( \{p_{WW}(v)\}^\ast_{M+D} \), and monthly wet day count time series for November at Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988).
the \( \{ p^*_{w w}(v) \}_D \) series coincided with maxima in the count series. The observed phasing of
the parameter and count trends is reasonable, because a decrease in the number of wet
days should result in an increase in the probability of dry day to dry day transitions, while
the probability of wet day to wet day transitions should decrease.

Minor discrepancies in phase between the \( \{ p^*_{w w}(v) \}_D \) series and the count series
were observed on rare occasions. These discrepancies were more pronounced during dry
seasons, and increased from north to south. As discussed above, discrepancies of this
type are not unexpected, and are related to the small sample size of wet day counts, and
mixture of rainfall processes occurring during dry seasons. Consider for example, the
April season at Ft. Myers (Fig. 6.5 and 6.8), where discrepancies were observed. From
1963 to 1967 there was a short-period ( < 10 years) oscillation in wet day counts with a
maximum at 1966. This short-period oscillation occurred during a prolonged downward
trend (1950 to 1970) in wet day counts. In the \( \{ p^*_{w w}(v) \}_D \) series, there was a similar
short-period oscillation at this time, which peaked in 1965 rather than 1966. These
short-period oscillations in the \( \{ p^*_{w w}(v) \}_D \) and count series were both removed by the
low-pass filter, however phase shifts in the unfiltered series can affect the filtered series.
In April of 1965, an unusual 5 day rainfall event occurred, while in April of 1966, 4
separate 2 day events occurred. Wrigth (1977) reported a major reversal in Equatorial
winds during the 1965 to 1966 period which may account for the unusual rainfall event of
1965. The possibility of unusual short period, large amplitude fluctuations such as these
are accounted for in the stochastic component of the parameter model.

The high degree of fit observed between the filtered \( \{ p^*_{d d}(v) \}_D \) and \( \{ p^*_{w w}(v) \}_D \)
series and the filtered wet day count series suggested that the deterministic component,
\( \{ P^*_{v, Deterministic} \} \), retains long-term trends present in the historical rainfall series.
Comparison of daily historical rainfall data with daily simulated rainfall, generated from
Figure 6.8 Comparison of observed and low-pass filtered \( \{p_{ww}(v)\} \) series, with observed and low-pass filtered monthly wet day count time series for April at Ft. Myers, Florida (1900 through 1988).
a stochastic rain model coupled with the proposed parameter simulation model, will constitute a more rigorous test of this hypothesis (Chapter 7).

Based on the results of long-term trend analyses presented in Chapter 4 and above, multiple seasonal analyses are recommended over a single annual analysis. These trend analyses suggested that different physical processes dominated during different seasons, resulting in different types of seasonal trends. However, after the seasonal analyses are completed, a single annual time series can be formed by combining the seasonal series. Annual \( \{P_{DD}(vS + \tau)\}_{M+D} \) and \( \{P_{WD}(vS + \tau)\}_{M+D} \) series composed of the mean, \( \{P^r_{\text{Mean}}\} \), and deterministic, \( \{P^r_{\text{Deterministic}}\} \), components are given in Figures 6.9 to 6.12 for the four locations considered. In addition, a 30-year expanded view of these series for Gainesville (1900 to 1929) is given in Figure 6.13. These figures show the effect of long-term trends, \( \{P^r_{\text{Deterministic}}\} \), on the time invariant (homogeneous) mean parameter cycle, \( \{P^r_{\text{Mean}}\} \), given in Figures 6.1 and 6.2.

**Stochastic.** After calculating the mean, \( P^r_{\text{Mean}} \), and deterministic, \( \{P^r_{\text{Deterministic}}\} \), components of equation 6.4, the remainder of the parameter time series, \( \{P^r - P^r_{\text{Mean}} - P^r_{\text{Deterministic}}\} \), was considered to be the stochastic component. The stochastic component, \( \{P^r_{\text{Stochastic}}\} \), contains periodicities or time correlations of less than 10 years remaining after subtracting the monthly means and low-pass filtering. The remainder time series has the form

\[
\{p_{DD}^\tau(v)\}_s = \{p_{DD}^\tau(v)\} - \overline{p_{DD}} - \overline{p_{DD}^\tau(v)}_D
\]

\[
\{p_{WD}^\tau(v)\}_s = \{p_{WD}^\tau(v)\} - \overline{p_{WD}} - \overline{p_{WD}^\tau(v)}_D
\]

were \( \tau \) is the season index (\( \tau = 1, 2, \ldots, 12 \)) and \( v \) is the year index (\( v = 0, 1, 2, \ldots, 88 \)). The two time series of equation 6.8 were fit to the bivariate autoregressive time series model of equation 6.5, using the method describe above and in Appendix F.
Figure 6.9  Time series of Markov-chain transition probabilities, \( \{ p_{DD}(vS + \tau) \}_{M+D} \) and \( \{ p_{WD}(vS + \tau) \}_{M+D} \), for Gainesville, Florida from 1900 through 1988.
Figure 6.10  Time series of Markov-chain transition probabilities, \( \{p_{DD}(vS + \tau)\}_{M+D} \) and \( \{p_{wD}(vS + \tau)\}_{M+D} \), for Orlando, Florida from 1900 through 1988.
Figure 6.11 Time series of Markov-chain transition probabilities, \( \{ p_{DD}(vS + \tau) \}_{M+D} \) and \( \{ p_{WD}(vS + \tau) \}_{M+D} \), for Tampa, Florida from 1900 through 1988.
Figure 6.12 Time series of Markov-chain transition probabilities, \( \{p_{DD}(vS + \tau)\}_{M+D} \) and \( \{p_{WD}(vS + \tau)\}_{M+D} \), for Ft. Myers, Florida from 1900 through 1988.
Figure 6.13 Time series of Markov-chain transition probabilities, \( \{p_{DD}(vS + \tau)\}_{M+D} \) and \( \{p_{WD}(vS + \tau)\}_{M+D} \), for Gainesville, Florida from 1900 through 1929.
The bivariate autoregressive model of equation 6.5 is based on the assumption that the time series used in model fitting (equation 6.8) are normally distributed. If these time series are not normally distributed, a suitable mathematical transformation is required to obtain normal distributions. The Kolmogorov-Smirnov goodness of fit test (Appendix B) was used to test the hypothesis that the time series of equation 6.8 were drawn from normally distributed populations. The Kolmogorov-Smirnov test statistic, D, for the \( p_{DD}(v) \) and \( p_{WD}(v) \) time series are given in Tables 6.1 and 6.2. The observed D values were small for all monthly seasons and locations considered offering no evidence to reject the hypothesis of normality. Based on these tests, the \( p_{DD}(v) \) and \( p_{WD}(v) \) time series were considered to be normally distributed, with no data transformation required.

The bivariate AR(p) model of equation 6.5 was fit to the time series of equation 6.8, using the methods given in Appendix F. Based on these methods, the model parameters and model orders of each seasonal model were determined. The orders of each seasonal model are given in Table 6.3. The model orders are indicative of the time dependence (years) contained in the \( p_{DD}(v) \) and \( p_{WD}(v) \) time series. Model orders ranged from 2 to 6 years with orders of 3 years occurring frequently. This time lag or correlation of 3 years may be related to oscillations of winds and temperatures in the lower stratosphere at the Equator as discussed in Chapter 4. The individual model parameters are not given, as they are extensive, and provide little insight into the rainfall process or the goodness of fit of the proposed AR(p) model.

The goodness of fit of the stochastic component, \( P^{*}_{v, Stochastic} \), of the parameter simulation model was assessed by analyzing the residuals of the AR(p) models. If the residuals of the AR(p) are stationary white noise, i.e., independent and normally distributed, then the model is considered to fit the data (Salas et al., 1985). The hypothesis that the model residuals are normally distributed was tested using the
Table 6.1  Kolmogorov-Smirnov test statistics, D, for testing the hypothesis that the \( \{ \rho_{DD}(v) \} \) time series are normally distributed for Gainesville, Orlando, Tampa, and Ft. Myers, Florida.

<table>
<thead>
<tr>
<th>Month</th>
<th>Gainesville</th>
<th>Orlando</th>
<th>Tampa</th>
<th>Ft. Myers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.067</td>
<td>0.082</td>
<td>0.052</td>
<td>0.065</td>
</tr>
<tr>
<td>2</td>
<td>0.090</td>
<td>0.061</td>
<td>0.079</td>
<td>0.083</td>
</tr>
<tr>
<td>3</td>
<td>0.081</td>
<td>0.113</td>
<td>0.078</td>
<td>0.064</td>
</tr>
<tr>
<td>4</td>
<td>0.080</td>
<td>0.087</td>
<td>0.064</td>
<td>0.102</td>
</tr>
<tr>
<td>5</td>
<td>0.093</td>
<td>0.100</td>
<td>0.090</td>
<td>0.114</td>
</tr>
<tr>
<td>6</td>
<td>0.058</td>
<td>0.051</td>
<td>0.090</td>
<td>0.061</td>
</tr>
<tr>
<td>7</td>
<td>0.075</td>
<td>0.082</td>
<td>0.071</td>
<td>0.061</td>
</tr>
<tr>
<td>8</td>
<td>0.067</td>
<td>0.105</td>
<td>0.078</td>
<td>0.047</td>
</tr>
<tr>
<td>9</td>
<td>0.080</td>
<td>0.068</td>
<td>0.071</td>
<td>0.124</td>
</tr>
<tr>
<td>10</td>
<td>0.095</td>
<td>0.096</td>
<td>0.086</td>
<td>0.120</td>
</tr>
<tr>
<td>11</td>
<td>0.065</td>
<td>0.066</td>
<td>0.084</td>
<td>0.102</td>
</tr>
<tr>
<td>12</td>
<td>0.105</td>
<td>0.106</td>
<td>0.112</td>
<td>0.094</td>
</tr>
</tbody>
</table>

NOTE: The critical values of the Kolmogorov-Smirnov test statistic (Haan, 1977) for a specified level of significance using a sample size of \( n = 89 \) are: \( D = 0.129 \) for \( \alpha = 0.10 \); \( D = 0.144 \) for \( \alpha = 0.05 \); and \( D = 0.173 \) for \( \alpha = 0.01 \).
Table 6.2  Kolmogorov-Smirnov test statistics, $D$, for testing the hypothesis that the $\{\beta_{wp}(v)\}$ time series are normally distributed for Gainesville, Orlando, Tampa, and Ft. Myers, Florida.

<table>
<thead>
<tr>
<th>Month</th>
<th>Gainesville</th>
<th>Orlando</th>
<th>Tampa</th>
<th>Ft. Myers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.062</td>
<td>0.066</td>
<td>0.067</td>
<td>0.087</td>
</tr>
<tr>
<td>2</td>
<td>0.071</td>
<td>0.062</td>
<td>0.081</td>
<td>0.085</td>
</tr>
<tr>
<td>3</td>
<td>0.080</td>
<td>0.060</td>
<td>0.078</td>
<td>0.098</td>
</tr>
<tr>
<td>4</td>
<td>0.065</td>
<td>0.068</td>
<td>0.061</td>
<td>0.059</td>
</tr>
<tr>
<td>5</td>
<td>0.054</td>
<td>0.048</td>
<td>0.094</td>
<td>0.076</td>
</tr>
<tr>
<td>6</td>
<td>0.059</td>
<td>0.081</td>
<td>0.055</td>
<td>0.056</td>
</tr>
<tr>
<td>7</td>
<td>0.069</td>
<td>0.090</td>
<td>0.113</td>
<td>0.116</td>
</tr>
<tr>
<td>8</td>
<td>0.062</td>
<td>0.072</td>
<td>0.087</td>
<td>0.083</td>
</tr>
<tr>
<td>9</td>
<td>0.069</td>
<td>0.053</td>
<td>0.120</td>
<td>0.059</td>
</tr>
<tr>
<td>10</td>
<td>0.068</td>
<td>0.075</td>
<td>0.097</td>
<td>0.074</td>
</tr>
<tr>
<td>11</td>
<td>0.065</td>
<td>0.074</td>
<td>0.077</td>
<td>0.085</td>
</tr>
<tr>
<td>12</td>
<td>0.062</td>
<td>0.061</td>
<td>0.066</td>
<td>0.104</td>
</tr>
</tbody>
</table>

NOTE: The critical values of the Kolmogorov-Smirnov test statistic (Haan, 1977) for a specified level of significance using a sample size of $n = 89$ are: $D = 0.129$ for $\alpha = 0.10$; $D = 0.144$ for $\alpha = 0.05$; and $D = 0.173$ for $\alpha = 0.01$. 
Table 6.3  Multivariate AR(p) model orders for the \( \{P_{v,Stochastic}\} \) time series.

<table>
<thead>
<tr>
<th>Month</th>
<th>Gainesville</th>
<th>Orlando</th>
<th>Tampa</th>
<th>Ft. Myers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>12</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>
Kolmogorov-Smirnov fit test (Appendix B). The Kolmogorov-Smirnov test statistics, D, for the \( \{p^{\uparrow}_{BD}(v)\}_s \) and \( \{p^{\uparrow}_{WD}(v)\}_s \) series are given in Tables 6.4 and 6.5 respectively. The D values were small for all months and locations offering no evidence to reject the hypothesis of normality of the residuals. The hypothesis that the model residuals are independent was tested using the Porte Manteau test. The correlation of the residuals of both the \( \{p^{\uparrow}_{BD}(v)\}_s \) and \( \{p^{\uparrow}_{WD}(v)\}_s \) were calculated for lags up to 30 years (L=30). The Porte Manteau test statistic Q (Appendix F) was calculated and given in Tables 6.6 and 6.8 along with the associated degrees of freedom in Tables 6.7 and 6.9. In all seasons, the Q values were low and did not offer evidence to reject the hypothesis of independence of residuals. Based on the results of these test, the residuals were shown to be independent and normally distributed, indicating a good fit between the bivariate AR(p) model and the \( \{p^{\uparrow}_{BD}(v)\}_s \) and \( \{p^{\uparrow}_{WD}(v)\}_s \) series.

In Chapter 5.1.3, the north to south increase in \( p_{DD} = 1 \) (\( p_{WW} = 0 \)) parameter values was discussed and summarized in Figures 5.10 to 5.13. It was suggested that this latitudinal increase may be related to a southerly decrease in the frequency of rainfall associated with the passage of continental fronts, especially during winter. The possibility of using separate rainfall occurrence models for rainfall associated with convective thunderstorms and passing fronts was proposed in this discussion. After removing the seasonal means and deterministic trends, the remaining parameter series were highly Gaussian (Tables 6.1 and 6.2). Separate occurrence models were not developed, based on the good fit obtained from the bivariate autoregressive model of the \( \{p^{\uparrow}_{BD}(v)\}_s \) and \( \{p^{\uparrow}_{WD}(v)\}_s \) series discussed above.

6.1.5 Summary and Conclusions

A parameter simulation methodology for the occurrence component of a stochastic rainfall model was developed. The proposed model was a composite model consisting of a time invariant (homogeneous) component, and a time variant (nonhomogeneous) component. The nonhomogeneous component contained a deterministic segment to
Table 6.4  Kolmogorov-Smirnov test statistics, $D$, for testing the hypothesis that the residuals of the $\{p_{\bar{D}}(v)\}_s$ time series are normally distributed for Gainesville, Orlando, Tampa, Ft. Myers, Florida.

<table>
<thead>
<tr>
<th>Month</th>
<th>Gainesville</th>
<th>Orlando</th>
<th>Tampa</th>
<th>Ft. Myers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.070</td>
<td>0.063</td>
<td>0.048</td>
<td>0.070</td>
</tr>
<tr>
<td>2</td>
<td>0.056</td>
<td>0.076</td>
<td>0.074</td>
<td>0.073</td>
</tr>
<tr>
<td>3</td>
<td>0.072</td>
<td>0.107</td>
<td>0.072</td>
<td>0.059</td>
</tr>
<tr>
<td>4</td>
<td>0.074</td>
<td>0.054</td>
<td>0.088</td>
<td>0.089</td>
</tr>
<tr>
<td>5</td>
<td>0.088</td>
<td>0.051</td>
<td>0.097</td>
<td>0.081</td>
</tr>
<tr>
<td>6</td>
<td>0.082</td>
<td>0.044</td>
<td>0.093</td>
<td>0.050</td>
</tr>
<tr>
<td>7</td>
<td>0.052</td>
<td>0.101</td>
<td>0.058</td>
<td>0.044</td>
</tr>
<tr>
<td>8</td>
<td>0.064</td>
<td>0.076</td>
<td>0.072</td>
<td>0.057</td>
</tr>
<tr>
<td>9</td>
<td>0.057</td>
<td>0.048</td>
<td>0.064</td>
<td>0.063</td>
</tr>
<tr>
<td>10</td>
<td>0.079</td>
<td>0.069</td>
<td>0.080</td>
<td>0.116</td>
</tr>
<tr>
<td>11</td>
<td>0.055</td>
<td>0.103</td>
<td>0.092</td>
<td>0.103</td>
</tr>
<tr>
<td>12</td>
<td>0.050</td>
<td>0.071</td>
<td>0.089</td>
<td>0.049</td>
</tr>
</tbody>
</table>

NOTE: The critical values of the Kolmogorov-Smirnov test statistic (Haan, 1977) for a specified level of significance using a sample size of $n = 89$ are: $D = 0.129$ for $\alpha = 0.10$; $D = 0.144$ for $\alpha = 0.05$; and $D = 0.173$ for $\alpha = 0.01$. 
Table 6.5  Kolmogorov-Smirnov test statistics, D, for testing the hypothesis that the residuals of the \( \{ \tilde{p}_{wD}(v) \} \) time series are normally distributed for Gainesville, Orlando, Tampa, Ft. Myers, Florida.

