STOCHASTIC DAILY PRECIPITATION MODELING AND DAILY STREAMFLOW TRANSFER PROCESSES

by

Tiao J. Chang
M. Levent Kavvas
Jacques W. Delleur

June 1982

PURDUE UNIVERSITY
WATER RESOURCES RESEARCH CENTER
WEST LAFAYETTE, INDIANA
STOCHASTIC DAILY PRECIPITATION MODELING AND
DAILY STREAMFLOW TRANSFER PROCESSES

by

Tiao J. Chang
M. Levent Kavvas
Jacques W. Delleur

The work upon which this report is based
was supported in part by funds provided by
the Office of Water Research and Technology
Project No. OWRT-B-112-IND, U.S. Department
of the Interior, Washington, D.C., as
authorized by the Water Research and
Development Act of 1978 (PL95-467)

Period of Investigation: October 1978- June 1982
Matching Fund Agreement: 14-34-0001-0219

PURDUE UNIVERSITY WATER RESOURCES RESEARCH CENTER
TECHICAL REPORT NO. 146
JUNE 1982
Contents of this paper do not necessarily reflect the views and policies of the Office of Water Research and Technology, U.S. Department of the Interior, nor does mention of trade names or commercial products constitute their endorsement or recommendation for use in the U.S. Government.
ACKNOWLEDGEMENTS

The authors gratefully acknowledge the financial support from the Office of Water Resources Research, Department of the Interior, under the grant OWRT-B-112-IND and from Purdue University. The authors also wish to express their appreciation to Dr. Dan Wiersma, Director of the Purdue Water Resources Research Center, for his assistance in the administration of the project.
ABSTRACT

This report constructs two newly defined stochastic processes: the Binary Discrete Autoregressive Moving Average modeling the wet-dry precipitation sequence mixed with an Exponential distribution to express the magnitude of the precipitation (B-DARMA-E) and the Multi-state Discrete Autoregressive Moving Average process (M-DARMA). Both processes are used to model the daily precipitation time series in Indiana. A three-step procedure of identification, Estimation, and Diagnostic Checking, is formulated for the modeling by the B-DARMA-E and the M-DARMA processes. The identification by means of the autocorrelation function is convenient from an engineering point of view, while the estimation through the preservation of the autocorrelations is quite suitable for a stochastic process. The diagnostic checking by means of the run length distributions is shown to be statistically efficient. The binary run length distribution is defined and derived for the former, while the multi-state run length distributions are constructed for the latter. A criterion for the selection of the best model is discussed separately and makes use of the run length property. These schemes of diagnostic checking and of best model selection are shown very effective in the illustrative example which uses Indiana data. Finally the Transfer Discrete Autocoregressive Moving Average model (T-DARMA) is
conceptually constructed for the daily precipitation-streamflow transfer process based on the known properties of the B-DARMA-E or the M-DARMA precipitation model. The statistical relationships between the input and the output series are Resources Research Center, for his assistance in the administration of the project. Formulated and used to estimate the model parameters so that the T-DARMA model preserves the first and the second order statistical properties as well as the conceptual water balance. The residual series are investigated and show that the model is very satisfactory.
TABLE OF CONTENTS

LIST OF TABLE ------------------------------- ix

LIST OF FIGURES ----------------------------- xi

LIST OF SYMBOLS ----------------------------- xvii

CHAPTER I-INTRODUCTION ---------------------- 1

I.1. The place of this research in the study of water resources ----------------- 1

I.2. A brief history of stochastic models in daily hydrological series ------- 2

I.3. Introduction and objectives of the report ------------------------------- 7

I.4. Research data ----------------------------- 11

CHAPTER II-DISCRETE AUTOREGRESSIVE MOVING AVERAGE MODELS ------------------ 14

II.1. Introduction ------------------------------ 14

II.2. General DARMA models ------------------- 16

II.3. Properties of models of particular interest ----------------------------- 20

II.4. Run lengths of B-DARMA and M-DARMA models ---------------------------- 28

CHAPTER III-ANALYSES AND RESULTS OF B-DARMA-E MODELING ------------------ 44

III.1. Introduction ----------------------------- 44

III.2. The BDARMA-E model --------------------- 47
III.2.1. Definitions and assumptions ---- 47
III.2.2. Verifications of assumptions ---- 49
III.3. DARMA modeling --------------- 51
   III.3.1. Model identification --------- 52
   III.3.2. Model estimation ---------- 55
   III.3.3. Model diagnostic checking --- 61
   III.3.4. Selection of the best model --- 67
III.4. The nonzero precipitation sequence --- 68
III.5. Summary ---------------------- 69
III.6. Supplement --------------------- 71

CHAPTER IV- ANALYSES AND RESULTS OF W-DARMA MODELING --------------------- 123

IV.1. Introduction --------------------- 123
IV.2. The W-DARMA daily precipitation model -
   IV.2.1. Definition ------------------- 127
   IV.2.2. Physical reality of the W-DARMA process ------------------- 129
IV.3. Multi-state DARMA modeling ------- 130
   IV.3.1. Model identification --------- 131
   IV.3.2. Model estimation ------------- 133
   IV.3.3. Model diagnostic checking----- 134
   IV.3.4. Selection of the best model --- 138
IV.4. Summary ------------------------ 139
IV.5. Supplement ---------------------- 141

CHAPTER V- TRANSFER DISCRETE AUTOREGRESSIVE MOVING AVERAGE MODEL ----------- 199
V.1. Introduction -------------------------- 199
V.2. Daily precipitation-streamflow processes 202
   V.2.1. Stochastic transfer processes ------ 202
   V.2.2. The T-DARMA process ------------- 204
V.3. Particular T-DARMA models ----------- 205
   V.3.1. Statistical properties ---------- 206
   V.3.2. Model estimation --------------- 212
   V.3.3. The checking by residual series -- 214
V.4. Summary ----------------------------- 215

CHAPTER VI-DISCUSSIONS ---------------------- 229
VI.1. BDARMAE modeling ------------------- 229
VI.2. M-DARMA modeling ------------------- 232
VI.3. T-DARMA modeling ------------------- 234

CHAPTER VII- CONCLUSIONS AND RECOMMENDATIONS ---- 236
VII.1. Conclusions ------------------------ 236
VII.2. Recommendations ------------------- 239

BIBLIOGRAPHY --------------------------------- 241
APPENDICES ----------------------------------- 248
   A. Covariance structures of particular DARMA models ------------------- 248
   B. Binary multi-state run lengths ---------------- 255
   C. Estimation of parameters in the DARMA(1,1) model ------------------- 268
<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-1</td>
<td>Watershed identification</td>
<td>12</td>
</tr>
<tr>
<td>3-1</td>
<td>Statistics for the verification of independences; station: 12-1229</td>
<td>73</td>
</tr>
<tr>
<td>3-1a</td>
<td>Chi-square values of contingency tables for independence verification</td>
<td>74</td>
</tr>
<tr>
<td>3-2</td>
<td>Estimated parameters for B-DARMA-E models; station: 12-1229</td>
<td>75</td>
</tr>
<tr>
<td>3-2a</td>
<td>Estimated parameters for B-DARMA-E models; season 1</td>
<td>76</td>
</tr>
<tr>
<td>3-2b</td>
<td>Estimated parameters for B-DARMA-E models; season 2</td>
<td>77</td>
</tr>
<tr>
<td>3-2c</td>
<td>Estimated parameters for B-DARMA-E models; season 3</td>
<td>78</td>
</tr>
<tr>
<td>3-2d</td>
<td>Estimated parameters for B-DARMA-E models; season 4</td>
<td>79</td>
</tr>
<tr>
<td>3-3</td>
<td>Statistics of Goodness-of-Fit for B-DARMA-E models; station: 12-1229</td>
<td>80</td>
</tr>
<tr>
<td>3-3a</td>
<td>Judgements by the test of goodness-of-fit; season 1</td>
<td>81</td>
</tr>
<tr>
<td>3-3b</td>
<td>Judgements by the test of goodness-of-fit; season 2</td>
<td>82</td>
</tr>
<tr>
<td>3-3c</td>
<td>Judgements by the test of goodness-of-fit; season 3</td>
<td>83</td>
</tr>
<tr>
<td>3-3d</td>
<td>Judgements by the test of goodness-of-fit; season 4</td>
<td>84</td>
</tr>
<tr>
<td>3-4</td>
<td>Selection of the best models in B-DARMA-E processes; station: 12-1229</td>
<td>85</td>
</tr>
<tr>
<td>3-4a</td>
<td>Selection of the best B-DARMA-E model</td>
<td>86</td>
</tr>
<tr>
<td>3-5</td>
<td>K-S statistics for the exponential fit of precipitation magnitudes; station: 12-1229</td>
<td>87</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>3-5a</td>
<td>Exponential fits of non-zero precipitations by K-S test.</td>
<td>88</td>
</tr>
<tr>
<td>4-1</td>
<td>Initial parameters for M-DARMA models; station: 12-1229</td>
<td>142</td>
</tr>
<tr>
<td>4-2</td>
<td>Estimated parameters for M-DARMA models; station: 12-1229</td>
<td>143</td>
</tr>
<tr>
<td>4-2a</td>
<td>Estimated parameters for M-DARMA models: season 1.</td>
<td>144</td>
</tr>
<tr>
<td>4-2b</td>
<td>Estimated parameters for M-DARMA models: season 2.</td>
<td>145</td>
</tr>
<tr>
<td>4-2c</td>
<td>Estimated parameters for M-DARMA models: season 3.</td>
<td>146</td>
</tr>
<tr>
<td>4-2d</td>
<td>Estimated parameters for M-DARMA models: season 4.</td>
<td>147</td>
</tr>
<tr>
<td>4-3</td>
<td>Statistics of goodness-of-fit for M-DARMA models; station: 12-1229</td>
<td>148</td>
</tr>
<tr>
<td>4-3a</td>
<td>Judgements by the test of goodness-of-fit: season 1.</td>
<td>149</td>
</tr>
<tr>
<td>4-3b</td>
<td>Judgements by the test of goodness-of-fit: season 2.</td>
<td>150</td>
</tr>
<tr>
<td>4-3c</td>
<td>Judgements by the test of goodness-of-fit: season 3.</td>
<td>151</td>
</tr>
<tr>
<td>4-3d</td>
<td>Judgements by the test of goodness-of-fit: season 4.</td>
<td>152</td>
</tr>
<tr>
<td>4-4</td>
<td>Selection of the best model in the M-DARMA process; station: 12-1229</td>
<td>153</td>
</tr>
<tr>
<td>4-4a</td>
<td>Selection of the best model.</td>
<td>154</td>
</tr>
<tr>
<td>5-1</td>
<td>Estimated parameters of transfer models; station: 327500</td>
<td>218</td>
</tr>
<tr>
<td>5-2</td>
<td>Statistics of generated and observed series; station: 327500</td>
<td>219</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figure
1-1 Watersheds used for analysis --------------- Page 13

3-1a Number of dry days, N0, in 20-year record; station: 12-1229---------------------- 88

3-1b Number of wet days, N1, in 20-year record; station: 12-1229 ------------------- 89

3-1c Seasonal means; station: 12-1229----------------- 89

3-2 Relationship between the nonzero precipitation and its occurrence time; station: 12-1229-------- 90

3-3 Theoretical autocorrelations for some candidate B-DARMA models; station: 12-1229-- 91

3-4 The estimated autocorrelation functions for different seasons; station: 12-1229-------- 32

3-5a Empirical and B-DARMA(0,1) model autocorrelation; station: 12-1229------------ 93

3-5b Empirical and B-DARMA(1,0) model autocorrelation functions; station: 12-1229 94

3-5c Empirical and B-DARMA(1,1) model autocorrelation functions; station: 12-1229 95

3-5d Empirical and B-DARMA(1,2) model autocorrelation functions; station: 12-1229 96

3-6a Probability distributions of run length 0 for B-DARMA(0,1) models; station: 12-1229---- 97

3-6b Probability distributions of run length 0 for B-DARMA(1,0) models; station: 12-1229---- 98

3-6c Probability distributions of run length 0 for B-DARMA(1,1) models; station: 12-1229--- 99

3-6d Probability distributions of run length 0 for B-DARMA(1,2) models; station: 12-1229--- 100

3-7a Probability distributions of run length 1 for B-DARMA(0,1) models; station: 12-1229--- 101

3-7b Probability distributions of run length 1 for B-DARMA(1,0) models; station: 12-1229--- 102
3-7c Probability distributions of run length 1 for B-DARMA(1,1) models; station: 12-1229
103
3-7d Probability distributions of run length 1 for B-DARMA(1,2) models; station: 12-1229
104
3-8 Autocorrelation functions of the nonzero precipitation sequences; station: 12-1229
105
3-9 Exponential distributions of the nonzero precipitation sequences; station: 12-1229
106
3-10a Probability distributions of run length 0 for B-DARMA-E models, station: 12-0676
107
3-10b Probability distributions of run length 1 for B-DARMA-E models, station: 12-0676
108
3-11a Probability distributions of run length 0 for B-DARMA-E models, station: 12-0831
109
3-11b Probability distributions of run length 1 for B-DARMA-E models, station: 12-0831
110
3-12a Probability distributions of run length 0 for B-DARMA-E models, station: 12-1734
111
3-12b Probability distributions of run length 1 for B-DARMA-E model, station: 12-1734
112
3-13a Probability distributions of run length 0 for B-DARMA-E models, station: 12-5337
113
3-13b Probability distributions of run length 1 for B-DARMA-E models, station: 12-5337
114
3-14a Probability distributions of run length 0 for B-DARMA-E models, station: 12-6018
115
3-14b Probability distributions of run length 1 for B-DARMA-E models, station: 12-6018
116
3-15a Probability distributions of run length 0 for B-DARMA-E models, station: 12-7362
117
3-15b Probability distributions of run length 1 for B-DARMA-E models, station: 12-7362
118
3-16a Probability distributions of run length 0 for B-DARMA-E models, station: 12-7747
119
3-16b Probability distributions of run length 1 for B-DARMA-E models, station: 12-7747
120
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-17a</td>
<td>Probability distributions of run length 0 for B-DARMA-E models, station: 12-9138-</td>
<td>121</td>
</tr>
<tr>
<td>3-17b</td>
<td>Probability distributions of run length 1 for B-DARMA-E models, station: 12-9138-</td>
<td>122</td>
</tr>
<tr>
<td>4-1a</td>
<td>Autocorrelation functions of observed daily precipitations; station: 12-1229-</td>
<td>155</td>
</tr>
<tr>
<td>4-1b</td>
<td>Autocorrelation functions of 100-state daily precipitations; station: 12-1569-</td>
<td>156</td>
</tr>
<tr>
<td>4-1c</td>
<td>Autocorrelation functions of 10-state daily precipitations; station: 12-1229-</td>
<td>157</td>
</tr>
<tr>
<td>4-1d</td>
<td>Autocorrelation functions of 3-state daily precipitations; station: 12-1229-</td>
<td>158</td>
</tr>
<tr>
<td>4-2a</td>
<td>Autocorrelation functions of M-DARMA(0,1;3) models; station: 12-1229-</td>
<td>159</td>
</tr>
<tr>
<td>4-2b</td>
<td>Autocorrelation functions of M-DARMA(1,0;3) models; station: 12-1229-</td>
<td>160</td>
</tr>
<tr>
<td>4-2c</td>
<td>Autocorrelation functions of M-DARMA(1,1;3) models; station: 12-1229-</td>
<td>161</td>
</tr>
<tr>
<td>4-2d</td>
<td>Autocorrelation functions of M-DARMA(1,2;3) models; station: 12-1229-</td>
<td>162</td>
</tr>
<tr>
<td>4-3a</td>
<td>Probability distributions of run length 0 for M-DARMA(0,1;3) models; station: 12-1229-</td>
<td>163</td>
</tr>
<tr>
<td>4-3b</td>
<td>Probability distributions of run length 1 for M-DARMA(0,1;3) models; station: 12-1229-</td>
<td>164</td>
</tr>
<tr>
<td>4-3c</td>
<td>Probability distributions of run length 2 for M-DARMA(0,1;3) models; station: 12-1229-</td>
<td>165</td>
</tr>
<tr>
<td>4-4a</td>
<td>Probability distributions of run length 0 for M-DARMA(1,0;3) models; station: 12-1229-</td>
<td>166</td>
</tr>
<tr>
<td>4-4b</td>
<td>Probability distributions of run length 1 for M-DARMA(1,0;3) models; station: 12-1229-</td>
<td>167</td>
</tr>
<tr>
<td>4-4c</td>
<td>Probability distributions of run length 2 for M-DARMA(1,0;3) models; station: 12-1229-</td>
<td>168</td>
</tr>
<tr>
<td>4-5a</td>
<td>Probability distributions of run length 0 for M-DARMA(1,1;3) models; station: 12-1229-</td>
<td>169</td>
</tr>
<tr>
<td>4-5b</td>
<td>Probability distributions of run length 1 for M-DARMA(1,1;3) models; station: 12-1229-</td>
<td>170</td>
</tr>
<tr>
<td>Page</td>
<td>Title</td>
<td>Station</td>
</tr>
<tr>
<td>------</td>
<td>----------------------------------------------------------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>4-5c</td>
<td>Probability distributions of run length 2 for M-DARMA(1,1:3) models;</td>
<td>12-1229-</td>
</tr>
<tr>
<td></td>
<td>station: 12-1229-</td>
<td></td>
</tr>
<tr>
<td>4-6a</td>
<td>Probability distributions of run length 0 for M-DARMA(1,2:3) models;</td>
<td>12-1229-</td>
</tr>
<tr>
<td></td>
<td>station: 12-1229-</td>
<td></td>
</tr>
<tr>
<td>4-6b</td>
<td>Probability distributions of run length 1 for M-DARMA(1,2:3) models;</td>
<td>12-1229-</td>
</tr>
<tr>
<td></td>
<td>station: 12-1229-</td>
<td></td>
</tr>
<tr>
<td>4-6c</td>
<td>Probability distributions of run length 2 for M-DARMA(1,2:3) models;</td>
<td>12-1229-</td>
</tr>
<tr>
<td></td>
<td>station: 12-1229-</td>
<td></td>
</tr>
<tr>
<td>4-7a</td>
<td>Probability distributions of run length 0 for M-DARMA models, station:</td>
<td>12-0676-</td>
</tr>
<tr>
<td></td>
<td>12-0676-</td>
<td></td>
</tr>
<tr>
<td>4-7b</td>
<td>Probability distributions of run length 1 for M-DARMA models, station:</td>
<td>12-0676-</td>
</tr>
<tr>
<td></td>
<td>12-0676-</td>
<td></td>
</tr>
<tr>
<td>4-7c</td>
<td>Probability distributions of run length 2 for M-DARMA models, station:</td>
<td>12-0676-</td>
</tr>
<tr>
<td></td>
<td>12-0676-</td>
<td></td>
</tr>
<tr>
<td>4-8a</td>
<td>Probability distributions of run length 0 for M-DARMA models, station:</td>
<td>12-0831-</td>
</tr>
<tr>
<td></td>
<td>12-0831-</td>
<td></td>
</tr>
<tr>
<td>4-8b</td>
<td>Probability distributions of run length 1 for M-DARMA models, station:</td>
<td>12-0831-</td>
</tr>
<tr>
<td></td>
<td>12-0831-</td>
<td></td>
</tr>
<tr>
<td>4-8c</td>
<td>Probability distributions of run length 2 for M-DARMA models, station:</td>
<td>12-0831-</td>
</tr>
<tr>
<td></td>
<td>12-0831-</td>
<td></td>
</tr>
<tr>
<td>4-9a</td>
<td>Probability distributions of run length 0 for M-DARMA models, station:</td>
<td>12-1734-</td>
</tr>
<tr>
<td></td>
<td>12-1734-</td>
<td></td>
</tr>
<tr>
<td>4-9b</td>
<td>Probability distributions of run length 1 for M-DARMA models, station:</td>
<td>12-1734-</td>
</tr>
<tr>
<td></td>
<td>12-1734-</td>
<td></td>
</tr>
<tr>
<td>4-9c</td>
<td>Probability distributions of run length 2 for M-DARMA models, station:</td>
<td>12-1734-</td>
</tr>
<tr>
<td></td>
<td>12-1734-</td>
<td></td>
</tr>
<tr>
<td>4-10a</td>
<td>Probability distributions of run length 0 for M-DARMA models, station:</td>
<td>12-5337-</td>
</tr>
<tr>
<td></td>
<td>12-5337-</td>
<td></td>
</tr>
<tr>
<td>4-10b</td>
<td>Probability distributions of run length 1 for M-DARMA models, station:</td>
<td>12-5337-</td>
</tr>
<tr>
<td></td>
<td>12-5337-</td>
<td></td>
</tr>
<tr>
<td>4-10c</td>
<td>Probability distributions of run length 2 for M-DARMA models, station:</td>
<td>12-5337-</td>
</tr>
<tr>
<td></td>
<td>12-5337-</td>
<td></td>
</tr>
<tr>
<td>4-11a</td>
<td>Probability distributions of run length 0 for M-DARMA models, station:</td>
<td>12-6018-</td>
</tr>
<tr>
<td></td>
<td>12-6018-</td>
<td></td>
</tr>
<tr>
<td>4-11b</td>
<td>Probability distributions of run length 1 for M-DARMA models, station:</td>
<td>12-6018-</td>
</tr>
<tr>
<td></td>
<td>12-6018-</td>
<td></td>
</tr>
</tbody>
</table>
4-11c Probability distributions of run length 2 for M-DARMA models, station: 12-8018------- 180
4-12a Probability distributions of run length 0 for M-DARMA models, station: 12-7362------- 190
4-12b Probability distributions of run length 1 for M-DARMA models, station: 12-7362------- 191
4-12c Probability distributions of run length 2 for M-DARMA models, station: 12-7362------- 192
4-13a Probability distributions of run length 0 for M-DARMA models, station: 12-7747-------- 193
4-13b Probability distributions of run length 1 for M-DARMA models, station: 12-7747-------- 194
4-13c Probability distributions of run length 2 for M-DARMA models, station: 12-7747-------- 195
4-14a Probability distributions of run length 0 for M-DARMA models, station: 12-9138-------- 196
4-14b Probability distributions of run length 1 for M-DARMA models, station: 12-9138-------- 197
4-14c Probability distributions of run length 2 for M-DARMA models, station: 12-9138-------- 198
5-1 Empirical autocorrelations of daily streamflows; station: 3275000--------------------- 220
5-2 Empirical and T-DARMA(0,1,1,0,1) model autocorrelation functions; station: 3275000-- 221
5-3 Empirical and T-DARMA(0,1,1,1,1) model autocorrelation functions; station: 3275000-- 222
5-4 Autocorrelations of the residual series based on T-DARMA(0,1,1,0,1) models; station: 3275000 ----------------------------- 223
5-5 Autocorrelations of the residual series based on T-DARMA(0,1,1,1,1) models; station: 3275000 ----------------------------- 224
5-6 Generated daily flows by the T-DARMA(0,1,1,0,1) model and its observed counterparts; station: 3275000--------------- 225
5-7 Generated daily flows by the T-DARMA(0,1,1,1,1) model and its observed counterparts; station: 3275000------------- 226
5-8 Autocorrelations of $T$-DARMA($0,1,1,0,1$) generated and observed series; station: 3275000

6-9 Autocorrelations of $T$-DARMA($0,1,1,1,1$) generated and observed series; station: 3275060
LIST OF SYMBOL

Latin Symbols

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_n$</td>
<td>an operating sequence in the DARMA process</td>
</tr>
<tr>
<td>$\bar{A}$</td>
<td>matrix in the nonlinear least squares derivation</td>
</tr>
<tr>
<td>$\bar{A}^0$</td>
<td>the correction of $\bar{A}$</td>
</tr>
<tr>
<td>$a_{ij}$</td>
<td>the ij-th element of $\bar{A}$</td>
</tr>
<tr>
<td>$a_n$</td>
<td>the sequence for the derivation of the run length 0 in the DARMA(0,1) model</td>
</tr>
<tr>
<td>$b_n$</td>
<td>a sequence for the derivation of run length 1 in the DARMA(0,1) model</td>
</tr>
<tr>
<td>$\mathbf{C}$</td>
<td>the vector $(c_1, c_2, \ldots, c_m)$</td>
</tr>
<tr>
<td>$\mathbf{C}^0$</td>
<td>the correction of $\mathbf{C}$</td>
</tr>
<tr>
<td>$c_k$</td>
<td>the k-th correction in the linear expansion of nonlinear normal equation</td>
</tr>
<tr>
<td>Cov</td>
<td>covariance function</td>
</tr>
<tr>
<td>Corr</td>
<td>correlation function</td>
</tr>
<tr>
<td>D</td>
<td>Kolmogorov-Smirnov statistic</td>
</tr>
<tr>
<td>$D^+$</td>
<td>the upper of K-S statistic</td>
</tr>
<tr>
<td>$D_-$</td>
<td>the lower limit of K-S statistic</td>
</tr>
<tr>
<td>E</td>
<td>expectation function</td>
</tr>
<tr>
<td>$E_i$</td>
<td>expected number in the i-th sample class</td>
</tr>
<tr>
<td>F</td>
<td>cumulative distribution function</td>
</tr>
<tr>
<td>$F_{na}$</td>
<td>empirical cumulative distribution function</td>
</tr>
<tr>
<td>H</td>
<td>bivariate transition matrix in the B-DARMA or M-DARMA model</td>
</tr>
<tr>
<td>f</td>
<td>mathematical function</td>
</tr>
</tbody>
</table>
I: identical matrix
Lj: the transfer-process loss in the j-th season
M: discrete K-S statistic
Oi: observed number in i-th sample class
P: probability function
Qk: streamflow in the k-th day
Pk: precipitation in the k-th day
Rj: the mean value of precipitation in the j-th season
r: the class of discrete precipitation series
S: sum of squared errors
Ti: run length i in a discrete series
U: random variable for defining general DARMA process
V: random variable for defining general DARMA process
W: the magnitude variable of precipitation in the B-DARMA-E model
X: discrete variable in the DARMA process
Y: the i.i.d. variable in defining DARMA model
Z: the variable of nonzero precipitation sequences

Greek Symbol

αi: the probabilities in defining the DAR process
βi: the probabilities in defining the DMA process
δ: the operator function in the general transfer process
ϕi: parameters in the general autoregressive process
γ: the parameters for the DARMA(1,2) model
μ: autoregressive parameters for particular T-DARMA model
\( v \): moving-average parameters for particular T-DARMA models
\( \pi \): probabilities for the discrete-state precipitations
\( \rho \): parameters for particular DARMA models
\( \delta_j \): the standard deviation of precipitation series in the \( j \)-th season
\( \Theta \): the operator function of the moving average process
\( \Sigma \): summation sign
\( \phi \): the operator function of the general autoregressive process
CHAPTER I

INTRODUCTION

I.1. The place of this research in the study of water resources

It is the purpose of this report to study the behaviors of both floods and droughts and to provide the characteristics of the run lengths of wet and dry periods, which are used to physically interpret the properties of floods and droughts. This is done through the modelling of daily hydrological series and the models are applied to several locations in Indiana. Precipitation being the only source of fresh water the earth receives, floods and droughts are the results of precipitation excesses or deficiencies. Changnon(1980) studied the floods and droughts in Illinois and concluded that the most efficient way to define floods and droughts is through precipitation informations which are also applicable to streamflows. Therefore this study is devoting itself to the modelling of daily precipitations as well as the daily streamflow through the former.
Floods and droughts are dependent, in part, on the periods of occurrence of precipitations. The same amount of precipitation could result in a flood if it happens in a short period and may bring a drought if it occurs over an extremely long time interval. The amounts of precipitations and their occurrence periods are very uncertain and must therefore be analyzed by probabilistic methods. This approach requires an adequately long record of hydrological data, including precipitation and runoff series, which, fortunately, are available in the area of interest. Through these hydrological sequences, the uncertainties of the occurrences of rainfall or flood or drought events are studied by means of stochastic models which are mathematical developments governed by probabilistic laws. Uncertain phenomena of different hydrological series can be represented by different simple stochastic models. Hydrological data are observed mostly at discrete times, forming the time series, or are recorded continuously along the time axis averaged over discrete time intervals, Δt, forming time series, which will be used repeatedly in this report.

I.2. A brief history of stochastic models in daily hydrological series

Stochastic models have been used to fit hydrological series since the early 1900s. Grant (1938) fitted a simple
Poisson model to the number of excessive rainfall occurrences in any single year in the Midwest. Gabriel and Neumann (1957) formulated a homogeneous, first-order Markov chain for the daily precipitation occurrences in Tel-Aviv by assuming that the probability of precipitations in any day depended on whether the previous day was wet or dry. Later Feyerherm et al. (1965) used a non-homogeneous first-order Markov chain to model the daily rainfalls in Indiana, while Todorovic and Yevjevich (1969) treated the daily rainfall occurrences with the non-homogeneous Poisson model at Fort Collins in Colorado. Smith and Schreiber (1973) tested the hypothesis of the sequential independence against a first-order Markov chain alternative for daily precipitation sequences in the southwestern U.S. concluded that the transition probabilities vary throughout the seasons of the year. Kavvas and Delleur (1975) investigated the precipitation series in Indiana by the analysis of trends and the homogenization of the daily precipitation counts and introduced the concept of the trigger model to the stochastic modelling of hydrological processes. Waymire and Gupta (1981) tried to generalize the trigger model to the spatial distribution of precipitation. Meanwhile Cervantes (1981) used and extended the trigger model to study the flood events, which were defined by daily streamflow sequences.
Many of the above studies required elaborate mathematical derivations which overshadowed the physical realities and often puzzled engineers and made them hesitant in using these models. Furthermore, the above studies on the occurrence of rainfalls or floods did not provide continuous time series of hydrological quantities. These continuous time series are extremely important in the study of the rainfall-runoff relation. Therefore, the following paragraphs will give a brief review of stochastic time series models used in hydrology.

The studies of the stochastic time series models in hydrology did not appear until the beginning of the 1960's. As pioneers, Matalas (1963) and Yevjevich (1963), respectively, used the moving average method to analyze annual runoff and annual precipitations and obtained somewhat reasonable results. About a decade later Box-Jenkins (1972) constructed the autoregressive integrated moving average model (ARIMA) and set up a simple procedure for model building of stochastic time series. Since then the ARIMA model has been very popular in hydrology. It has been used, for example, by Hipel (1975), O'Connell (1977), Delleur and Kavvas (1978). In general, annual hydrological time series are stationary and approximately normally distributed so that the ARIMA model works reasonably well in modelling. Monthly series can often be transformed so that they are approximately normally distributed, and the
periodicity in the mean and standard derivation can be removed by subtraction of the appropriate monthly mean and division by the corresponding standard deviation, as shown by Kavvas and Delleur (1975). Unfortunately the daily hydrological time series are far from being normally distributed that makes the fitting of ARIMA model usually unsuccessful. Furthermore daily precipitation series include long persistances of zero quantities, which are not properly simulated by ARIMA models. The study of the flood-drought requires the investigation of the characteristics of wet-dry run lengths, which the ARIMA model is unable to provide. In late 1970's Buishand (1978) analyzed wet-dry spells in the Netherlands by means of the Discrete Autoregressive Moving Average (DARMA) model, which was originally formulated by Jacobs and Lewis (1977). Buishand simply forced all the observed series to be fitted by one single model, which may not be the case for different climatological watersheds, or may not be the most suitable model in that area. Moreover no discussion was made on the distributions and statistical structure of run lengths which are very important in the investigation of the phenomena of the flood-drought under varied climatological conditions. In addition the previous work did not include any discussion of the nonzero sequences of precipitation, which is indispensable for the point of view of simulation. Also the fitting procedure and the estimation of parameters of the ARIMA model are no longer simple and new procedures need to
be constructed.

The shortcomings of the previous works on the stochastic modelling of daily hydrological series are summarized as follows:

1. The property of normal distribution, which is generally true for long-duration hydrological time series for which the ARIMA model is reasonably well fitted, no longer holds for daily hydrological sequences.

2. The persistence of long runs of zero quantities of hydrological sequences such as precipitations is not preserved by ARIMA models.

3. Inadequate procedure for the selection of the most suitable DARMA model when several competitive models are available.

4. Lack of specification of the distributions of run lengths for different DARMA models, as different watersheds may require different distributions because of climate variations.

5. Inability to interpret the complete characteristics of the flood-drought events in single model.

6. Lack of discussion of the non-zero precipitation series. These play important roles in expressing the intensities of precipitations that are indispensible for the
simulation of hydrological time series and fulfilling the relationship between precipitation and runoff.

7. Lack of the simplicity for engineering applications. Appropriate procedures of parameter estimations and of model building need to be developed.

I.3. Introduction and objectives of the report

Based on the awareness of the disadvantages of the previous studies, it is the goal of this report to develop three new stochastic models; two for the simulation of the daily precipitation sequences and one transfer model used to produce daily runoff series from the above two daily precipitation models. At the same time it is desired to preserve the spirit of the simplicity in ARIMA model and to develop the construction of an adequate procedure to select the best model for daily hydrological time series and also to interpret the properties of floods and droughts.

First a Binary-Discrete Autoregressive Moving Average-Mixed with Exponential model (B-DARMA-E(p,q)) will be analyzed in chapter 3, where p and q are the orders of the autoregressive and of the moving average components, respectively. The model can be simply expressed by

$$W_t = X_t Z_t $$

(1-1)

$\{X_t\}$ is a sequence of 0 and 1 and will be modelled by a
DARMA(p, q) process which is generally discussed in chapter
2. In the above definition, 0 means the quantity of the
precipitation in that day is less than 0.01" and 1 is
otherwise. \( \{Z_t\} \) is a sequence of precipitation quantities
whenever there is a wet day, denoted by 1 in \( \{X_t\} \), and \( \{Z_t\} \)
has an exponential distribution. The procedure of model
building and analysis are given in chapter 3, while results
and discussions are shown in chapter 6.

Secondly, it is shown that the precipitation quantity can
be classified into \( r \) discrete states. The new discrete
sequence of the precipitation can be fitted by a DARMA(p, q)
model whose general properties are discussed in chapter 2.
This second model is designated as M-DARMA(p, q; r), where \( p \)
is the order of the autoregressive components, \( q \) is the
order of the moving average components, and \( r \) is the number
of states in which the precipitation amounts are classified
and \( M \) stands for multistate. The procedures for the
estimation of parameters and model construction of the M-
DARMA(p, q; r) are discussed in chapter 4. After fitting the
daily precipitation data by M-DARMA(p, q; r) models, a general
discussion of the results is presented in chapter 6.

Finally with above two precipitation models as inputs, a
Transfer Autoregressive Moving Average model , T-
DARMA(p, q, m, n, 1), will be used to produce daily runoff
series. A T-DARMA(p, q, m, n, 1) model includes \( p \) orders of
autoregressive components and \( q \) orders of moving average
components in the input process, and \(m\) orders of the autoregressive components and \(n\) orders of the moving average component in the transfer process, and can be denoted by

\[
\phi(B)Q(k) = \Theta(B)R(k) - L_j
\]  \hspace{1cm} (1-2)

where

\[
\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 - \ldots - \phi_mB^n
\]  \hspace{1cm} (1-2a)

and

\[
\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3 - \ldots - \theta_n B^n
\]  \hspace{1cm} (1-2b)

\(Q(k)\): Runoff output at time \(k\),

\(R(k)\): Precipitation input at time \(k\)

\(L_j\): The loss in the \(j\)-th season.

The precipitation models, B-DARMA-E or M-DARMA with \(p\) and \(q\) autoregressive and moving average components, respectively, are used as input, \(R(k)\), to the T-DARMA model. The detailed analysis of the transfer model is given in chapter 5 and the results as well as their discussions are presented in chapter 6.

In addition to the formulation of two new stochastic models for the simulation of daily precipitations and of a new daily rainfall-runoff transfer model, the following secondary objectives are also part of this research:

1. By the newly constructed stochastic time series models, B-DARMA-E and M-DARMA this study will attempt to
overcome the difficulties existing in the ARIMA model for fitting daily precipitation time series.

2. A simple method of parameter estimation of the DARMA model will be formulated and shown efficient by fitting the models to the data of several stations located in Indiana. Simple procedures of model building for both B-DARMA-E and M-DARMA will be obtained and proper criteria for the selection of the best model are also to be presented.

3. The persistences of wet and dry runs in the B-DARMA-E model and of multiple runs in the M-DARMA model are to be preserved, where the multiple runs will be defined in chapter 2. The theoretical run length distributions of different competitive models will be derived for both B-DARMA-E and M-DARMA models and will serve as different criteria to check the preservation of the persistence in wet-dry day spells as well as in multiple day spells and will be moreover used to find the best fit.

4. The properties of floods and droughts will be conceptually interpreted through the comparison of the theoretical and estimated run length distributions.

5. The daily precipitation series simulated by the B-DARMA-E or the M-DARMA model will be used as inputs to a transfer function to obtain the daily runoff series through the T-DARMA system.
I.4. Research data

The daily precipitation series observed at 9 stations located over 5 watersheds in Indiana are used in the analysis of the B-DARMA-E and M-DARMA models. The station names and their Index numbers are listed in Table 1.1. They are also shown on Map 1.1. The daily streamflow series observed in 5 watersheds in Indiana area are analyzed and are used for precipitation-runoff transfer model. Table 1.1 gives the identification numbers of the watersheds and their drainage areas. The names and the identification numbers of different precipitation gage stations and the Thiessen polygon weights used to obtain the precipitation over the watershed are also listed in the same table. The precipitation data were obtained from the the National Weather Service(NOAA) and the runoff data were obtained from the U.S. Geological Survey. The locations of the stream gaging stations are shown in Map 1.1, where G3 corresponds to the USGS Identification number of gaging station, 327500; G5 to 3303000; G7 to 3324500; G8 to 3326500; G9 to 3328000.
Table 1.1 WATERSHED IDENTIFICATION

<table>
<thead>
<tr>
<th>Watershed Id. No.</th>
<th>Area (sq.mi.)</th>
<th>P.G.S.¹</th>
<th>Index No.</th>
<th>Name</th>
<th>T.F.W.²</th>
</tr>
</thead>
<tbody>
<tr>
<td>3275000³</td>
<td>529</td>
<td>12-1229</td>
<td>Cambridge</td>
<td>Richmond</td>
<td>0.722</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12-7362</td>
<td></td>
<td></td>
<td>0.278</td>
</tr>
<tr>
<td>3303000⁴</td>
<td>461</td>
<td>12-7747</td>
<td>Salomonia</td>
<td></td>
<td>1.000</td>
</tr>
<tr>
<td>3324500⁵</td>
<td>557</td>
<td>12-0676</td>
<td>Berne</td>
<td></td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12-0831</td>
<td>Bluffton</td>
<td></td>
<td>0.168</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12-5337</td>
<td>Marion College</td>
<td></td>
<td>0.335</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12-7747</td>
<td>Salomonia</td>
<td></td>
<td>0.465</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12-9138</td>
<td>Wabash</td>
<td></td>
<td>0.030</td>
</tr>
<tr>
<td>3326500⁶</td>
<td>682</td>
<td>12-5337</td>
<td>Marion College</td>
<td></td>
<td>0.337</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12-6018</td>
<td>Muncie</td>
<td></td>
<td>0.209</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12-7747</td>
<td>Salomonia</td>
<td></td>
<td>0.434</td>
</tr>
<tr>
<td>3328000⁷</td>
<td>417</td>
<td>12-1734</td>
<td>Columbia</td>
<td></td>
<td>1.000</td>
</tr>
</tbody>
</table>

¹: Precipitation Gage Station
²: Thiessen Polygon Weight
³: Whitewater River near Alpine, IN
⁴: Blue River near White Cloud, IN
⁵: Salamonie River at Dora, IN
⁶: Mississinewa River at Marion, IN
⁷: Eel River at North Manchester, IN.
Map. 1. Watersheds used for analysis
CHAPTER II

DISCRETE AUTOREGRESSIVE MOVING AVERAGE MODEL

II.1. Introduction

In this chapter a broad Discrete Autoregressive Moving Average (DARMA) process is introduced as a simple model for a stationary sequence of discrete random variables having a specified marginal distribution. The general definition of the DARMA(p,q) model is given and particular models, DAR(p), DMA(q), and DARMA(1,q), respectively, are discussed by Jacobs and Lewis (1978a, 1978b, 1978c). The DAR(p) process which is a particular p-th order Markov chain, has the physical structure and correlation function of an autoregressive process of order p. The process and its transition matrix are completely determined by the specified marginal distribution and by several model parameters that are independent of the marginal distribution. On the other hand the DMA(q) process which is a special q-th order moving average, has its correlation function governed by the q parameters of the model which are also independent of the specified marginal distribution. The DARMA(1,q) process is a model of particular interest in the autoregressive of the first order and the moving average of the q-th order.

Due to the desired parsimony of parameters in building a model and because of the characteristics of the regional
climatology, several DARMA submodels are especially useful and will be shown to be suitable for the area studied in this research. Therefore the second part in this chapter discusses in detail these models, namely: DARMA(0,1), DARMA(1,0), DARMA(1,1), and DARMA(1,2), to prepare the statistical background for the B-DARMA-E and the M-DARMA models in the following chapters. The definitions of these four submodels are given and their correlation structures, which play a role in the preliminary model fitting and in the first stage of parameter estimation, are derived and discussed. Some statistical properties such as their Markovian or non-Markovian behavior are noted for the purpose of discussions on their run length properties.

The last part of this chapter concentrates on the properties of run lengths, which are specifically defined here for the building of the B-DARMA-E and the M-DARMA as well as for the hydrological interpretations of floods and droughts. It is believed to be indispensible for irrigation systems and reservoir operations to obtain the knowledge of extreme wet and dry spells which could be physically explained through the properties of run lengths derived in this section. Therefore, in this part, the theoretical distributions of run lengths are derived for the four particular submodels. These distributions of run lengths will be used for comparisons with those estimated from observed sequences. Furthermore, the run length
characteristics will serve to select the best among the competitive models.

II.2. General DARMA models

(A) DARMA(p,q)

The DARMA(1,q), DARMA(p,0), and DARMA(0,q) families of models were originally mathematically developed by Jacobs and Lewis (1978). This section gives a new definition of the pool of the DARMA(p,q) processes in order to obtain the best model for the study of fitting the daily precipitation time series. After the investigations of the data in Indiana area, only a few models with low autoregressive and moving average orders are found to be useful. Therefore the next section concentrates on the discussions of these models. A DARMA(p,q) model is a sequence, \( \{X_n\} \), formed by a probabilistic combination of elements of a sequence, \( \{Y_n\} \), which is identically and independently distributed, where \( p \) is the order of autoregressive process and \( q \) is the order of the moving average. Let \( \{Y_n\} \) have a common probability distribution, \( \pi \), i.e.

\[
\pi(i) = P(Y_n = i), \quad i = 0, 1, 2, \ldots
\]  \hspace{1cm} (2-1)

Let \( \{U_n\} \) and \( \{V_n\} \) be independent sequences of independent random variables taking values only in \{0,1\} such that
\[ P(U_n=1) = \beta = 1 - P(U_n=0), \; 0 \leq \beta \leq 1 \]  \hspace{1cm} (2-2a)

\[ P(V_n=1) = \rho = 1 - P(V_n=0) \; 0 \leq \rho < 1 \]  \hspace{1cm} (2-2b)

Let

\( m_1 \): be an index independently taking values in \( \{0, 1, \ldots, q-1\} \)

\( m_2 \): be an index independently taking values in \( \{1, 2, 3, \ldots, p\} \)

Then the DARMA\((p,q)\) model can be defined by

\[ X_n = U_n Y_{n-m_1} + (1-U_n) A_{n-q} \]  \hspace{1cm} (2-3)

where

\[ A_n = V_n A_{n-m_2} + (1-V_n) Y_n . \]

Assuming that the process starts with \( A_{-q} \) which has the same probability distribution as \( \{Y_n\} \), but being independent of \( \{Y_n\} \), then the sequence of \( \{X_n\} \) is formed in such a way that its marginal distribution is the same as \( \{Y_n\} \) and behaves as a stationary sequence of dependent discrete random variables. This kind of process has the advantage that the marginal distribution and correlation structure are specified separately. In addition the DARMA process provides a simple scheme for obtaining models with which to analyze stationary sequences of dependent random variables and it is suitable for modelling both Markovian and non-Markovian sequences of discrete random variables. The family of DARMA\((1,q)\) processes was defined and discussed by Jacobs and Lewis\((1978a),(1978b),(1978c)\).

(B) DARMA\((p,0)\)
The DARMA(*p*,0) model is a particular sequence \{X_n\} with a \(p\)-th order Markovian dependence and with a given marginal distribution as specified in \{Y_n\}. It is thus a process which has the physical and statistical properties of a \(p\)-th order autoregressive process. Therefore the process depends explicitly on \(X_{n-1}, ..., X_{n-\text{\textup{p}}}\), and the \(p\) parameters, namely the \(p\) probabilities, \(\alpha_1, \alpha_2, ..., \alpha_p\) defined below. The model can be expressed by

\[ X_n = V_n X_{n-m_2} + (1-V_n) Y_n \]  \hspace{1cm} (2-4)

where \(V_n\) takes values in \{0,1\} independently with probabilities

\[ P(V_n=1) = \rho = 1 - P(V_n=0) \]  \hspace{1cm} (2-4a)

\(m_2\) takes values in \{1,2, ..., \text{p}\} independently with probabilities \(\alpha_1, \alpha_2, ..., \alpha_p\), i.e.

\[ P(m_2=j) = \alpha_j \geq 0 \text{ and } \alpha_1 + ... + \alpha_p = 1 \]  \hspace{1cm} (2-4b)

Assuming \(P(X_{-\text{\textup{p}}} = k) = \pi(k)\) and that the process starts at \(X_{-\text{\textup{p}}}\), then \{\(X_n\)\} has the same marginal distribution as \{\(Y_n\)\}. Moreover (2-4) shows that the dependence of the process results from the random selection procedure by which \{\(X_n\)\} chooses one element of the sequence \{\(Y_n\)\} so that \{\(X_n\)\} can be written as a random index model over \{\(Y_n\)\}. From (2-4) it is possible to derive the correlation function of the process which ends up with a set of Yule-Walker equations, which are stated in equations (2-5, 2-5a), the derivation of which is given in Appendix A.
\[
\text{Corr}(1) = \rho \alpha_1 \text{Corr}(0) + \rho \alpha_2 \text{Corr}(1) + \ldots + \rho \alpha_p \text{Corr}(p-1)
\]
\[
\text{Corr}(2) = \rho \alpha_1 \text{Corr}(1) + \rho \alpha_2 \text{Corr}(0) + \ldots + \rho \alpha_p \text{Corr}(p-2)
\]
\[
\vdots
\]
\[
\text{Corr}(p) = \rho \alpha_1 \text{Corr}(p-1) + \rho \alpha_2 \text{Corr}(p-2) + \ldots + \rho \alpha_p \text{Corr}(0)
\]  
for \( k \geq 1 
\]
\[
\text{Corr}(p+k) = \rho \alpha_1 \text{Corr}(p+k-1) + \ldots + \rho \alpha_p \text{Corr}(k) \quad (2-5a)
\]

(C) DARMA(0, q)

Let \( \{m_1\} \) take values only in \( \{0, 1, 2, \ldots, q-1\} \) with probabilities \( \beta_0, \beta_1, \beta_2, \ldots, \beta_{q-1}; \quad \beta_i \geq 0, \quad 0 \leq i \leq q-1; \beta_0 + \beta_1 + \beta_2 + \ldots + \beta_{q-1} = 1 \). \( \{Y_n\} \) is defined as a sequence of identically and independently distributed random variables with a common distribution \( \pi \) as before. Then a DARMA(0, q) model can be expressed by equation (2-6)

\[
X_n = Y_{n-m_1} \quad (2-6)
\]

From equation (2-6) it is obvious that \( \{X_n\} \) is simply the random selection of a sequence of random variables with a specified marginal distribution. Therefore the process of a DARMA(0, q) is a particular q-th order moving average model with q parameters governing its covariance structure. The derivation of the covariance function is demonstrated in Appendix A and its corresponding correlation structure is briefly given as follows:
For $1 \leq j \leq q - 1$,

$$\text{Corr}(j) = \sum_{k}^{q-1} \beta_k \beta_{k-j}, \ k \geq j \quad (2-7)$$

For $j > q - 1$,

$$\text{Corr}(j) = 0 \quad (2-7a)$$

It can be seen from (2-7) that the correlation coefficients of the process has a cut-off at the $q$-th time lag. This is a phenomenon similar to a regular moving average model MA($q$) in the Box-Jenkins (1976) terminology. It is an important property to be used in the preliminary identification of the models, the procedures of which will be described in the following chapters. Moreover, one of the advantages of this process, shown by Jacobs and Lewis (1978a), is that the model can still have more long-term dependence than a sequence of independent, identically distributed random variables which are used in renewal processes.

II.3. Properties of models of particular interest

This section gives the second moment properties of the models which are particularly useful in this research while the run-length characteristics are described in the next section.

(A) DARMA(0,1)

Let $\{X_n\}$ be a DARMA(0,1) process. Then it can be defined by
\[ X_n = \begin{cases} Y_{n-1} & \text{with probability } 1-\beta, \\ Y_n & \text{with probability } \beta, \end{cases} \quad (2-8) \]

where \( \{Y_n\} \sim \text{i.i.d. } \pi\text{-distribution.} \)

The probability distribution of \( \{X_n\} \) is the same as that of \( \{Y_n\} \) since \( \{X_n\} \) is just a random selection of \( \{Y_n\} \). Thus both \( \{X_n\} \) and \( \{Y_n\} \) are stationary sequences with the same marginal distribution and are not Markovian processes. From the derivation in Appendix A, the covariance structure can be written as follows:

\[
\begin{align*}
\text{Cov}(1) &= \beta(1-\beta) \text{Var}(Y_n) = \beta(1-\beta) \text{Var}(X_n) \quad (2-9) \\
\text{Cov}(k) &= 0, \text{ if } k > 1 \quad (2-9a)
\end{align*}
\]

Hence the correlation function for DARMA(0,1) can be expressed by

\[
\text{Corr}(k) = \begin{cases} 
\beta(1-\beta), & \text{if } k=1 \\
0, & \text{if } k>1 
\end{cases} \quad (2-10)
\]

This correlation function shows that there is a cut-off after the first lag. It is however noted that the DARMA(0,1) has the property of the time reversibility.

(B) DARMA(1,0)

A DARMA(1,0) is defined by a sequence of \( \{X_n\} \) such that

\[ X_n = \begin{cases} X_{n-1} & \text{with probability } \rho \\ Y_n & \text{with probability } 1-\rho \end{cases} \quad (2-11) \]

where \( \{Y_n\} \sim \text{i.i.d. } \text{with } \pi\text{-distribution} \)
Assuming the process begins at $X_{-1}$ having the same probability distribution as $\{Y_n\}$, i.e. $P(X_{-1}=k)=\pi(k)$, the marginal distribution of $\{X_n\}$ can be derived by induction as shown in Appendix A and is just the same as that of $\{Y_n\}$.