<table>
<thead>
<tr>
<th>Month</th>
<th>Gainesville</th>
<th>Orlando</th>
<th>Tampa</th>
<th>Ft. Myers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.066</td>
<td>0.057</td>
<td>0.063</td>
<td>0.099</td>
</tr>
<tr>
<td>2</td>
<td>0.064</td>
<td>0.074</td>
<td>0.068</td>
<td>0.090</td>
</tr>
<tr>
<td>3</td>
<td>0.037</td>
<td>0.061</td>
<td>0.057</td>
<td>0.091</td>
</tr>
<tr>
<td>4</td>
<td>0.053</td>
<td>0.069</td>
<td>0.052</td>
<td>0.058</td>
</tr>
<tr>
<td>5</td>
<td>0.062</td>
<td>0.045</td>
<td>0.060</td>
<td>0.055</td>
</tr>
<tr>
<td>6</td>
<td>0.092</td>
<td>0.094</td>
<td>0.049</td>
<td>0.082</td>
</tr>
<tr>
<td>7</td>
<td>0.076</td>
<td>0.071</td>
<td>0.052</td>
<td>0.109</td>
</tr>
<tr>
<td>8</td>
<td>0.054</td>
<td>0.055</td>
<td>0.049</td>
<td>0.087</td>
</tr>
<tr>
<td>9</td>
<td>0.063</td>
<td>0.086</td>
<td>0.057</td>
<td>0.083</td>
</tr>
<tr>
<td>10</td>
<td>0.065</td>
<td>0.040</td>
<td>0.082</td>
<td>0.085</td>
</tr>
<tr>
<td>11</td>
<td>0.097</td>
<td>0.085</td>
<td>0.058</td>
<td>0.087</td>
</tr>
<tr>
<td>12</td>
<td>0.061</td>
<td>0.082</td>
<td>0.051</td>
<td>0.051</td>
</tr>
</tbody>
</table>

NOTE: The critical values of the Kolmogorov-Smirnov test statistic (Haan, 1977) for a specified level of significance using a sample size of \( n = 89 \) are: \( D = 0.129 \) for \( \alpha = 0.10 \); \( D = 0.144 \) for \( \alpha = 0.05 \); and \( D = 0.173 \) for \( \alpha = 0.01 \).
Table 6.6 Chi-square test statistics, $Q$, for the Porte Manteau goodness of fit test of the monthly $\{p_{DD}(v)\}_s$ series residuals.

<table>
<thead>
<tr>
<th>City</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>GVL</td>
<td>17.8</td>
<td>13.9</td>
<td>20.9</td>
<td>16.3</td>
<td>22.2</td>
<td>15.4</td>
<td>38.5</td>
<td>20.5</td>
<td>27.1</td>
<td>16.5</td>
<td>21.3</td>
<td>15.1</td>
</tr>
<tr>
<td>ORL</td>
<td>31.4</td>
<td>11.2</td>
<td>21.2</td>
<td>27.1</td>
<td>13.7</td>
<td>19.2</td>
<td>13.4</td>
<td>32.6</td>
<td>20.1</td>
<td>16.2</td>
<td>17.8</td>
<td>23.5</td>
</tr>
<tr>
<td>TAM</td>
<td>20.3</td>
<td>13.1</td>
<td>14.9</td>
<td>18.1</td>
<td>17.2</td>
<td>17.1</td>
<td>17.9</td>
<td>18.0</td>
<td>13.6</td>
<td>19.5</td>
<td>27.8</td>
<td>26.8</td>
</tr>
<tr>
<td>FTM</td>
<td>22.1</td>
<td>14.2</td>
<td>18.6</td>
<td>22.6</td>
<td>17.2</td>
<td>17.5</td>
<td>35.1</td>
<td>18.7</td>
<td>21.7</td>
<td>17.0</td>
<td>17.1</td>
<td>32.2</td>
</tr>
</tbody>
</table>

Table 6.7 Chi-square degrees of freedom (L-p) for the Porte Manteau goodness of fit test of the monthly $\{p_{DD}(v)\}_s$ series residuals.

<table>
<thead>
<tr>
<th>City</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>GVL</td>
<td>28</td>
<td>26</td>
<td>27</td>
<td>27</td>
<td>28</td>
<td>28</td>
<td>27</td>
<td>28</td>
<td>28</td>
<td>26</td>
<td>27</td>
<td>26</td>
</tr>
<tr>
<td>ORL</td>
<td>27</td>
<td>26</td>
<td>28</td>
<td>27</td>
<td>25</td>
<td>28</td>
<td>26</td>
<td>27</td>
<td>27</td>
<td>27</td>
<td>28</td>
<td>27</td>
</tr>
<tr>
<td>TAM</td>
<td>27</td>
<td>25</td>
<td>27</td>
<td>26</td>
<td>28</td>
<td>25</td>
<td>26</td>
<td>28</td>
<td>26</td>
<td>27</td>
<td>27</td>
<td>27</td>
</tr>
<tr>
<td>FTM</td>
<td>26</td>
<td>24</td>
<td>27</td>
<td>27</td>
<td>27</td>
<td>28</td>
<td>28</td>
<td>28</td>
<td>26</td>
<td>28</td>
<td>27</td>
<td>26</td>
</tr>
</tbody>
</table>

NOTE: The critical values of the Porte Manteau test statistic (Ott, 1984) for a specified level of significance with the number of degrees of freedom, df = 27, are: $Q = 36.7$ for $\alpha = 0.10$; $Q = 40.1$ for $\alpha = 0.05$; and $Q = 47.0$ for $\alpha = 0.01$. The critical values of $Q$ change slowly for degrees of freedom ranging from 25 to 30.
Table 6.8  Chi-square test statistics, $Q$, for the Porte Manteau goodness of fit test of the monthly $\{p_{\omega}(v)\}_{s}$ series residuals.

<table>
<thead>
<tr>
<th>City</th>
<th>Month</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>GVL</td>
<td></td>
<td>29.1</td>
<td>13.6</td>
<td>30.8</td>
<td>16.6</td>
<td>16.8</td>
<td>18.1</td>
<td>18.9</td>
<td>16.1</td>
<td>28.8</td>
<td>18.5</td>
<td>23.9</td>
<td>15.3</td>
</tr>
<tr>
<td>ORL</td>
<td></td>
<td>18.5</td>
<td>21.1</td>
<td>27.5</td>
<td>23.1</td>
<td>16.3</td>
<td>20.8</td>
<td>12.1</td>
<td>10.1</td>
<td>22.8</td>
<td>18.6</td>
<td>12.3</td>
<td>26.5</td>
</tr>
<tr>
<td>TAM</td>
<td></td>
<td>20.6</td>
<td>13.4</td>
<td>17.0</td>
<td>11.8</td>
<td>20.5</td>
<td>17.1</td>
<td>19.1</td>
<td>18.5</td>
<td>30.2</td>
<td>14.2</td>
<td>16.8</td>
<td>17.6</td>
</tr>
<tr>
<td>FTM</td>
<td></td>
<td>18.0</td>
<td>15.7</td>
<td>17.0</td>
<td>20.5</td>
<td>14.3</td>
<td>18.6</td>
<td>28.3</td>
<td>24.9</td>
<td>19.8</td>
<td>19.3</td>
<td>21.7</td>
<td>19.2</td>
</tr>
</tbody>
</table>

Table 6.9  Chi-square degrees of freedom (L-p) for the Porte Manteau goodness of fit test of the monthly $\{p_{\omega}(v)\}_{s}$ series residuals.

<table>
<thead>
<tr>
<th>City</th>
<th>Month</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>GVL</td>
<td></td>
<td>28</td>
<td>26</td>
<td>27</td>
<td>27</td>
<td>28</td>
<td>28</td>
<td>27</td>
<td>28</td>
<td>28</td>
<td>26</td>
<td>27</td>
<td>26</td>
</tr>
<tr>
<td>ORL</td>
<td></td>
<td>27</td>
<td>26</td>
<td>28</td>
<td>27</td>
<td>25</td>
<td>28</td>
<td>26</td>
<td>27</td>
<td>27</td>
<td>28</td>
<td>27</td>
<td>27</td>
</tr>
<tr>
<td>TAM</td>
<td></td>
<td>27</td>
<td>25</td>
<td>27</td>
<td>26</td>
<td>28</td>
<td>25</td>
<td>26</td>
<td>28</td>
<td>26</td>
<td>27</td>
<td>27</td>
<td>27</td>
</tr>
<tr>
<td>FTM</td>
<td></td>
<td>26</td>
<td>24</td>
<td>27</td>
<td>27</td>
<td>28</td>
<td>28</td>
<td>28</td>
<td>28</td>
<td>26</td>
<td>28</td>
<td>27</td>
<td>26</td>
</tr>
</tbody>
</table>

NOTE: The critical values of the Porte Manteau test statistic (Ott, 1984) for a specified level of significance with the number of degrees of freedom, df = 27, are: $Q = 36.7$ for $\alpha = 0.10$; $Q = 40.1$ for $\alpha = 0.05$; and $Q = 47.0$ for $\alpha = 0.01$. The critical values of $Q$ change slowly for degrees of freedom ranging from 25 to 30.
account for long-term trends, and a stochastic segment to account for short-term variability. The proposed parameter simulation model was fit to 89 year time series of Markov-chain transition probabilities, estimated using the methodology of Chapter 5. A preliminary evaluation of the model was made by comparison with the results of historical data analyses made in Chapter 4. This preliminary evaluation demonstrated a good fit between the parameter simulation model, and the observed variability and long-term trends of the historical data. A more rigorous evaluation of the proposed methodology was made in Chapter 7, where a stochastic rainfall model coupled to the proposed parameter simulation model was evaluated in detail.

6.2 Amount Model Parameter Simulation

6.2.1 Parameter Simulation Method

In Chapter 5, a parameter estimation method was proposed and implemented for the gamma-distribution rainfall amount model (equation 5.21). This estimation method used maximum likelihood point parameter estimates of $\alpha(\tau)$ and $\beta(\tau)$, with parametric uncertainty quantified by confidence intervals. The confidence intervals were constructed using the standard deviations, $\sigma_\alpha(\tau)$ and $\sigma_\beta(\tau)$, derived from large sample properties.

The formulation and implementation (fitting) of a parameter simulation model for the rainfall amount process was fundamentally different from the approach used in the rainfall occurrence process. Simple estimates of rainfall occurrence probabilities (Markov-chain model parameters) were made for each month of each year of record, resulting in parameter time series that were analyzed and modeled. Single monthly samples of this type were considered inappropriate for wet day amount model parameter estimation. For any given month in a particular year, the number of wet days can vary from a minimum of 0 to a maximum of 31, with 5 to 10 wet days typical. A statistical sample of this size is inadequate for parameter estimation of a rainfall amount probability distribution model. This sample size limitation hindered the use of time series filtering
and modeling techniques applied to the rainfall amount model parameters, and required
the development of an alternative parameter simulation approach.

The analyses of daily rainfall data conducted in Chapter 4 suggested considerable
seasonal differences in the degree of variability (Figs. 4.4 to 4.7) and nature of long-term
trends (Figs. 4.13 to 4.20) in rainfall amounts. These observations of the rainfall amount
process, similar to those of the rainfall occurrence process, suggested that seasonal
variability and long-term trends were significant, and should be included in the parameter
simulation model. To account for seasonal differences, the historical rainfall data were
sorted into S seasons prior to model development.

Based on these analyses of historical rainfall records, a parameter time series
model for the gamma distribution was developed to retain the statistical properties and
trends of the rainfall amount process. A composite model, similar to equation 6.4, was
developed having the form

\[
\begin{align*}
\alpha(\tau, v) &= \alpha(\tau, v)_{\text{Deterministic}} + \alpha(\tau, v)_{\text{Stochastic}} \\
\beta(\tau, v) &= \beta(\tau, v)_{\text{Deterministic}} + \beta(\tau, v)_{\text{Stochastic}}
\end{align*}
\]

where \( \tau \) is the seasonal time index, and \( v \) is the annual time index.

The deterministic component was included to account for long-term trends in
seasonal rainfall amounts. A compromise approach was used to investigate the presence
of long-term trends in the gamma-distribution parameters, \( \alpha(\tau) \) and \( \beta(\tau) \), balancing the
need for an adequate sample size, with the need to estimate a trend time series. A
moving data-window sampling technique combined with the maximum likelihood
parameter estimation technique (Appendix A) was developed for the deterministic
component. This technique was used directly on the seasonal subsamples of observed
wet day amounts of rainfall. For each season, \( \tau, W \) years of data are sampled and
maximum likelihood parameter estimates are made. The data window is then advanced
one year and the maximum likelihood parameter estimation repeated. This approach results in a time series of parameters, \( \alpha(\tau, \nu)_D \) and \( \beta(\tau, \nu)_D \), which can be used for trend identification and analysis.

The stochastic component was used to include short-term variability in the simulated parameter time series. A simple zero-order moving-average model, MA(0), was used for the stochastic component. This component has the form

\[
\begin{align*}
\alpha(\tau, \nu)_{\text{Stochastic}} &= \sigma_a(\tau)\epsilon_v \\
\beta(\tau, \nu)_{\text{Stochastic}} &= \sigma_b(\tau)\epsilon_v
\end{align*}
\]

where \( \epsilon_v \) is a normally distributed random variable with zero mean and unit variance, and \( \sigma_a(\tau) \) and \( \sigma_b(\tau) \) are the seasonal standard deviations associated with the seasonal maximum likelihood parameter estimates \( \alpha(\tau) \) and \( \beta(\tau) \). The details of implementing or fitting the proposed parameter model are given below.

6.2.2 Implementation of Parameter Simulation Method

Each component of the proposed parameter simulation model (equation 6.9) is fit directly to the daily historical rainfall data, on a seasonal basis.

**Deterministic.** The deterministic components, \( \alpha(\tau, \nu)_D \) and \( \beta(\tau, \nu)_D \), are estimated by implementing the moving data-window sampling technique described above. The data window is centered at \( \nu = (W - 1)/2 + i \) \( i = 1, 2, \ldots, n - (W - 1)/2 \) where \( W \) is an odd integer, and \( n \) is the number of years of observed data. This results in a time series of parameter estimates given by

\[
\left\{ \begin{array}{c}
\alpha\left(\tau, \frac{(W - 1)}{2} + 1\right), \alpha\left(\tau, \frac{(W - 1)}{2} + 2\right), \ldots, \alpha\left(\tau, n - \frac{(W - 1)}{2}\right) \\
\beta\left(\tau, \frac{(W - 1)}{2} + 1\right), \beta\left(\tau, \frac{(W - 1)}{2} + 2\right), \ldots, \beta\left(\tau, n - \frac{(W - 1)}{2}\right)
\end{array} \right\}
\]
where \((W - 1)/2\) time points are lost from the beginning and end of the time series as a result of the sampling technique.

Selecting the data-window width, \(W\), for this method is analogous to selecting the cutoff frequency used in conjunction with the low-pass filtering technique (Chapter 6.2.1) used to identify long-term trends in the occurrence model parameter time series. The data-window width, \(W\), is selected based on the time scale or period of the physical processes included in the deterministic component of the parameter model. However, selecting \(W\) is not as straightforward as selecting the cutoff frequency, and the techniques required to select and evaluate \(W\) are discussed in the following.

**Stochastic.** To implement the stochastic component (equation 6.10), it is only necessary to estimate the standard deviations, \(\sigma_\alpha(\tau)\) and \(\sigma_\beta(\tau)\), from the historical data. In Chapter 5.2, large sample properties of maximum likelihood parameter estimates were used to derive estimates of the standard deviations (equations 5.36 and 5.37) associated with the gamma-distribution parameters \(\alpha(\tau)\) and \(\beta(\tau)\). These estimates of \(\sigma_\alpha(\tau)\) and \(\sigma_\beta(\tau)\) are used to implement the stochastic component (equation 6.10) of the parameter model.

### 6.2.3 Evaluation of Parameter Simulation Method

Similar to the rainfall occurrence model (Chapter 6.1.3), evaluating a stochastic rainfall amount model coupled to the proposed parameter simulation method is the primary method of evaluation, and is considered in detail in Chapter 7. However, a preliminary evaluation of the proposed parameter simulation model can be made, based on the statistical properties of the gamma distribution and the results of historical data analyses discussed in Chapter 4.

In this investigation, the gamma probability density function, \(f_Y(y;\alpha,\beta)\), was selected to model the wet day amount of rainfall, \(Y\). If the random variable \(Y\) is distributed according to the gamma density function, then the mean of this distribution
can be expressed as the product \( \alpha \cdot \beta \). For a given sample size, \( n \), if the sample mean \( \frac{1}{n} \sum_{i=1}^{n} y_i \) increases (decreases), then the sample total, \( \sum_{i=1}^{n} y_i \), will increase (decrease) in a proportionate manner.

Based on these properties, it is suggested that an observed increase (decrease) in the total amount of rainfall should result in an increase (decrease) in the mean amount of rainfall. If the gamma density is a reasonably good model for the wet day amount of rainfall (shown in Chapter 5.2), then the theoretical mean \( \alpha \cdot \beta \) should increase (decrease) as the observed rainfall total increases (decreases).

These basic properties of the gamma distribution combined with observed rainfall totals (Chapter 4.2) provide the basis for evaluating the deterministic component of the rainfall amount parameter model. Implementing the moving data-window method results in time series estimates of the parameters \( \alpha(\tau, v) \) and \( \beta(\tau, v) \) (equation 6.11). Multiplying these time series, term by term, results in the time series \( \alpha \beta(\tau, v) \), which should indicate long-term changes in the theoretical mean of the gamma density model. In Chapter 4.2, time series of observed seasonal rainfall totals were calculated and low-pass filtered to identify long-term trends in rainfall amounts. Comparison of the \( \alpha \beta(\tau, v) \) time series and the filtered seasonal rainfall totals is the basis of the proposed evaluation method. If these two time series are similar, then the observed long-term trends found in the filtered time series of monthly total amounts of rainfall are preserved by the data-window estimation method for \( \alpha(\tau, v) \) and \( \beta(\tau, v) \).

To illustrate this evaluation method, \( \alpha(\tau, v) \), \( \beta(\tau, v) \), and \( \alpha \beta(\tau, v) \) are plotted in Figure 6.14 for February at Gainesville, with \( W = 7 \) years. The \( \alpha \beta(\tau, v) \) time series is

---

4. If the random variable \( Y \) is distributed according to the gamma distribution, then the mean or expected value of \( Y \) is

\[
E[Y] = \int_{0}^{\infty} y f_y dy = \alpha \beta
\]

The details of this derivation are given in most texts on mathematical statistics (Mendenhall et al., 1981).
Figure 6.14 Comparison of $\alpha(\tau, v)_D$, $\beta(\tau, v)_D$, $\alpha\beta(\tau, v)_D$ and low-pass filtered monthly total amount time series for February at Gainesville, Florida (1900 through 1988).
compared with the low-pass \( f_c = 1/10 \) years) filtered series of monthly total rainfall amounts. The long-term trends observed in the low-pass filtered monthly total amount time series are clearly present in the \( \alpha \beta(\tau, v) \) time series.

6.2.4 Results of Parameter Simulation Method

The proposed parameter simulation model (equation 6.9) was fit directly to historical wet day amount data (Chapter 3) using the methods proposed above. Daily rainfall data from 1900 through 1988, \( (v = 0, 1, 2, \cdots, 88) \), for Gainesville, Orlando, Tampa, and Ft. Myers, Florida were used. Monthly seasons \( (\tau = 1, 2, \cdots, 12) \) were selected following the approach used by Richardson and Wright (1985).

Deterministic. To calculate the deterministic components, \( \alpha(\tau, v) D \) and \( \beta(\tau, v) D \), of equation 6.9, the moving data-window sampling method described above was used. Using this method to estimate a time series of amount model parameters required the selection of the sampling-window width, \( W \). Similar to the cutoff frequency used in low-pass filtering (Chapter 6.1.2), the window width was selected to retain solar cycles of approximately 11 to 12 years.

To identify long-term trends with periodicities of 10 years or greater, \( W \) values less than 10 years are required. Considering only odd integer values, it was anticipated that a \( W \) value of 9 years would be appropriate. To evaluate the choice of window-width, \( W \) values of 5, 7, 9, and 11 years were used to estimate \( \alpha(\tau, v) D \), \( \beta(\tau, v) D \). After multiplication term by term, \( \alpha \beta(\tau, v) D \) were compared to the low-pass filtered series of monthly total amounts of rainfall. An illustrative example for February at Gainesville is given in Figure 6.15. Based on these tests, a \( W \) value of 7 rather than 9 years was selected. The major long-period features of the filtered monthly total amount series were observed in all \( \alpha \beta(\tau, v) D \) series considered, however the \( W=11 \) years series were considered too smooth (lacked resolution), while the \( W=5 \) years were considered too rough (too much resolution). The \( W=7 \) year series was selected over the \( W=9 \) year series, due to slightly better agreement with the filtered amount series. A final evaluation
Figure 6.15  Comparison of $\alpha\beta(\tau, v)_d$ for $W = 5, 7, 9,$ and 11 years with low-pass filtered monthly total amount time series for February at Gainesville, Florida (1900 through 1988).
of sampling-window width will be made after implementing a stochastic rainfall model using the proposed parameter simulation model.

After selecting the sampling-window width, a preliminary evaluation of the deterministic component of equation 6.9 was made using the monthly total amount of rainfall comparison method discussed above. Rather than using annual or seasonal rainfall totals (Chapter 4), monthly total rainfall time series were calculated. Each monthly total amount time series was then low-pass filtered using a cutoff frequency of 1/10 years. To investigate the possible presence of long-term trends in the amount model parameter time series (equation 6.11) and evaluate the agreement with observed trends, the estimated \( \alpha \beta(\tau, v)_D \) time series, using \( W = 7 \) years, was plotted with the filtered monthly total rainfall amounts (Figures 6.16 to 6.19).

Similar long-term trends were found in both the \( \alpha \beta(\tau, v)_D \) time series and the filtered monthly total amount time series. These trends were cyclic, rather than linear, with each cycle having a slightly different period. Pronounced periods in the 10 to 12 year range were present, and similar to the trends discussed in Chapter 6.1.4 for the rainfall occurrence model parameters. The \( \alpha \beta(\tau, v)_D \) series were generally in phase with the monthly total amount series at all seasons and locations. Minor discrepancies between the \( \alpha \beta(\tau, v)_D \) and total amount series were observed, particularly during the dry season (April) at the southern-most location of Ft. Myers. This may be related to small sample size and a mixture of rainfall generating processes discussed in Chapter 6.1.4.

The high degree of fit observed between the \( \alpha \beta(\tau, v)_D \) time series and the low-pass filtered monthly total amount time series suggested that the moving data-window method of parameter estimation retains long-term trends present in the historical rainfall data. In Chapter 7, a more rigorous evaluation of this parameter estimation method was made by comparing observed rainfall time series with generated rainfall time series using the proposed parameter simulation method.
Figure 6.16 Comparison of $\alpha\beta(\tau, v)_p$ and low-pass filtered monthly total amount time series for February at Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988).
Figure 6.17 Comparison of $\alpha \beta(\tau, \nu)\theta$ and low-pass filtered monthly total amount time series for April at Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988).
Figure 6.18  Comparison of \( \alpha \beta(\tau, v) \) and low-pass filtered monthly total amount time series for July at Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988).
Figure 6.19 Comparison of $\alpha\beta(\tau, v)_D$ and low-pass filtered monthly total amount time series for November at Gainesville, Orlando, Tampa, and Ft. Myers, Florida (1900 through 1988).
Similar to the rainfall occurrence parameter model (Chapter 6.1), seasonal time series rather than annual time series were used to investigate long-term trends in amount model parameters. After seasonal analyses are completed, a single annual time series can be formed by combining the seasonal series. Annual $\alpha(vS + \tau)_D$ and $\beta(vS + \tau)_D$ series are given in Figures 6.20 to 6.23. In addition, a 30-year expanded view of these series for Gainesville (1903 to 1932) is given in Figure 6.24.

**Stochastic.** To implement or fit the stochastic component of equation 6.9 to the historical data, it is only necessary to estimate the standard deviations $\sigma_a$ and $\sigma_b$ of equation 6.10. The methodology to estimate these standard deviations was considered in detail in Appendix E and the results of applying this method were discussed in detail in Chapter 5.2. In Chapter 6.1.4, a rigorous evaluation of the stochastic component of the occurrence parameter model was made. As discussed above, the small sample sizes associated with wet day amount data precluded the development of high-order time series models which can be rigorously evaluated. Therefore, a formal evaluation of the stochastic component considered here was made (Chapter 7) by comparing observed rainfall time series with generated rainfall time series using the parameter simulation method presented above.

### 6.2.5 Summary and Conclusions

A parameter simulation methodology for the wet day amount component of a stochastic rainfall model was developed. The proposed model contained a deterministic component to account for long-term trends, and a stochastic component to account for the short-term variability. The proposed parameter simulation model was fit to 89 years of historical rainfall data, for four Florida locations (Chapter 3). A preliminary evaluation of the deterministic component was made by comparison with observed long-term trends in monthly total amounts of rainfall. This preliminary evaluation suggested a high degree of fit between the deterministic component of the parameter model and the observed
Figure 6.20  Time series of gamma-distribution parameters, $\alpha(\nu S + \tau)_D$ and $\beta(\nu S + \tau)_D$, for Gainesville, Florida from 1903 through 1985.
Figure 6.21 Time series of gamma-distribution parameters, $\alpha(\nu S + \tau)_D$ and $\beta(\nu S + \tau)_D$, for Orlando, Florida from 1903 through 1985.
Figure 6.22  Time series of gamma-distribution parameters, $\alpha(v + \tau)_{\sigma}$ and $\beta(v + \tau)_{\sigma}$, for Tampa, Florida from 1903 through 1985.
Figure 6.23  Time series of gamma-distribution parameters, $\alpha(vS + \tau)_D$ and $\beta(vS + \tau)_D$, for Ft. Myers, Florida from 1903 through 1985.
Figure 6.24  Time series of gamma-distribution parameters, $\alpha(v S + \tau)_D$ and $\beta(v S + \tau)_D$, for Gainesville, Florida from 1903 through 1932.
historical trends. A rigorous evaluation of the proposed methodology was made in
Chapter 7, where a stochastic rainfall model coupled to the proposed parameter
simulation model was evaluated for the rainfall occurrence and amount processes.
CHAPTER 7
RESULTS AND DISCUSSION

The mathematical description or modeling of daily rainfall at a point has used the theory of stochastic processes extensively. In all research efforts and engineering applications reviewed (Chapter 2), homogeneous parameter estimation (time invariant) methods were used to fit models to rainfall observations. The analyses (Chapter 4) of long historical time series (89 years) of daily rainfall data suggested that considerable interannual variability and long-term trends have occurred in natural rainfall processes. The hypothesis that stochastic model parameters may be nonhomogeneous (time variant) was proposed to address the need for rainfall models which preserve the variability and trends found in historical data. In this chapter an experimental design and evaluation methodology were developed to test this hypothesis.

7.1 Model Implementation

The mathematical formulation of the proposed hypothesis involves the sequential implementation of several methods and simulation models resulting in the generation of long time series of daily rainfall. An experimental design was developed to evaluate this sequence of subexperiments individually and collectively. Preliminary experiments were conducted to evaluate the parameter estimation methods (Chapter 5) and parameter simulation models (Chapter 6). The results of these preliminary experiments indicated that observed (Chapter 4) interannual variability and long-term trends in rainfall were preserved by the proposed methods and models. In this section, a collective experimental design was implemented which coupled a stochastic rainfall model to the proposed parameter estimation and simulation methodologies. The stochastic rainfall model used
was a first-order Markov-chain occurrence model, and a gamma-distribution wet day amount model (MC/G).

In any scientific investigation, a carefully controlled experimental procedure is essential. In the proposed experimental procedure, there are several variables, in addition to parametric uncertainty, that will influence the outcome of the experiment, thus weakening the control. The most important variable, in addition to parametric uncertainty, is the adequacy of the MC/G rainfall model. The MC/G model was selected due to its widespread use in the WGEN weather generation model (Richardson and Wright, 1984). It may not be the best possible model for the Florida locations considered. Separating the inadequacies of the proposed parameter estimation methodologies and the MC/G model was not straightforward, and was the primary weakness of the experimental design used. In addition, there are several other variables which will affect the experiment, however they can be controlled more readily than selection of the rainfall model. They include:

1. Adequacy of random number generators;
2. Effect of seed numbers used in random number generators;
3. Effect of model initial conditions;
4. Adequacy of number of simulation runs;

The model implementation procedure and experimental design given below will attempt to address each of these problems.

7.1.1 Rainfall Generation

A first-order Markov-chain rainfall occurrence model and a gamma-distribution wet day amount model, coupled to the proposed parameter estimation and simulation methods were implemented for Gainesville, Orlando, Tampa, and Ft. Myers, Florida. The steps used in implementing the model were:

1. Nonhomogeneous parameter estimates were made from historical data (Chapter 3) using the methods of Chapter 5.
2. Parameter simulation models (Chapter 6) were fit to the parameter time series resulting from step 1. Each parameter simulation model was composed of a stochastic and deterministic component.

3. The seed numbers for all random number generators were initialized (reinitialized).

4. The Markov-chain model was initialized (reinitialized) by setting the first day to dry (wet and dry alternating).

5. One year of monthly parameters were simulated.

6. Using the simulated parameters, one year of daily rainfall was generated.

7. Each subsequent year of daily rainfall was generated by returning to step 5.

8. After generating a specified number of years of daily rainfall data, the initial simulation was stopped. Subsequent simulation runs were made by returning to step 3.

The nonhomogeneous parameter estimates of step 1 were made using 89 years of daily rainfall data (1900 to 1988). In step 2, stochastic and deterministic components were identified and modeled for both the occurrence and amount model parameters. A moving data-sampling window (Chapter 6.2.2), similar to a centered moving average, was used to estimate the deterministic component or long-term trend of the gamma-distribution parameters, $\alpha(v, \tau)_d$ and $\beta(v, \tau)_d$. The use of a 7-year sampling window resulted in the amount model parameter trends beginning in 1903 and ending in 1985. These observed trends, rather than trend scenarios, were used in all simulation runs for model evaluation. For this reason, the rainfall simulation model was used to generate 83 years of daily rainfall beginning in 1903 and ending in 1985. If a climate change scenario were implemented using an artificial trend for the deterministic component, then the rainfall simulation model could be run for any desired period of years.

After completing an 83 year simulation run of the model, the seed numbers for the random number generators were changed and the next simulation run started (step 3). In addition, the first day of each simulation run alternated between the wet and dry state.
(step 4). Multiple simulation runs were made in an attempt to eliminate the effect of seed numbers and initial conditions from the model evaluation analyses. Ten separate simulation runs were made at each of the four locations considered.

Multiple simulation runs of stochastic rainfall models were made by Haan (1976, 6 runs), however no justification for an adequate number of runs was given. An adequate number of runs might be considered to be a sample large enough to allow the relative frequency of some statistical descriptor to attain stability. Further, it is suggested that the specific statistical descriptor of interest will influence the number of runs required during rainfall model evaluation. A descriptor such as an extreme event (1 event per run) would require more runs than a descriptor such as an average of hundreds of events per run. A rough estimate for an adequate number of runs was made using standard techniques (Mendenhall et al., 1981) from statistical estimation theory and sample size selection.¹

A primary objective of this experiment was to determine if the proposed model preserved the distribution of events found in the historical data. A histogram was the technique used to evaluate this objective, and the statistical descriptors of interest were the integer count and relative frequency of events contained in a particular bin of the

---

¹ Consider the problem of selecting a sample size \( n \) for an experiment where it is desired to estimate the mean, \( \mu \), plus or minus some bound, \( B \), with a probability of \( 1 - \alpha \). In repeated sampling, the mean will lie within \( \pm z_{\alpha/2} \) approximately 95% of the time. Assume that \( z_{\alpha/2} \sigma = B \) and approximate the population standard deviation, \( \sigma \), by the sample standard deviation \( \sigma_s \sqrt{n} \). From the above, a sample size for this experiment can be estimated using

\[
z_{\alpha/2} \frac{\sigma_s}{\sqrt{n}} = B
\]

Since \( \sigma_s \) is unknown, make an estimate of the possible range of \( \mu \) based on experience or previous sampling. The range, \( R \), is approximately four times \( \sigma_s \). Based on the above, an estimate of the sample size is obtained from

\[
n = \left( \frac{z_{\alpha/2}(R/4)}{B} \right)^2
\]
histogram. Consider as an illustrative example, the distribution of wet day amounts of rainfall received during a particular month. To obtain a rough estimate for an adequate number of runs, the following bound and range for an individual bin were specified. A very stringent bound of plus or minus 1 day, with a confidence level of 0.95 was placed on the average number of days per bin across all runs used. To allow for the possibility of considerable variability between runs, the possible range of bin counts was set at 2 to 8 days. These restrictions are most pertinent to the tail of the histogram where a very limited number of wet days with large amounts of rainfall are binned. Where the number of wet days counted in a bin are small, the relative frequency (bin count/total count) will be strongly affected by deviations of several days (2 to 8 days). Using the sample size selection method described above, \( (1 - \alpha = 0.95, z_{0.02} = 1.96, B = 1, \text{ and } R = 6) \), an adequate number of runs \( (n) \) was found to be approximately 8.6. Thus the 10 runs used should be more than adequate for evaluating the distribution of wet days with small amounts (< 25.4 mm) of rainfall and marginally adequate for evaluating the distribution of wet days with large amounts (> 25.4 mm). If 10 runs are adequate for evaluating small sample size variables such as extreme wet day amounts, they are also adequate for evaluating large sample variables such as monthly totals and long-term trends.