Since $\{X_n\}$ and $\{Y_n\}$ are stationary with the same mean, it is thus possible to perform the derivation of the covariance function for the process as follows:

For $k\geq 1$,

$$\text{Cov}(k)=\text{Cov}(X_n,X_{n-k})$$

$$=E(X_n X_{n-k})-E(X_n)E(X_{n-k})$$

$$=\rho E(X_{n-1} X_{n-k})+(1-\rho)E(Y_n X_{n-k})$$

$$-\rho E(X_{n-1})E(X_{n-k})-(1-\rho)E(Y_n)E(X_{n-k})$$

$$=\rho \{E(X_{n-1} X_{n-k})-E(X_{n-1})E(X_{n-k})\}$$

$$=\rho \text{Cov}(k-1) \quad (2-12)$$

By induction method the corresponding correlation function is obtained as:

$$\text{Corr}(k)=\rho^k, \quad k\geq 0 \quad (2-13)$$

$\{X_n\}$ is a Markov chain process of the first order and its transition probability is derived in the following;

$$P(i,j)=P(X_{k+1}=j|X_k=i)$$

$$=\rho P(X_{k+1}=j|X_k=i) + (1-\rho)P(Y_{k+1}=j|X_k=i)$$

$$=\rho \delta_{ij} + (1-\rho)\pi_j$$

where

$$\delta_{ij}=\begin{cases} 1, & \text{if } i=j \\ 0, & \text{if } i\neq j. \end{cases}$$
thus,  \[ P(i,j) = \begin{cases} 
\rho + (1-\rho)\pi_i, & \text{if } i=j \\
(1-\rho)\pi_j, & \text{if } i\neq j 
\end{cases} \] (2-14)

It is noted that the process has a Markovian dependency structure specified by a single parameter \( \rho \) and by a specified stationary distribution \( \pi \) associated with the chain. Furthermore, changing \( \rho \) does not affect the distribution \( \pi \). These properties simplify many calculations, such as run length characteristics, which will be discussed in the following section.

(C) DARMA(1,1)

Let \( \{Y_n\} \) be a sequence of independent random variables with an identical \( \pi \)-distribution. Then \( \{X_n\} \) defined in (2-15) is a DARMA(1,1) process:

\[ X_n = \begin{cases} 
Y_n, \text{ with probability } \beta, \\
A_{n-1}, \text{ with probability } 1-\beta. 
\end{cases} \] (2-15)

where

\[ A_n = \begin{cases} 
A_{n-1}, \text{ with probability } \rho, \\
Y_n, \text{ with probability } 1-\rho. 
\end{cases} \] (2-15a)

Assuming that the process starts with \( A_{-1} \) which has the same distribution as \( \{Y_n\} \) and is independent of \( \{Y_n\} \), then the marginal distribution of \( \{X_n\} \) is also specified by \( \pi \), as derived in detail in Appendix A. Therefore, \( \{X_n\}, \{A_n\} \), and \( \{Y_n\} \) are stationary with the same mean. This fact simplifies the derivation of the covariance structure of the
process, details of which are given in Appendix A, and which is expressed as (2-16).

\[ \text{Cov}(k) = (1-\beta)(\rho+\beta-2\rho\beta)\rho^{k-1}\text{Var}(Y), \text{ if } k \geq 1. \] (2-16)

The corresponding correlation function is

\[ \text{Corr}(k) = c \rho^{k-1} \] (2-17)

where

\[ c = (1-\beta)(\rho+\beta-2\rho\beta) \] (2-17a)

\{X_n\} is itself not Markovian, but \{X_n, A_n\} forms a bivariate first order Markov chain. If the \( \pi \)-distribution is assumed to be a Bernouilli distribution, which is the case used for the B-DARMA-E models, then the transition probabilities are \( 2 \times 2 \) matrices. On the other hand if a \( r \)-dimensional multistate trial is given for \( \pi \), then the transition probabilities are \( r \times r \) matrices, and only a \( 3 \)-dimensional multistate trial is considered in this study. Since these properties are very important in the discussion of the run-length characteristics, they will be delayed until the run-length definitions are given in next section.

(D) DARMA(1,2)

If the sequence \{Y_n\} is the same as defined before, then a DARMA(1,2) can be defined by

\[ X_{n+1} = \begin{cases} Y_n, & \text{with probability } \gamma_1, \\
Y_1, & \text{with probability } \gamma_2, \\
A_{n-1}, & \text{with probability } 1-\gamma_1-\gamma_2. \end{cases} \] (2-18)
where

\[ Y_n, \text{ with probability } 1-\rho, \]
\[ A_n = \{ \]
\[ A_{n-1}, \text{ with probability } \rho. \]

(2-18a)

and

\[ 0 < \gamma_1 < 1, \quad 0 < \gamma_2 < 1, \quad 0 < \rho < 1, \quad \gamma_1 + \gamma_2 \geq 1. \]

It can be proved in exactly the same way as in the DARMA(1,1) case that \( \{X_n\} \), \( \{A_n\} \), and \( \{Y_n\} \) have the same marginal distribution and mean. This result is convenient for the derivation of the covariance structure of the process. The detailed derivations are given in Appendix A and the corresponding correlation function is expressed as follows:

if \( k = 1 \),

\[ \text{Corr}(1) = (1-\gamma_1-\gamma_2) \{ \gamma_2 (1-\rho) + (1-\gamma_1-\gamma_2) \rho \} + \gamma_1 \gamma_2 \]

(2-19)

if \( k \geq 2 \)

\[ \text{Corr}(k) = (1-\gamma_1-\gamma_2) \rho^{k-2} \{ (1-\gamma_1-2\gamma_2) \rho^2 - (\gamma_1-\gamma_2) \rho + \gamma_1 \} \]

(2-19a)

As in the DARMA(1,1) process, \( \{X_n\} \) is not a Markov chain. However \( \{X_{n+1}, A_n\} \) can be considered as a bivariate Markov chain. The transition probability matrices of this Markov chain will be derived for the binary DARMA model and for the multi-state DARMA model, respectively, in the next section, where the former is used for the B-DARMA-E process and the latter is used for the M-DARMA process.
II.4. Run lengths of B-DARMA and M-DARMA models

There have been several different definitions of the run lengths, both in hydrology (Yevjevich, 1972) and in mathematics (Buishand, 1979). The former was concerned with practical discussions of hydrological phenomena, while the latter used run lengths for the estimation of parameters in the DARMA(1,1) process. The sequence of dry days and wet days, or of flood and no flood stages, or of drought or no drought stages, are of particular interest and therefore binary DARMA processes are of great utility. If instead of considering the presence or absence of hydrologic variables, a number of levels or stages of those variables are of interest, then multi-state DARMA processes are useful. Accordingly, definitions of run lengths for both the Binary DARMA (B-DARMA) and the Multi-state DARMA (M-DARMA) models are given in this section. These definitions not only serve mathematical purposes, but can also be used to reflect hydrological effects such as the durations of floods and of droughts, which are of great engineering importance. Thus this section is divided into two parts; first, the run lengths for the four models mentioned in Section 3 are defined for the purpose of B-DARMA processes; secondly, the run lengths of the M-DARMA processes are defined and the probability distributions of these run lengths are derived.

(A) Binary run length
In the Binary Discrete Autoregressive Moving Average model, which is designated as B-DARMA in this paper, \( \{Y_n\} \) is assumed to take only \( \{0,1\} \) with probabilities \( \pi_0 \) and \( \pi_1 \), where
\[
\pi_0 + \pi_1 = 1 \tag{2-20}
\]
From the derivations shown in Section 2, \( \{X_n\} \) and \( \{Y_n\} \) have the same marginal distribution. This is, of course, true for the four particular models discussed in Section 4. Thus the sequence of \( \{X_n\} \) is simply a sequence consisting only of \( \{0,1\} \). Therefore, the run lengths in a B-DARMA include the run length of 0, designated by \( T_0 \), and the run length of 1, denoted by \( T_1 \). The mathematical definitions of the run lengths are as follows:

\[
(T_0 = n) = \{X_0 = 1, X_1 = 0, \ldots, X_n = 0, X_{n+1} = 1\} \tag{2-21}
\]
\[
(T_1 = n) = \{X_0 = 0, X_1 = 1, \ldots, X_n = 1, X_{n+1} = 0\}. \tag{2-21a}
\]

Then the probability distributions of run lengths can be obtained from the above definitions as follows:

for \( n = 1, 2, \ldots, \infty \)

\[
\begin{align*}
P(T_0 = n) &= P\{X_0 = 1, X_1 = 0, \ldots, X_n = 0, X_{n+1} = 1\} | X_0 = 1, X_1 = 0 \\
&= P\{X_0 = 1, X_1 = 0, \ldots, X_n = 0, X_{n+1} = 1\} / P\{X_0 = 1, X_1 = 0\} \tag{2-22}
\end{align*}
\]
similarly,

\[
P(T_1 = n) = P\{X_0 = 0, X_1 = 1, \ldots, X_n = 1, X_{n+1} = 0\} / P\{X_0 = 0, X_1 = 1\}. \tag{2-22a}
\]

In the following the probability distributions of binary run lengths for the four special DARMA models are briefly given and their detailed derivations are performed in Appendix B.
(i) Binary run length of DARMA(0,1)

Define

\[ a_0 = P(X_0 = 1) \]
\[ a_1 = P(X_0 = 1, X_1 = 0) \]
\[ a_2 = P(X_0 = 1, X_1 = 0, X_2 = 0) \]
\[ \vdots \]
\[ a_n = P(X_0 = 1, X_1 = 0, \ldots, X_n = 0) \]
\[ \vdots \]

Since \( \{X_n\} \) takes only \( \{0,1\} \),

\[ P(X_0 = 1, X_1 = 0, \ldots, X_n = 0, X_{n+1} = 1) \]
\[ = P(X_0 = 1, X_1 = 0, \ldots, X_n = 0) \]
\[ - P(X_0 = 1, X_1 = 0, \ldots, X_n = 0, X_{n+1} = 0) \]
\[ = a_n - a_{n+1} \quad (2-23) \]

From equations (2-22) and (2-23), the probability distribution of the run length, \( T_0 \), can be expressed by

\[ P(T_0 = n) = (a_n - a_{n+1}) / a_1 \quad (2-24) \]

By means of the definitions (2-6) and (2-20), it is possible to obtain expression (2-25) for \( \{a_n\} \). The derivation is shown in Appendix B. Equation (2-25) is a general equation for \( a_n \) which is convenient for programming in a digital computer.

For \( n \geq 1 \)

\[ a_n = \beta \pi_0 a_{n-1} + \beta (1-\beta) \pi_0 a_{n-2} + \ldots \]
\[ + \beta (1-\beta)^{n-1} \pi_0 a_0 + (1-\beta)^{n+1} \pi_0 \pi_1 \quad (2-25) \]

and

\[ a_0 = \pi_1 \quad (2-25a) \]
In a similar way, a sequence of \( \{b_n\} \) can be defined and the probability distribution of the run length, \( T_1 \), is expressed in terms of \( b_n \) as

\[
P(T_1 = n) = (b_n - b_{n+1}) / b_1
\]  \hspace{1cm} (2-26)

where,

\[
b_1 = P(X_0 = 0)
\]
\[
b_1 = P(X_0 = 0, X_1 = 1)
\]
\[
b_2 = P(X_0 = 0, X_1 = 1, X_2 = 1)
\]
\[
\ldots
\]
\[
b_n = P(X_0 = 0, X_1 = 1, \ldots, X_n = 1)
\]

The expression of \( \{b_n\} \) can also be derived in the same way as for the \( \{a_n\} \). Further details can be found in Appendix B.

(ii) Binary run length of DARMA(1,0)

A binary DARMA(1,0) is a Markov chain and it has two states, 0 and 1, if \( \{Y_n\} \) is assumed to take only \( \{0, 1\} \). From equation (2-14) the transition probabilities which are used for the discussion of the run length can be given as follows:

\[
P(0, 0) = \rho + (1-\rho)\pi_0
\]  \hspace{1cm} (2-27)
\[
P(1, 1) = \rho + (1-\rho)\pi_1 \hspace{1cm} (2-27a)
\]
The probability distributions of run lengths, \( T_0 \) and \( T_1 \), are obtained from equations (2-22) and (2-22a) as

\[
P(T_0=n) = P(X_{n-1}|X_n=0)P(X_n=0|X_{n-1}=0) \cdots P(X_2=0|X_1=0) = \prod_{i=0}^{n-1} (1-p(i,0)) .
\]

(2-28)

Similarly,

\[
P(T_1=n) = \prod_{i=1}^{n-1} (1-p(i,1)) .
\]

(2-28a)

(iii) Binary run length of DARMA(1,1)

From (2-15) a B-DARMA(1,1) sequence can form a bivariate Markov chain, \( \{X_n, A_n\} \), whose transition probabilities are defined by,

\[
H_k(i,j) = P(X_{n+1}=k, A_{n+1}=j | A_n=i, X_n=m) \quad (2-29)
\]

where

\[
i, j, k, m \text{ take values in } \{0,1\}
\]

And equation (2-29) becomes

\[
H_k(i,j) = \prod_{i=1}^{n-1} (1-p(i,1)) .
\]

(2-29a)

Appendix B provides the detailed derivations of the transition probability matrices of the binary DARMA(1,1) process, which are denoted by \( H_0 \) and \( H_1 \) and are briefly given in (2-30) and (2-30a)

\[
H_0 = \begin{bmatrix}
\rho(1-\beta)+(1-\rho)(1-\beta) & (1-\beta)(1-\rho)
\end{bmatrix} \begin{bmatrix}
\pi_0 \\
\beta(1-\rho)\pi_0
\end{bmatrix} = \begin{bmatrix}
\rho \pi_0 \\
\beta \pi_0
\end{bmatrix} \quad (2-30)
\]
Let $H_{k(n)}(i,j)$ be the $n$-step transition probability matrix of a binary DARMA(1,1) process, where $i,j,k$ take values in $\{0,1\}$.

Then,

$$P(X_0=1, X_1=0, \ldots, X_n=0) = P(X_0=1, A_0=0) \{H_{0(n)}(0,0)+H_{0(n)}(0,1)\}$$
$$+ P(X_0=1, A_0=1) \{H_{0(n)}(1,0)+H_{0(n)}(1,1)\}$$

(2-31)

Let

$$H_{0(n)}(0,0)=H_{0(n)}(0,0)+H_{0(n)}(0,1)$$

(2-32)

$$H_{0(n)}(1,0)=H_{0(n)}(1,0)+H_{0(n)}(1,1)$$

(2-32a)

Therefore, (2-31) becomes

$$P(X_0=1, X_1=0, \ldots, X_n=0) = P(X_0=1, A_0=0)H_{0(n)}(0,e)$$
$$+ P(X_0=1, A_0=1)H_{0(n)}(1,e).$$

(2-33)

From (2-22) and (2-23) the probability distribution of the run length of 0 is obtained as follows:

$$P(T_0=n) = \{P(X_0=1, A_0=0) \{H_{0(n)}(0,e)-H_{0(n+1)}(0,e)\}$$
$$+ P(X_0=1, A_0=1) \{H_{0(n)}(1,e)-H_{0(n+1)}(1,e)\}\}$$
$$P(X_1=0, X_2=0).$$

(2-34)
When it is assumed that the process starts at $A_{-1}$, then

$$
P(X_0=1, A_0=0) = P(X_0=1, A_0=0 | A_{-1}=0) P(A_{-1}=0) + P(X_0=1, A_0=1 | A_{-1}=1) P(A_{-1}=1)
$$

$$
= H_1(0,0) \pi_0 + H_1(1,0) \pi_1.
$$

(2-34a)

In a similar way

$$
P(X_0=1, A_0=1) = H_1(0,1) \pi_0 + H_1(1,1) \pi_1.
$$

(2-34b)

Since $\{X_n\}$ has the same marginal distribution as $\{Y_n\}$,

$$
P(X_0=1, X_1=0) = P(X_1=0 | X_0=1) P(X_0=1) = \pi_0 \pi_1.
$$

(2-34c)

For the derivation of the probability distribution of the run length of 1, the same procedure can be followed with the definitions

$$
H_{1(n)}(0,e) = H_{1(n)}(0,0) + H_{1(n)}(0,1)
$$

(2-35)

$$
H_{1(n)}(1,e) = H_{1(n)}(1,0) + H_{1(n)}(1,1)
$$

(2-35a)

Therefore the probability distribution of $T_1$ can be expressed by

$$
P(T_1=n) = \{ P(X_0=0, A_0=0) H_{1(n)}(0,e) - H_{1(n+1)}(0,e) \\
+ P(X_0=0, A_0=1) H_{1(n)}(1,e) - H_{1(n+1)}(1,e) \} / P(X_0=0, A_0=1)
$$

(2-36)
where with the assumption that the process starts at $A_{-1}$,

\[
P(X_0=0, A_0=0) = P(X_0=0, A_0=0 | A_{-1}=0)P(A_{-1}=0) \\
+ P(X_0=0, A_0=0 | A_{-1}=1)P(A_{-1}=1) \\
= H_0(0,0)\pi_0 + H_0(1,0)\pi_1 \quad (2-36a)
\]

\[
P(X_0=0, A_0=1) = H_0(0,1)\pi_0 + H_0(1,1)\pi_1 . \quad (2-36b)
\]

(iv) Binary run length of DARMA(1,2)

It is possible from the definition of the DARMA(1,2) in (2-18) to form $(X_{n+1}, A_n)$ as a bivariate Markov process since

\[
P(X_{n+1}=1, A_n=j | X_n=k, A_{n-1}=i) = P(X_{n+1}=1, A_n=j | A_{n-1}=i). \quad (2-37)
\]

where $i, j, k, l$ take values in $\{0,1\}$.

Let $H_1(i,j)$ be the transition probability matrix of the process, i.e.

\[
H_1(i,j) = P(X_{n+1}=1, A_n=j | A_{n-1}=i) \quad (2-38)
\]

since $(X_{n+1}=1, A_n=j)$ is independent of $X_n$.

Like the case in binary DARMA(1,1), there are two transition probability matrices, $H_0$ and $H_1$, defined for binary DARMA(1,2) since $\{X_n\}$ takes values in $\{0,1\}$. The reader may be referred to Appendix B for the further derivations of $H_0$ and $H_1$. In the following $H_0$ and $H_1$ are given for the DARMA(1,2) model as follows,
\[ H_0 = \{ \begin{array}{l}
(1-\gamma_1-\gamma_2)\rho + \gamma_1(1-\rho)\pi_0^2 + (1-\rho-\gamma_1+2\gamma_1\rho+\gamma_2\rho)\pi_0 + (1-\gamma_1-\gamma_2)^2 \gamma_1(1-\rho)\pi_0\pi_1 + (1-\gamma_1-\gamma_2)^2 \gamma_1(1-\rho)\pi_0^2 \\
\gamma_1(1-\rho)\pi_0^2 + \gamma_2(1-\rho)\pi_0 \\
(1-\gamma_1-\gamma_2)^2 \gamma_1(1-\rho)\pi_0\pi_1 \\
\gamma_1(1-\rho)\pi_0\pi_1 \\
\gamma_1(1-\rho)\pi_0\pi_1 \\
\end{array} \} \quad (2-39) \]

\[ H_1 = \{ \begin{array}{l}
\rho(\gamma_1+\gamma_2) + \gamma_1(1-\rho)\pi_0\pi_1 + (1-\gamma_1-\gamma_2)^2 \gamma_1(1-\rho)\pi_0\pi_1^2 \\
\gamma_2(1-\rho)\pi_1 + \gamma_1(1-\rho)\pi_1^2 \\
\gamma_1(1-\rho)\pi_0\pi_1 \\
(1-\gamma_1-\gamma_2)^2 \rho + \gamma_1(1-\rho)\pi_1^2 \\
(1-\rho-\gamma_1+2\gamma_1\rho+\gamma_2\rho)\pi_1 \\
\end{array} \} \quad (2-39a) \]

Let the n-step transition probability matrix be defined as in binary DARMA(1,1) except that \( H_0 \) and \( H_1 \) are obtained from (2-39) and (2-39a). Then the probability distributions of the run lengths for the BDARMA(1,2) process can be formulated as follows:

\[
P(T_0=n) = \{ P(X_0=1, A_1=0) (H_0(n)(0,e)-H_0(n+1)(0,e)) + P(X_0=1, A_1=1) (H_0(n)(1,e)-H_0(n+1)(1,e)) \}
\]

\[
P(X_0=1, X_1=0) \quad (2-40) \]

If the process is assumed to start at \( A_{-2} \), then

\[
P(X_0=1, A_{-1}=0) = P(X_0=1, A_{-1}=0|A_{-2}=0)P(A_{-2}=0)
\]

\[
+ P(X_0=1, A_{-1}=0|A_{-2}=1)P(A_{-2}=1)
\]

\[
= H_1(0,0)\pi_0 + H_1(1,0)\pi_1 \quad (2-40a)
\]

similarly,

\[
P(X_0=1, A_{-1}=1) = H_1(0,1)\pi_0 + H_1(1,1)\pi_1 \quad (2-40b)
\]

\[
P(X_0=1, X_1=0) = \pi_0\pi_1 \quad (2-40c)
\]

In exactly the same way shown above, the probability distribution of the run length, \( T_1 \), is obtained as
\[ P(T_1=n) = \{ P(X_0 = 0, A_{-1} = 0) \left( H_1(n, 0, e) - H_1(n+1, 0, e) \right) + P(X_0 = 0, A_{-1} = 1) \left( H_1(n, 1, e) - H_1(n+1, 1, e) \right) \} / P(X_0 = 0, X_1 = 1) \]  
\[ (2-41) \]

Assuming the process starts at \( A_{-2} \), then

\[ P(X_0 = 0, A_{-1} = 0) = P(X_0 = 0, A_{-1} = 0 | A_{-2} = 0) P(A_{-2} = 0) + P(X_0 = 0, A_{-1} = 0 | A_{-2} = 1) P(A_{-2} = 1) \]

\[ = H_0(0, 0) \pi_0 + H_0(1, 0) \pi_1 \]  
\[ (2-41a) \]

and,

\[ P(X_0 = 0, A_{-1} = 1) = H_0(0, 1) \pi_0 + H_0(1, 1) \pi_1. \]  
\[ (2-41b) \]

(B) Multi-state run length

Let \( \{ Y_n \} \) take values only in \( \{ 0, 1, 2, \ldots, m-1 \} \) with probabilities, \( \pi_0, \pi_1, \pi_2, \ldots, \pi_{m-1} \), such that

\[ \pi_0 + \pi_1 + \pi_2 + \ldots + \pi_{m-1} = 1 \]  
\[ (2-42) \]

Since \( \{ X_n \} \) has the same distribution as \( \{ Y_n \} \), it can be seen that \( \{ X_n \} \) has \( m \) discrete states, which are used to classify the daily precipitation quantity to form a new precipitation sequence that can be adequately modeled by the M-DARMA process in Chapter 4. For the purpose of model building in the M-DARMA process, a multi-state run length of \( i \), denoted by \( T_i \), is defined as follow:

\[ \{ T_i = n \} = \{ X_0 \neq i, X_1 = i, \ldots, X_n = i, X_{n+1} \neq i | X_0 \neq i, X_i = i \} \]  
\[ (2-43) \]

where \( i = 0, 1, 2, \ldots, m-1. \)
Therefore the probability distribution of $T_i$ is expressed by

$$
P(T_i=n) = P(X_0 \neq i, X_1 = i, \ldots, X_n = i, X_{n+1} \neq i | X_0 = i, X_1 = i)$$

$$= P(X_0 \neq i, X_1 = i, \ldots, X_n = i, X_{n+1} \neq i) / P(X_0 = i, X_1 = i)$$  \hspace{1cm} (2-44)

For the convenience of derivations in the following. (2-44) can be represented by two equivalent expressions, (2-44a) and (2-44b) as follows:

$$P(T_i = n) = \{ P(X_0 = i, X_1 = i, \ldots, X_n = i) - P(X_0 = i, X_1 = i, \ldots, X_{n+1} = i) \} / P(X_0 = i, X_1 = i)$$  \hspace{1cm} (2-44a)

and,

$$P(T_i = n) = P(X_{n+1} \neq i | X_n = i) P(X_n = i | X_{n-1} = i) \ldots P(X_2 = i | X_1 = i)$$  \hspace{1cm} (2-44b)

where

$$i = 0, 1, 2, \ldots, m-1 ; \quad n = 1, 2, 3, \ldots$$

In order to illustrate the probability distribution of the multi-state run lengths for the four particularly useful models, $m$ is assumed to equal to 3 in the following discussions.

(i) Multi-state run length of DARMA$(0,1)$

Let

$$c_0 = P(X_0 \neq i)$$

$$c_1 = P(X_0 \neq i, X_1 = i)$$

$$c_2 = P(X_0 \neq i, X_1 = i, X_2 = i)$$

$$\ldots$$

$$c_n = P(X_0 \neq i, X_1 = i, \ldots, X_n = i)$$
By the definitions in (2-6) and (2-42), the \( \{c_n\} \) can be expressed by (2-45) whose derivation is shown in Appendix B, as

for \( n \geq 1 \),

\[
c_n = \beta \pi_1 c_{n-1} + \beta (1-\beta) \pi_1 c_{n-2} + \ldots + \beta(1-\beta) \pi_1 c_1 + (1-\beta) \pi_1 (1-\pi_1)
\]

and it is assumed that

\[
c_0 = 1 - \pi_1
\]

(ii) Multi-state run length of DARMA(1,0)

Since \( m = 3 \), \( \{Y\} \) takes values in \( \{0, 1, 2\} \) with probabilities \( \pi_0, \pi_1, \) and \( \pi_2 \), where \( \pi_0 + \pi_1 + \pi_2 = 1 \). This results in that the DARMA(1,0) process becomes a Markov process such that \( \{V_n\} \) has a \( 3 \times 3 \) transition probability matrix as shown in (2-14). The particular probabilities of interest for the run length discussions are \( P_{00}, P_{11}, \) and \( P_{22} \), where

\[
P_{00} = \rho + (1-\rho)\pi_0
\]

\[
P_{11} = \rho + (1-\rho)\pi_1
\]

\[
P_{22} = \rho + (1-\rho)\pi_2
\]

From (2-44b) the probability distribution of the run length of 0 is obtained as

\[
P(T_0 = n) = P_{00} (1-P_{00})^n
\]

(2-43)
Similarly,

$$P(T_1=n) = p_{11}^{n-1} (1-p_{11}) \quad (2-48a)$$

$$P(T_2=n) = p_{22}^{n-1} (1-p_{22}) \quad (2-48b)$$

and the general case with m states is

$$P(T_i=n) = p_{ii}^{n-1} (1-p_{ii}) \quad (2-49)$$

where $i=0,1,2,\ldots,m-1$,

$n=1,2,3,\ldots,\infty$.

(iii) Multi-state run length of DARMA(1,1)

Based on the same assumption that $\{Y_n\}$ takes values in
{0,1,2}, the bivariate Markov process derived in (2-29) has
its transition probability matrices $H_0,H_1,$ and $H_2$ shown in
(2-50), (2-50a), and (2-50b) below. They are 3x3 matrices
whose derivations are given in Appendix B.

$$H_0 = \begin{pmatrix}
  \rho(1-\beta) & (1-\beta)(1-\beta)\pi_1 & (1-\beta)(1-\rho)\pi_2 \\
  + (1-\rho(1-\beta))\pi_0 & \beta(1-\rho)\pi_0 & \beta\rho\pi_0 \\
  \beta(1-\rho)\pi_0 & 0 & \beta(1-\rho)\pi_0
\end{pmatrix} \quad (2-50)$$

$$H_1 = \begin{pmatrix}
  \beta(1-\rho)\pi_0 & \rho(1-\beta) & (1-\beta)(1-\rho)\pi_2 \\
  + (1-\rho(1-\beta))\pi_1 & \beta(1-\rho)\pi_1 & 0 \\
  \beta(1-\rho)\pi_1 & 0 & \beta\rho\pi_1
\end{pmatrix} \quad (2-50a)$$
\[
H_2 = \begin{cases} 
\beta \rho \pi_2 & 0 & \beta (1-\rho) \pi_2 \\
0 & \beta \rho \pi_2 & \beta (1-\rho) \pi_2 \\
(1-\beta)(1-\rho)\pi_0 & (1-\beta)(1-\rho)\pi_1 & \rho (1-\beta) + (1-\rho(1-\beta)) \pi_2 
\end{cases} \tag{2-50b}
\]

Since \( m = 3 \),

\[
P(X_0 \neq 0, X_1 = 0, \ldots, X_n = 0) \\
= P(A_0 = 0, X_0 \neq 0) \{ H_0(n)(0,0) + H_0(n)(0,1) + H_0(n)(0,2) \} \\
+ P(A_0 = 1, X_0 \neq 0) \{ H_0(n)(1,0) + H_0(n)(1,1) + H_0(n)(1,2) \} \\
+ P(A_0 = 2, X_0 \neq 0) \{ H_0(n)(2,0) + H_0(n)(2,1) + H_0(n)(2,2) \}. \tag{2-51}
\]

From (2-44a) and (2-51), the probability distribution of \( T_0 \) is obtained as,

\[
P(T_0 = n) = \{ P(X_0 \neq 0, A_0 = 0)(H_0(n)(0,e) - H_0(n+1)(0,e)) \\
+ P(X_0 \neq 0, A_0 = 1)(H_0(n)(1,e) - H_0(n+1)(1,e)) \\
+ P(X_0 \neq 0, A_0 = 2)(H_0(n)(2,e) - H_0(n+1)(2,e)) \} / P(X_0 = 0, X_1 = 0), \tag{2-52}
\]

where

\[
H_0(n)(0,e) = H_0(n)(0,0) + H_0(n)(0,1) + H_0(n)(0,2) \tag{2-53}
\]

\[
H_0(n)(1,e) = H_0(n)(1,0) + H_0(n)(1,1) + H_0(n)(1,2) \tag{2-53a}
\]

\[
H_0(n)(2,e) = H_0(n)(2,0) + H_0(n)(2,1) + H_0(n)(2,2) \tag{2-53a}
\]

Assuming that the process starts at \( A_{-1} \),

\[
P(X_0 \neq 0, A_0 = 0) = P(X_0 \neq 0, A_0 = 0 | A_{-1} = 0)P(A_{-1} = 0) \\
+ P(X_0 \neq 0, A_0 = 0 | A_{-1} = 1)P(A_{-1} = 1) \\
+ P(X_0 \neq 0, A_0 = 0 | A_{-1} = 2)P(A_{-1} = 2)
\]

\[
= \{ H_1(0,0) + H_2(0,0) \} \pi_0 \\
+ \{ H_1(1,0) + H_2(1,0) \} \pi_1 \\
+ \{ H_1(2,0) + H_2(2,0) \} \pi_2 \tag{2-54a}
\]
Similarly,

\[ P(X_0=0, A_0=1) = \{H_1(0,1) + H_2(0,1)\} \pi_0 + \{H_1(1,1) + H_2(1,1)\} \pi_1 + \{H_1(2,1) + H_2(2,1)\} \pi_2 \]  
\[ (2-54b) \]

\[ P(X_0=0, A_0=2) = \{H_1(0,2) + H_2(0,2)\} \pi_0 + \{H_1(1,2) + H_2(1,2)\} \pi_1 + \{H_1(2,2) + H_2(2,2)\} \pi_2 \]  
\[ (2-54c) \]

and,

\[ P(X_0=0, X_1=0) = \pi_0 (1-\pi_0) \]  
\[ (2-55) \]

Let the general form of (2-53) be

\[ H_i(n)(j,e) = H_i(n)(j,0) + H_i(n)(j,1) + H_i(n)(j,2) \]  
\[ (2-56) \]

where \( i = 0, 1, 2 \); \( j = 0, 1, 2 \).

Then in exactly the same way as before, the probability distributions of \( T_1 \) and \( T_2 \) can be obtained as follows:

\[ P(T_1=n) = \{P(X_0=1, A_0=0)H_1(n)(0,e) - H_1(n+1)(0,e)\} + P(X_0=1, A_0=1)H_1(n)(1,e) - H_1(n+1)(1,e)\} + P(X_0=1, A_0=2)H_1(n)(2,e) - H_1(n+1)(2,e)\} / P(X_0=1, X_1=1) \]  
\[ (2-57) \]

where

\[ P(X_0=1, A_0=0) = \{H_0(0,0) + H_2(0,0)\} \pi_0 + \{H_0(1,0) + H_2(1,0)\} \pi_1 + \{H_0(2,0) + H_2(2,0)\} \pi_2 \]  
\[ (2-57a) \]

\[ P(X_0=1, A_0=1) = \{H_0(0,1) + H_2(0,1)\} \pi_0 + \{H_0(1,1) + H_2(1,1)\} \pi_1 + \{H_0(2,1) + H_2(2,1)\} \pi_2 \]  
\[ (2-57b) \]
\[ P(X_0 \neq 1, A_0 = 2) = \{ H_0(0, 2) + H_2(0, 2) \} \pi_0 + \{ H_0(1, 2) + H_2(1, 2) \} \pi_1 + \{ H_0(2, 2) + H_2(2, 2) \} \pi_2 , \]  \hspace{1cm} (2-57c) \\
\[ P(X_0 \neq 1, X_1 = 1) = \pi_1 (1 - \pi_1) , \]  \hspace{1cm} (2-57d) \\
and \\
\[ P(T_2 = n) = \{ P(X_0 \neq 2, A_0 = 0) (H_2(n), 0, e0 - H_2(n + 1)(0, e)) + P(X_0 \neq 2, A_0 = 1) (H_2(n), 1, e0 - H_2(n + 1)(1, e)) + P(X_0 \neq 2, A_0 = 2) (H_2(n), 2, e0 - H_2(n + 1)(2, e)) / P(X_0 \neq 2, X_1 = 2) \} , \]  \hspace{1cm} (2-58) \\
where \\
\[ P(X_0 \neq 2, A_0 = 0) = \{ H_0(0, 0) + H_1(0, 0) \} \pi_0 + \{ H_0(1, 0) + H_1(1, 0) \} \pi_1 + \{ H_0(2, 0) + H_1(2, 0) \} \pi_2 , \]  \hspace{1cm} (2-58a) \\
\[ P(X_0 \neq 2, A_0 = 1) = \{ H_0(0, 1) + H_1(0, 1) \} \pi_0 + \{ H_0(1, 1) + H_1(1, 1) \} \pi_1 + \{ H_0(2, 1) + H_1(2, 1) \} \pi_2 , \]  \hspace{1cm} (2-58b) \\
\[ P(X_0 \neq 2, A_0 = 2) = \{ H_0(0, 2) + H_1(0, 2) \} \pi_0 + \{ H_0(1, 2) + H_1(1, 2) \} \pi_1 + \{ H_0(2, 2) + H_1(2, 2) \} \pi_2 , \]  \hspace{1cm} (2-58c) \\
\[ P(X_0 \neq 2, X_1 = 2) = \pi_2 (1 - \pi_2) . \]  \hspace{1cm} (2-58d) 

(iv) Multi-state run length of DARMA(1,2) 

As discussed in the binary DARMA(1,2), a bivariate Markov process, \((X_{n+1}, A_n)\), can be formed in the multi-state DARMA(1,2). The transition probabilities are given in equations (2-37) and (2-38). Moreover assuming \(\{Y_n\}\) is
assumed to take values in \{0,1,2\}, the transition probability matrices, \(H_0\), \(H_1\), and \(H_2\), which are derived in Appendix B, can be given as follows:

\[
H_0 = \begin{cases}
(1-\gamma_1-\gamma_2)\rho + (1-\rho - \gamma_1 + 2\gamma_1\rho + \gamma_2\rho) & \gamma_1(1-\rho)\pi_0\pi_1 + (1-\gamma_1-\gamma_2)(1-\rho) \\
\pi_0 + \gamma_1(1-\rho)\pi_0^2 & \pi_1 \\
\gamma_1(1-\rho)\pi_0^2 + \gamma_2(1-\rho)\pi_0 & \gamma_2(1-\rho)\pi_0 \\
\end{cases}
\]

\[
H_1 = \begin{cases}
\gamma_1(1-\rho)\pi_1\pi_0 + (1-\gamma_1-\gamma_2)\rho + (1-\rho - \gamma_1 + 2\gamma_1\rho + \gamma_2\rho) & \gamma_1(1-\rho)\pi_1\pi_2 + (1-\gamma_1-\gamma_2)(1-\rho) \\
\pi_0 + \gamma_2(1-\rho)\pi_1 & \pi_2 \\
\gamma_1(1-\rho)\pi_1^2 + \gamma_2(1-\rho)\pi_1 & \gamma_2(1-\rho)\pi_1 \\
\end{cases}
\]

\[
H_2 = \begin{cases}
\gamma_1(1-\rho)\pi_2\pi_0 + \rho(\gamma_1 + \gamma_2)\pi_2 & \gamma_1(1-\rho)\pi_2\pi_1 + \gamma_2(1-\rho)\pi_2 \\
\pi_0 + \gamma_2(1-\rho)\pi_2 & \pi_2 \\
\gamma_1(1-\rho)\pi_1\pi_0 + (1-\gamma_1-\gamma_2)\rho + (1-\rho - \gamma_1 + 2\gamma_1\rho + \gamma_2\rho) & \gamma_1(1-\rho)\pi_2^2 + (1-\gamma_1-\gamma_2)(1-\rho) \\
\pi_0 + \gamma_1(1-\rho)\pi_2^2 & \pi_2 \\
\end{cases}
\]

Defining the n-step transition probability as in the case of binary DARMA(1,2), the probability distribution of the run length of 0 for multi-state DARMA(1,2) is obtained as follows:

\[
P(T_0 = n) = \frac{P(X_0 = 0, A_{-1} = 0)(H_0(n)(0,e) - H_0(n+1)(0,e))}{P(X_0 = 0, X_1 = 0)} \quad (2-60)
\]
where, \( H_0(n)(0,e), H_0(n)(1,e), H_0(n)(2,e) \) are defined in the same way as given in equation (2-53), except that the transition probability matrices are replaced by those shown in (2-59), (2-59a), and (2-59b).

Assume that the process starts at \( A_{-2} \), then

\[
P(X_0=0, A_{-1}=0 | A_{-2}=0) = P(X_0=0, A_{-1}=0 | A_{-2}=0) P(A_{-2}=0) \\
+ P(X_0=0, A_{-1}=0 | A_{-2}=1) P(A_{-2}=1) \\
+ P(X_0=0, A_{-1}=0 | A_{-2}=2) P(A_{-2}=2) \\
= \{H_1(0,0) + H_2(0,0)\} \pi_0 + \{H_1(1,0) + H_2(1,0)\} \pi_1 \\
+ \{H_1(2,0) + H_2(2,0)\} \pi_2 \tag{2-61}
\]

Similarly,

\[
P(X_0=0, A_{-1}=1) = \{H_1(0,1) + H_2(0,1)\} \pi_0 + \{H_1(1,1) + H_2(1,1)\} \pi_1 \\
+ \{H_1(2,1) + H_2(2,1)\} \pi_2 \tag{2-61a}
\]

\[
P(X_0=0, A_{-1}=2) = \{H_1(0,2) + H_2(0,2)\} \pi_0 + \{H_1(1,2) + H_2(1,2)\} \pi_1 \\
+ \{H_1(2,2) + H_2(2,2)\} \pi_2 \tag{2-61b}
\]

Following the same procedure shown from (2-57) to (2-58d), by replacing \( H_1(i,j) \) with those given in (2-59), (2-59a), and (2-59b), the probability distributions of \( T_1 \) and \( T_2 \) can be obtained by assuming that the process starts at \( A_{-2} \).
CHAPTER III

ANALYSES AND RESULTS OF B-DARMA-E MODELING

III.1. Introduction

In this chapter the B-DARMA-E process defined in equation (1-1) is broadly analyzed with a newly formulated procedure of model building which is applied to the daily precipitations in Indiana. The daily precipitation data under consideration have four distinctly different seasons. Since the DARMA process used to in the discussion of the B-DARMA-E model is stationary, it is assumed that the daily precipitations in each of the four seasons are from the same four populations and thus stationary within the seasons. This will be graphically discussed in the following section.

Furthermore the daily precipitation time series can be transformed into dry-wet sequence which have been studied by many authors (Richardson, 1981; Buishand, 1978; Chin, 1977). Unfortunately the precipitation quantities are no longer distinguished in the transformed wet-dry day sequences. Therefore in many applications such as the construction of
rainfall-runoff process, the above models are very limited and cannot be used to reflect the extreme hydrological events such as floods and droughts. In order to preserve the daily quantities of precipitation, it is postulated in this study that the precipitation quantities occurring in the wet days can form a new independent series, constructed from the nonzero precipitation sequence. This assumption will be statistically verified in Section 2.

A simple procedure of model building for the DARMA process is discussed in section 3 and is shown to be very efficient in preserving the statistical properties of floods and droughts. The procedure starts with the observation of the autocorrelation function of the precipitation sequence. A plot of the autocorrelation function is used to determine what kinds of models may be adequate in the DARMA family. During this step the type of models is decided, it is thus named Identification. In the next step a nonlinear least squares method, developed by Marquardt (1966), is used to estimate the model parameters by fitting the model autocorrelation function to that of the precipitation sequences. With these computed parameters, the theoretical autocorrelations of several DARMA models are calculated and are compared to their empirical counterpart obtained from the observations. This gives the further verification to the identification step. As a result of the procedure several competitive models are determined with their
estimated parameters. This step is called Estimation. The final step is the Diagnosis in which the probability distributions of the run lengths derived in chapter 2 are used to check the persistence property of the wet and dry spells in each competitive model. Since the run lengths govern the persistence property of the binary DARMN process, it is desirable to check whether the probability distributions of the run lengths of 0 and of 1 are adequate. The run length distributions are discrete, so a discrete Kolmogrov-Smirnov goodness of fit is used to perform the test. It is required for both of the probability distributions of the run lengths of 0 and 1 to pass the test before a model can be selected.

After the three-step procedure of model building, there may still be more than one good model left. Thus it is desirable to develop a criterion for the selection of the best model among these remaining candidate models. Section III.3.4. devotes itself to the discussion of the construction of the criterion to find the best model. Conceptually it is required to preserve the run lengths of wet and dry days in order to preserve the long-term characteristics of floods and droughts. Hence the estimated probability distribution of the binary run length is compared with its theoretical counterpart from each competitive model. Then the best model is selected by the criterion of the best match, i.e. the one that has the
smallest sum of squared errors between the estimated and the theoretical probability distributions. An example is given to illustrate the procedure of model building.

Finally, the nonzero precipitation sequence is discussed in section 4. The Kolmogorov-Smirnov test is used to investigate the exponential distributions of nonzero precipitations in different seasons. Moreover the plot of the accumulated probability function is used for the graphical comparison between the empirical and the theoretical distributions.

III.2. The B-DARMA-E model

III.2.1. Definitions and assumptions

The B-DARMA-E model is defined in equation (1-1). Let \( W_i = X_i Z_i \), where \( \{X_i\} \) denotes the discrete-state sequence of the daily precipitation time series defined by

\[
X_i = \begin{cases} 
1, & \text{if } R_i \geq 0.01'' \\
0, & \text{if } R_i < 0.01'' 
\end{cases}
\]

(3-1)

where \( R_i \) is the observed daily precipitation quantity ranging from 0.01" to 99.99". The sequence \( \{Z_i\} \) is formed by the daily precipitation quantities whenever \( \{X_i\} \) takes the value 1. Therefore \( \{W_i\} \) represents the complete daily precipitation time series.
\{X_i\} which includes only integers 0 and 1 is called the binary discrete time series, in which 1 represents a wet day and 0 a dry day. Therefore \{X_i\} is also called a wet-dry day time series and can be modeled by the binary discrete autoregressive moving average process. Since the B-DARMA processes described in chapter 2 are stationary and the daily precipitations are seasonal, the data need to be separated into seasons before they can be fitted by the B-DARMA models. It has been assumed by many previous authors (Buishand, 1978; Chin, 1977) that there are four seasons in daily precipitations. This study will statistically show by means of a graphical method that the daily precipitations in Indiana have four distinct seasons. Therefore the nonstationarity within the year can be removed by fitting different models to the different seasons. Hence four different populations from four separate seasons are assumed in this study.

\{Z_i\} denotes the sequence of precipitation quantities if there are wet days. \{Z_i\} is called the nonzero precipitation sequence and is shown to be exponentially distributed.

Combining the \{X_i\} and \{Z_i\} sequences, the time series, \{W_i\}, is formed and is modeled by the Binary Discrete Autoregressive Moving-Average mixed with Exponential model (B-DARMA-E). In this model \{X_i\} and \{Z_i\} are assumed to be independent of each other. This assumption will be
verified in the following.

III.2.2. Verification of assumptions

(A) Seasonality

The seasonality in the daily precipitation time series has been assumed by previous researchers (Buishand, 1978; Woolhiser and Pegram, 1978; Hann, 1976). However, no statistical verification of the assumption has been given so far. In this research, the different seasons in the daily precipitations in Indiana are statistically detected. Two graphical methods are used in which different statistics of the observations are plotted. First, since the B-DARMA-E model formulates the wet-dry days sequence, the numbers of wet and dry days in a 30-year record are respectively counted for each day of the year. It can be seen from Figure (3-1a) and Figure (3-1b) that there are four different behaviors existing within the year. Starting from January first, the year can be roughly separated into 90-day intervals based on the frequency plots of wet and dry days. Therefore four 90-days seasons are used in the calendar year. Secondly, based on the 90-day seasons, the plot of the average precipitations during the seasons are used to check the differences of the precipitation quantities among the different seasons. Figure (3-1c) clearly shows the distinction among the seasons. The above phenomena
similarly happen to all the gaging stations studied.

(B) Independence

The independence between the two series \( \{X_t\} \) and \( \{Z_t\} \), means that the occurrence time of a wet day, \( t \), and the precipitation quantity of the wet day, \( z \), are statistically independent. First, a graphical method is used to see the statistical relation between these two variables. The correlation coefficient of two variables is also computed. From Figure (3-2) it can be seen that the time-precipitation points are very scattered for the four different seasons. Table 3.1. shows that the correlation coefficients are very small with their absolute values ranging from 0.036 to 0.079. All the gaging stations shown in Table 1.1. are studied and give the similar results. Consequently, \( \{X_t\} \) and \( \{Z_t\} \) can be assumed to be statistically independent.

The second method to test the independence of the two variables is the contingency table, which conducts a Chi-square test. Since the test is equivalent to checking that two variables, \( t \) and \( z \), are independent, it can compare the observed frequencies with the frequencies that are expected if the variables are independent. Then a Chi-square statistic is computed from the contingency table formed by the above two compared frequencies. When the Chi-square statistic is small, it indicates that the variables
are independent; if it is large, the data support the conclusion that the variables are not independent. In this study a 3x3 contingency table is respectively constructed for each season. The mean values of \( z_i \) and \( t_i \) are calculated and 0.8 and 1.2 times each mean are used to classify and group the contingency table. Following the above procedure the precipitation series of Station 12-1229 in Indiana is used to illustratively form the contingency table and to compute the Chi-square statistic as an example. Table (3-1) lists the results of the Chi-square statistics for 4 respective seasons. If a 5 percent significance level is assumed for the test, all of the cases pass the test. All the 9 gaging stations studied in this research give the satisfactory results. Therefore, by the above verifications it can be concluded that these two variables are statistically independent and it is proper to form the B-DARMA-E mixed stochastic model.

III.3. DARMA modeling

The B-DARMA-E process includes the submodel DARMA which plays the very important role of preserving the persistence of wet and dry spells in the whole process. The DARMA model was first developed by Jacobs and Lewis (1977), but a clear procedure of the model building has not appeared. Striving for engineering simplicity in hydrologic modeling, this paper develops a three-step procedure made up of the
Identification, Estimation, and Diagnosis steps to fit the B-DARMA-E model to the daily precipitation time series in Indiana. Based on the methodology developed in chapter 2, this section gives the analysis of each step of model building by an illustrative example which is chosen from one of the precipitation gaging stations used in this study.

III.3.1. Model Identification

The identification step is applied to a set of data to indicate the kind of representational model which is worthy of further investigations. The goal of this step is to obtain an idea of the model orders, $p$ and $q$, that are needed in the DARMA($p,q$) process, where $p$ and $q$ are the orders of the autoregressive and of the moving average components, respectively. Therefore, the preliminary identification selects a tentative class of models which will later be effectively fitted and checked. For the precipitation time series studied there are four candidate models selected by the identification step. As a part of this step the initial trial values of the model parameters are roughly determined from the estimated autocorrelations and are used as inputs to the estimation step.

For the identification of the models the graphical methods are particularly useful and the judgement must be exercised. The principal tool for the identification of
DARMA models is the autocorrelation function. It is used not only to help guess the form of the model, but also to obtain the initial trial values of the parameters. These initial parameter values are useful at the estimation step as the starting values for iterative calculations in the nonlinear least squares method. In order to obtain clues about the choice of the orders $p$ and $q$ for the autoregressive and moving average operators, the theoretical autocorrelation characteris of different DARMA models are compared to the sample autocorrelation function. Briefly, the autocorrelation function of a discrete autoregressive process of order $p$ tails off. On the other hand the autocorrelation of a discrete moving average process of order $q$ has a cutoff after the $q$-th lag. Therefore, if the estimated autocorrelations are mixed with both of the above behaviors, a discrete mixed DARMA process is suggested.

Due to the large variation between the estimated and the theoretical autocorrelations (Kendall, 1945), the detailed adherence to the theoretical autocorrelation function cannot be expected from the estimated function. Hence in employing the estimated autocorrelations as a tool for identification, it is usually possible to be fairly sure about the broad characteristics. More subtle indications may or may not represent real effects and a group of related models may need to be entertained and investigated further at the estimation and the diagnostic steps.
The initial trial values of the parameters for the models selected are guessed in the final stage of the identification step. Since the parameters \( p \) and \( q \) have been guessed, where \( p \) is the order of the autoregressive component and \( q \) is the order of the moving average process, the initial trial values of the autoregressive parameters, \( \rho_1, \rho_2, \ldots, \rho_p \), are obtained as follows:

\[
\rho_i = \frac{\{\text{Corr}(p+i+1) + \text{Corr}(p+i+2) + \text{Corr}(p+i+3)\}}{\{\text{Corr}(p+i) + \text{Corr}(p+i+1) + \text{Corr}(p+i+2)\}}
\]

(3-2)

where \( i = 1, 2, \ldots, p \).

since it is preferable to smooth the geometrically decreasing tail of the serial correlation function (Jacobs and Lewis, 1973b). For the trial values of the moving average parameters, \( \beta_0, \beta_1, \beta_2, \ldots, \beta_{q-1} \), they are assigned as 0.5 in the first trial since the process is a random selection of random variables. Following the above procedure an illustrative example is presented.

A fitted example of the daily precipitations in Indiana is demonstrated in the following. The autocorrelation function of the DARMA(0,1) model, given in equation (2-10), has a cutoff after the first lag. Conversely that of a DARMA(1,0) process decays exponentially and tails off. In the DARMA(1,1) and DARMA(1,2) processes, the autocorrelations are decaying exponentially after the first
and after the second lag, respectively, as given in equations (2-17) and (2-19). Figure (3-3) shows the theoretical autocorrelation plots of the various candidate DARMA models for the second season of the daily precipitation time series at Cambridge, Indiana. The models with $p$ larger than 1 and $q$ larger than 2 are discarded in order to maintain the parsimony of the parameters. Figure (3-4) gives the estimated autocorrelations for each season, in which the models, DARMA(3,1), DARMA(1,0), DARMA(1,1), and DARMA(1,2) can be seen to be equally competitive. Therefore these four models are worthy of further investigation.

III.3.2. Model Estimation

The identification procedure having led to several candidate models, it is then necessary to obtain efficient estimates of the parameters. After the parameters are estimated, the fitted models will be subjected to the diagnostic checking to test the goodness of fit. Since the daily precipitation data are not normally distributed, the maximum likelihood estimation is no longer valid in estimating the parameters of the B-DARMA-E model. Thus the procedure of the model estimation employed by Box-Jenkins(1974) cannot be used here.