7.1.2 Random Number Generation

To implement the parameter and rainfall simulation models it was necessary to generate sequences of random numbers from the uniform, normal and gamma distributions. Uniform deviates are used in the occurrence model and to generate random numbers from other distributions, normal deviates are used in the parameter simulation model, and gamma deviates are used to generate the wet day amount of rainfall. The development and evaluation of random number generators used has received considerable attention by Knuth (1981). These random number generators were shown by Knuth (1981) to be sequentially uncorrelated, with a period that was effectively infinite in relation to the number of deviates generated during an individual run. For this reason, the
generators used were not evaluated in detail, however a histogram was used to verify the subroutines given by Press et al. (1986) based on Knuth's (1981) work.

**Uniform deviates.** Uniformly distributed random numbers (R) between 0 and 1 were generated using a linear congruential generator of the form

\[ I_{j+1} = aI_j + c \mod m \]  

where \( \{I_j\} \) is a sequence of integers, and \( a, c, \) and \( m \) are constants. The generator is initiated or "seeded" by setting \( I_0 \) to any negative integer value. This sequence of integers is converted to real numbers (between 0 and 1) by dividing \( I_{j+1} \) by \( m \). Dahlquist et al. (1974), Forsythe and Malcolm (1977) and Knuth (1981) have studied and evaluated random number generators of this type in considerable detail. Knuth (1981) proposed and evaluated a combination of three linear congruential generators, which were found to be uncorrelated and of infinite period (uniform deviates were not repeated). Knuth's method, as implemented by Press et al. (1986), was used to generate uniform deviates in the daily rainfall simulation model used in this research.

In addition to Knuth's rigorous evaluation of this method, a histogram of all uniform deviates (sample size = 4,204) generated during a single run of the rainfall model was constructed. This histogram was compared with the uniform probability density function (Fig. 7.1) and demonstrated that this method was generating uniformly distributed random numbers.

**Normal deviates.** Normally distributed random numbers (\( \xi \)) with mean zero and unit variance were generated using the Box-Muller method and the uniform random number generator discussed above. The derivation and evaluation of this method are given in Ahrens and Dieter (1972) and Box and Muller (1958). The method was implemented following Press et al. (1986). In addition, a histogram of normal deviates (sample size = 3,336) generated during a single run of the rainfall model was constructed.
Figure 7.1 Comparison of generated uniform deviates, R, with uniform probability density function.
This histogram was compared with a normal probability density function with zero mean and unit variance (Fig. 7.2) and demonstrated that this method was generating normally distributed random numbers.

**Gamma deviates.** Random numbers with a gamma distribution ($\Gamma$) were generated using a method developed and evaluated by Tadikamalla (1978 a,b). In application, the gamma random number generator is used to generate wet day amounts of rainfall, with nonhomogeneous (time variant) parameters which change slightly during each year of a simulation run. To evaluate this generator, homogeneous (time invariant) parameters estimated for August at Ft. Myers (Table 5.9) ($\alpha = 0.792$, $\beta = 16.478$) were used to generate 2000 gamma deviates. A histogram of the generated deviates was then compared to the gamma probability density function (Fig. 7.3). This comparison demonstrated that the method was generating random numbers with a gamma distribution.

### 7.2 Evaluation of Occurrence Model

The evaluation of stochastic rainfall models has emphasized accumulated rainfall amounts for annual and seasonal periods. Accumulated or total amounts of rainfall are influenced by both the amount of rainfall received per day (amount model) and the number of wet days per accumulation period (occurrence model). Combined evaluation methods of this type were considered in detail in Chapter 7.3. Independent evaluations of occurrence models have used seasonal wet day counts, and dry or wet run lengths as evaluation criteria.

To compare observed and generated time series of wet and dry days, the average number of wet days per season has been widely used. In general, good agreement was found between the observed and generated average number of wet days per season for all types of occurrence models studied (Chapter 2). This is not surprising, as seasonal wet day counts are used to estimate occurrence model parameters. To a lesser extent, the standard deviations of wet day counts have been used for model evaluation. Kline and
Figure 7.2  Comparison of generated normal deviates, $\xi_t$, with normal probability density function ($\mu = 0, \sigma = 1$).
Figure 7.3  Comparison of generated normal deviates, $\Gamma$, with gamma probability density ($\alpha = 0.792, \beta = 16.478$).
McFarland (1988) found good agreement between the observed and generated average number of wet days, but poor agreement between the standard deviations. The mean and standard deviation of a sample give little insight into the distribution of the sample. For this reason, it is suggested that the distribution of wet day counts is a useful criteria that can be used to evaluate occurrence models.

In addition to wet day counts, run lengths of wet and dry days have been used for model evaluation. Integer counts of run length are usually presented in tabular form, however Jones et al. (1972) used a histogram to provide a clear graphical summary of the distribution of run lengths. Empirical cumulative distributions and the chi-square goodness of fit tests were used to compare the distribution of observed and generated run lengths by Buishand (1978). As Haan (1977) suggested, this approach is insensitive to extreme events in the tails of the distributions, and was not used in this work as an evaluation method. Further inadequacies of this approach were discussed in Chapters 5.2.3 and in Chapter 7.3. Roldan and Woolhiser (1982) used a combination of likelihood functions and the Akaike Information Criteria to compare the ability of competing occurrence models to reproduce the distribution of run lengths. However, Roldan and Woolhiser (1982) found this approach to be sensitive to sample size and the number of model parameters, and was not used in this work.

To evaluate the proposed model, several methods were considered which focused on the ability of the occurrence model to preserve the observed distributions and long-term trends in wet day counts and dry run lengths.

7.2.1 Distribution of Number of Wet Days

**Method.** To evaluate the ability of the proposed occurrence model to preserve the long-term variability of observed rainfall events, the distribution of the number of wet days per month was considered. In Chapter 4.3.4, histograms of the observed number of wet days per month were constructed and described mathematically using the discrete binomial probability mass function. The observed histograms were generally bell
shaped, however significant deviations from the bell shape were also observed. The binomial mass function did not fit the irregularly shaped observed histograms, thus it was rejected as an inadequate model. A pronounced annual cycle in the shape and location of the monthly observed histogram was observed. From winter into summer, the observed bell-shaped histograms broadened (standard deviation increased) and shifted to the right (mean increased), while from summer into winter, the cycle reversed.

The ability of the proposed model to preserve the bell shape and annual cycle in the monthly wet day count histograms was investigated. Discrete histograms of the number of wet days were constructed in the same manner as those in Chapter 4.3.4. Monthly histograms with 31 bins, corresponding to the maximum possible number of wet days, were constructed for the observed and generated rainfall time series and combined on the same graph. Monthly wet day counts from all simulation runs were combined into a single histogram.

**Results.** Histograms of monthly wet day counts were constructed for all months and locations considered using ten separate simulation runs. A representative month (February, April, July, November) from each season (winter wet, spring dry, summer wet, autumn dry) was selected for graphical presentation (Figs. 7.4 to 7.7). For the four locations considered the general shape and location of the observed histograms was preserved in the generated histograms. The distributions of moderate to large (upper tail of histogram) wet day counts were well preserved for all seasons, however there were deviations in the low to moderate (lower tail of histogram) counts of wet days, primarily during winter and spring.

From autumn through winter into spring, there was a definite over prediction of completely dry months (0 wet days), combined with a slight tendency to over predict months with low wet day counts (< 5 wet days). If the Markov-chain transition probability, $p_{DD}$, for any particular month is 1.0 and the last day of the previous month was dry, then the occurrence model will generate a completely dry month. The over
Figure 7.4 Comparison of histograms for observed and generated number of wet days for February at Gainesville, Orlando, Tampa, and Ft. Myers, Florida.
Figure 7.5  Comparison of histograms for observed and generated number of wet days for April at Gainesville, Orlando, Tampa, and Ft. Myers, Florida.
Figure 7.6  Comparison of histograms for observed and generated number of wet days for July at Gainesville, Orlando, Tampa, and Ft. Myers, Florida.
Figure 7.7  Comparison of histograms for observed and generated number of wet days for November at Gainesville, Orlando, Tampa, and Ft. Myers, Florida.
prediction of completely dry months suggested that the Markov-chain parameter simulation model may be over predicting $p_{DD}$ values of 1.0. Recall that this parameter simulation model (Chapter 6.1) is a combination of a deterministic (long-term trend) and stochastic component. Also, there is annual cycle in the $p_{DD}$ parameter with low values during summer and higher values from autumn into spring. It is suggested that on occasion, a high value from the stochastic component occurs in conjunction with a relative maxima in the deterministic (trend) component, resulting in an over prediction in the number of months with $p_{DD}$ values of 1.0.

The parameter simulation model is a complex bivariate autoregressive time series model, in which the $p_{DD}$ and $p_{WD}$ time series are mathematically coupled at each time step. To account for the slight over prediction of completely dry months, the parameter simulation model could be completely reformulated as a quatrivariate model. The correlation structure of this model would sense the presence of maxima in the deterministic trends and adjust the stochastic components to avoid an over abundance of months with $p_{DD}$ values of 1.0. This reformulation would increase the complexity of the parameter simulation model, and increase the amount of computer time\(^2\) required to fit and use the model. Reformulation of the parameter simulation model should be considered, however users of the proposed rainfall model should balance their objectives against the considerable increase in complexity and implementation time required to attain a relatively small improvement in the distribution of completely dry months.

7.2.2 Distribution of Dry Runs

Method. To further evaluate the ability of the proposed occurrence model to preserve the long-term variability of observed rainfall events, the monthly distribution of

\[2. \text{To fit the proposed parameter simulation model, it was necessary to repeatedly take the inverse of a matrix using LU (lower triangular/upper triangular) matrix decomposition (Press et al., 1986) which requires } N^3 \text{ operations. Reformulating the parameter simulation model from bivariate (N=2) to quatrivariate (N=4) increases the number of operations by } 4^3 - 2^3 = 64 - 8 = 56.\]
dry run lengths was investigated. The observed and generated time series of daily rainfall were sorted by month and the dry run lengths counted. The run length counting started on the first day of each month and ended on the last day. Dry runs continuing from one month into the next were not considered and thus lost in this method. Discrete histograms with 31 bins were constructed for the observed and generated data. The generated histograms combined counts from all simulation runs made.

**Results.** Discrete histograms of dry run lengths were constructed for all months and locations considered using ten simulation runs. Similar to the approach used for the distribution of wet day counts, the representative months of February, April, July and November were presented graphically (Figs. 7.8 to 7.11). Observed and generated histograms for the same month were combined in the same figure for visual comparison.

The observed histograms of dry runs all had an "exponential" type shape with a high incidence of short runs and a low incidence of long runs. The histograms from autumn through winter into spring were very similar with a significant number of dry runs greater than 5 days in length, while the summer histograms showed a much lower incidence of runs greater than 5 days. This may be related to the winter dominance of frontally generated rainfall and high incidence of convective thunder storms in summer. The general shape of the observed histograms was preserved in the generated histograms at all seasons and locations.

In general, the observed distribution of dry runs was preserved in the generated rainfall data. For dry runs of less than 5 days, there were isolated deviations, however for most months and locations there was no consistent over or under predictions in this portion of the histogram. The most important deviation between the observed and generated histograms occurred from autumn through spring in the 28 to 31 day bins in the extreme tails of the distributions. These deviations occurred in the bin representing the total number of days in the particular month considered (28, 29, 30, 31 days). The proposed model slightly over predicted the number of months that were completely dry.
Figure 7.8  Comparison of histograms for observed and generated dry run lengths for February at Gainesville, Orlando, Tampa, and Ft. Myers, Florida.
Figure 7.9  Comparison of histograms for observed and generated dry run lengths for April at Gainesville, Orlando, Tampa, and Ft. Myers, Florida.
Figure 7.10 Comparison of histograms for observed and generated dry run lengths for July at Gainesville, Orlando, Tampa, and Ft. Myers, Florida.
Figure 7.11 Comparison of histograms for observed and generated dry run lengths for November at Gainesville, Orlando, Tampa, and Ft. Myers, Florida.
This over prediction was discussed in Chapter 7.2.1 above, and was related to the Markov-chain parameter simulation model. If the maximum possible run length for each month were excluded, the distribution of long dry runs, greater than 10 days in length, were well preserved at all locations and seasons considered. Simple Markov-chain occurrence models have been criticized for their inability to preserve the distribution of long dry runs. These results suggested that, with the inclusion of parametric uncertainty, Markov-chain models can accurately reproduce the observed distribution of long dry spells.

7.2.3 Long-term Trend of Number of Wet Days

Method. The possible presence of long-term trends in seasonal wet day counts was investigated in Chapter 4.3.5. The presence of pronounced periodicities of 11 to 12 years were identified by spectral analyses of seasonal wet day count time series. Based on the results of the spectral analyses, the time series of seasonal wet day counts were low-pass filtered (Appendix D) using a cutoff frequency of 1/10 years (Figs. 4.32, 4.34, 4.36, 4.38). The period and amplitude of the low-passed filtered data varied from cycle to cycle, however periodicities of 11 to 12 years were pronounced during all seasons. The literature on long-term trends in rainfall (Chapter 4.2.5) suggested that 10 to 12 year periodicities in rainfall total amounts are related to sunspots with similar periods. In the literature reviewed, considerably less attention was given to trends in the rainfall occurrence process. The analyses given in Chapter 4.3.5 suggested that long-term trends were present in the wet day count time series, which were similar to those found in the seasonal amount time series. Based on these findings, the long-term trend used in this evaluation method was defined as the low-pass filtered data using a cutoff frequency of 1/10 years.

To evaluate the ability of the proposed occurrence model to preserve observed long-term trends, time series of monthly wet day counts were constructed using the observed data and each individual simulation run. Each monthly time series was
low-passed filtered to remove the long-term trend. The filtered time series of wet day counts from each simulation run were averaged to give a single generated time series for each month. The low-pass filtered time series for the observed and generated data were then compared graphically to determine if the proposed model preserved the observed long-term trends.

Results. Time series of monthly wet day counts were constructed for the observed rainfall data (1900 to 1988) and generated rainfall data (1903 to 1985) for all months at the four locations considered. Each of these time series was low-pass filtered using a cutoff frequency of 1/10 years. A representative month for the winter wet season (February), the spring dry season (April), the summer wet season (July), and the autumn dry season (November) were plotted in Figures 7.12 to 7.15.

During all seasons at the four locations considered, the amplitude and phase of the observed long-term trends were retained in the generated trends. The period and amplitude of the generated trends were not constant, but varied from cycle in a manner very similar to the observed trends. There were no major deviations in phase between the observed and generated trends, but there were slight deviations in amplitude. Differences in amplitude between the observed and generated trends were not constant, but varied from no difference to 1 or 2 wet days per month.

7.3 Evaluation of Amount Model

The primary goal of stochastic modeling is to develop models which preserve the statistical properties of observed time series. In addition, it is desirable to use statistical properties or descriptors which are not used in fitting the model to the observed time series. The primary statistical descriptor used to evaluate wet day amount models has been an average amount. Average wet day amounts, average seasonal total amounts and average annual total amounts have also been used. In all literature reviewed (Chapter 2), comparisons of observed and generated averages were made, with good to very good agreement reported. To fit a continuous distribution, such as the exponential, gamma, or
Figure 7.12 Comparison of observed and generated low-pass filtered time series of number of wet days for February at Gainesville, Orlando, Tampa, and Ft. Myers, Florida.
Figure 7.13  Comparison of observed and generated low-pass filtered time series of number of wet days for April at Gainesville, Orlando, Tampa, and Ft. Myers, Florida.
Figure 7.14  Comparison of observed and generated low-pass filtered time series of number of wet days for July at Gainesville, Orlando, Tampa, and Ft. Myers, Florida.
Figure 7.15  Comparison of observed and generated low-pass filtered time series of number of wet days for November at Gainesville, Orlando, Tampa, and Ft. Myers, Florida.
mixed exponential, to daily rainfall amounts, parameters are estimated which are the
average wet day amount or some function thereof. If a stochastic model of this type is
implemented and used to generate a long time series of wet day amounts, it is suggested
that the average may be preserved, while the generated distribution may or may not agree
with the observed distribution of amounts.