In this study the method of moment estimation is used. Moment estimations are important because they are commonly
used in practice and are easily accepted by engineers (Salas, Boes, and Smith; 1981). As it is seen in the identification step, the autocorrelation function describes the second moment property of the observations. Therefore, it is required to obtain the closest possible match between the estimated and the theoretical autocorrelations. As shown in chapter 2, the theoretical autocorrelation functions for the DARMA family of models, are nonlinear in their parameters. Hence the method used to estimate the parameters here is a nonlinear least squares scheme, which was derived by Marquardt (1963, 1965).

Let a nonlinear function, $f$ be fitted to the data of the dependent variable $Y$, the autocorrelation function in this study, with one independent variable, $x$, the time lag (day), and with the parameters, $v_1, v_2, \ldots, v_m$, of the DARMA process. Then the equation can be formed as,

$$Y = f(x; v_1, v_2, \ldots, v_m) + \epsilon$$

(3-3)

where $\epsilon$ is the error component. The theory of the least squares method assumes that $E(\epsilon) = 0$, and $\text{Var}(\epsilon) = a$ constant, and the errors, $\epsilon_i$, are uncorrelated. Therefore, equation (3-3) becomes

$$E(Y) = f(x; v_1, v_2, \ldots, v_m)$$

(3-3a)
If there are \( n \) observations, then \( n \) equations of the form of equation (3-2) are obtained

\[
Y_i = f(x_i; v_1, v_2, \ldots, v_m) + \epsilon_i, \quad i = 1, 2, \ldots, n \tag{3-4a}
\]

The least squares method estimates \( v_1, v_2, \ldots, v_m \), so as to minimize the parameters,

\[
S(\mathbf{v}) = \sum_{i}^{n} (Y_i - Y_i(\mathbf{v}(0)))^2 \tag{3-5}
\]

where \( Y_i(\mathbf{v}) \) is the value of \( y \) predicted by equation (3-3a) at the \( i \)-th data point; \( \mathbf{v} = (v_1, v_2, \ldots, v_m) \). Equation (3-5) is differentiated with respect to \( v_1, v_2, \ldots, v_m \) and the derivatives are set equal to zero so that \( v_1 \) can be obtained by solving the following simultaneous equations, called the normal equations:

\[
\sum_{i}^{n} \frac{\partial f_i}{\partial v_j} [Y_i - Y_i(\mathbf{v}(0))] = 0, \quad j = 1, 2, \ldots, m \tag{3-6}
\]

Since the normal equations are nonlinear, they are linearized through the Taylor series expansion with \( \mathbf{v} = (v_1, v_2, \ldots, v_m) \) which is a small perturbation of \( \mathbf{v} \). Then,

\[
(Y_i(x_i; \mathbf{v} + \mathbf{c})) = f_i(x_i; \mathbf{v}) + \sum_{j}^{m} \frac{\partial f_i}{\partial v_j} \frac{c_j}{c_j} \tag{3-7}
\]

or

\[
(Y) = \mathbf{F} + \mathbf{R} \mathbf{c} \tag{3-7a}
\]

where (\( \cdot \)) means the predicted value; \( \mathbf{Y} = (Y_1, Y_2, \ldots, Y_n) \);
\[ F = (f_1, f_2, \ldots, f_n); \hat{R} = \partial f_i / \partial v_j \quad i=1,2,\ldots,n; \quad j=1,2,\ldots,m. \]

Since the normal equations are now linear in \( \hat{C} \) after the Taylor expansion, \( \hat{C} \) can be found by the standard least squares method by setting \( \partial \langle S \rangle / \partial c_j = 0 \), for all \( j \), where

\[ \langle S \rangle = \sum_i^n (Y_i - \langle Y_i \rangle) \]  

(3-8)

Thus \( \hat{C} \) is obtained by solving

\[ \hat{A} \hat{C} = \hat{G} \]  

(3-9)

where

\[ \hat{A} = \hat{R}^T \hat{R} \]  

(3-9a)

and \( \hat{G} = (g_1, g_2, \ldots, g_m) \), \( \hat{R}^T \) is the transpose of \( \hat{R} \), with

\[ g_j = \sum_i^n (Y_i - f_i) \frac{\partial f_i}{\partial v_j} \]  

(3-9c)

for \( j=1,2,\ldots,m \).

In general the methods to solve the above nonlinear problem have centered around two approaches. The first approach is to expand the dependent function as a Taylor series and to calculate the parameters at each iteration on the assumption of the local linearity. The second approach uses a modification of the steepest descend method. These methods either have the problem of divergence in the successive iteration or of agonizingly slow convergence. Therefore, the maximum neighborhood method developed by Marquardt (1965) will be used in this study. The method performs an optimum interpolation between the Taylor series
method and the gradient method. The interpolation is based on the maximum neighborhood in which the truncated Taylor series gives an adequate representation of the nonlinear function.

In equation (3-9) the direction of the correction vector, \( \vec{C} \), obtained by the linearization procedure has the best angle between \( 80^0 \) and \( 90^0 \). So the best scheme to solve the problem is to scale the matrix \( \overline{A} \) and the vector \( \overline{g} \) in equation (3-9) by

\[
\overline{a}_{ij}^{0}=\overline{a}_{ij}/\sqrt{\overline{a}_{ij}\overline{a}_{jj}} \quad \text{and} \quad \overline{g}_{i}^{0}=\overline{g}_{i}/\sqrt{\overline{a}_{jj}} \quad (3-10)
\]

where \( \overline{a}_{ij}^{0} \) is the correction of \( \overline{a}_{ij} \), \( \overline{a}_{ij} \) is the \( i,j \)-th element of \( \overline{A} \), \( \overline{a}^{0} \) is the correction of \( \overline{A} \); \( \overline{g}_{i}^{0} \) is the correction of \( \overline{g}_{i} \), \( g_{j} \) is the \( j \)-th element of \( \overline{g} \), \( \overline{g}^{0} \) is the correction of \( \overline{g} \). Then in the \( k \)-th iteration the equation (3-8) becomes,

\[
(\overline{A}_{k}^{0}+\lambda_{k}I)\overline{g}_{k}^{0}=\overline{g}_{k}^{0} \quad (3-11)
\]

where \( I \) is the identity matrix, and \( \lambda_{k} \) is a positive constant that satisfies the condition, \( S(\overline{V}_{k+1})<S(\overline{V}_{k}) \).

where \( \overline{V}_{k} \) is the \( k \)-th iterated \( \overline{V} \) values in equation (3-5) The \( \overline{C}_{k}^{0} \) solved from (3-11) is rescaled to obtain the correction vector \( \overline{C} \) by
\[ \bar{c}_j = c_0 j / \sqrt{a_{jj}}, \quad j = 1, 2, \ldots, m, \]  

(3-12)

The revised parameters for the new iteration are

\[ v_{k+1} = v_k + \bar{c}_k \]

(3-13)

By this procedure the nonlinear least squares method continues the iteration until the convergence is achieved or the iteration number is attained.

For the purpose of estimation of the parameters in the DARMA models, a computer program was designed by the nonlinear least squares method described above. Appendix C provides an illustrative example for obtaining the parameters in the DARMA(1,1) model by the program developed in this study. Table 3.2 shows the estimated parameters for the competitive models selected at the identification step for the different seasons. The theoretical autocorrelation functions of the models for each season can be computed with the parameters in Table 3.2. Figures (3-5a) to (3-5d) show the plots of the theoretical and estimated autocorrelations which give further check of the identification step.

The estimation of the parameters, \( \pi_0 \) and \( \pi_1 \), makes use of the run length property discussed in chapter 2. The alternate dry and wet spells can be expressed by the run lengths, \( T_0 \) and \( T_1 \). As defined in equations (2-21) and (2-21a), the mean value of the run length of 0's, \( T_0' \), and
the mean value of the run length of 1's, \( T^0_1 \), can be computed directly from the observed data. Thus, \( \pi_0 \) is given by

\[
\pi_0 = \frac{T^0_0}{T^0_0 + T^0_1} \quad (3-14)
\]

and

\[
\pi_1 = 1 - \pi_0. \quad (3-14a)
\]

It is noted that the estimations of \( \pi_0 \) and \( \pi_1 \) based on the occurrence probabilities of the wet and dry spells in the sequence are conceptually plausible since the wet and the dry spells govern the characteristics of the time series through the run lengths of 0 and 1.

III.3.3. Model Diagnostic Checking

Having estimated parameters, the fitted models are subjected to the procedure of diagnostic checking, which diagnoses whether the model is adequate. The procedure has to discover in what way a model is adequate or inadequate. Since no model form ever represents the truth absolutely, it follows that, given sufficient data, statistical methods can discredit models which could nevertheless be entirely adequate for the purpose at hand. Conversely, methods can fail to indicate serious departures from assumptions because these methods are insensitive to the type of discrepancy that occurs. Thus the best policy is to devise the most
sensitive statistical scheme to detect the important properties of the observed data. No system of diagnostic checks can ever be comprehensive. This research as an engineering study in hydrology seeks a method with important physical characteristics which are useful for the engineering purposes.

One useful method of checking a model is the goodness-of-fit test, that is, to check whether the model behavior based on the estimated parameters is close enough to that existing in the observed data. The method used here is to investigate the binary run length properties which are defined in chapter 2. It is extremely important in a daily precipitation time series to preserve the long persistence of the wet and dry spells in order to interpret the duration characteristics of the floods and the droughts. Changnon (1980) discussed floods and droughts and concluded that the precipitation time series can most efficiently reflect the dry and the wet spells which govern the phenomena of the floods and the droughts. The binary run lengths defined in chapter 2 are physically exactly the dry and wet spells. That is to say, the precipitation time series models preserve the property of the persistances in the wet and the dry spells if the characteristics of the run lengths are preserved.

Hence the scheme of diagnostic checking used in this research is to check whether the binary run length property
in the observations is preserved by the calibrated theoretical models. The theoretical probability distributions for the binary DARMA models, derived in chapter 2, are used while the estimated probability distribution is computed by equations (2-21) and (2-21a). Then the method is to check both probability distributions that of the run length of 0's and that of the run length of 1's. It is required that both run lengths pass the test of the discrete Kolmogrov-Smirnov goodness of fit (Pettitt and Stephens, 1977) since the two alternate run lengths equally govern the persistence property in the time series. The diagnostic checking method by the probability distributions of the run lengths is hydrologically adequate because the latter contains the information about the durations of the droughts and the floods. On the other hand it is statistically efficient because the probability distribution is the first-hand information among all the statistics, i.e. to preserve the probability distribution is equivalent to preserving all the non-product moments (mean, variance, skewness, etc.) of the run lengths.

If there is more than one model passing both the tests of the 0 and 1 run lengths, then the sum of squares errors between the theoretical and the estimated probability distributions are computed. Therefore, the sums of square errors from the probability distributions of the 0 and 1 run lengths, respectively, for each model are added and the
least value is used as the criterion for the best model. It is noted that the errors of the run length 0 and the run length 1 in the same model equally discredit the model so that they are added up to a single criterion.

III.3.3.1. Discrete K-S Goodness-of-Fit test

The Kolmogorov-Smirnov (K-S) statistic, usually denoted by D, is employed to test the hypothesis that a random sample comes from a specified continuous distribution and is a well-defined statistical method. Since the probability distribution of the run length is a discrete distribution, a discrete K-S method is used for the test of the goodness-of-fit in the run length distributions. The outcomes are divided by days into k different groups to test the null hypothesis:

\[ H_0: P(\text{an observation falls in group } i) = p_i, \ i = 1, 2, \ldots, k. \]

Suppose \( N \) independent observations are taken, and let \( O_i \) be the observed number and \( E_i \) be the expected number \( (E_i = NP_i) \) in the group. Define the statistic \( M \) by

\[ M = \max_{1 \leq i \leq k} \left| \sum_{j=1}^{i} (O_j - E_j) \right|. \]  

(3-15)

There is a straightforward parallel between \( M \) and the Kolmogorov-Smirnov D. For continuous observations, say \( X_1, X_2, \ldots, X_n \), from a continuous cumulative function \( F(x) \), D
is defined

\[ D = \text{Sup} | F_n(x) - F(x) | \quad (3-16) \]

where \( F_n(x) \) is the empirical distribution of the observations. Therefore the relation can be obtained as

\[ N F_n(x_j) = \sum_{i} O_i ; j = 1, 2, \ldots, k. \quad (3-17) \]

\[ N F(x_j) = \sum_{i} E_i ; j = 1, 2, \ldots, k. \quad (3-17a) \]

where \( F_n(x) \) is the cumulative histogram of the data and \( F(x) \) is the expected cumulative grouped distribution function. Then \( M \) can be written as

\[ M = N \text{ Sup} | F_n(x) - F(x) | \quad (3-18) \]

This parallel property provides the evidence that the Kolmogrov-Smirnov test can be used for the discrete sample as in the case of the run lengths of 0 and 1 in this study.

III.3.3.2. Test of the binary run length of 0

The run length of 0 defined in equation (2-21) govern the key property of the dry spell in the B-DARMA-E process. It is desirable to obtain a model which is able to preserve the statistical properties of the run length of 0. Thus in this section it is intended to check by the discrete K-S test the
B-DARMA-E models which were constructed through the steps of identification and estimation.

The run length of 0 is discrete in terms of days which serve as the building blocks of the groups in the discrete K-S goodness-of-fit. After grouping, the expected probability can be obtained from equation (2-22) which is derived respectively for each DARMA model in chapter 2. Then the expected cumulative grouped distribution function, \( F \), as shown in (3-18) is obtained by adding up the expected probabilities. On the other hand, the observed cumulative distribution function, \( F_n(y) \), is computed directly from the observed data. Therefore the \( M \) value is computed from (3-18) and the corresponding \( D \) can be obtained from the parallel property described in section III.3.3.1. Table(3-3) lists the test of the discrete K-S goodness-of-fit for the four competitive B-DARMA models, where a 5 percent significance level is used for the judgement.

III.3.3.3. Test of the binary run length of 1

Equation (2-21a) defines the binary run length of 1 which represents the wet spell in the B-DARMA-E process. Since the run lengths of 0 and 1 alternate in the binary DARMA model, they are equally important for the persistence property. Following the procedure of the goodness-of-fit test of the run length of 0, it is needed to test the distribution of
the run length of 1 by the discrete K-S method.

From (2-21), the discrete run length of 1 is counted by days and is grouped so that it is applicable to the discrete K-S scheme as shown in (3-18). The expected probability which is used for obtaining the expected cumulative distribution function is computed by (2-22a). In the meantime the observed cumulative grouped distribution can be calculated by the observed frequency in each group. Consequently, (3-18) is used to obtain the statistics of the K-S goodness-of-fit test. The statistics for the selected competitive models are given in Table (3-3) where a 5 percent significance level is set to perform the test.

III.3.4. Selection of the best model

Through the discrete Kolmogorov-Smirnov goodness-of-fit test, the diagnostic checking step may qualify more than one model. This section intends to set a criterion to find the most suitable model among the models which pass the test. The criterion used here is the sum of squared errors between the estimated and the theoretical probability distributions of the run lengths of 0 and 1. Since the run lengths of 0 and 1 alternate in each model, they are equally important in reflecting the characteristics of the model. Therefore, the sum of squared errors from both of the probability distributions of the run lengths of 0 and 1 are added up for
the comparison. The best model is defined to be the one having the smallest error. Table 3.4. gives the result of the sum of squared errors in the run length distributions and the judgements are based on the criterion set above.

As a further aid in selection, Figures (3-6a) to (3-6d) show the plots of the probability distributions of the binary run length of 0 for all the competitive models; these figures also provide a comparison between the estimated and the theoretical probability distributions of the run length of 0. Figures (3-7a) to (3-7d) show the probability distributions of the run length of 1, which are respectively compared to their empirical counterparts that are estimated from the observations in different seasons. It is noted that the judgement in the selection of the best model is based on both the discrete Kolmogorov-Smirnov test of goodness-of-fit and the criterion of the smallest sum of squared errors in the probability distributions. Plots of the theoretical and the empirical run length probability distributions also help the consistancy of the best model selection.

III.4. The nonzero precipitation sequence

The nonzero precipitation magnitudes sequence with exponential distributions have been discussed by several authors (Richardson, 1981; Woolhisser, 1979; Buishand, 1978b).
In order to fulfill the B-DARMA-E process the exponential distributions are used for the modeling of the nonzero precipitation magnitude sequences in the different seasons. The autocorrelation plots of the nonzero precipitation sequences, as in Figure (3-8), have shown that the autocorrelation coefficients of the sequences are very small. This means that the precipitation quantities in the nonzero precipitation sequences are uncorrelated and may be assumed to be independent. The Kolmogorov-Smirnov method is used to test the goodness of the exponential fit to the nonzero precipitation sequences. Table 3.5. shows that the sequence in the first season has a weakly exponential distribution, while the sequences in the other three seasons are exponentially distributed at the 5 percent significance level. On the other hand the plots of the empirical and the theoretical cumulative distributions are constructed for comparison in Figure (3-9). This figure shows that the exponential distributions are quite satisfactory for the sequences in the different seasons. It is noted that the sequence in the first season is during the snowfall period when the precipitation observations are recorded less precisely.

III.6. Summary

(1) The newly developed stochastic process, B-DARMA-E, consisting of a submodel, Binary Discrete Autoregressive
Moving Average (B-DARMA) for the wet and dry days sequence, and an independent Exponential distribution for the nonzero precipitation sequence, is used to fit the daily precipitation time series in Indiana.

(2) Based on the verification in chapter 2, the B-DARMA-E stochastic model is shown both statistically and physically suitable for the daily precipitation time series.

(3) A three-step procedure of model building for the B-DARMA, is developed and is seen to be simple and effective in modeling the daily precipitation time series.

(4) The step of identification based on the autocorrelation function is quite suitable since the second moment is easily obtained and is well understood by engineers. Moreover, the estimation step can serve to recheck the identification as well as to compute the parameters of the identified models.

(5) The diagnostic checking step based on the discrete Kolmogorov-Smirnov method is used to test the goodness-of-fit for the probability distributions of the run lengths of 0 and 1, which govern the characteristics of the B-DARMA process. It is statistically efficient to test the probability distribution functions of the run lengths since the probability distribution is the first-hand information among all the statistics.
(6) Since the probability distributions of the run lengths of 0 and 1 govern the duration characteristics of the floods and the droughts, it is desirable to have a model with its probability distributions of the run lengths of 0 and 1 as close to its empirical counterpart as possible. Hence the criterion for the selection of the best model is to choose the one with the smallest error between the estimated and the theoretical probability distributions.

(7) The precipitation quantities in the nonzero precipitation sequences are, in general, exponentially distributed and may be assumed to be independent of the B-DARMA process in forming the B-DARMA-E model.

(8) The persistence of the wet and the dry spells can effectively be preserved in the models by preserving the run lengths of 0 and 1. The behaviors of floods and droughts can be interpreted through the probability distributions of the run lengths of 0 and 1 respectively since the latter are physically and statistically defined.

III.6. Supplement

The Chi-square values of independence verification for 9 daily precipitation stations are tabulated in Table 3.1a. The estimated parameters for each model are given in tables 3.2a, 3.2b, 3.2c and 3.2d. The goodness-of-fit tests of modeling are listed in Tables 3.3a, 3.3b, 3.3c and 3.3d,
while Table 3.4a shows the selected best model for each station. Table 3.5a provides the exponential fit of each nonzero precipitation series. In the meantime Figures (3-10a) to (3-17b) give the run length distributions of the best model for each station except those for station 12-1229.
Table 3.1. Statistics for the Verification of Independence Station: 12-1229

<table>
<thead>
<tr>
<th>Season</th>
<th>$z_i$</th>
<th>$t_i$</th>
<th>Corr.</th>
<th>Contingency Table</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>St.De</td>
<td>Mean</td>
<td>St.De</td>
</tr>
<tr>
<td>1</td>
<td>0.30</td>
<td>0.39</td>
<td>1.9</td>
<td>2.8</td>
</tr>
<tr>
<td>2</td>
<td>0.37</td>
<td>0.44</td>
<td>1.7</td>
<td>2.8</td>
</tr>
<tr>
<td>3</td>
<td>0.38</td>
<td>0.47</td>
<td>2.4</td>
<td>3.3</td>
</tr>
<tr>
<td>4</td>
<td>0.31</td>
<td>0.39</td>
<td>2.1</td>
<td>3.2</td>
</tr>
</tbody>
</table>

$z_i$: the precipitation quantity in the wet day (inch).

$t_i$: interarrival time between consecutive wet days.

$\chi^2$: Chi-square statistic; $\chi^2(.95; 4) = 9.49$. 
TABLE 3.1a. $\chi^2$ Values of Contingency Tables for Independence Verification

<table>
<thead>
<tr>
<th>Season</th>
<th>Station 12-0676</th>
<th>12-0831</th>
<th>12-1734</th>
<th>12-5337</th>
<th>12-6018</th>
<th>12-7362</th>
<th>12-7747</th>
<th>12-9318</th>
<th>12-1229</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5.56</td>
<td>1.54</td>
<td>1.65</td>
<td>5.23</td>
<td>9.50</td>
<td>1.77</td>
<td>6.82</td>
<td>2.38</td>
<td>4.03</td>
</tr>
<tr>
<td>2</td>
<td>3.45</td>
<td>4.33</td>
<td>4.69</td>
<td>7.03</td>
<td>10.54*</td>
<td>7.14</td>
<td>5.50</td>
<td>3.19</td>
<td>8.60</td>
</tr>
<tr>
<td>3</td>
<td>1.95</td>
<td>9.45</td>
<td>3.87</td>
<td>2.87</td>
<td>2.06</td>
<td>4.78</td>
<td>3.83</td>
<td>2.15</td>
<td>2.94</td>
</tr>
<tr>
<td>4</td>
<td>7.37</td>
<td>1.77</td>
<td>3.20</td>
<td>9.15</td>
<td>4.2</td>
<td>15.96*</td>
<td>8.87</td>
<td>3.04</td>
<td>9.25</td>
</tr>
</tbody>
</table>

1: Degree of freedom: 4
2: $\chi^2(.95; 4) = 9.49$
*: fail to pass
Table 3.2. Estimated Parameters for B-DARMA-E Models.
Station: 12-1229

<table>
<thead>
<tr>
<th>Season</th>
<th>Model</th>
<th>Estimated by Autocorrelation Function</th>
<th>Estimated by Run Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(0,1)</td>
<td>$\beta = .79296$</td>
<td>$\pi_0 = .62183$</td>
</tr>
<tr>
<td></td>
<td>(1,0)</td>
<td>$\rho = .15633$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1,1)</td>
<td>$\rho = .00006$, $\beta = .79294$</td>
<td>$\pi_1 = .37837$</td>
</tr>
<tr>
<td></td>
<td>(1,2)</td>
<td>$\rho = .99901$, $\gamma_1 = .63717$, $\gamma_2 = .23629$</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>(0,1)</td>
<td>$\beta = .71156$</td>
<td>$\pi_0 = .59760$</td>
</tr>
<tr>
<td></td>
<td>(1,0)</td>
<td>$\rho = .21281$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1,1)</td>
<td>$\rho = .32722$, $\beta = .62688$</td>
<td>$\pi_1 = .40250$</td>
</tr>
<tr>
<td></td>
<td>(1,2)</td>
<td>$\rho = .77301$, $\gamma_1 = .47216$, $\gamma_2 = .35763$</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>(0,1)</td>
<td>$\beta = .71548$</td>
<td>$\pi_0 = .69447$</td>
</tr>
<tr>
<td></td>
<td>(1,0)</td>
<td>$\rho = .19562$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1,1)</td>
<td>$\rho = .09758$, $\beta = .68705$</td>
<td>$\pi_1 = .30553$</td>
</tr>
<tr>
<td></td>
<td>(1,2)</td>
<td>$\rho = .99930$, $\gamma_1 = .49614$, $\gamma_2 = .38940$</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>(0,1)</td>
<td>$\beta = .68237$</td>
<td>$\pi_0 = .65154$</td>
</tr>
<tr>
<td></td>
<td>(1,0)</td>
<td>$\rho = .20186$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1,1)</td>
<td>$\rho = .00669$, $\beta = .68024$</td>
<td>$\pi_1 = .34846$</td>
</tr>
<tr>
<td></td>
<td>(1,2)</td>
<td>$\rho = .93481$, $\gamma_1 = .40508$, $\gamma_2 = .47844$</td>
<td></td>
</tr>
</tbody>
</table>

1: Models are the competitive DARMA models.
### TABLE 3.2a. Estimated Parameters for B-DARMA-E Models; Season 1.

<table>
<thead>
<tr>
<th>Model</th>
<th>Station</th>
<th>12-0676</th>
<th>12-0831</th>
<th>12-1734</th>
<th>12-5337</th>
<th>12-6018</th>
<th>12-7362</th>
<th>12-7747</th>
<th>12-9138</th>
<th>12-1229</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0,1)</td>
<td>$\beta$</td>
<td>.83057</td>
<td>.72764</td>
<td>.80109</td>
<td>.84026</td>
<td>.81850</td>
<td>.82941</td>
<td>.80685</td>
<td>.83575</td>
<td>.79296</td>
</tr>
<tr>
<td>(1,0)</td>
<td>$\rho$</td>
<td>.13569</td>
<td>.18502</td>
<td>.15498</td>
<td>.12982</td>
<td>.14329</td>
<td>.13645</td>
<td>.14900</td>
<td>.13249</td>
<td>.15633</td>
</tr>
<tr>
<td>(1,1)</td>
<td>$\rho$</td>
<td>.00007</td>
<td>.00005</td>
<td>.06351</td>
<td>.00007</td>
<td>.00008</td>
<td>.00007</td>
<td>.00006</td>
<td>.00007</td>
<td>.00006</td>
</tr>
<tr>
<td></td>
<td>$\beta$</td>
<td>.83056</td>
<td>.72762</td>
<td>.78794</td>
<td>.84025</td>
<td>.81849</td>
<td>.82940</td>
<td>.80684</td>
<td>.83574</td>
<td>.79294</td>
</tr>
<tr>
<td>(1,2)</td>
<td>$\rho$</td>
<td>.99878</td>
<td>.96579</td>
<td>.99999*</td>
<td>.99999*</td>
<td>.99999*</td>
<td>.99999*</td>
<td>.99999*</td>
<td>.99999*</td>
<td>.99999*</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>.73605</td>
<td>.49289</td>
<td>.62685</td>
<td>.74845</td>
<td>.66163</td>
<td>.68341</td>
<td>.65089</td>
<td>.51601</td>
<td>.63717</td>
<td></td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>.18206</td>
<td>.40050</td>
<td>.22193</td>
<td>.17126</td>
<td>.19365</td>
<td>.18328</td>
<td>.22150</td>
<td>.35839</td>
<td>.23629</td>
<td></td>
</tr>
<tr>
<td>All**</td>
<td>$\pi_0$</td>
<td>.64957</td>
<td>.68507</td>
<td>.68568</td>
<td>.66485</td>
<td>.62636</td>
<td>.62398</td>
<td>.62079</td>
<td>.67325</td>
<td>.62163</td>
</tr>
<tr>
<td></td>
<td>$\pi_1$</td>
<td>.35043</td>
<td>.31493</td>
<td>.31432</td>
<td>.33515</td>
<td>.37364</td>
<td>.37602</td>
<td>.37921</td>
<td>.32675</td>
<td>.37837</td>
</tr>
</tbody>
</table>

*: Unrealistic estimation  
**: All four models
TABLE 3.2b. Estimated Parameters for B-DARMA-E Models; Season 2.

<table>
<thead>
<tr>
<th>Model</th>
<th>Station</th>
<th>12-0676</th>
<th>12-0831</th>
<th>12-1734</th>
<th>12-5337</th>
<th>12-6018</th>
<th>12-7362</th>
<th>12-7747</th>
<th>12-9138</th>
<th>12-1229</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0,1)</td>
<td>$\beta$</td>
<td>.72381</td>
<td>.72837</td>
<td>.54533</td>
<td>.68641</td>
<td>.69577</td>
<td>.75266</td>
<td>.75642</td>
<td>.77821</td>
<td>.71156</td>
</tr>
<tr>
<td>(1,0)</td>
<td>$\rho$</td>
<td>.19834</td>
<td>.30716</td>
<td>.25405</td>
<td>.21831</td>
<td>.29944</td>
<td>.18186</td>
<td>.18583</td>
<td>.17551</td>
<td>.21281</td>
</tr>
<tr>
<td>(1,1)</td>
<td>$\rho$</td>
<td>.18057</td>
<td>.36924</td>
<td>.31411</td>
<td>.24925</td>
<td>.42378</td>
<td>.17241</td>
<td>.20715</td>
<td>.27095</td>
<td>.32722</td>
</tr>
<tr>
<td></td>
<td>$\beta$</td>
<td>.67271</td>
<td>.35891</td>
<td>.51128</td>
<td>.61347</td>
<td>.42972</td>
<td>.72227</td>
<td>.70187</td>
<td>.71561</td>
<td>.62688</td>
</tr>
<tr>
<td>(1,2)</td>
<td>$\rho$</td>
<td>.30130</td>
<td>.86236</td>
<td>.76896</td>
<td>.24489</td>
<td>.79941</td>
<td>.72201</td>
<td>.89035</td>
<td>.81986</td>
<td>.77301</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>.64729</td>
<td>.32837</td>
<td>.37355</td>
<td>.36025</td>
<td>.32056</td>
<td>.66978</td>
<td>.48808</td>
<td>.61497</td>
<td>.47216</td>
<td></td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>.28625</td>
<td>.37393</td>
<td>.42828</td>
<td>.33020</td>
<td>.38310</td>
<td>.26653</td>
<td>.38111</td>
<td>.24412</td>
<td>.35763</td>
<td></td>
</tr>
<tr>
<td>All*</td>
<td>$\pi_1$</td>
<td>.61980</td>
<td>.59887</td>
<td>.64977</td>
<td>.61994</td>
<td>.56365</td>
<td>.61207</td>
<td>.60862</td>
<td>.63000</td>
<td>.59750</td>
</tr>
<tr>
<td></td>
<td>$\pi_2$</td>
<td>.36020</td>
<td>.40113</td>
<td>.35023</td>
<td>.38006</td>
<td>.43635</td>
<td>.38793</td>
<td>.39138</td>
<td>.37000</td>
<td>.40250</td>
</tr>
</tbody>
</table>

*: All four models
TABLE 3.2c. Estimated Parameters for B-DARMA-E Models; Season 3.

<table>
<thead>
<tr>
<th>Model</th>
<th>Station</th>
<th>12-0676</th>
<th>12-0831</th>
<th>12-1734</th>
<th>12-5337</th>
<th>12-6018</th>
<th>12-7362</th>
<th>12-7747</th>
<th>12-9138</th>
<th>12-1229</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0,1)</td>
<td>$\beta$</td>
<td>.81559</td>
<td>.79929</td>
<td>.82109</td>
<td>.76477</td>
<td>.61321</td>
<td>.75234</td>
<td>.77454</td>
<td>.81728</td>
<td>.71548</td>
</tr>
<tr>
<td>(1,0)</td>
<td>$\rho$</td>
<td>.14421</td>
<td>.15323</td>
<td>.14419</td>
<td>.16994</td>
<td>.24546</td>
<td>.17746</td>
<td>.17058</td>
<td>.14327</td>
<td>.19562</td>
</tr>
<tr>
<td></td>
<td>$\beta$</td>
<td>.00007</td>
<td>.62344</td>
<td>.08086</td>
<td>.00006</td>
<td>.22389</td>
<td>.04176</td>
<td>.10259</td>
<td>.00007</td>
<td>.09758</td>
</tr>
<tr>
<td>(1,1)</td>
<td>$\rho$</td>
<td>.81559</td>
<td>.79928</td>
<td>.80577</td>
<td>.76476</td>
<td>.50971</td>
<td>.74178</td>
<td>.75009</td>
<td>.81726</td>
<td>.68705</td>
</tr>
<tr>
<td></td>
<td>$\gamma_1$</td>
<td>.99999*</td>
<td>.99999*</td>
<td>.99999*</td>
<td>.99999*</td>
<td>.97705</td>
<td>.99999*</td>
<td>.99999*</td>
<td>.99999*</td>
<td>.99999*</td>
</tr>
<tr>
<td></td>
<td>$\gamma_2$</td>
<td>.71892</td>
<td>.65312</td>
<td>.70771</td>
<td>.61722</td>
<td>.41183</td>
<td>.64562</td>
<td>.59457</td>
<td>.66297</td>
<td>.49814</td>
</tr>
<tr>
<td>(1,2)</td>
<td>$\rho$</td>
<td>.20802</td>
<td>.22517</td>
<td>.19501</td>
<td>.27594</td>
<td>.39930</td>
<td>.27987</td>
<td>.27489</td>
<td>.19999</td>
<td>.38940</td>
</tr>
<tr>
<td>All**</td>
<td>$\pi_0$</td>
<td>.70142</td>
<td>.70410</td>
<td>.73702</td>
<td>.72548</td>
<td>.70153</td>
<td>.70273</td>
<td>.70603</td>
<td>.71688</td>
<td>.69447</td>
</tr>
<tr>
<td></td>
<td>$\pi_1$</td>
<td>.29858</td>
<td>.29590</td>
<td>.26298</td>
<td>.27452</td>
<td>.29847</td>
<td>.29727</td>
<td>.29397</td>
<td>.29312</td>
<td>.30553</td>
</tr>
</tbody>
</table>

*: Unrealistic estimation  
**: All four models
TABLE 3.2d. Estimated Parameters for B-DARMA-E Models: Season 4.

<table>
<thead>
<tr>
<th>Model</th>
<th>Station</th>
<th>12-0676</th>
<th>12-0831</th>
<th>12-1734</th>
<th>12-5337</th>
<th>12-6018</th>
<th>12-7362</th>
<th>12-7747</th>
<th>12-9138</th>
<th>12-1229</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0,1)</td>
<td>$\beta$</td>
<td>.76563</td>
<td>.68669</td>
<td>.75349</td>
<td>.72067</td>
<td>.66468</td>
<td>.66492</td>
<td>.62734</td>
<td>.75556</td>
<td>.68237</td>
</tr>
<tr>
<td>(1,0)</td>
<td>$\rho$</td>
<td>.16941</td>
<td>.25949</td>
<td>.18016</td>
<td>.19482</td>
<td>.20503</td>
<td>.20819</td>
<td>.21557</td>
<td>.17959</td>
<td>.20183</td>
</tr>
<tr>
<td>(1,1)</td>
<td>$\rho$</td>
<td>.00005</td>
<td>.00063</td>
<td>.08876</td>
<td>.10789</td>
<td>.00005</td>
<td>.02371</td>
<td>.00005</td>
<td>.09386</td>
<td>.00669</td>
</tr>
<tr>
<td></td>
<td>$\beta$</td>
<td>.76562</td>
<td>.45194</td>
<td>.73079</td>
<td>.68996</td>
<td>.66467</td>
<td>.65453</td>
<td>.62735</td>
<td>.73174</td>
<td>.68024</td>
</tr>
<tr>
<td>(1,2)</td>
<td>$\rho$</td>
<td>.99999*</td>
<td>.00100</td>
<td>.95107</td>
<td>.99999*</td>
<td>.99999*</td>
<td>.99999*</td>
<td>.99999*</td>
<td>.99999*</td>
<td>.90129</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td></td>
<td>.63279</td>
<td>.45507</td>
<td>.58762</td>
<td>.52689</td>
<td>.43584</td>
<td>.44642</td>
<td>.47436</td>
<td>.57654</td>
<td>.40508</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td></td>
<td>.27136</td>
<td>.50158</td>
<td>.24776</td>
<td>.35774</td>
<td>.43848</td>
<td>.43953</td>
<td>.40182</td>
<td>.28230</td>
<td>.47844</td>
</tr>
<tr>
<td>All**</td>
<td>$\pi_0$</td>
<td>.66798</td>
<td>.70447</td>
<td>.70815</td>
<td>.69136</td>
<td>.68213</td>
<td>.68456</td>
<td>.65443</td>
<td>.69238</td>
<td>.65154</td>
</tr>
<tr>
<td></td>
<td>$\pi_1$</td>
<td>.33202</td>
<td>.29553</td>
<td>.29185</td>
<td>.30864</td>
<td>.31787</td>
<td>.31544</td>
<td>.34557</td>
<td>.30717</td>
<td>.34846</td>
</tr>
</tbody>
</table>

*: Unrealistic estimation
**: All four models
Table 3.3. Statistics of Goodness of Fit¹
Station: 12-1223

<table>
<thead>
<tr>
<th>Season</th>
<th>Model</th>
<th>$D_0$</th>
<th>P&amp;NP</th>
<th>$D_1$</th>
<th>P&amp;NP</th>
<th>Judge</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(0,1)</td>
<td>0.03</td>
<td>P</td>
<td>0.04</td>
<td>P</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>(1,0)</td>
<td>0.02</td>
<td>P</td>
<td>0.04</td>
<td>P</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>(1,1)</td>
<td>0.17</td>
<td>NP</td>
<td>0.16</td>
<td>NP</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td>(1,2)</td>
<td>0.07</td>
<td>P</td>
<td>0.18</td>
<td>NP</td>
<td>B</td>
</tr>
<tr>
<td>2</td>
<td>(0,1)</td>
<td>0.05</td>
<td>P</td>
<td>0.03</td>
<td>P</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>(1,0)</td>
<td>0.03</td>
<td>P</td>
<td>0.02</td>
<td>P</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>(1,1)</td>
<td>0.21</td>
<td>NP</td>
<td>0.09</td>
<td>NP</td>
<td>E</td>
</tr>
<tr>
<td></td>
<td>(1,1)</td>
<td>0.06</td>
<td>P</td>
<td>0.13</td>
<td>NP</td>
<td>E</td>
</tr>
<tr>
<td>3</td>
<td>(0,1)</td>
<td>0.02</td>
<td>P</td>
<td>0.03</td>
<td>P</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>(1,0)</td>
<td>0.02</td>
<td>P</td>
<td>0.03</td>
<td>P</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>(1,1)</td>
<td>0.21</td>
<td>NP</td>
<td>0.05</td>
<td>P</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td>(1,2)</td>
<td>0.11</td>
<td>NP</td>
<td>0.20</td>
<td>NP</td>
<td>B</td>
</tr>
<tr>
<td>4</td>
<td>(0,1)</td>
<td>0.05</td>
<td>P</td>
<td>0.02</td>
<td>P</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>(1,0)</td>
<td>0.04</td>
<td>P</td>
<td>0.06</td>
<td>P</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>(1,1)</td>
<td>0.22</td>
<td>NP</td>
<td>0.13</td>
<td>NP</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td>(1,2)</td>
<td>0.08</td>
<td>NP</td>
<td>0.21</td>
<td>NP</td>
<td>B</td>
</tr>
</tbody>
</table>

¹Discrete Kolmogorov-Smirnov Goodness of Fit.

$D_0$: Statistic for the run length of 0.

$D_1$: Statistic for the run length of 1.

P: Pass; NP: No Pass. ; G: Good ; B: Bad.
TABLE 3.3a. Judgements by the Test of Goodness-of-Fit; Season 1.

<table>
<thead>
<tr>
<th>Model</th>
<th>Station</th>
<th>12-0676</th>
<th>12-0831</th>
<th>12-1734</th>
<th>12-5337</th>
<th>12-6018</th>
<th>12-7362</th>
<th>12-7747</th>
<th>12-9138</th>
<th>12-1229</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0,1)</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>(1,0)</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>(1,1)</td>
<td>P</td>
<td>NP</td>
<td>NP</td>
<td>P</td>
<td>NP</td>
<td>P</td>
<td>P</td>
<td>F</td>
<td>P</td>
<td>NP</td>
</tr>
<tr>
<td>(1,2)</td>
<td>P</td>
<td>NP</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>NP</td>
<td>--</td>
<td>--</td>
<td>NP</td>
<td>NP</td>
</tr>
</tbody>
</table>

P: All of run lengths of 0, 1, and 2 pass the test
NP: One or more of run lengths of 0, 1, and 2 fail the test
-: Unrealistic estimation
TABLE 3.3b. Judgements by the Test of Goodness-of-Fit; Season 2.

<table>
<thead>
<tr>
<th>Model</th>
<th>Station 12-0676</th>
<th>12-0831</th>
<th>12-1734</th>
<th>12-5337</th>
<th>12-6018</th>
<th>12-7362</th>
<th>12-7747</th>
<th>12-9138</th>
<th>12-1229</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0,1)</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>(1,0)</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>(1,1)</td>
<td>P</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>(1,2)</td>
<td>P</td>
<td>NP</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>NP</td>
<td>NP</td>
</tr>
</tbody>
</table>

P: All of run lengths of 0, 1, and 2 pass the test
NP: One or more of run lengths of 0, 1, and 2 fail the test
TABLE 3.3c. Judgements by the Test of Goodness-of-Fit; Season 3.

<table>
<thead>
<tr>
<th>Model</th>
<th>Station 12-0676</th>
<th>12-0831</th>
<th>12-1734</th>
<th>12-5337</th>
<th>12-6018</th>
<th>12-7362</th>
<th>12-7747</th>
<th>12-9138</th>
<th>12-1229</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0,1)</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>(1,0)</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>(1,1)</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>NP</td>
<td>NP</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>NP</td>
</tr>
<tr>
<td>(1,2)</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>NP</td>
<td>P</td>
<td>--</td>
<td>--</td>
<td>NP</td>
</tr>
</tbody>
</table>

P: All of run lengths of 0, 1, and 2 pass the test
NP: One or more of run lengths of 0, 1, and 2 fail the test
—: Unrealistic estimation
TABLE 3.3d. Judgements by the Test of Goodness-of-Fit; Season 4.

<table>
<thead>
<tr>
<th>Model</th>
<th>Station</th>
<th>12-0676</th>
<th>12-0831</th>
<th>12-1734</th>
<th>12-5337</th>
<th>12-6018</th>
<th>12-7362</th>
<th>12-7747</th>
<th>12-9138</th>
<th>12-1229</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0,1)</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>(1,0)</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>(1,1)</td>
<td>NP</td>
<td>NP</td>
<td>P</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
</tr>
<tr>
<td>(1,2)</td>
<td>--</td>
<td>NP</td>
<td>P</td>
<td>--</td>
<td>NP</td>
<td>--</td>
<td>--</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
</tr>
</tbody>
</table>

P: All of run lengths of 0, 1, and 2 pass the test
NP: One or more of run lengths of 0, 1, and 2 fail the test
-: Unrealistic estimation
Table 3.4. Selection Of the Best Model  
Station: 12-1223

<table>
<thead>
<tr>
<th>Season</th>
<th>Model</th>
<th>$S_0$</th>
<th>$S_1$</th>
<th>$S_0+S_1$</th>
<th>Judge</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(0,1)</td>
<td>.0007</td>
<td>.0071</td>
<td>.0078</td>
<td>(0,1)</td>
</tr>
<tr>
<td></td>
<td>(1,0)</td>
<td>.0012</td>
<td>.0152</td>
<td>.0164</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1,1)</td>
<td>.0071</td>
<td>.0137</td>
<td>.0203</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1,2)</td>
<td>.0057</td>
<td>.0404</td>
<td>.0461</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>(0,1)</td>
<td>.0035</td>
<td>.0027</td>
<td>.0062</td>
<td>(1,0)</td>
</tr>
<tr>
<td></td>
<td>(1,0)</td>
<td>.0010</td>
<td>.0012</td>
<td>.0022</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1,1)</td>
<td>.0086</td>
<td>.0034</td>
<td>.0160</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1,2)</td>
<td>.0041</td>
<td>.0182</td>
<td>.0223</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>(0,1)</td>
<td>.0029</td>
<td>.0018</td>
<td>.0047</td>
<td>(1,0)</td>
</tr>
<tr>
<td></td>
<td>(1,0)</td>
<td>.0014</td>
<td>.0029</td>
<td>.0043</td>
<td>or</td>
</tr>
<tr>
<td></td>
<td>(1,1)</td>
<td>.0066</td>
<td>.0016</td>
<td>.0102</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1,2)</td>
<td>.0079</td>
<td>.0426</td>
<td>.0499</td>
<td>(0,1)</td>
</tr>
<tr>
<td>4</td>
<td>(0,1)</td>
<td>.0021</td>
<td>.0008</td>
<td>.0029</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1,0)</td>
<td>.0012</td>
<td>.0037</td>
<td>.0099</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1,1)</td>
<td>.0112</td>
<td>.0036</td>
<td>.0148</td>
<td>(0,1)</td>
</tr>
<tr>
<td></td>
<td>(1,2)</td>
<td>.0064</td>
<td>.0478</td>
<td>.0542</td>
<td></td>
</tr>
</tbody>
</table>

$S_0$: Sum of the square errors for the run length 0.

$S_1$: Sum of the square errors for the run length 1.

Judge: The model has the smallest error.
<table>
<thead>
<tr>
<th>Station</th>
<th>12-0676</th>
<th>12-0831</th>
<th>12-1734</th>
<th>12-5337</th>
<th>12-6018</th>
<th>12-7362</th>
<th>12-7747</th>
<th>12-9138</th>
<th>12-1229</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(0,1)</td>
<td>(0,1)</td>
<td>(1,0)</td>
<td>(1,0)</td>
<td>(0,1)</td>
<td>(0,1)</td>
<td>(0,1)</td>
<td>(0,1)</td>
<td>(0,1)</td>
</tr>
<tr>
<td>2</td>
<td>(1,0)</td>
<td>(1,0)</td>
<td>(1,0)</td>
<td>(1,0)</td>
<td>(1,0)</td>
<td>(0,1)</td>
<td>(1,0)</td>
<td>(1,0)</td>
<td>(1,0)</td>
</tr>
<tr>
<td>3</td>
<td>(0,1)</td>
<td>(0,1)</td>
<td>(1,0)</td>
<td>(0,1)</td>
<td>(1,0)</td>
<td>(1,0)</td>
<td>(1,0)</td>
<td>(0,1)</td>
<td>(1,0)</td>
</tr>
<tr>
<td>4</td>
<td>(0,1)</td>
<td>(0,1)</td>
<td>(0,1)</td>
<td>(0,1)</td>
<td>(0,1)</td>
<td>(0,1)</td>
<td>(0,1)</td>
<td>(0,1)</td>
<td>(0,1)</td>
</tr>
</tbody>
</table>

*The criterion is the minimum of the sums of squared errors of the theoretical and estimated probability distributions of the run lengths of 0 and 1.*
Table 3.5. K-S statistics for exponential fits of nonzero precipitations

Station: 12-1229

<table>
<thead>
<tr>
<th>Season</th>
<th>Mean</th>
<th>St.De.</th>
<th>D_</th>
<th>D_+</th>
<th>D(0.95)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.30</td>
<td>0.39</td>
<td>0.18</td>
<td>0.04</td>
<td>0.13</td>
</tr>
<tr>
<td>2</td>
<td>0.37</td>
<td>0.44</td>
<td>0.12</td>
<td>0.03</td>
<td>0.13</td>
</tr>
<tr>
<td>3</td>
<td>0.38</td>
<td>0.47</td>
<td>0.14</td>
<td>0.05</td>
<td>0.15</td>
</tr>
<tr>
<td>4</td>
<td>0.31</td>
<td>0.39</td>
<td>0.14</td>
<td>0.06</td>
<td>0.15</td>
</tr>
</tbody>
</table>

D_ : K-S lower bound.
D_+ : K-S upper bound.
D(0.95): D value with 5 percent significance level.
TABLE 3.5a. Exponential Fits of Non-zero Precipitations by Kolmogrov-Smirnov Test.

<table>
<thead>
<tr>
<th>Station</th>
<th>12-0676</th>
<th>12-0831</th>
<th>12-1734</th>
<th>12-5337</th>
<th>12-6018</th>
<th>12-7362</th>
<th>12-7747</th>
<th>12-9138</th>
<th>12-1229</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>NP</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>NP</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>NP</td>
</tr>
<tr>
<td>2</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>3</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>4</td>
<td>P</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>NP</td>
</tr>
</tbody>
</table>

*: With 5 percent significance level
P: Pass the test
NP: Fail to pass the test
Figure 3-1a. Number of dry days, NO, in 20-year record;
station: 12-1229

Figure 3-1b. Number of wet days, N1, in 20-year record,
Station: 12-1229

Figure 3-1c. Seasonal means, Station: 12-1229
Figure (3-2) Relationship between the nonzero precipitation and its occurrence time; station: 12-1229
Figure 3-3. Theoretical autocorrelations for some candidate B-DARMA models; station: 12-1229
Figure 3-4. The estimated autocorrelation functions for different seasons; station: 12-1229
Figure 3-5a. Empirical and B-DARMA(0,1) model autocorrelation; station: 12-1229
Figure (3-5b) Empirical and B-DARMA(1,0) model autocorrelation functions: station: 12-1229
Figure (3-5c) Empirical and B-DARMA(1,1) model autocorrelation functions; station: 12-1229
Figure (3-5d) Empirical and B-DARMA(1,2) model autocorrelation functions; station: 12-1229
Figure (3-6b) Probability distributions of run length 0 for B-DARMA(1,0) models; station: 12-1228
Figure (3-8c) Probability distributions of run length of B-DARMA(1,1) models; station: 12-1228
Figure (3-6d) Probability distributions of run length 0 for B-DARMA(1,2) models; station: 12-1229
Figure 3-7a. Probability distributions of run length 1 for B-DARMA(0,1) models; station: 12-1229
Figure (3-7b) Probability distributions of run length 1 for B-DARMA(1.0) models; station: 12-1229
Figure (3-7c) Probability distributions of run length 1 for B-DARMA(1,1) models; station: 12-1229
Figure(3-7d) Probability distributions of run length 1 for B-DARMA(1,2) models; station: 12-1229
Figure 3-8. Autocorrelation functions of the nonzero precipitation sequences; station: 12-1229
Figure(3-9) Exponential distributions of the nonzero precipitation sequences; station: 12-1229
Figure (3-10a) Probability distributions of run length 0 for B-DARMA-E models, station: 12-0676
Figure (3-10b): Probability distributions of run length 1 for B-DARMA-E models, station: 12-0676
Figure (3-11a) Probability distributions of run length 0 for B-DARMA-E models, station: 12-0831
Figure (3-11b) Probability distributions of run length 1 for B-ARMA-E models, station: 12-0831
Figure (S-12a) Probability distributions of run length 0 for B-DARNA-E models, station: 12-1734
Figure (3-12b) Probability distributions of run length 1 for E-DARMA-E model, station: 12-1734
Figure (3-13a) Probability distributions of run length 0 for E-DARMA-E models, station: 12-6337
Figure (3-13b) Probability distributions of run length 1 for
B-Darma-E models, station: 12-5337
Figure (3-14a) Probability distributions of run length 0 for E-DARMA-E models, station: 12-6018
Figure (S-16a) Probability distributions of run length 0 for E-DARMA-E models, station: 12-7362
Figure(S-15b) Probability distributions of run length 1 for B-DARMA-E models, station: 12-7382
Figure 3-13a: Probability distributions of run length 0 for E-DARMA-E models, station: 12-7747
Figure (3-16b) Probability distributions of run length 1 for B-DARMA-E models, station: 12-7747
Figure (S-17a) Probability distributions of run length 0 for E-CARMA-E models, station: 12-9138
Figure (C-17b) Probability distributions of run length 1 for B-DARMA models, station: 12-9138.
CHAPTER IV

ANALYSES AND RESULTS OF M-DARMA MODELING

IV.1. Introduction

The submodel B-DARMA in the B-DARMA-E process analyzed in chapter 3 has yielded promising results in preserving the persistances of the wet and the dry spells, in which the wet days cover the precipitation quantities ranging from 0.01" to 9.99" and the dry days include any day with the precipitation less than 0.01". The quantities of precipitation are no distinguished in the binary DARMA process so that the B-DARMA-E process has to fit the exponential distribution to the nonzero precipitation sequence. Since one can transform the daily precipitation time series into a wet and dry days sequence including only 0 and 1, it is also possible to transform this series into a multi-state precipitation sequence in which the precipitation quantities can be represented in magnitude classes. The Multi-state Discrete Autoregressive Moving Average (M-DARMA) process is applied to such a sequence.
In this chapter the multi-state precipitation is first defined conceptually by classifying the precipitation quantity into different discrete states. It is desirable to preserve the properties of the original data as much as possible. The autocorrelation function is used to check the suitability of the transformed sequence. The sample autocorrelation function of the originally observed daily precipitation time series is plotted to compare with those estimated from the different discrete multi-state precipitation sequences. By the adequate classification of states the covariance structure of the daily precipitation can be preserved. It is therefore possible to analyze the daily precipitation time series by a DARMA process if the transformed series has a suitable covariance structure.