To a lesser extent, the standard deviation of amounts has also been used as a
statistical descriptor for model evaluation. The standard deviations of monthly and
annual total amounts were used for model evaluation by Zucchini and Adams (1984) and
Hanson et al. (1988). They found that the standard deviations were generally not well
preserved by stochastic rainfall models. They attributed this to the model’s inability to
preserve extreme rainfall events. If samples are drawn from two different distributions,
the standard deviations may be similar, but the shape of the distribution may be
significantly different. For this reason, it is suggested that the standard deviation may not
be a useful descriptor for model evaluation.

In addition to the average and standard deviation, empirical histograms and
empirical cumulative distributions have been used to compare observed and generated
rainfall amounts. To quantify the goodness of fit between observed and generated
distributions, the Kolmogorov-Smirnov goodness of fit test has been used routinely. In
Chapter 5.2.3, this test was discussed in relation to wet day amount models. Haan (1977)
suggested that this goodness of fit test is insensitive to the extreme events in the tails of
the distributions considered. This test uses the maximum vertical distance (D) between
the observed and generated distributions for evaluation. Very small vertical differences
(D) between the tails of two distributions can indicate large differences in wet day
amounts, while large D values in the middle of a distribution may not be indicative of
large discrepancies. In addition, two histograms may be very similar throughout, with a
significant difference over a limited range of data (1 bin). Integrating these two
histograms to obtain empirical cumulative distributions and applying the K-S test would
result in a large D value suggesting a low degree of fit. A similar problem exists with histograms of wet day amounts. All observed wet day amounts of rainfall have some type of "exponentially" shaped distribution with a high incidence of small amounts and a low incidence of large amounts. If standard histograms with equal bin widths are constructed, most of the sample will be contained in lowest bin, while the remaining few sample points will be allocated over a large number of bins. It is thus very difficult to determine if the observed distribution of extreme events (very large or small) is preserved.

To evaluate the proposed model, several alternative methods were considered. These evaluation methods focused on the ability of the wet day amount model to preserve the observed distribution and long-term trends of rainfall amounts.

7.3.1 Distribution of Wet Day Amounts

**Method.** To evaluate the distribution of wet day amounts, histograms with variable bin widths were used. This method was adopted to provide a concise, but detailed assessment of the "exponential-type" distribution of amounts, where a high incidence of small amounts and low incidence of large amounts was typical. The selection of a range of bin widths will depend on the needs of the user, similar to the selection of a microscope objective.

In nature, daily amounts of rainfall are continuous, however observed amounts (in the USA) are recorded in inches at increments of 0.01 in with a lower threshold of 0.01 in. This recording procedure results in a discrete subsample for amounts less than 0.1 in (0.01, 0.02,...,0.08, 0.09 in).

To account for the high incidence and discrete nature of small amounts, bin widths of 0.01 inches (0.254 mm) were used for amounts less than 0.105 in (2.667 mm). For amounts between 0.105 and 0.95 in (2.667 and 24.13 mm), bin widths of 0.1 in (0.254 mm) were used, and for amounts greater than 0.95 in (24.13 mm), bin widths of 1.0 in (25.4 mm) were used. Slightly different bin widths resulted during transitions between
groups of bins with different widths (Table 7.1). Bin 1 was set to contain only amounts of 0.01 in (0.254 mm). In this work, historical observations recorded in inches were converted to millimeters before analysis and model fitting. Daily rainfall amounts were then generated in millimeters using the continuous gamma model. Bin limits in millimeters were used to sort both the observed and generated data. After sorting, the histograms were presented as bar graphs with the midpoint of each bin labeled in inches. The height of each bin in the histogram was presented as the probability density which is the relative frequency of events in each bin divided by the bin width. Each bin was equally spaced along the abscissa of the histogram, rather than using arithmetic or logarithmic scales.\textsuperscript{3} To evaluate the distribution of wet day amounts, separate histograms were constructed for the observed and generated data, and plotted together on the same set of coordinates.

**Results.** Variable bin width histograms of wet day rainfall amounts were constructed for all months for the four locations considered. In Chapter 4.2.5, four rainfall seasons were defined (wet winter, dry spring, wet summer, dry autumn) for Florida. A representative month for each season (February, April, July, November) was selected for presentation. Separate histograms were constructed using observed and generated rainfall data, with all ten simulation runs combined to construct a single histogram for the generated data. For each month and location considered, histograms for the observed and generated rainfall amounts were combined into a single figure (Figs. 7.16 to 7.19). To compare evaluation methods, empirical cumulative probability distributions were included with each composite histogram.

For all months at all locations considered, the distribution of wet day amounts was well preserved for amounts greater than 0.1 in (2.54 mm), however there were discrepancies for amounts less than 0.1 in. The wet day threshold amount of 0.01 in

\textsuperscript{3} Bar graphs of this type are easily made using a spreadsheet.
Table 7.1 Bin limits, midpoint of bin, and bin width for variable bin width histograms.

<table>
<thead>
<tr>
<th>Bin Number</th>
<th>Lower Bin Limit (mm)</th>
<th>Bin Label (in) (midpoint)</th>
<th>Upper Bin Limit (mm)</th>
<th>Bin Width (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0000</td>
<td>0.01</td>
<td>0.2541</td>
<td>0.2541</td>
</tr>
<tr>
<td>2</td>
<td>0.2541</td>
<td>0.02</td>
<td>0.635</td>
<td>0.3809</td>
</tr>
<tr>
<td>3</td>
<td>0.635</td>
<td>0.03</td>
<td>0.889</td>
<td>0.254</td>
</tr>
<tr>
<td>4</td>
<td>0.889</td>
<td>0.04</td>
<td>1.143</td>
<td>0.254</td>
</tr>
<tr>
<td>5</td>
<td>1.143</td>
<td>0.05</td>
<td>1.397</td>
<td>0.254</td>
</tr>
<tr>
<td>6</td>
<td>1.397</td>
<td>0.06</td>
<td>1.651</td>
<td>0.254</td>
</tr>
<tr>
<td>7</td>
<td>1.651</td>
<td>0.07</td>
<td>1.905</td>
<td>0.254</td>
</tr>
<tr>
<td>8</td>
<td>1.905</td>
<td>0.08</td>
<td>2.159</td>
<td>0.254</td>
</tr>
<tr>
<td>9</td>
<td>2.159</td>
<td>0.09</td>
<td>2.413</td>
<td>0.254</td>
</tr>
<tr>
<td>10</td>
<td>2.413</td>
<td>0.10</td>
<td>2.667</td>
<td>0.254</td>
</tr>
<tr>
<td>11</td>
<td>2.667</td>
<td>0.20</td>
<td>6.35</td>
<td>3.683</td>
</tr>
<tr>
<td>12</td>
<td>6.35</td>
<td>0.30</td>
<td>8.89</td>
<td>2.54</td>
</tr>
<tr>
<td>13</td>
<td>8.89</td>
<td>0.40</td>
<td>11.43</td>
<td>2.54</td>
</tr>
<tr>
<td>14</td>
<td>11.43</td>
<td>0.50</td>
<td>13.97</td>
<td>2.54</td>
</tr>
<tr>
<td>15</td>
<td>13.97</td>
<td>0.60</td>
<td>16.51</td>
<td>2.54</td>
</tr>
<tr>
<td>16</td>
<td>16.51</td>
<td>0.70</td>
<td>19.05</td>
<td>2.54</td>
</tr>
<tr>
<td>17</td>
<td>19.05</td>
<td>0.80</td>
<td>21.59</td>
<td>2.54</td>
</tr>
<tr>
<td>18</td>
<td>21.59</td>
<td>0.90</td>
<td>24.13</td>
<td>2.54</td>
</tr>
<tr>
<td>19</td>
<td>24.13</td>
<td>1.00</td>
<td>38.10</td>
<td>13.97</td>
</tr>
<tr>
<td>20</td>
<td>38.10</td>
<td>2.00</td>
<td>63.5</td>
<td>25.4</td>
</tr>
<tr>
<td>21</td>
<td>63.5</td>
<td>3.00</td>
<td>88.9</td>
<td>25.4</td>
</tr>
<tr>
<td>22</td>
<td>88.9</td>
<td>4.00</td>
<td>114.3</td>
<td>25.4</td>
</tr>
<tr>
<td>23</td>
<td>114.3</td>
<td>5.00</td>
<td>139.7</td>
<td>25.4</td>
</tr>
<tr>
<td>24</td>
<td>139.7</td>
<td>6.00</td>
<td>165.1</td>
<td>25.4</td>
</tr>
</tbody>
</table>

NOTE: The wet day amount of rainfall, Y, was sorted between bins using

\[ L_i \leq Y < L_{i+1} \]

where \( L_i \) is bin limit \( i \). Rainfall amounts exceeding 6.5 in were included in the last bin.
Figure 7.16 Comparison of histograms and cumulative distributions for observed and generated wet day amounts of rainfall for February at Gainesville, Orlando, Tampa, and Ft. Myers, Florida.
Figure 7.17 Comparison of histograms and cumulative distributions for observed and generated wet day amounts of rainfall for April at Gainesville, Orlando, Tampa, and Ft. Myers, Florida.
Figure 7.18 Comparison of histograms and cumulative distributions for observed and generated wet day amounts of rainfall for July at Gainesville, Orlando, Tampa, and Ft. Myers, Florida.
Figure 7.19  Comparison of histograms and cumulative distributions for observed and generated wet day amounts of rainfall for November at Gainesville, Orlando, Tampa, and Ft. Myers, Florida.
(0.254 mm) was in general well preserved with occasional over or under predictions. This may be related to inadequacies in the generation of uniform deviates less than 0.01. The threshold amount was not generated from the gamma distribution model, but was generated using a uniform random number generator. A uniformly distributed random number (R) between 0 and 1 was generated each time a wet day occurred. If R was less than \( \delta(\tau) \) (Equation 5.18), the amount was the threshold, while if R was greater than \( \delta(\tau) \), the amount was drawn from the gamma distribution. The estimated parameter \( \delta(\tau) \) was small (0.05 to 0.1) and the uniform deviate R was slightly under predicted in this range (Figure 7.1).

Wet day amounts in the second bin of the histogram (0.01 in < \( Y \) ≤ 0.025 in) were consistently over predicted at all times, while amounts in bins 3 through 9 (0.025 in < \( Y \) ≤ 0.095 in) were generally under predicted. This is related to the continuous gamma distribution used to generate wet day amounts. In the range of amounts from \( Y > 0.01 \) in to \( Y < 0.025 \) in, or bin 2 of the histogram, the gamma model will generate real numbers throughout this range. More values are generated between \( Y > 0.01 \) in and \( Y = 0.02 \) in, than between \( Y > 0.02 \) in and \( Y < 0.025 \) in. In any simulation, the occurrence model generates a finite number of wet days to be filled from the gamma distribution. If this finite number of wet days is over filled with amounts \( Y < 0.025 \) in, this results in an under prediction of amounts from 0.025 in to 0.09 in. It is suggested that this problem will occur, regardless of the parameter estimation method or continuous model used. These continuous distributions maybe be accurately reproducing naturally occurring rainfall amounts, but they can not reproduce the discretized recorded amounts less than 0.09 in. This suggests that a three state (DRY, TRACE, WET) occurrence model may be appropriate, with the trace state modeled by a discrete distribution.

For research purposes such as model development, histograms with a large number of bins provide a detailed evaluation of the wet day amount model. For applications such
as model comparison, a smaller number of variable width bins may be useful, similar to changing objectives on a microscope. To overview the information in Figures 7.16 to 7.19, summary histograms (Figs. 7.20 to 7.23) were constructed. Bins labeled 0.01 to 0.09 in were combined into a single bin labeled 0.05 in, bins labeled 0.1 to 0.9 in were combined into a single bin labeled 0.5 in, bins labeled 1 to 3 in were not changed, and the bin labeled 4 in included all data greater than 3.5 in. For all seasons and locations considered, amounts in the bin labeled 0.05 in were slightly under predicted, while amounts in the bins labeled 0.5, 1, 2, 3, and 4 in were well preserved. For these reasons, it is suggested that the variable bin width or "microscope" histogram may be more useful for model development and model comparison than the cumulative distribution of wet day amounts of rainfall.

Empirical cumulative distributions were included in Figures 7.16 to 7.19 for comparison purposes. These cumulative distributions were constructed by integrating the variable bin width histograms using the same scale for the abscissa. Deviations between the observed and generated data are accumulated as the histogram is integrated. If a model consistently over (under) predicts wet day amounts, the observed and generated distributions will diverge, resulting in a large K-S D value. The empirical distributions will also diverge if the model fits well throughout most of the range of amounts, but over (under) predicts amounts in a limited range. If a model over and under predicts in different ranges of wet day amounts, the empirical distributions may converge, resulting in a low D value and incorrectly indicating a good fit.

7.3.2 Distribution of Monthly Total Amounts

Method. To further evaluate the ability of the proposed rainfall model to preserve the long-term variability of historical observations, the distribution of monthly total amounts of rainfall was considered. In Chapter 4.2.4, the distribution of monthly total rainfall was investigated in detail and described successfully using a gamma-probability-density model. These histograms and gamma model showed a
Figure 7.20  Comparison of summary histograms for observed and generated wet day amounts of rainfall for February at Gainesville, Orlando, Tampa, and Ft. Myers, Florida.
Figure 7.21  Comparison of summary histograms for observed and generated wet day amounts of rainfall for April at Gainesville, Orlando, Tampa, and Ft. Myers, Florida.
Figure 7.22  Comparison of summary histograms for observed and generated wet day amounts of rainfall for July at Gainesville, Orlando, Tampa, and Ft. Myers, Florida.
Figure 7.23 Comparison of summary histograms for observed and generated wet day amounts of rainfall for November at Gainesville, Orlando, Tampa, and Ft. Myers, Florida.
pronounced annual cycle (Figs. 4.4 to 4.7) ranging from an "extreme exponential" shape in winter to a "skewed-bell" shape in summer. The ability of the proposed rainfall model to preserve the observed annual cycle of histogram shape was investigated.

Histograms of generated monthly rainfall totals were constructed in the same manner as those in Chapter 4.2.4. Twenty bins with a bin width of 25 mm were used, and the relative frequency divided by the bin width, or probability density, was plotted so that the area under each histogram was unity. Monthly rainfall totals from all simulation runs were combined in each monthly histogram. Visual inspection and the Kolmogrov-Smirnov test (Tables 4.1 to 4.4) indicated that the gamma-density model fit the observed histograms of monthly totals quite well. For this reason, histograms of generated monthly totals were compared with the gamma-density model, rather than directly to the observed histogram. This approach was used to simplify the graphical presentation and focus attention on the annual cycle in histogram shape. Empirical cumulative distributions and the classical Kolmogrov-Smirnov test were not used for the reasons discussed above.

Results. Histograms of monthly total amount of rainfall were constructed for all months and locations using ten separate simulation runs. Each monthly histogram, representing the generated data, was plotted with the gamma-density model fitted to the observed data (Figs. 7.24 to 7.27) in Chapter 4. For the four locations considered, the annual cycle in histogram shape, or distribution of monthly total rainfall, was preserved. The distributions of large to extreme monthly totals (> 300 mm) were well preserved, while the distributions of low to moderate totals showed small deviations from the gamma density fit to the observed data.

In the first bin (0 to 25 mm), generated amounts were slightly over predicted from autumn through winter into spring, but were well preserved during the summer and autumn. This is related to the slight over prediction of completely dry months, during this part of the year, discussed in the evaluation of the occurrence model above. The
Figure 7.24  Comparison of generated histograms and observed gamma density functions for monthly total amounts of rainfall at Gainesville, Florida.
Figure 7.25 Comparison of generated histograms and observed gamma density functions for monthly total amounts of rainfall at Orlando, Florida.
Figure 7.26 Comparison of generated histograms and observed gamma density functions for monthly total amounts of rainfall at Tampa, Florida.
Figure 7.27  Comparison of generated histograms and observed gamma density functions for monthly total amounts of rainfall at Ft. Myers, Florida.
distributions of moderate monthly totals (100 to 300 mm) were well preserved from autumn through winter into spring, but deviated from the observed distribution during the wet summer season. There was a slight shift or skewness in the generated distributions during July and August, which did not occur during the rest of the year. This slight distortion is related to both the amount and occurrence models. During July, the amount model slightly under predicts small amounts of rainfall (Fig. 7.22), and months with a moderate number of wet days (15 to 20 days) are under predicted by the occurrence model (Fig. 7.6). The distribution of large monthly totals (> 300 mm) was well preserved during all seasons and locations considered.

As a statistical indicator for model evaluation, monthly total amounts of rainfall are influenced by both the amount of rainfall received per day (amount model) and the number of wet days per month (occurrence model). With an "exponential-type" density function for an amount model, the probability of generating large wet day amounts increases as the number of wet days per month increases, resulting in larger monthly totals. Similarly, as the number of wet days per month decreases, the probability of generating small daily amounts increases, resulting in smaller monthly totals. During July, the amount model slightly under predicted small amounts of rainfall (Fig. 7.22), and the occurrence model under predicted months with a moderate number (15 to 20 days) of wet days (Fig. 7.6). These shifts in the generated distributions of monthly total amounts may be considered as a measure of the combined goodness of fit of the amount model, occurrence model, and associated parameter simulation models.

It is suggested that visual inspection of the histograms used in this evaluation method is far more informative and useful in model development than the single test statistic or p-value obtained from a classical goodness of fit test. A single test statistic does not give an indication of where the lack of fit occurs over a broad range of values
contained in a distribution. A particular model may perform quite well over a broad range of conditions, however, a deviation or lack of fit in a very limited range would result in a test statistic suggesting a poor fit of the model in general.

7.3.3 Long-term Trend of Monthly Total Amounts

**Method.** In Chapter 4.2.5, the possible presence of long-term trends in monthly rainfall totals was investigated. Spectral analyses of seasonal rainfall total time series indicated the presence of pronounced periodicities of 11 to 12 years in the historical data (1900 to 1988). Based on these findings, time series of seasonal rainfall totals were low-pass filtered (Appendix D) using a cutoff frequency of 1/10 years (Figs. 4.13, 4.14, 4.15, 4.16). The period and amplitude of the low-pass filtered data varied from cycle to cycle, however periodicities of 11 to 12 years were pronounced during all seasons. The literature on long-term trends in rainfall (Chapter 4.2.5) suggested that 10 to 12 year periodicities in rainfall are related to sunspot cycles with similar period. Based on these findings, the long-term trend used in this evaluation method was defined as low-pass filtered data using a cutoff frequency of 1/10 years.

To evaluate the ability of the proposed model to preserve observed long-term trends, times series of monthly total amounts were constructed for the observed data and each individual simulation run. Each monthly time series was low-passed filtered to remove the long-term trend. The filtered time series of monthly totals, resulting from each simulation run were averaged to give a single time series for each month. The low-pass filtered time series for the observed and generated data were then compared graphically to determine if the proposed model preserved the observed long-term trend.

**Results.** Monthly total time series of rainfall amounts were constructed for the observed data (1900 to 1988) and generated data (1903 to 1985) for all months at the four locations considered. Each of these time series was low-pass filtered using a cutoff
frequency of 1/10 years. A representative month for the winter wet season (February),
the spring dry season (April), the summer wet season (July), and the autumn dry season
(November) were plotted in Figures 7.28 to 7.31.

During all seasons at the four locations considered, the amplitude and phase of the
observed long-term trends were preserved. The period and amplitude of the generated
trends were not constant, but varied from cycle to cycle in a manner very similar to the
observed trends. The phasing of the observed trend was preserved in the generated trends
at most seasons with only minor time shifts of approximately 2 to 3 years. Differences in
amplitude between the observed and generated trends were not constant, but varied from
no difference to isolated maxima of approximately 25 mm in summer and 10 mm in the
other seasons. These differences in amplitude were approximately 0 to 5% of average
monthly total amounts in most instances and approximately 10% at extremes.

In a evaluation of the weather simulation model CLIMATE, developed by
Woolhiser (1985, 1986, and 1988), Hanson concluded that:

The results of this validation support the findings of Woolhiser et al.
(1988) which suggests that one should use caution when using simulated
daily precipitation data to study annual phenomena. (1989, p. 873)

The above evaluation of the proposed model suggested that a stochastic rainfall model
can be designed which will preserve the distribution and long-term trends in observed
rainfall. In addition, the observed trends used in this implementation could be replaced
with climate change scenarios to study long-term effects of interannual or interdecade
phenomena.
Figure 7.28  Comparison of observed and generated low-pass filtered time series of monthly total amounts of rainfall for February at Gainesville, Orlando, Tampa, and Ft. Myers, Florida.
Figure 7.29  Comparison of observed and generated low-pass filtered time series of monthly total amounts of rainfall for April at Gainesville, Orlando, Tampa, and Ft. Myers, Florida.
Figure 7.30 Comparison of observed and generated low-pass filtered time series of monthly total amounts of rainfall for July at Gainesville, Orlando, Tampa, and Ft. Myers, Florida.
Figure 7.31  Comparison of observed and generated low-pass filtered time series of monthly total amounts of rainfall for November at Gainesville, Orlando, Tampa, and Ft. Myers, Florida.
CHAPTER 8
SUMMARY AND CONCLUSIONS

The mathematical description or modeling of daily rainfall at a point has used the
theory of stochastic processes extensively. Research activities in stochastic rainfall
modeling have consisted of selecting a classical stochastic process, fitting the selected
model to an observed data sample, generating a simulated time series of daily rainfall,
and evaluating the goodness of fit of the model considered. Increased interest in
interannual variability and long-term trends in rainfall has led to concern over the ability
of existing models to reproduce the variability found in historical rainfall records on a
time scale of decades. In an evaluation of the weather simulation model, CLIMATE,
Woolhiser and his colleagues cautioned users of existing stochastic rainfall models
saying:

We found that the MCME {Markov-chain/mixed exponential} model
preserved the important statistics within a year, but that caution should be
taken when using it to study annual phenomena. (Hanson et al., 1989, p.
873)

The problem considered in this research was to quantitatively investigate the nature
of long-term variability in historical rainfall records and to determine the ability of
stochastic rainfall models to preserve this variability on time scales of decades. The
hypothesis tested in this investigation was that long-term variability in stochastic rainfall
model parameters existed. This hypothesis was tested by including parametric
uncertainty in stochastic rainfall model fitting methods, and evaluating the ability of a
fitted model to preserve observed interannual variability and long-term trends.

In the literature reviewed, all stochastic rainfall models used annually time
invariant or annually homogeneous parameter estimates. The largest possible data
samples were selected and the best possible fit of the models to observations was sought.
To account for intraannual cycles in meteorological conditions, data samples were sorted seasonally and model parameters estimated for each season. To generate a long (decades) time series of daily rainfall using any particular model, the estimated parameter set was used repeatedly for each year of rainfall generated.

In contrast to this approach, annually time variant or annually nonhomogeneous parameter estimation methods were developed. Two different approaches were used to develop nonhomogeneous parameter estimation methods. Parametric uncertainty was quantified using statistical estimation theory to make estimates of the standard deviations associated with parameters and to construct confidence intervals. To investigate the possible presence of long-term trends, time dependent parameter estimation methods were developed. Parameter time series were obtained using maximum likelihood estimation methods combined with data sampling methods similar to a moving average.

To investigate the long-term variability of rainfall and implement the proposed parameter estimation methods, the longest possible time series of daily rainfall was sought as a statistical sample. From historical weather records, 89-year (1900 through 1988) time series of daily rainfall data were compiled for four locations in Florida (Gainesville, Orlando, Tampa, and Ft. Myers). Using this data base, time series of monthly total amounts and monthly number of wet days were constructed. Descriptive statistics (average, standard deviation, and quartile ranges), histograms, linear regression, and Fourier-domain analysis were used to quantitatively assess and describe long-term variability and trends in the amount and wet day count time series. Both the standard deviations and quartile ranges for the monthly amount and wet day count samples were large, indicating significant long-term variability. The histograms of monthly amounts and wet day counts showed a pronounced annual cycle in shape. The gamma probability density function was successfully fit to the monthly amount histograms, while the binomial mass function was found to be an inadequate model for the count histograms. Linear regression and Fourier-domain analyses were used to investigate the presence of
long-term trends in annual and seasonal (winter, spring, summer, and autumn) time series of amounts and counts. Spectral analysis and low-pass filtering indicated the presence of pronounced trends in both the seasonal amount and counts time series during all seasons and at all locations considered. The observed trends were cyclic in nature, but varied in period and amplitude from cycle to cycle. Statistically significant, long-term linear trends were not identified in the regression analyses made.

To illustrate the proposed parameter estimation methodology, a first-order Markov-chain was used for the rainfall occurrence model and a gamma distribution was used for the amount model. The Markov-chain / gamma-distribution rainfall simulation model was selected for illustrative purposes because of its widespread use in the WGEN weather generation model (Richardson and Wrigth, 1984). The proposed parameter estimation methodology is not model specific, thus other rainfall simulation models could have been used. Nonhomogeneous Markov-chain transition probabilities and gamma-distribution shape and scale parameters were estimated using the 89-year data base of daily rainfall data for Gainesville, Orlando, Tampa and Ft. Myers, Florida. Significant long-term trends and variability were found in the resulting parameter time series and interval parameter estimates.

Parameter simulation models were developed to permit the generation of long time series of daily rainfall based on a limited sample of historical data. Parameter simulation models were formulated and fit to the nonhomogeneous parameter estimates using time series analysis and modeling techniques. Each parameter simulation model was composed of a deterministic and a stochastic component. The deterministic component was included to describe any long-term trends contained in the parameter estimates. The deterministic or long-term trend was defined as the output of a low-pass filter, retaining periodicities of 10 years or greater. This filter was selected to retain possible solar cycles with periodicities of approximately 11 to 12 years. Other definitions of "trend" are also possible and could be substituted by adjusting the filter. A climate change scenario or
physically based predictive model could also be substituted for the trend component of the parameter simulation models. The stochastic component was included to describe any short-term correlation or variability in the parameter time series. Markov-chain transition probabilities were modeled using a seasonal multivariate autoregressive model, and the gamma-distribution parameters were modeled using a seasonal moving-average model.

The proposed annually nonhomogeneous parameter estimation and simulation methodologies were evaluated in two steps. A preliminary evaluation of the methodology was made by comparing long-term trends found in the historical rainfall data with long-term trends found in the estimated parameter time series. Time series of monthly total amounts and monthly number of wet days were filtered and compared with the long-term trends found in the amount and occurrence model parameter time series. During all seasons and locations considered, both the amount and occurrence trends found in the historical data were also found in the parameter trends. The amplitude and phase of the historical trends were accurately preserved in the parameter trends. This finding demonstrated that long-term trends in stochastic model parameters exist, and accurately reflect observed trends in historical rainfall data. A more rigorous evaluation of the proposed methodology was made by coupling a Markov-chain / gamma-distribution rainfall simulation model to the proposed parameter simulation model.

The proposed annually nonhomogeneous parameter estimation and simulation methodology was coupled to a first-order Markov-chain / gamma-distribution rainfall model and used to generate long time series of daily rainfall. Ten separate simulation runs were made using different seeds for the random number generators contained in the simulation models. Evaluation emphasized the ability of the proposed model to preserve the interannual variability and long-term trends found in historical rainfall records. To evaluate for the preservation of long-term variability of rainfall, monthly histograms of
wet day amounts, total amounts, number of wet days and dry run lengths were constructed from the observed and generated rainfall time series and compared. To evaluate for the presence of long-term trends in rainfall, observed and generated time series of monthly total amounts and monthly number of wet days were low-pass filtered and compared. During all seasons and at all locations considered, both the distributions and long-term trends found in observed rainfall records were preserved in the rainfall time series generated using the proposed parameter simulation model.

The implementation and evaluation of the proposed annually nonhomogeneous parameter estimation methodology demonstrated that stochastic rainfall models can be designed which will preserve the distribution and long-term trends found in historical rainfall records. In addition, the observed trends used in this implementation could be replaced with climate change scenarios or physically based models to study long-term effects of interdecade phenomena.
APPENDIX A
GAMMA DISTRIBUTION PARAMETER ESTIMATION

The gamma probability density is given by

\[
f(y) = \begin{cases} 
  y^{\alpha-1}e^{-y/\beta} \\ \frac{1}{\beta^\alpha\Gamma(\alpha)} 
\end{cases}, \quad \alpha, \beta > 0; \quad 0 < y < \infty \quad \text{A.1}
\]
\[y \leq 0\]

The method of maximum likelihood was used to estimate the parameters \(\alpha\) and \(\beta\) of the gamma density function. The likelihood function, \(L\), was defined by

\[
L = \prod_{i=1}^{n} f(y_i) \quad \text{A.2}
\]

where \(n\) is the sample size. Substituting A.1 into A.2 gives the likelihood function for the gamma density function given by

\[
L = \frac{y^{n(\alpha-1)}\exp\left[-\frac{1}{\beta}\sum_{i=1}^{n} y_i\right]}{\beta^n\Gamma(n\alpha)} \quad \text{A.3}
\]

To facilitate analysis the natural logarithm of A.3 was taken to give

\[
\ln L = (\alpha - 1) \sum_{i=1}^{n} \ln y_i - \frac{1}{\beta} \sum_{i=1}^{n} y_i - n \{\alpha \ln \beta + \ln \Gamma(\alpha)\} \quad \text{A.4}
\]
The maximum likelihood estimates of the parameters $\alpha$ and $\beta$ were found by maximizing the log likelihood function (A.4) using a downhill simplex method of function optimization (Nedler and Mead, 1965). To start the downhill simplex method, initial estimates of $\alpha$ and $\beta$ were required. Initial maximum likelihood estimates (Choi and Wette, 1969) were made using

\[
\hat{\alpha} = \left\{ 2 \left[ \ln \bar{y} - \frac{1}{n} \sum_{i=1}^{n} \ln y_i \right] \right\}^{-1} \tag{A.5}
\]
\[
\hat{\beta} = \frac{\bar{y}}{\alpha} \tag{A.6}
\]

where $\bar{y}$ is the sample mean.

To optimize the likelihood function, it was necessary to make repeated numerical evaluations of the gamma function in a computationally efficient and accurate manner. An approximation derived by Lanczos (1964) was used for this purpose and is given by

\[
\Gamma(y + 1) = \left( y + \gamma + \frac{1}{2} \right)^{y + \frac{1}{2}} e^{-\left( y + \gamma + \frac{1}{2} \right)} \] 
\[
\times \sqrt{2\pi} \left[ c_0 + \frac{c_1}{y + 1} + \frac{c_2}{y + 2} + \cdots + \frac{c_N}{y + N} + \varepsilon \right] \quad (y > 0) \tag{A.7}
\]

where $\gamma = 5$, $N = 6$, and the $c_i$'s are predetermined constants. The error of this approximation is $|\varepsilon| < 2 \times 10^{-10}$. 

APPENDIX B
GOODNESS OF FIT TESTS

Consider two sets of data or information. Let set A contain a sample of observed data, such as the wet day amount of rainfall recorded by a rain gauge. Secondly, let set B contain one of the following types of information:

i. a set of points obtained by evaluating a function, such as a theoretical cumulative probability distribution.

ii. a set of points generated by a model, such as the wet day amount of rainfall generated by a stochastic rainfall model.

iii. a second set of observed data, such as wet day rainfall amounts recorded at a second geographical location.

The problem is to determine the "goodness of fit" between sets A and B, or to determine if sets A and B are drawn from the same population. By constructing histograms or plotting sets A and B it should be visually apparent whether sets A and B "fit" very well or very badly. However, in many circumstances, the degree or goodness of fit is uncertain and a more rigorous statistical test is required.

When there is uncertainty about the goodness of fit between sets A and B, a decision making procedure or formal statistical test is often used. To make a decision about the goodness of fit between sets A and B, a research hypothesis is proposed. The research hypothesis (H₁) is supported by offering evidence to reject a null hypothesis (H₀) which is the converse of the research hypothesis. This is a proof by contradiction, rather than a direct proof that sets A and B are from different distributions. The evidence to test the hypothesis comes from the observed sample set A. To make a decision about the goodness of fit between sets A and B, the following hypotheses are tested:

1. The term data set is used for observed or measured values, while the term information set is used for values generated from a mathematical function or model.
\[ H_0: \text{ two data sets are drawn from the same distribution.} \]
\[ H_a: \text{ two data sets are drawn from different distributions.} \]

To test these hypotheses, it is necessary to obtain evidence from observations summarized in the form of a test statistic. Test statistics with known sampling distributions are available for many widely used statistical tests.

A test statistic for goodness of fit tests is usually a measure of the deviation between the observed distribution of set A, and the distribution of set B. A large test statistic suggests that the fit is poor, while a small test statistic suggests that the fit is good. If for example, the sampling distribution of the test statistic is bell shaped with zero mean, then a large test statistic lying in either tail of this distribution offers evidence to reject \( H_0 \), while a small statistic lying close to the mean does not offer evidence to reject \( H_0 \). To specify when the null hypothesis is to be rejected, a rejection region for the test is set. When the test statistic falls in the rejection region, the null hypothesis is rejected. The rejection region is set by specifying a significance level or total tail area (\( \alpha \)) of the sampling distribution. This tail area is the probability of rejecting the null hypothesis when it is true and is known as the Type I error of the statistical test. Rather than specifying a rejection region for a specific level of significance (\( \alpha \)), an attained significance level or p-value can be specified. The p-value is the tail area of the sampling distribution associated with the observed test statistic. Small p-values are associated with large test statistics and offer evidence to reject \( H_0 \).