In this study a 3-state daily precipitation sequence is constructed for the illustration of the M-DARMA model building. Based on the preservation of the autocorrelation criterion, the daily precipitation time series considered in this research are found to be suitable under a classification into 3 states in which 0 represents the precipitation quantity less than \( R_j \), 1 represents it between \( R_j \) and \( R_j + \delta_j \), and 2 represents it greater than \( R_j + \delta_j \), where \( R_j \) is the mean value in the j-th season and \( \delta_j \) is the standard deviation in the j-th season. For the four different seasons the 3-state discrete daily precipitation time series are statistically verified and constructed in
The discrete autoregressive moving average process is applicable to the daily precipitation series which have been transformed into a discrete multi-state sequence. In section 3 the 3-step procedure of model building, based on the methodology of the second moment property and the concept of the multi-state run lengths which were developed in chapter 2, is established. Furthermore, a criterion for selecting the best model from the models which passed the test in the diagnostic step is constructed using the probability distributions of the multi-state run lengths. The same data described in chapter 3 is used to illustrate the procedure of the M-DARMA model building.

Model identification is a step to tentatively guess the possible type of the models based on the sample autocorrelation function. The order of the autoregressive process, $p$, and the order of the moving average process, $q$, are obtained in this step. Those models whose theoretical autocorrelation functions are behave in a similar way as the sample autocorrelation functions are selected for further investigations. In order to have an efficient estimation in the next step, the preliminary trial values of the parameters for each selected model are discussed and are given at the end of the identification step.
Model estimation is used to precisely compute the values of the parameters as well as to give a further check of the identification step. The nonlinear least squares method described in chapter 3 is also applicable to this case since the autocorrelation function of the M-DARMA process is nonlinear in the parameters that need to be estimated. After the parameters are estimated, the theoretical autocorrelations are plotted together with the corresponding sample autocorrelations for a further visual checking.

Model diagnostic checking is used to check whether the models whose parameters have been estimated are adequate. The multi-state run lengths introduced and defined in chapter 2 play the most important role in the persistence properties of the multi-state M-DARMA process which govern the daily precipitation time series. It is desired to investigate whether the model preserves the persistence properties of the multiple states. For this purpose, it is suitable to check whether the multi-state run length properties are preserved in the model. Since the probability distributions of the multi-state run lengths are the first-hand informations of the run lengths, the scheme of the diagnostic checking is to check whether the probability distributions of the multi-state run lengths from the model and from the data match sufficiently well. The multi-state run lengths are counted by days which are discrete. Thus the discrete Kolmogorov-Smirnov test is used
to perform the goodness-of-fit checking. It is required to have all the probability distributions of the multi-state run lengths in the same model pass the discrete Kolmogorov-Smirnov test at a certain significance level in order to credit that model as a good model.

Selection of the best model is a step to be used if there is more than one model passing the test in the diagnostic checking step. Since it is preferable to have the best preservation of the multi-state run length property, the sum of squared errors between the estimated and the theoretical probability distributions of the multi-state run lengths is used as a criterion to select the best model.

A summary, in section 4, lists the results which are obtained through the fit of the multi-state discrete autoregressive moving average model to the daily precipitation time series in Indiana.

IV.2. The M-DARMA daily precipitation model

IV.2.1. Definitions

The Discrete Autoregressive Moving Average (DARMA) process discussed in chapter 2 is shown to be very capable of handling the persistance in a time series. The persistances existing in the daily precipitation sequence
may be properly modeled by the DARMA process if the quantity of the precipitation can be transformed into discrete states. Since the B-DARMA-E model constructed in chapter 2 cannot distinguish among the precipitation quantities by its wet and dry spell sequence, it has to model the nonzero precipitation sequence by mean of an exponential distribution. If the daily precipitation quantities are distinguished by transforming the original precipitation series into a sequence made up of discrete precipitation states, then it is possible to avoid the exponential distribution assumption in the B-DARMA-E model. In order to achieve the above goal, the multi-state precipitation time series is defined in the following way.

The daily precipitation quantity is classified into several states, namely \( r \), and the multi-state precipitation sequence can be formed by assigning the states, 0,1,2,.....\( r-1 \), to each precipitation quantity. Therefore the \( r \)-state discrete precipitation time series is a sequence consisting of the integer values in \{0,1,2,.....\( r-1 \}\}. If a DARMA(p,q) process, where \( p \) is the order of the autoregressive component and \( q \) is the order of the moving average component, is used to model the \( r \)-state daily precipitation time series, then a multi-state discrete autoregressive moving average process is defined and denoted by M-DARMA(p,q;r), where \( r \) is the number of discrete states of the precipitation quantity and \( p, q \) are defined in the same
way as in the DARMA\((p,q)\) process.

IV.2.2. The physical reality of the M-DARMA process

It is desired to preserve the statistical properties of the original daily precipitation time series after the transformation into discrete states. Strictly speaking the model can never preserve the nature absolutely. In this section the autocorrelation function is used in order to preserve of the second moment properties in of the transformed daily precipitation time series. First the sample autocorrelations of the observed daily precipitations, i.e. the infinite-state sequence, is plotted in Figure (4-1a). Next a 1000-state transformation with 0.1" interval separation is used to form a 1000-state discrete daily precipitation time series whose autocorrelation function is shown in Figure (4-1b). The 100-state discrete precipitation sequence is obtained by the 1" interval separation and its autocorrelation function is plotted in Figure (4-1c). Finally the 3-state discrete daily precipitation time series is constructed by

\[
X_i = \begin{cases} 
0 & , R_i < \bar{R}_j \\
1 & , \bar{R}_j \leq R_i < \bar{R}_j + \delta_j \\
2 & , R_i \geq \bar{R}_j + \delta_j 
\end{cases} \tag{4-1}
\]

where \(R_i\) is the \(i\)-th day precipitation quantity; \(\bar{R}_j\) is the mean value of the precipitation in the \(j\)-th season; \(\delta_j\) is
the standard deviation of precipitation in the j-th season. The autocorrelation function of the 3-state precipitation time series is shown in Figure (4-1d) respectively for different seasons. Compared with the autocorrelations of the infinite-state daily precipitation series and those of the 1000-state and the 100-state transformed series, it can be seen that the 3-state daily precipitation sequence can reasonably preserve the second moment property of the observed daily precipitation time series. Therefore, the 3-state transformed daily precipitation time series is used to demonstrate the M-DARMA modeling in this study.

IV.3. The multi-state DARMA Modeling

The M-DARMA model established in Sections IV.1. and IV.2. is based on the concept of the multi-state run lengths, which is defined in chapter 2. As in the B-DARMA modeling, the M-DARMA model is used to preserve both the second moment and the multi-state run length properties of the discrete time series. In accordance with this goal, a three-step procedure, identification, estimation, and diagnostic checking, is developed to fit the M-DARMA model to the multi-state daily precipitation time series. Then a criterion for selecting the best model is based on the least sum of the squared errors in the probability distributions of the multi-state run lengths. The same data used in building the B-DARMA model is also used for demonstrating
the modeling of the M-DARMA process in this section.

IV.3.1. Model Identification

The model identification is a step used to get a general idea of the representational model for a set of discrete time series. The goal of this procedure is to obtain the possible values of p and q in the preliminary M-DARMA models, where p is the order of the autoregressive component and q is the order of the moving average component. Thus the initial identification step provides us tentatively a class of candidate models which are worthy of further investigation. It is also a step to guess the initial values of the parameters, after the orders p and q are decided. The initial trial values of the parameters give more efficient estimations in the estimation step.

The tool in the identification of the M-DARMA modeling is the autocorrelation function. As discussed in chapter 2 the theoretical autocorrelation function of the autoregressive process with any order p tails off, while the moving average process of an order q has a cutoff after the q-th lag. It is possible to guess the orders p and q by studying the sample autocorrelation function. Because of the large sampling variation of the estimated autocorrelations, it is suggested that several combinations of p and q be tried in the selection process of the candidate models. It is
usually possible to determine the general characteristics of the suitable models in the identification step, but the detailed properties need further studies.

After the orders $p$ and $q$ are decided, the initial trial values of the model parameters are given in the following way. It is desired to smooth the geometrically decreasing tail of the serial correlations (Jacobs and Lewis, 1978b). Therefore the initial values of the parameters of the autoregressive components are obtained from the equation (3-2). On the other hand the initial values of parameters of the moving average process are assumed to be 0.5 since the process is a random selection of a sequence of random variables.

The same data of the daily precipitation time series used, in the B-DARMA modeling, is used to illustrate the M-DARMA modeling. Figure (4-1d) shows that the sample autocorrelation functions for the four seasons decay very fast after either the first or the second lag and then tail off. Hence the orders $p$ and $q$ are small and the M-DARMA models $(0,1)$, $(1,0)$, $(1,1)$, and $(1,2)$ are selected as the tentative models in the identification step. Moreover the initial values of the parameters are determined by the same method discussed in the identification step of the B-DARMA modeling. Then they are used for the iterative computation in the next step of the estimation. Table 4.1. lists the initial trial values for the further iterative estimation.
IV.3.2. Model Estimation

Since the tentative M-DARMA models and the initial trial values of their parameters are determined by the identification step, the fine estimation of the parameters needs to be performed before the diagnostic checking. As discussed in the binary DARMA modeling, the conventional maximum likelihood estimation is not available because the multi-state daily precipitation is not normally distributed.

In order to preserve the second moment property, the nonlinear least squares method is used to obtain the model parameters by fitting the theoretical autocorrelation function to the estimated autocorrelation function. The program described for the estimation of the parameters in the B-DARMA modeling is also applicable and is used in this case.

By inputting the initial values given by the identification step, the parameters are estimated by the above described procedure and are given in Table 4.2. for each season. After the parameters are obtained, the autocorrelation function of each model can be computed for further checking. Figures (4-2a) to (4-2d) show the theoretical autocorrelation functions of the DARMA(0,1), DARMA(1,0), DARMA(1,1), and DARMA(1,2) models, as well as the corresponding sample autocorrelation functions for each season. It can be seen that all the candidate models are
quite competitive in fitting the sample autocorrelation functions so that they deserve for the further investigation in the diagnostic checking.

Concerning the estimation of the parameters \( \pi_0, \pi_1, \) and \( \pi_2 \) in the 3-state M-DARMA model of the daily precipitation time series, the multi-state run lengths are used. The multi-state runs in the multi-state daily precipitation sequence are successively followed by one after another and govern the persistence property in the series. By using the definitions of the multi-state run lengths in equations (2-43) to (2-44a), the mean run lengths \( T_{0^0}, T_{1^0}, \) and \( T_{2^0} \) can be computed directly from the observed data. Then \( \pi_0, \pi_1, \) and \( \pi_2 \) are estimated as

\[
\pi_0 = \frac{T_{0^0}}{(T_{0^0} + T_{1^0} + T_{2^0})} \quad (4-2)
\]

\[
\pi_1 = \frac{T_{1^0}}{(T_{0^0} + T_{1^0} + T_{2^0})} \quad (4-2a)
\]

\[
\pi_2 = \frac{T_{2^0}}{(T_{0^0} + T_{1^0} + T_{2^0})} \quad (4-2b)
\]

Table 4.2. shows the estimated \( \pi_0, \pi_1, \) and \( \pi_2 \) for each season.

IV.3.3. Model Diagnostic Checking

The model diagnostic checking is a step to verify the model adequacy. If there is any evidence of serious inadequacy, it is needed to know how the model should be modified in the next iterative cycle. The verification in
this step is done by the goodness of fit test. It is very important to have a suitable technique which can test whether the properties of nature that are of engineering interest are preserved in the model. No man-made checking criterion can be comprehensive, while no model can represent the truth absolutely. This study establishes the concept of the multi-state run length for the diagnostic checking based on the observed persistence of the daily precipitation time series.

IV.3.3.1. Probability distributions of the multi-state run lengths

The concept of the run length had been discussed by several authors in hydrology (Saldarriaga, 1969; Llamas and Siddequi, 1989; Saldarriaga and Yevjevich, 1970; Yevjevich, 1972; Sen, 1976; Buishand, 1978). In order to construct a procedure for the B-DARMA-E modeling and to interpret the duration properties of the floods and the droughts, a newly defined binary run length which was developed in chapter 2, is applied to the daily precipitation data in this chapter.

Following the way of constructing binary run length, the multi-state run length, which is the general case of the former, is defined in equation (2-44). For instance, the 3-state run lengths $T_0$, $T_1$, and $T_2$ are given as follows:
\[ T_0=\{X_0=0, X_1=0, \ldots, X_n=0, X_{n+1} \neq 0 | X_0 \neq 0, X_1=0 \} \]  \hspace{1cm} (4-3)

\[ T_1=\{X_0=1, X_1=1, \ldots, X_n=1, X_{n+1} \neq 1 | X_0 \neq 1, X_1=1 \} \]  \hspace{1cm} (4-3a)

\[ T_2=\{X_0=2, X_1=2, \ldots, X_n=2, X_{n+1} \neq 2 | X_0 \neq 2, X_1=2 \} \]  \hspace{1cm} (4-3b)

Thus the multi-state run lengths describe the persistancies in the different levels of the precipitation quantity. Since the persistance is the most important property in the daily precipitation time series, it is desirable to preserve this characteristic through preserving the multi-state run length properties. The preservation of the sample probability distribution function automatically preserves all the sample moments. Hence the verification of the M-DARMA model is performed through the probability distribution of the multi-state run length. The theoretical probability distributions of the multi-state run lengths are given in equation (2-44). Then the checking technique is to investigate whether the theoretical probability distribution of the multi-state run length in the model is close to its empirical counterparts. It is required that all the run length distributions of a model pass the test before the model is selected.

IV.3.3.2. Goodness-of-Fit for the multi-state run lengths

Since the parameters of the model have been obtained from the estimation step, the theoretical probability distribution of the multi-state run length can be computed
by equations (2-44a) and (2-44b). On the other hand the multi-state run length is counted by days and has then a discrete distribution. Hence the discrete Kolmogorov-Smirnov test discussed in Section III.3.3.1. is applicable to the goodness of fit test of the multi-state run length distribution. It is required that all the run length distributions have to pass the test at a 5 percent significance level in order that the model is accepted.

The observed probability of the run length of 0 can be computed by the conditional formulation given in equation (2-43), while the observed cumulative distribution function, \( F_n(y) \), is simply accumulated from the former. The expected cumulative distribution function, \( F(y) \), is obtained from the theoretical probability distribution of each model. The \( D_0 \) value, which is the K-S statistic, can be obtained from equations (3-16) and (3-17). Table 4.3. lists the \( D_0 \) values for the competitive models in each season and their test results at a 5 percent significance level.

The probability distributions of the run lengths of 1 and 2 defined in (4-3a) and (4-3b) can, in a same way as above, be used to estimate the sample cumulative distribution functions. In the meantime the expected cumulative distribution functions are obtained respectively from their theoretical probability distributions for each model. Therefore the test of the discrete Kolmogorov-Smirnov goodness-of-fit is applied at the 5 percent significance
level as before. The test results are given in Table 4.3.

IV.3.4. the selection of the best model

In the diagnostics step a model is selected if all of its multi-state run length distributions pass the test. For the case where there are more than one model which pass the test, a criterion to find the best model among the selected models is going to be established in this section. The scheme used is the sum of the squared errors between the theoretical and the estimated probability distributions. The squared error sums of all the multi-state run length distributions are added for each model and the smallest total is the criterion used to select the best model, since all the run lengths are equally important for the persistence in the model. For instance, there are three squared error sums for the distributions of $T_0$, $T_1$, and $T_2$ for a 3-state M-DARMA model. Table 4.4. lists the sums of squared error totals for the various candidate models which pass the test and the selected models. Furthermore, Figures (4-3a) to (4-6c) respectively show the comparisons between the theoretical and the estimated probability distributions of the multi-state run lengths of 0, 1, and 2 for each model. These figures give a further check for the selection of the best model.
IV.4. Summary

(1) A Multi-state Discrete Autoregressive Moving Average (M-DARMA) model is developed for the modeling of the daily precipitation time series which are transformed into a multi-state discrete sequence. The transformation is based on the preservation of the second moment property of the observed series. By the comparison of the sample autocorrelation functions under different classifications, a 3-state daily precipitation series is a reasonably acceptable multi-state discrete series and are used to illustrate the modeling of the M-DARMA process.

(2) A three-step procedure including the identification, the estimation, and the diagnostic checking, is developed for the M-DARMA model building and is demonstrated to be effective by an illustrative example.

(3) The identification step uses the autocorrelation function as a main tool. The autocorrelation function is easy to obtain and to interpret in engineering applications.

(4) The estimation scheme by the nonlinear least squares method through the autocorrelation function not only makes the fine estimation of the parameters but is also used as a further check of the suitability of the candidate models indentified in the identification step.
(6) The concept of the multi-state run length is established for the purpose of diagnostic checking in the M-DARMA model building. It is physically important to preserve the persistences of the daily precipitation time series. Therefore this diagnostic checking scheme which uses the multi-state run lengths is a suitable and sensitive method for testing the model.

(6) The selection of the best model based on the best preservation of the probability distributions of the multi-state run lengths is a reasonable method since it is conceptually suitable to preserve the persistence of the observed precipitation time series which are magnified in the run lengths.

(7) One of the advantages for the multi-state runs is that it brings wet the dry periods into a more understandable form, i.e. the highest-state run represents floods while the lowest-state run represents droughts. Since the probability distributions of the different run lengths can be obtained by the conditional probability expressions in chapter 2, these distributions are useful in applications such as irrigation or reservoir operation where the informations about the floods and the droughts are important.
IV.5. Supplement

The estimated parameters of M-DARMA modeling for 9 stations are given in Tables 4.2a, 4.2b, 4.2c and 4.2d. The goodness-of-fit tests of modeling for separated seasons are provided in Tables 4.3a, 4.3b, 4.3c and 4.3d. Tables 4.4a shows the selected best model for each station. Figures (4-7a) to (4-14c) give run length distributions of the best model for each station except for station 12-1229.
Table 4.1. Initial parameters for the candidate M-DARMA models
Station: 12-1229

<table>
<thead>
<tr>
<th>Season</th>
<th>Model</th>
<th>Initial Trial Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0,1</td>
<td>$\beta = 0.500$</td>
</tr>
<tr>
<td></td>
<td>1,0</td>
<td>$\rho = 0.118$</td>
</tr>
<tr>
<td></td>
<td>1,1</td>
<td>$\rho = 0.118, \beta = 0.500$</td>
</tr>
<tr>
<td></td>
<td>1,2</td>
<td>$\rho = 0.118, \gamma_1 = 0.500, \gamma_2 = 0.500$</td>
</tr>
<tr>
<td>2</td>
<td>0,1</td>
<td>$\beta = 0.500$</td>
</tr>
<tr>
<td></td>
<td>1,0</td>
<td>$\rho = 0.121$</td>
</tr>
<tr>
<td></td>
<td>1,1</td>
<td>$\rho = 0.121, \beta = 0.500$</td>
</tr>
<tr>
<td></td>
<td>1,2</td>
<td>$\rho = 0.121, \gamma_1 = 0.500, \gamma_2 = 0.500$</td>
</tr>
<tr>
<td>3</td>
<td>0,1</td>
<td>$\beta = 0.500$</td>
</tr>
<tr>
<td></td>
<td>1,0</td>
<td>$\rho = 0.095$</td>
</tr>
<tr>
<td></td>
<td>1,1</td>
<td>$\rho = 0.095, \beta = 0.500$</td>
</tr>
<tr>
<td></td>
<td>1,2</td>
<td>$\rho = 0.095, \gamma_1 = 0.500, \gamma_2 = 0.500$</td>
</tr>
<tr>
<td>4</td>
<td>0,1</td>
<td>$\beta = 0.500$</td>
</tr>
<tr>
<td></td>
<td>1,0</td>
<td>$\rho = 0.126$</td>
</tr>
<tr>
<td></td>
<td>1,1</td>
<td>$\rho = 0.126, \beta = 0.500$</td>
</tr>
<tr>
<td></td>
<td>1,2</td>
<td>$\rho = 0.126, \gamma_1 = 0.500, \gamma_2 = 0.500$</td>
</tr>
</tbody>
</table>

1: The candidate M-DARMA models.
Table 4.2. Estimated parameters for the candidate M-DARMA models

Station: 12-1229

<table>
<thead>
<tr>
<th>Season</th>
<th>Model $^1$</th>
<th>Estimated by Autocorrelation Function</th>
<th>Estimated by Run Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(0,1) $\beta=0.68327$</td>
<td>$\pi_0=53762$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1,0) $\rho=0.33022$</td>
<td>$\pi_1=0.26532$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1,1) $\rho=0.97371$, $\beta=0.59737$</td>
<td>$\pi_2=0.19706$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1,2) $\rho=0.99723$, $\gamma_1=0.31841$, $\gamma_2=0.31976$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>(0,1) $\beta=0.58419$</td>
<td>$\pi_0=0.53081$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1,0) $\rho=0.29037$</td>
<td>$\pi_1=0.25973$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1,1) $\rho=0.86585$, $\beta=0.60689$</td>
<td>$\pi_2=0.20946$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1,2) $\rho=0.94015$, $\gamma_1=0.33223$, $\gamma_2=0.34285$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>(0,1) $\beta=0.71718$</td>
<td>$\pi_0=0.61411$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1,0) $\rho=0.20590$</td>
<td>$\pi_1=0.21971$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1,1) $\rho=0.25129$, $\beta=0.64739$</td>
<td>$\pi_2=0.16872$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1,2) $\rho=0.9999^2$, $\gamma_1=0.33339$, $\gamma_2=0.44673$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>(0,1) $\beta=0.68596$</td>
<td>$\pi_0=0.56881$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1,0) $\rho=0.31718$</td>
<td>$\pi_1=0.23636$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1,1) $\rho=0.93903$, $\beta=0.60413$</td>
<td>$\pi_2=0.19481$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1,2) $\rho=0.97985$, $\gamma_1=0.33433$, $\gamma_2=0.33040$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^1$: The candidate M-DARMA models.

$^2$: .9999 means an unrealistic estimation.
<table>
<thead>
<tr>
<th>Model</th>
<th>Station</th>
<th>12-0676</th>
<th>12-0831</th>
<th>12-1734</th>
<th>12-5337</th>
<th>12-6018</th>
<th>12-7362</th>
<th>12-7747</th>
<th>12-9138</th>
<th>12-1229</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0,1)</td>
<td>$\beta$</td>
<td>.57773</td>
<td>.70073</td>
<td>.66493</td>
<td>.64194</td>
<td>.64946</td>
<td>.58427</td>
<td>.68237</td>
<td>.70305</td>
<td>.68327</td>
</tr>
<tr>
<td>(1,0)</td>
<td>$\rho$</td>
<td>.29571</td>
<td>.22303</td>
<td>.25495</td>
<td>.25146</td>
<td>.28957</td>
<td>.31375</td>
<td>.34761</td>
<td>.24146</td>
<td>.33322</td>
</tr>
<tr>
<td>(1,1)</td>
<td>$\rho$</td>
<td>.97432</td>
<td>.92005</td>
<td>.98958</td>
<td>.97079</td>
<td>.98923</td>
<td>.97697</td>
<td>.96895</td>
<td>.97708</td>
<td>.97371</td>
</tr>
<tr>
<td></td>
<td>$\beta$</td>
<td>.60691</td>
<td>.68457</td>
<td>.64887</td>
<td>.65487</td>
<td>.60957</td>
<td>.59827</td>
<td>.59378</td>
<td>.65165</td>
<td>.59737</td>
</tr>
<tr>
<td>(1,2)</td>
<td>$\rho$</td>
<td>.99782</td>
<td>.97763</td>
<td>.99999*</td>
<td>.99999*</td>
<td>.99577</td>
<td>.99887</td>
<td>.99999*</td>
<td>.99729</td>
<td>.99999*</td>
</tr>
<tr>
<td></td>
<td>$\gamma_1$</td>
<td>.32211</td>
<td>.38900</td>
<td>.28375</td>
<td>.35452</td>
<td>.31269</td>
<td>.31760</td>
<td>.35349</td>
<td>.31100</td>
<td>.31841</td>
</tr>
<tr>
<td></td>
<td>$\gamma_2$</td>
<td>.32748</td>
<td>.36617</td>
<td>.42581</td>
<td>.34954</td>
<td>.33217</td>
<td>.31805</td>
<td>.28850</td>
<td>.37605</td>
<td>.31976</td>
</tr>
<tr>
<td>All**</td>
<td>$\pi_0$</td>
<td>.54211</td>
<td>.59200</td>
<td>.57539</td>
<td>.58637</td>
<td>.52969</td>
<td>.54200</td>
<td>.51877</td>
<td>.57294</td>
<td>.53762</td>
</tr>
<tr>
<td></td>
<td>$\pi_1$</td>
<td>.26601</td>
<td>.24282</td>
<td>.24709</td>
<td>.22440</td>
<td>.27561</td>
<td>.26249</td>
<td>.27556</td>
<td>.24615</td>
<td>.26532</td>
</tr>
<tr>
<td></td>
<td>$\pi_2$</td>
<td>.19188</td>
<td>.16518</td>
<td>.17752</td>
<td>.18923</td>
<td>.19470</td>
<td>.19551</td>
<td>.20567</td>
<td>.18091</td>
<td>.19706</td>
</tr>
</tbody>
</table>

*: Unrealistic estimation
**: All four models
TABLE 4.2b. Estimated Parameters for M-DARMA Models; Season 2.