Two widely used goodness of fit tests are the chi-square test and the Kolmogorov-Smirnov test. The chi-square test is used with binned data, such as the number of wet days per month. The Kolmogorov-Smirnov test is used with continuous data, such as the wet day amount of rainfall. It is possible to use the chi-square test for continuous data, if the data are binned before testing. However, sorting continuous data

---

2. If a statistic is estimated by repeatedly sampling a population, then the distribution of this statistic is known as the sampling distribution.
into bins is not recommended, as the bin width used can distort the shape of the observed distribution. The details of these tests are discussed below closely following von Mise (1964), along with analytical details and necessary numerical approximations for computer implementation following Abramowitz and Stegun (1964).

**B.1 Chi-square Test**

If set A contains observed data and set B contains information on the expected number of events from some theoretical distribution, then the chi-square test statistic is

\[ \chi^2_{\text{observed}} = \sum_{i=1}^{b} \frac{(O_i - E_i)^2}{E_i} \]

where \(O_i\) is the observed number of events in bin \(i\), \(E_i\) is the expected number of events in bin \(i\) from some known distribution, and \(b\) is the number of bins. If set A and set B both contain observed data, then the chi-square test statistic is

\[ \chi^2_{\text{observed}} = \sum_{i=1}^{b} \frac{(O_i^A - O_i^B)^2}{O_i^A + O_i^B} \]

where \(O_i^A\) is the observed number of events in bin \(i\) from set A, and \(O_i^B\) is the observed number of events in bin \(i\) from set B. \(^3\)

---

3. The denominator of B.2 is not the average of \(O_i^A\) and \(O_i^B\), but the sum. Consider the square of the standardized random variable, \(Z_i^2 = \frac{(O_i^A - O_i^B)^2}{\sigma_i^2}\), where \(\sigma_i^2 = \sigma_A^2 + \sigma_B^2\). The variances for sets A and B reduce to \(\sigma_A^2 = \sum_i O_i^A\) and \(\sigma_B^2 = \sum_i O_i^B\) respectively. After substitution, the standardized random variable is

\[ Z_i^2 = \frac{(O_i^A - O_i^B)^2}{\sum_i (O_i^A + O_i^B)} \]

If the squared standardized random variable, \(Z_i^2\), is summed over the number of bins, then the result

\[ \sum_i Z_i^2 = \sum_i \frac{(O_i^A - O_i^B)^2}{O_i^A + O_i^B} \]

is the test statistic given in B.2.
The test statistic given in either B.1 or B.2 is the sum of squared random variables. If each random variable in this sum is normally distributed, then the sum has a chi-square cumulative distribution. In application, each term in this sum may not be normally distributed, however, if either the number of bins is large or the number of events in each bin is large, then the chi-square distribution is considered an adequate approximation for this test. The sampling distribution of $\chi^2$ has the form

$$f(x) = \frac{1}{2^{v/2} \Gamma(v/2)} x^{(v/2)-1} e^{-x/2}$$

where $v$ is the number of degrees of freedom. If $b$ is the number of bins and $p$ is the number of parameters estimated from the observed data, then $v = b - p - 1$. Integrating B.3 results in the chi-square distribution given by

$$Pr\{\chi^2 \leq x\} = \frac{1}{2^{v/2} \Gamma(v/2)} \int_0^x t^{v/2-1} e^{-t/2} dt$$

To calculate the p-value for an observed value of $\chi^2$, it is necessary to take the compliment of B.4 which is

$$p\text{-value} = 1 - Pr\{\chi^2 \leq \chi^2_{\text{observed}}\} = Pr\{\chi^2 > \chi^2_{\text{observed}}\}$$

$$= \frac{1}{2^{v/2} \Gamma(v/2)} \int_{\chi^2_{\text{observed}}}^{\infty} t^{v/2-1} e^{-t/2} dt$$

Tables can be used to evaluate B.5, however for large samples analyzed on a digital computer this approach is not practical, and a numerical technique is needed. Equation B.5 can be evaluated numerically by using a continued fraction given by
where the gamma function, \( \Gamma(v/2) \), is evaluated using the Lanczos' approximation (Appendix A.1).

From B.6 above, it is clear that a large estimated test statistic, \( \chi^2 \), results in a small p-value. Large \( \chi^2 \) values and small p-values then offer evidence to reject the null hypothesis and suggest that sets A and B are drawn from different distributions. Tabled values of \( \chi^2 \) at several different significance levels (\( \alpha \)) are available in most statistics texts for comparison with calculated p-values.

**B.2 Kolmogorov-Smirnov Test**

To avoid sorting continuous data into bins before making a goodness of fit test, the Kolmogorov-Smirnov test is regularly used. This test compares a theoretical cumulative distribution, \( P(z) \), with an observed distribution, point by point. Consider the continuous random variable \( Z \) for which the observed sample set \( A \) is \( \{ z_1, z_2, \ldots, z_n \} \). If the sample set is first sorted into ascending order, then an empirical cumulative distribution, \( S_n(z) \), is determined by calculating the fraction of data points less than any given value of \( z \). This empirical distribution has the form

\[
S_n(z) = \begin{cases} 
0, & z < z_1 \\
\frac{j}{n}, & z_j \leq z < z_{j+1}, \quad j = 1, 2, \ldots n - 1 \\
1, & z \geq z_n
\end{cases}
\]

The Kolmogorov-Smirnov test statistic is the absolute value of the maximum deviation between the observed and theoretical distributions, given by

\[
D_{\text{observed}} = \max_{-\infty < z < \infty} |S_n(z) - P(z)|
\]
If set B is an observed or simulated sample rather than a theoretical function, then the test statistic for comparing two observed distributions is

\[ D_{\text{observed}} = \max_{- \infty < z < \infty} |S_{n_A}(y) - S_{n_B}(y)| \]

where \( n_A \) and \( n_B \) are the sample sizes for sets A and B respectively.

Using the absolute value of the deviation between the distribution of sets A and B results in a two sided test where the p-value of some value, \( \lambda \), was found by Kolmogorov (1933) to be

\[ \text{p-value} = Q_{KS}(\lambda) = 2 \sum_{j=1}^{\infty} (-1)^{j-1} e^{-2j^2\lambda^2} \]

Using this function, the p-value of the observed value of D can be obtained from

\[ \text{p-value} = Pr\{ D > D_{\text{observed}} \} = Q_{KS}(\sqrt{n} D_{\text{observed}}) \]

for a single observed sample, or from

\[ \text{p-value} = Pr\{ D > D_{\text{observed}} \} = Q_{KS}\left( \sqrt{\frac{n_A n_B}{n_A + n_B}} D_{\text{observed}} \right) \]

if two observed samples are compared.

Large D values and associated small p-values indicate a significant deviation between the distributions of sets A and B, and offer evidence to reject the null hypothesis.
Tabled D values for several levels of significance are not routinely available in engineering statistical texts, however Birnbaum (1952), Haan (1977) and von Mise (1964) provide suitable tables.
APPENDIX C
NUMERICAL EVALUATION OF THE STUDENT’S T DISTRIBUTION

The student’s $t$ distribution is frequently used in hypothesis testing. It is often necessary to evaluate the attained significance level or p-value for a particular value of $t$. For this application, tabled values of the student’s $t$ distribution were not appropriate, and it was necessary to integrate the student’s distribution directly. The integration of the student’s distribution is not straightforward and approximations were necessary for computer implementation. The methods and approximations (Abramowitz and Stegun, 1984) used are given below.

The student’s $t$ distribution has the form

$$S(t, v) = \frac{1}{\sqrt{v} B\left(\frac{1}{2}, \frac{v}{2}\right)} \int_{-t}^{t} \left(1 + \frac{x^2}{v}\right)^{-\frac{v+1}{2}} dx$$

where $v$ is the number of degrees of freedom and $B(\cdot, \cdot)$ is the beta function. The significance level of the statistical test is desired which is the tail area $1 - S(t, v)$. This tail area is related to the incomplete beta function, $I_x(a, b)$ by

$$1 - S(t, v) = I_x\left(\frac{v}{2}, \frac{1}{2} \right)$$

C.1

C.2
To evaluate the significance level, it is only necessary to evaluate the incomplete beta function, \( I_x(a, b) \), for \( x = \frac{v}{v + t^2} \) with \( a = \frac{v}{2} \), and \( b = 1/2 \). The incomplete beta function is defined by

\[
I_x(a, b) = \frac{B_x(a, b)}{B(a, b)} = \frac{1}{B(a, b)} \int_0^x t^{a-1}(1-t)^{b-1} \, dt \quad (a, b > 0)
\]

where \( B(a, b) \) is the beta function defined by

\[
B(a, b) = \int_0^1 t^{a-1}(1-t)^{b-1} \, dt
\]

To numerically evaluate the incomplete beta function (C.3) a continued fraction approximation was used with the form

\[
I_x(a, b) = \frac{x^a(1-x)^b}{a B(a, b)} \left[ \frac{1}{1 + \frac{d_1}{1 + \frac{d_2}{1 + \cdots}} \right]
\]

where

\[
d_{2m+1} = \frac{(a + m)(a + b + m)x}{(a + 2m)(a + 2m + 1)}
\]

\[
d_{2m} = \frac{m(b - m)x}{(a + 2m - 1)(a + 2m)}
\]

The continued fraction approximation converges rapidly with the number of iterations on the order of \( O(\sqrt{\text{max}(a, b)}) \). A stopping criteria of \( O(10^{-7}) \) was used in the iteration. The beta function, \( B(a, b) \), was evaluated using the gamma function as
\[ B(a, b) = \frac{\Gamma(a) \Gamma(b)}{\Gamma(a + b)} \]  

where \( \Gamma(\cdot) \) is the gamma function evaluated using the Lanczos' approximation (Appendix A).
APPENDIX D
SPECTRAL ANALYSIS AND FOURIER DOMAIN FILTERING METHODS

Spectral analysis and Fourier-domain methods are well established and widely used approaches to time series analysis. Originally developed in electrical engineering, this approach has not been as widely applied to hydrologic problems as classical statistical approaches and time-domain analysis (Box and Jenkins, 1976). Many references on spectral methods of time series analysis exist, including Bendat and Piersol (1971), Jenkins and Watts (1968), and Priestley (1981). The objective of this appendix was to document the particular details of the methods used, rather than give a detailed theoretical development.

D.1 Fourier Series Approximation

Consider the observed time series of \( N \) discrete points \( \{Z_0, \ldots, Z_{\tau}, \ldots, Z_{N-1}\} \). This time series can be approximated by a function such as a polynomial, a cubic spline or a Fourier series composed of \( N \) cosine terms. This Fourier series can be written as

\[
Z_\tau = \frac{1}{N} \sum_{k=0}^{N-1} C_k \cos \left( \frac{2\pi \tau k}{N} \right)
\]

D.1

To use the Fourier series approximation, it is necessary to determine the Fourier coefficients, \( C_k \). The discrete Fourier transform given by

\[
C_k = \sum_{\tau=0}^{N-1} Z_\tau e^{2\pi ik/N} \quad k = 0, \ldots, N - 1
\]

D.2
expresses the Fourier coefficients as a function of the N discrete observed points. For long time series, the direct computation of the Fourier coefficients is prohibitively time consuming. Widespread use of the Fourier series and spectral methods occurred with the development of the fast Fourier transform (FFT) popularized by J. W. Cooley and J. W. Tukey of IBM. A Cooley-Tukey type FFT implemented by Brenner (Press et al., 1986) was used to take the Fourier transform of the time series \( \{Z_t\} \).

\[ D.2 \text{ Spectral Energy Estimation} \]

A simple periodogram estimate of the spectral energy, \( P \), of the time series \( \{Z_t\} \) was used. For a discrete time series with \( N \) points and an equally spaced time interval of \( \Delta \), the periodogram estimate has the form

\[
P(0) = P(f_0) = \frac{1}{N^2} |C_0|^2
\]

\[
P(f_k) = \frac{1}{N^2} [ |C_k|^2 + |C_{N-k}|^2 ] \quad k = 1, 2, \ldots, (N/2 - 1)
\]

\[
P(f_c) = P(f_{N/2}) = \frac{1}{N^2} |C_{N/2}|^2
\]

where \( f_k \) is frequency in cycles per time interval, \( \Delta \), defined by

\[
f_k = \frac{k}{N\Delta} = 2f_c \frac{k}{N} \quad k = 0, 1, \ldots, N/2
\]

with \( f_c \) the critical frequency equal to \( 1/2\Delta \).

To use this method, it is necessary to begin with a time series whose length is an integer power of 2. To obtain a time series of this length, the raw observed time series was padded with trailing zeroes. After zero padding, the Fourier transform of the time
series was taken using the FFT to obtain the Fourier coefficients, $C_k$. A one-sided spectral energy estimate was then calculated using equation D.3. It is important to note, that $P(f_k)$ does not represent the portion of the spectral energy at exactly $f_k$, but rather that portion of the spectrum contained in the frequency "bin"

$$f_k \pm \left( \frac{f_{k+1} - f_k}{2} \right)$$

The estimated spectral energy was plotted as discrete spikes centered in each frequency "bin". In the analysis of many types of hydrologic time series, the period of the $k$th cosine component, $T_k = 1/f_k$, in equation D.1 is more meaningful than the frequency, $f_k$. Because of this, the spectral energy was plotted against period rather than frequency. Using these methods and graphical presentation, large spectral peaks indicate a range of periods or "bin" which contains a relatively large proportion of the total signal. The time series analyzed in this work were relatively short (< 100 points). To obtain the maximum "resolution" or minimum "bin width", the entire time series was analyzed in one segment rather than in windowed segments. Data windowing (Jenkins and Watts, 1968) is commonly used to reduce frequency "bin" leakage and the uncertainty of the spectral estimate, however it does decrease the resolution and increase the bias of the estimate.

**D.3 Fourier-domain Filtering**

Fourier or frequency-domain filtering is perhaps the most straightforward, but least used method of filtering a hydrologic time series. To filter a time series: take the FFT of the entire time series, and then multiply the output of the FFT, $C_k$, by an appropriate filter function $H(f)$

$$C_k^F = H(f) \cdot C_k$$

D.6
where $C^F_t$ is the set of Fourier coefficients after filtering. After multiplying by the filter function, the inverse FFT or equation D.1 is used to recombine the $C^F_t$'s to obtain the filtered time series by.

To low-pass filter the time series using a cutoff frequency of $f_c$, the filter function

$$H(f) = \begin{cases} 
1 & f \leq f_c \\
0 & \text{elsewhere}
\end{cases} \quad \text{D.7}$$

was used. This filter passed low frequency (long period) components of the observed time series, and removed high frequency (short period) components. To high-pass filter, the filter function

$$H(f) = \begin{cases} 
1 & f \geq f_c \\
0 & \text{elsewhere}
\end{cases} \quad \text{D.8}$$

was used. This filter passed high frequency (short period) components of the observed time series and removed low frequency (long period) components. Finally, to band-pass filter using the frequency band $f_1$ to $f_2$, the filter function

$$H(f) = \begin{cases} 
1 & f_1 \leq f \leq f_2 \\
0 & \text{elsewhere}
\end{cases} \quad \text{D.9}$$

was used. This filter passed a predetermined range of frequencies, and removed very high and low frequencies.
APPENDIX E
EVALUATION OF LARGE SAMPLE VARIANCES FOR THE MAXIMUM LIKELIHOOD PARAMETER ESTIMATES OF THE GAMMA DISTRIBUTION PARAMETERS

To quantify the uncertainty associated with the maximum likelihood estimates of the gamma density function parameters, \( \hat{\alpha} \) and \( \hat{\beta} \), a simultaneous confidence region was constructed about these jointly estimated parameters in Chapter 5. Using large-sample\(^1\) properties, it was assumed that the joint distribution of the maximum likelihood parameter estimates \((\hat{\alpha}, \hat{\beta})\) are asymptotically Gaussian. Using the pivotal-quantity method a joint confidence region was constructed for \( \hat{\alpha} \) and \( \hat{\beta} \) (Mood et al., 1963).

To calculate the confidence region it is necessary to calculate the large-sample standard deviation of \( \alpha \), \( \sigma_\alpha \), and \( \beta \), \( \sigma_\beta \). In Theorem 5.2 of Chapter 5 the large-sample standard deviations of the gamma probability density \( f(x; \alpha, \beta) \) were given as

\[
\sigma_\alpha^2 = \frac{-E\left[ \frac{\partial}{\partial \beta} \log f(x; \alpha, \beta) \right] \Delta}{n \Delta} \quad \text{E.1}
\]

\[
\sigma_\beta^2 = \frac{-E\left[ \frac{\partial}{\partial \alpha} \log f(x; \alpha, \beta) \right] \Delta}{n \Delta} \quad \text{E.2}
\]

where

\[
\Delta = E\left[ \frac{\partial^2}{\partial \alpha^2} \log f(x; \alpha, \beta) \right] E\left[ \frac{\partial^2}{\partial \beta^2} \log f(x; \alpha, \beta) \right] - \left( E\left[ \frac{\partial^2}{\partial \beta \partial \alpha} \log f(x; \alpha, \beta) \right] \right)^2 \quad \text{E.3}
\]

\(^1\) A large sample is one for which the sample size tends to infinity.
To calculate \( \sigma_\alpha \) and \( \sigma_\beta \), it is necessary to take the first and second partial derivatives of

\[
\log f(x; \alpha, \beta) = (\alpha - 1) \log x - \frac{x}{\beta} - \alpha \log \beta - \log \Gamma(\alpha)
\]

where

\[
f(x; \alpha, \beta) = \frac{x^{\alpha-1} e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)} \quad y > 0
\]

\[
f(x; \alpha, \beta) = 0 \quad y \leq 0
\]

is the gamma probability density function. The seasonal index has been dropped from the gamma density for clarity, but the following applies to any season considered. The required first derivative with respect to \( \alpha \) is

\[
\frac{\partial}{\partial \alpha} \log f(x) = \log x - \log \beta - \frac{\partial}{\partial \alpha} \log \Gamma(\alpha)
\]

where the derivative of \( \log \Gamma(\alpha) \) is evaluated following Spiegel (1968) to give

\[
\frac{\partial}{\partial \alpha} \log \Gamma(\alpha) = \frac{1}{\Gamma(\alpha) \partial \alpha} \Gamma(\alpha)
\]

\[
= -\gamma + \left(1 - \frac{1}{\alpha}\right) + \left(\frac{1}{2} - \frac{1}{\alpha + 1}\right) + \ldots
\]

\[
+ \left(\frac{1}{n} - \frac{1}{\alpha + n - 1}\right) + \ldots
\]

where

\[
-\gamma = \Gamma'(1) = \int_0^\infty e^{-t} \log t \, dt
\]

resulting in the required derivative
\[ \frac{\partial}{\partial \alpha} \log f(x; \alpha, \beta) = \frac{\partial}{\partial \alpha} \left( \log x - \log \beta - \left[ -\gamma + \left( 1 - \frac{1}{\alpha} \right) + \left( \frac{1}{2} - \frac{1}{\alpha + 1} \right) + \cdots \right] \right) \]

\[ + \left( \frac{1}{n} - \frac{1}{\alpha + n - 1} \right) + \cdots \] \quad E.9

The second partial derivative with respect to \( \alpha \) is

\[ \frac{\partial^2}{\partial \alpha^2} \log f(x; \alpha, \beta) = \frac{\partial}{\partial \alpha} \left\{ \log x - \log \beta - \left[ -\gamma + \left( 1 - \frac{1}{\alpha} \right) \right. \right. \]
\[ \left. + \left( \frac{1}{2} - \frac{1}{\alpha + 1} \right) + \cdots \right. \]
\[ + \left. \left( \frac{1}{n} - \frac{1}{\alpha + n - 1} \right) + \cdots \right\} \]
\[ = -\frac{\partial}{\partial \alpha} \left[ -\gamma + \left( 1 - \frac{1}{\alpha} \right) + \left( \frac{1}{2} - \frac{1}{\alpha + 1} \right) + \cdots \right] \]
\[ + \left( \frac{1}{n} - \frac{1}{\alpha + n - 1} \right) + \cdots \] \quad E.10

\[ = -\left[ \alpha^{-2} + (\alpha + 1)^{-2} + (\alpha + 2)^{-2} + \cdots \right. \]
\[ \left. + (\alpha + n - 1)^{-2} + \cdots \right] \]
\[ \equiv -\Sigma_\alpha \]

The required derivatives with respect to \( \beta \) are

\[ \frac{\partial}{\partial \beta} \log f(x; \alpha, \beta) = x \beta^{-2} + \alpha \beta^{-1} \] \quad E.11

\[ \frac{\partial^2}{\partial \beta^2} \log f(x; \alpha, \beta) = -2x \beta^{-3} - \alpha \beta^{-2} \] \quad E.12

while the derivative with respect to \( \alpha \) and \( \beta \) is
\[
\frac{\partial^2}{\partial \beta \partial \alpha} \log f(x; \alpha, \beta) = -\frac{1}{\beta}
\]
E.13

Taking the expected value\(^2\) of E.10, E.12, and E.13 gives

\[
E\left[ \frac{\partial^2}{\partial \alpha^2} \log f(x; \alpha, \beta) \right] = E[-\Sigma_\alpha] = -\Sigma_\alpha
\]
E.14
\[
E\left[ \frac{\partial^2}{\partial \beta^2} \log f(x; \alpha, \beta) \right] = -\frac{\alpha}{\beta^2}
\]
E.15
\[
E\left[ \frac{\partial^2}{\partial \beta \partial \alpha} \log f(x; \alpha, \beta) \right] = -\frac{1}{\beta}
\]
E.16

and substituting E.14, E.15, and E.16 into E.3 gives

\[
\Delta = \frac{\alpha \Sigma_\alpha - 1}{\beta^2}
\]
E.17

Finally, the large-sample variances for the maximum likelihood estimates \((\hat{\alpha}, \hat{\beta})\) of the gamma distribution parameters are obtained by substituting E.14 through E.17 into E.1 and E.2. The associated standard deviations have the form

\[
\sigma_\alpha = \sqrt{\frac{\alpha}{n(\hat{\alpha} \Sigma_\alpha - 1)}}
\]
E.18
\[
\sigma_\beta = \sqrt{\frac{\Sigma_\alpha \beta^2}{n(\hat{\alpha} \Sigma_\alpha - 1)}}
\]
E.19

---

2. The expected value or mean of a continuous random variable \(X\) with a gamma probability density function, \(f(x)\), is (Mendenhall et al., 1981)

\[
E[X] = \int_{\infty}^{\infty} x f(x) dx = \alpha \beta
\]
APPENDIX F
APPLICATION OF MULTIVARIATE AUTOREGRESSIVE TIME SERIES MODELS

Multivariate time series models are similar in form to univariate time series models (Box and Jenkins, 1976), but far more complex in application. The theory of multivariate autoregressive time series models is given by Priestley (1981) and Salas (1980), while the application of multivariate time series models to hydrology and water resource management is reviewed by Salas (1980) and Salas et al. (1985). However, the multivariate literature reviewed focused on specific applications of low order models, with limited discussion on general models with variable order. For this reason, a detailed derivation and discussion of a general bivariate autoregressive time series model was developed.

F.1 Theory

Consider two time series, \( \{z_1,t\} \) and \( \{z_2,t\} \), \( t = 1, 2, \ldots, n \) which are dependent. A general bivariate autoregressive model (AR(p)) of order p which preserves the dependence in time between the two series is

\[
\begin{align*}
    z_{1,t} &= a_{11,1}z_{1,t-1} + a_{12,1}z_{2,t-1} + a_{11,2}z_{1,t-2} + a_{12,2}z_{2,t-2} + \cdots \\
    &+ a_{11,p}z_{1,t-p} + a_{12,p}z_{2,t-p} + b_{11}e_{1,t} + b_{12}e_{2,t} \\
    z_{2,t} &= a_{21,1}z_{1,t-1} + a_{22,1}z_{2,t-1} + a_{21,2}z_{1,t-2} + a_{22,2}z_{2,t-2} + \cdots \\
    &+ a_{21,p}z_{1,t-p} + a_{22,p}z_{2,t-p} + b_{21}e_{1,t} + b_{22}e_{2,t}
\end{align*}
\]

F.1
where $a_{i,k}$ and $b_{i,j}$ are model parameters estimated from observed data and $\varepsilon_{i,t}$ are stationary, uncorrelated white noise terms with unit variance. Equation F.1 can be more concisely expressed in matrix form by

$$
\begin{bmatrix}
  z_{1,t} \\
  z_{2,t}
\end{bmatrix}
= 
\begin{bmatrix}
  a_{11,1} & a_{12,1} \\
  a_{21,1} & a_{22,1}
\end{bmatrix}
\begin{bmatrix}
  z_{1,t-1} \\
  z_{2,t-1}
\end{bmatrix}
+ 
\begin{bmatrix}
  a_{11,2} & a_{12,2} \\
  a_{21,2} & a_{22,2}
\end{bmatrix}
\begin{bmatrix}
  z_{1,t-2} \\
  z_{2,t-2}
\end{bmatrix}
+ \ldots
$$

or

$$
\begin{bmatrix}
  z_{1,t} \\
  z_{2,t}
\end{bmatrix}
= 
\begin{bmatrix}
  a_{11,p} & a_{12,p} \\
  a_{21,p} & a_{22,p}
\end{bmatrix}
\begin{bmatrix}
  z_{1,t-p} \\
  z_{2,t-p}
\end{bmatrix}
+ 
\begin{bmatrix}
  b_{11,p} & b_{12,p} \\
  b_{21,p} & b_{22,p}
\end{bmatrix}
\begin{bmatrix}
  \varepsilon_{1,t} \\
  \varepsilon_{2,t}
\end{bmatrix}
$$

where $Z_t$ are 2x1 time series vectors, $A_i$ and $B$ are 2x2 parameter matrices, and $\varepsilon_t$ is a 2x1 vector. Note that $E[\varepsilon_1, \varepsilon_2] = 0$ and $E[\varepsilon_t, \varepsilon_t'] = I$ where $I$ is the identity matrix and ($'$) indicates the transpose of a matrix.

### F.2 Parameter Estimation

To estimate the parameters of the AR(p) model, it is first necessary to estimate the lag k (k=1, ..., p) auto-covariance and cross-covariance of the observed series, $Z_t$, given by

$$
R(k) = E[Z_t Z_{t-k}']
$$

$$
= 
\begin{bmatrix}
  r_{11,k} & r_{12,k} \\
  r_{21,k} & r_{22,k}
\end{bmatrix}
$$

where
\[ r_{ij,k} = \text{Cov}(z_i, z_{j+t-k}) \]
\[ = E[z_i z_{j+t-k}] - E[z_i] E[z_{j+t-k}] \]
\[ = E[z_i, z_{j+t-k}] \]
\[ = \frac{1}{n} \sum_{t=1}^{k+1} z_{i,t} z_{j,t-k} \quad (i,j = 1,2) \]

assuming that the observed series have zero means, i.e. \( E[z_{1,t}] = 0, E[z_{2,t}] = 0. \)

Postmultiplying equation F.3 by \( Z'_{t-k} \) and taking expected values taken gives the multivariate Yule-Walker equations. For lag \( k = 0 \), the resulting equation is

\[
\begin{align*}
R(0) &= A_1 R(-1) + A_2 R(-2) + \cdots + A_p R(-p) + BB' \\
&= A_1 R'(1) + A_2 R'(2) + \cdots + A_p R'(p) + BB'
\end{align*}
\]

and for lags \( k = 1, \ldots, p \), the set of equations

\[
\begin{align*}
R(1) &= A_1 R(0) + A_2 R(-1) + \cdots + A_p R(1-p) \\
R(2) &= A_1 R(1) + A_2 R(0) + \cdots + A_p R(2-p) \\
&\vdots \\
R(p) &= A_1 R(p-1) + A_2 R(p-2) + \cdots + A_p R(0)
\end{align*}
\]

results. Rewriting equation F.7 and noting that \( R(-k) = R'(k) \) gives

\[
[R(1) \quad R(2) \quad \cdots \quad R(p)] = [A_1 \quad A_2 \quad \cdots \quad A_p]
\]

\[
\begin{bmatrix}
R(0) & R(1) & \cdots & R(p-1) \\
R'(1) & R(0) & \cdots & R(p-2) \\
& & & \ddots \\
R'(p-1) & R'(p-2) & \cdots & R(0)
\end{bmatrix}
\]

F.8
Solving for the parameter matrix gives

$$[A_1, A_2, \ldots, A_p] = [R(1) \quad R(2) \quad \cdots \quad R(p)]$$

To determine the parameter matrix $B$ in equation F.3, the lag 0 Yule Walker equation (F.6) is rearranged to give

$$D \equiv BB' = R(0) - A_1R'(1) - A_2R'(2) - \cdots - A_pR'(p)$$

where the elements of matrix $B$ are derived by lower triangularization (Graybill, 1969)

$$b_{ij} = d_{ij}/b_{jj} \quad j = 1, i = 1, 2$$

$$b_{ij} = \left[ d_{ij} - \sum_{k=1}^{j-1} b_{jk}^2 \right]^{1/2} \quad j = 2, i = j$$

**F.3 Order Selection**

For multivariate models, the Akaike information criteria (AIC) modified by Jones (1974) is used to select the model order where

$$AIC(p) = n \log |D| + 8p$$

The matrix $D$ is the variance-covariance matrix of the model residuals (equation F.10). The AR(p) model order is determined by fitting models of increasing order, calculating the residual variance-covariance matrix, and evaluating the associated AIC for each order. A model with minimum AIC or low residual variance is then selected.
F.4 Goodness of Fit

The residuals of the AR(p) model \( (\varepsilon_{1,t}, \varepsilon_{2,t}) \) are assumed to be stationary white noise, or they are independent and normally distributed. Checking the validity of these assumptions constitutes the goodness of fit test for the selected model (Salas et al. 1985). The residuals of the fitted model are determined using the estimated parameters and the observed time series. At each time step the equation

\[
\tilde{\varepsilon}_t = B^{-1}[Z_t - A_1 Z_{t-1} - A_2 Z_{t-2} - \cdots - A_p Z_{t-p}]
\]

is solved, resulting in a time series of model residuals, \( \tilde{\varepsilon}_t \). These time series are then analyzed testing the hypotheses that they are normally distributed and independent. The Kolmogorov-Smirnov test (Appendix B) is used to test the hypotheses that the residuals are normally distributed, and the Porte Manteau test (Salas, 1980) is used to test the independence or lack of correlation of the residuals. The test statistic for the Porte Manteau test is

\[
Q_i = n \sum_{k=1}^{L} \rho_{i,k}^2(\varepsilon_i)
\]

where \( \rho_{i,k}(\varepsilon_i) = r_{i,k}(\varepsilon_i)/r_{i,0}(\varepsilon_i) \) is the correlation of the residual series \( \varepsilon_{i,t} \) (\( i = 1,2 \) for a bivariate model). This test statistic has a chi-square distribution with \( L - p \) degrees of freedom, where \( L \) is 10 to 30% of the sample size. If \( Q_i < \chi^2(L - p) \) for a specified significance level, then \( \varepsilon_{i,t} \) are independent series.
REFERENCES


Earth Info Inc. *Climatedata*. 90 Madison Street, Suite 200, Denver.


George Michael Schmidt III was born on 10 April 1948, in Baltimore, Maryland. He is the son of Ruth Glaeser Schmidt and George M. Schmidt, Jr.

He attended public schools in Baltimore and graduated from high school in 1966. During high school, he developed a strong interest in biology and mathematics. Through the Maryland Academy of Sciences Intern Program for high school students, he spent 1963-1964 studying the complex equilibria of biochemical reactions at the University of Maryland School of Medicine. During 1964-1965, he studied neuroscience at Johns Hopkins School of Medicine.

Schmidt attended Johns Hopkins University on scholarship from 1966 to 1970. He was awarded the degree of Bachelor of Engineering Science, with a concentration in geophysical fluid mechanics, in 1970.

From 1970 to 1973, Schmidt worked as an instrumentation design engineer at the Chesapeake Bay Institute of Johns Hopkins University.

In 1973, Schmidt enrolled for graduate work at the University of Alaska. He received the Master of Science degree in physical oceanography in 1977. During his studies he served as cruise leader on numerous research expeditions to the Gulf of Alaska, Prince William Sound and Port Valdez. In 1976 Schmidt augmented his master’s program by studying the design, construction and deployment of geophysical time series instrumentation at the Department of Ocean Engineering, Woods Hole Oceanographic Institution, Massachusetts Institute of Technology.

From 1979 to 1980, Schmidt served as an intern at the United Nations Environment Programme, Palais des Nations, Geneva, Switzerland. He has worked as a consultant to
the World Health Organization and the United Nations Environment Programme in the areas of water quality and water resource management.

Schmidt was a graduate student at the University of Florida from 1983 to 1992. During this time, he worked toward the Ph.D. degree in the Department of Agricultural Engineering. He majored in the soil and water area of agricultural engineering, and minored in stochastic processes and time series analysis.
I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Allen G. Smajstrla, Chair
Professor of Agricultural Engineering

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Pedro S. Zazueta, Cochair
Associate Professor of Agricultural Engineering

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Michael K. Ochi
Professor of Coastal and Oceanographic Engineering

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

James W. Jones
Professor of Agricultural Engineering

This dissertation was submitted to the Graduate Faculty of the College of Engineering and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

December 1992

Winfred M. Phillips
Dean, College of Engineering

Madelyn M. Lockhart
Dean, Graduate School