<table>
<thead>
<tr>
<th>Model</th>
<th>Station</th>
<th>12-0676</th>
<th>12-0831</th>
<th>12-1734</th>
<th>12-537</th>
<th>12-6018</th>
<th>12-7362</th>
<th>12-7747</th>
<th>12-9138</th>
<th>12-1229</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0,1) β</td>
<td>.57308</td>
<td>.75464</td>
<td>.64979</td>
<td>.53412</td>
<td>.81495</td>
<td>.55537</td>
<td>.58785</td>
<td>.72745</td>
<td>.56419</td>
<td></td>
</tr>
<tr>
<td>(1,0) ρ</td>
<td>.26934</td>
<td>.37136</td>
<td>.29330</td>
<td>.28600</td>
<td>.42055</td>
<td>.27892</td>
<td>.27447</td>
<td>.22148</td>
<td>.29037</td>
<td></td>
</tr>
<tr>
<td>(1,1) ρ β</td>
<td>.86874</td>
<td>.83390</td>
<td>.79689</td>
<td>.87811</td>
<td>.85619</td>
<td>.89124</td>
<td>.90598</td>
<td>.82075</td>
<td>.88585</td>
<td></td>
</tr>
<tr>
<td>(1,2) ρ</td>
<td>.95445</td>
<td>.87767</td>
<td>.89348</td>
<td>.95989</td>
<td>.90095</td>
<td>.94843</td>
<td>.96099</td>
<td>.90710</td>
<td>.94015</td>
<td></td>
</tr>
<tr>
<td></td>
<td>γ1</td>
<td>.33285</td>
<td>.25673</td>
<td>.32062</td>
<td>.34760</td>
<td>.27962</td>
<td>.33458</td>
<td>.34439</td>
<td>.45544</td>
<td></td>
</tr>
<tr>
<td></td>
<td>γ2</td>
<td>.36989</td>
<td>.33178</td>
<td>.36809</td>
<td>.36566</td>
<td>.29416</td>
<td>.35631</td>
<td>.35623</td>
<td>.27931</td>
<td></td>
</tr>
<tr>
<td>All*</td>
<td>π₀</td>
<td>.55152</td>
<td>.52873</td>
<td>.56510</td>
<td>.54045</td>
<td>.51036</td>
<td>.53070</td>
<td>.53050</td>
<td>.54699</td>
<td></td>
</tr>
<tr>
<td></td>
<td>π₁</td>
<td>.26019</td>
<td>.28857</td>
<td>.25782</td>
<td>.26171</td>
<td>.28787</td>
<td>.27140</td>
<td>.26402</td>
<td>.25212</td>
<td></td>
</tr>
<tr>
<td></td>
<td>π₂</td>
<td>.18829</td>
<td>.18270</td>
<td>.17708</td>
<td>.19784</td>
<td>.20177</td>
<td>.19790</td>
<td>.20548</td>
<td>.20089</td>
<td></td>
</tr>
</tbody>
</table>

*: All four models
### TABLE 4.2c. Estimated Parameters for M-DARMA Models; Season 3.

<table>
<thead>
<tr>
<th>Model</th>
<th>Station</th>
<th>12-0676</th>
<th>12-0831</th>
<th>12-1734</th>
<th>12-5337</th>
<th>12-6018</th>
<th>12-7362</th>
<th>12-7747</th>
<th>12-9138</th>
<th>12-1229</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0,1)</td>
<td>β</td>
<td>.79460</td>
<td>.86761</td>
<td>.85247</td>
<td>.78288</td>
<td>.57390</td>
<td>.74886</td>
<td>.78574</td>
<td>.79305</td>
<td>.71716</td>
</tr>
<tr>
<td>(1,0)</td>
<td>ρ</td>
<td>.15742</td>
<td>.12197</td>
<td>.12443</td>
<td>.16277</td>
<td>.24502</td>
<td>.18276</td>
<td>.17288</td>
<td>.15672</td>
<td>.20590</td>
</tr>
<tr>
<td>(1,1)</td>
<td>ρ</td>
<td>.02739</td>
<td>.00008</td>
<td>.08222</td>
<td>.01323</td>
<td>.31247</td>
<td>.09949</td>
<td>.32607</td>
<td>.00006</td>
<td>.25129</td>
</tr>
<tr>
<td></td>
<td>β</td>
<td>.78889</td>
<td>.86759</td>
<td>.83951</td>
<td>.77998</td>
<td>.53569</td>
<td>.72313</td>
<td>.71349</td>
<td>.79304</td>
<td>.64739</td>
</tr>
<tr>
<td>(1,2)</td>
<td>ρ</td>
<td>.99999*</td>
<td>.99999*</td>
<td>.82599</td>
<td>.99999*</td>
<td>.99999*</td>
<td>.99999*</td>
<td>.99999*</td>
<td>.99999*</td>
<td>.99999*</td>
</tr>
<tr>
<td>γ1</td>
<td></td>
<td>.63535</td>
<td>.72643</td>
<td>.99999*</td>
<td>.62127</td>
<td>.55123</td>
<td>.60294</td>
<td>.58860</td>
<td>.64004</td>
<td>.38389</td>
</tr>
<tr>
<td>γ2</td>
<td></td>
<td>.22909</td>
<td>.12896</td>
<td>.00073</td>
<td>.24755</td>
<td>.24635</td>
<td>.27578</td>
<td>.21624</td>
<td>.23221</td>
<td>.44673</td>
</tr>
<tr>
<td>All**</td>
<td>π₀</td>
<td>.61017</td>
<td>.60819</td>
<td>.65731</td>
<td>.62452</td>
<td>.62071</td>
<td>.60831</td>
<td>.60944</td>
<td>.61526</td>
<td>.61411</td>
</tr>
<tr>
<td></td>
<td>π₁</td>
<td>.20946</td>
<td>.21324</td>
<td>.19044</td>
<td>.20778</td>
<td>.21432</td>
<td>.21794</td>
<td>.21618</td>
<td>.20871</td>
<td>.21971</td>
</tr>
<tr>
<td></td>
<td>π₂</td>
<td>.18036</td>
<td>.17858</td>
<td>.15225</td>
<td>.16770</td>
<td>.16497</td>
<td>.17374</td>
<td>.17438</td>
<td>.17602</td>
<td>.16672</td>
</tr>
</tbody>
</table>

*: Unrealistic estimation

**: All four models
<table>
<thead>
<tr>
<th>Model</th>
<th>Station</th>
<th>12-0676</th>
<th>12-0831</th>
<th>12-1734</th>
<th>12-5337</th>
<th>12-6018</th>
<th>12-7362</th>
<th>12-7747</th>
<th>12-9138</th>
<th>12-1229</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0,1)</td>
<td>$\beta$</td>
<td>.56154</td>
<td>.52111</td>
<td>.67299</td>
<td>.56296</td>
<td>.64887</td>
<td>.66215</td>
<td>.71027</td>
<td>.53353</td>
<td>.69596</td>
</tr>
<tr>
<td>(1,0)</td>
<td>$\rho$</td>
<td>.27183</td>
<td>.24606</td>
<td>.24043</td>
<td>.27661</td>
<td>.29627</td>
<td>.30024</td>
<td>.33250</td>
<td>.26794</td>
<td>.31718</td>
</tr>
<tr>
<td>(1,1)</td>
<td>$\beta$</td>
<td>.94343</td>
<td>.23242</td>
<td>.89845</td>
<td>.91313</td>
<td>.92468</td>
<td>.93437</td>
<td>.93828</td>
<td>.90192</td>
<td>.93903</td>
</tr>
<tr>
<td></td>
<td>$\rho$</td>
<td>.64461</td>
<td>.51141</td>
<td>.63670</td>
<td>.63584</td>
<td>.61965</td>
<td>.63212</td>
<td>.59453</td>
<td>.63333</td>
<td>.60413</td>
</tr>
<tr>
<td>(1,2)</td>
<td>$\gamma_1$</td>
<td>.99441</td>
<td>.99999*</td>
<td>.94123</td>
<td>.97931</td>
<td>.97378</td>
<td>.99501</td>
<td>.97666</td>
<td>.96258</td>
<td>.97865</td>
</tr>
<tr>
<td></td>
<td>$\gamma_2$</td>
<td>.35471</td>
<td>.39780</td>
<td>.32999</td>
<td>.36427</td>
<td>.34778</td>
<td>.35212</td>
<td>.32426</td>
<td>.36163</td>
<td>.33435</td>
</tr>
<tr>
<td>All**</td>
<td>$\pi_0$</td>
<td>.56981</td>
<td>.63645</td>
<td>.61889</td>
<td>.59686</td>
<td>.59000</td>
<td>.60453</td>
<td>.56458</td>
<td>.60063</td>
<td>.56881</td>
</tr>
<tr>
<td></td>
<td>$\pi_1$</td>
<td>.24672</td>
<td>.20003</td>
<td>.22118</td>
<td>.22427</td>
<td>.23318</td>
<td>.23025</td>
<td>.25569</td>
<td>.22300</td>
<td>.23638</td>
</tr>
<tr>
<td></td>
<td>$\pi_2$</td>
<td>.18347</td>
<td>.16352</td>
<td>.15993</td>
<td>.17887</td>
<td>.17682</td>
<td>.16522</td>
<td>.17972</td>
<td>.17636</td>
<td>.19481</td>
</tr>
</tbody>
</table>

*: Unrealistic estimation

**: All four models
<table>
<thead>
<tr>
<th>Season</th>
<th>Model</th>
<th>$D_0$</th>
<th>P&amp;NP</th>
<th>$D_1$</th>
<th>P&amp;NP</th>
<th>$D_2$</th>
<th>P&amp;NP</th>
<th>Judge</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(0,1)</td>
<td>.04</td>
<td>P</td>
<td>.10</td>
<td>P</td>
<td>.33</td>
<td>P</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>(1,0)</td>
<td>.01</td>
<td>P</td>
<td>.17</td>
<td>NP</td>
<td>.34</td>
<td>P</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td>(1,1)</td>
<td>.16</td>
<td>NP</td>
<td>.13</td>
<td>P</td>
<td>.18</td>
<td>P</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td>(1,2)</td>
<td>.18</td>
<td>NP</td>
<td>.13</td>
<td>P</td>
<td>.16</td>
<td>P</td>
<td>B</td>
</tr>
<tr>
<td>2</td>
<td>(0,1)</td>
<td>.03</td>
<td>P</td>
<td>.25</td>
<td>NP</td>
<td>.36</td>
<td>NP</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td>(1,0)</td>
<td>.02</td>
<td>P</td>
<td>.14</td>
<td>P</td>
<td>.24</td>
<td>P</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>(1,1)</td>
<td>.13</td>
<td>NP</td>
<td>.09</td>
<td>P</td>
<td>.11</td>
<td>P</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td>(1,2)</td>
<td>.20</td>
<td>NP</td>
<td>.06</td>
<td>P</td>
<td>.07</td>
<td>P</td>
<td>B</td>
</tr>
<tr>
<td>3</td>
<td>(0,1)</td>
<td>.10</td>
<td>P</td>
<td>.12</td>
<td>P</td>
<td>.31</td>
<td>P</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>(1,0)</td>
<td>.10</td>
<td>P</td>
<td>.06</td>
<td>P</td>
<td>.24</td>
<td>P</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>(1,1)</td>
<td>.10</td>
<td>P</td>
<td>.06</td>
<td>P</td>
<td>.25</td>
<td>P</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>(1,2)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>B^1</td>
</tr>
<tr>
<td>4</td>
<td>(0,1)</td>
<td>.10</td>
<td>P</td>
<td>.11</td>
<td>P</td>
<td>.25</td>
<td>P</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>(1,0)</td>
<td>.10</td>
<td>P</td>
<td>.14</td>
<td>P</td>
<td>.26</td>
<td>P</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>(1,1)</td>
<td>.18</td>
<td>NP</td>
<td>.11</td>
<td>P</td>
<td>.11</td>
<td>P</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td>(1,2)</td>
<td>.20</td>
<td>NP</td>
<td>.09</td>
<td>P</td>
<td>.07</td>
<td>P</td>
<td>B</td>
</tr>
</tbody>
</table>

1: Unrealistic estimated model.
<table>
<thead>
<tr>
<th>Model</th>
<th>Station 12-0676</th>
<th>Station 12-0831</th>
<th>Station 12-1734</th>
<th>Station 12-5337</th>
<th>Station 12-6018</th>
<th>Station 12-7362</th>
<th>Station 12-7747</th>
<th>Station 12-9138</th>
<th>Station 12-1229</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0,1)</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>(1,0)</td>
<td>P</td>
<td>P</td>
<td>NP</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>NP</td>
</tr>
<tr>
<td>(1,1)</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
</tr>
<tr>
<td>(1,2)</td>
<td>NP</td>
<td>NP</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>NP</td>
<td>NP</td>
<td>--</td>
<td>NP</td>
</tr>
</tbody>
</table>

P: All of run lengths of 0, 1, and 2 pass the test
NP: One or more of run lengths of 0, 1, and 2 fail the test
-: Unrealistic estimation
TABLE 4.3b. Judgements by the Test of Goodness-of-Fit; Season 2.

<table>
<thead>
<tr>
<th>Model</th>
<th>12-0676</th>
<th>12-0831</th>
<th>12-1734</th>
<th>12-5337</th>
<th>12-6018</th>
<th>12-7362</th>
<th>12-7747</th>
<th>12-9138</th>
<th>12-1229</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0,1)</td>
<td>P</td>
<td>NP</td>
<td>P</td>
<td>NP</td>
<td>P</td>
<td>P</td>
<td>NP</td>
<td>P</td>
<td>NP</td>
</tr>
<tr>
<td>(1,0)</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>NP</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>(1,1)</td>
<td>NP</td>
<td>NP</td>
<td>P</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>P</td>
<td>NP</td>
<td>NP</td>
</tr>
<tr>
<td>(1,2)</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>P</td>
<td>NP</td>
</tr>
</tbody>
</table>

P: All of run lengths of 0, 1, and 2 pass the test
NP: One or more of run lengths of 0, 1, and 2 fail the test
TABLE 4.3c. Judgements by the Test of Goodness-of-Fit; Season 3.

<table>
<thead>
<tr>
<th>Model</th>
<th>Station</th>
<th>12-0676</th>
<th>12-0831</th>
<th>12-1734</th>
<th>12-5337</th>
<th>12-6018</th>
<th>12-7362</th>
<th>12-7747</th>
<th>12-9138</th>
<th>12-1229</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0,1)</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>(1,0)</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>(1,1)</td>
<td>NP</td>
<td>P</td>
<td>NP</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>(1,2)</td>
<td>--</td>
<td>--</td>
<td>NP</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

P: All of run lengths of 0, 1, and 2 pass the test
NP: One or more of run lengths of 0, 1, and 2 fail the test
-: Unrealistic estimation
TABLE 4.3d. Judgments by the Test of Goodness-of-Fit; Season 4.

<table>
<thead>
<tr>
<th>Model</th>
<th>Station 12-0676</th>
<th>12-0831</th>
<th>12-1734</th>
<th>12-5337</th>
<th>12-6018</th>
<th>12-7362</th>
<th>12-7747</th>
<th>12-9138</th>
<th>12-1229</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0,1)</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>NP</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>(1,0)</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>(1,1)</td>
<td>P</td>
<td>NP</td>
<td>P</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>P</td>
<td>P</td>
<td>NP</td>
</tr>
<tr>
<td>(1,2)</td>
<td>NP</td>
<td>--</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
</tr>
</tbody>
</table>

**P**: All of run lengths of 0, 1, and 2 pass the test

**NP**: One or more of run lengths of 0, 1, and 2 fail the test

**: Unrealistic estimation
Table 4.4. Selection of the best model  
Station: 12-1229

<table>
<thead>
<tr>
<th>Season</th>
<th>Model</th>
<th>$S_0$</th>
<th>$S_1$</th>
<th>$S_2$</th>
<th>SUM</th>
<th>Selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(0,1)</td>
<td>0.0009</td>
<td>0.0131</td>
<td>0.1586</td>
<td>0.1726</td>
<td>(0,1)</td>
</tr>
<tr>
<td>2</td>
<td>(1,0)</td>
<td>0.0005</td>
<td>0.0239</td>
<td>0.0697</td>
<td>0.0931</td>
<td>(1,0)</td>
</tr>
<tr>
<td></td>
<td>(0,1)</td>
<td>0.0074</td>
<td>0.0190</td>
<td>0.1378</td>
<td>0.1640</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>(1,0)</td>
<td>0.0073</td>
<td>0.0045</td>
<td>0.0737</td>
<td>0.0855</td>
<td>(1,0)</td>
</tr>
<tr>
<td></td>
<td>(1,1)</td>
<td>0.0036</td>
<td>0.0046</td>
<td>0.0862</td>
<td>0.0994</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>(0,1)</td>
<td>0.0028</td>
<td>0.0152</td>
<td>0.0986</td>
<td>0.1146</td>
<td>(1,0)</td>
</tr>
<tr>
<td></td>
<td>(1,0)</td>
<td>0.0009</td>
<td>0.0188</td>
<td>0.0883</td>
<td>0.1087</td>
<td></td>
</tr>
</tbody>
</table>

$SUM = S_0 + S_1 + S_2$ ,

where $S_0, S_1,$ and $S_2$ are the sums of the squared errors of the theoretical and estimated probability distributions of the run lengths 0, 1, and 2.
### TABLE 4.4a. Selection of the Best Model*

<table>
<thead>
<tr>
<th>Station</th>
<th>12-0676</th>
<th>12-0831</th>
<th>12-1734</th>
<th>12-5337</th>
<th>12-6018</th>
<th>12-7362</th>
<th>12-7747</th>
<th>12-9138</th>
<th>12-1229</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(0,1)</td>
<td>(1,0)</td>
<td>(1,0)</td>
<td>(0,1)</td>
<td>(1,0)</td>
<td>(1,0)</td>
<td>(1,0)</td>
<td>(1,0)</td>
<td>(0,1)</td>
</tr>
<tr>
<td>2</td>
<td>(1,0)</td>
<td>(1,0)</td>
<td>(1,1)</td>
<td>(1,0)</td>
<td>(0,1)</td>
<td>(1,0)</td>
<td>(1,0)</td>
<td>(1,0)</td>
<td>(1,0)</td>
</tr>
<tr>
<td>3</td>
<td>(0,1)</td>
<td>(0,1)</td>
<td>(1,0)</td>
<td>(0,1)</td>
<td>(1,0)</td>
<td>(1,0)</td>
<td>(1,0)</td>
<td>(1,0)</td>
<td>(1,0)</td>
</tr>
<tr>
<td>4</td>
<td>(1,1)</td>
<td>(0,1)</td>
<td>(1,0)</td>
<td>(1,0)</td>
<td>(1,0)</td>
<td>(1,0)</td>
<td>(1,0)</td>
<td>(1,1)</td>
<td>(1,0)</td>
</tr>
</tbody>
</table>

*The criterion of the selection is the minimum sum of squared errors of the theoretical and the estimated probability distributions of run lengths 0, 1, and 2.
Figure (4-1a) Autocorrelation functions of observed daily precipitations; station: 12-1229
Figure (4-1b) Autocorrelation functions of 100-state daily precipitations; station: 12-1229
Figure (4-1c) Autocorrelation functions of 10-state daily precipitations; station: 12-1229
Figure (4-1d) Autocorrelation functions of 3-state daily precipitations; station: 12-1229
Figure (4-2a) Autocorrelation functions of $M$-DARMA(0,1;0) models: station:12-1229
Figure (4-2b) Autocorrelation functions of M-DARMA(1, 0, ;3) models; station: 12-1229
Figure (4-2c) Autocorrelation functions of M-DARMA(1,1;3) models; station: 12-1229
Figure 4.2d: Autocorrelation functions of M-DARMA(1, 2, 3) models; station: 12-1229
Figure (4-9a) Probability distributions of run length for M-DARMA(0,1;3) models; station: 12-1229
Figure (4-Sb) Probability distributions of run length for M-DARMA(0, I; 3) models: station: 12-1229
Figure (4-3c) Probability distributions of run length 2 for M-DARMA(0,1;3) models; station: 12-1223
Figure (4-4a) Probability distributions of run length 0 for M-DARMA(1,0;3) models; station: 12-1229
Figure 4-4b) Probability distributions of run length for M-DARMA(1,0;3) models; station: 12-1229
Figure (4-4c) Probability distributions of run length 2 for M-DARMA(1,0;3) models; station: 12-1229
Figure (4-5a) Probability distributions of run length $\sigma$ for M-DARMA(1,1;3) models: station: 12-1229
Figure 4.b: Probability distributions of run length for M-DARMA(1,1,3) models; station: 12-1229
Figure (4-5o) Probability distributions of run length 2 for M-DARMA(1,1;3) models; station: 12-1229
Figure(4-6a) Probability distributions of run length 0 for M-DARMA(1,2;3) models; station: 12-1229
Figure (4-8b) Probability distributions of run length 1 for M-DARMA(1,2;3) models; station: 12-1229
Figure(4-6c) Probability distributions of run length 2 for M-DARMA(1,2;3) models; station: 12-1229
Figure (4-7a) Probability distributions of run length 0 for M-DARMA models, station: 12-0676
Figure (4-7b) Probability distributions of run length 1 for M-DARMA models, station: 12-0676
Figure (4-7c) Probability distributions of run length 2 for M-DARMA models, station: 12-0678
Figure (4-8a) Probability distributions of run length 0 for M-DARMA models, station: 12-0531
Figure (4-8b) Probability distributions of run length 1 for M-DARMA models, station: 12-0831
Figure (4-8c) Probability distributions of run length 2 for M-DARMA models, station: 12-0831
Figure (4-3a) Probability distributions of run length 0 for M-BARMA models, station: 12-1794
Figure (4-9b) Probability distributions of run length 1 for M-DARMA models, station: 12-1734
Figure 4-9c) Probability distributions of run length Z for M-DARMA models, station: 12-1734
Figure (4-10a) Probability distributions of run length 0 for N-DARMA models, station: 12-5337
Figure (4-10) Probability distributions of run length 1 for M-DARMA models, station: 12-5337
Figure 4-10c) Probability distributions of run length 2 for M-DARMA models, station: 12-5337.
Figure (4-11a) Probability distributions of run length 0 for M-DARMA models, station: 12-6018
Figure(4-11b) Probability distributions of run length 1 for M-DARMA models, station: 12-6018
Figure (4-11c) Probability distributions of run length 2 for M-DARMA models, station: 12-0018.
Figure (4-12a) Probability distributions of run length 0 for M-DARMA models, station: 12-7362
Figure(4-12b) Probability distributions of run length 1 for M-DARMA models, station: 12-7362
Figure 4-12c: Probability distributions of run length 2 for M-DARIMA models, station: 12-7362.
Figure (4-13a) Probability distributions of run length 0 for M-DARM6 models, station: 12-7747
Figure (4-13c) Probability distributions of run length 2 for M-DARMA models, station: 12-7747
Figure 4-14a) Probability distributions of run length 0 for M-DARMA models, station: 12-G138
Figure (4-14b) Probability distributions of run length 1 for M-DARMA models, station: 12-9138
Figure (4-14c) Probability distributions of run length 2 for M-DARMA models, station: 12-9138
CHAPTER V

TRANSFER DISCRETE AUTOREGRESSIVE MOVING AVERAGE MODELS

V.1. Introduction

Many stochastic models have been tried for the simulation of the daily streamflows over the past decade. Quimpo (1967) was the first one to use standardized data to model daily flows which were decomposed in a deterministic seasonal component and a random component modeled as an autoregressive process; he also tried to obtain the link between the stochastic and deterministic processes (1968, 1977). O'Donnell, Hall, and O'Connell (1972) used a similar method to simulate daily streamflows for testing the reliability of a water system. Weiss (1973) fitted the shot noise model to daily flows and pointed out that the Gaussian processes were unsuitable for daily flows. Treiber and Plate (1975, 1977) used a linear transfer system to produce daily flows by inputting generated daily pulses. Dooge (1972, 1977) developed the conceptual rainfall-runoff modeling based on black-box analysis. Pegram (1977, 1980) adapted the conceptual model by assuming that the effective precipitation is a compound Poisson process and the
catchment can be modeled by a combination of linear reservoirs in series and in parallel. Kelman (1977, 1980) modeled the rising and the falling limbs of hydrographs separately and conceptualized the watershed as two linear reservoirs with physical interpretations. Sugawara (1979) originally developed the tank model to simulate the daily rainfall runoff relationship. Kottegoda and Horder (1980) used pulses and a transfer function to relate the daily flow model parameters to catchment characteristics and extended the model to the multisite case.

There are two possible approaches to the stochastic modeling of the daily runoff; one is to build a stochastic model for the daily precipitation and use a transfer function to map the precipitation sequences into runoff sequences; the other approach is to build a daily runoff model without reference to the precipitation data (O'Connell and Jones, 1979; Kelman, 1980 and 1977; Weiss, 1973; Quimpo, 1967). The existing models mostly belong to the latter approach, while several studies partly belong to the former since the precipitation series are partly used (Fegran, 1981 and 1980; Kottegoda and Horder, 1980; Treiber and Plate, 1977 and 1976). The two new comprehensive daily precipitation models, B-DARMA-E and M-DARMA processes, which were developed in previous chapters, will be used in this chapter to construct a stochastic daily precipitation-streamflow model, namely the Transfer Discrete Autoregressive Moving Average (T-DARMA) model.
In a $T$-DARMA$(p,q,m,n,l)$ process, $p$ is the order of the autoregressive component in the daily precipitation series, $q$ is the order of the moving average component in the precipitation series, $m$ is the order of the autoregressive component in the transfer process, $n$ is the order of the moving average component in the transfer process, $l$ represents the number of parameters used to describe the loss during the transfer process from the observed precipitation to the observed streamflow. First in Section 2, a general transfer process is discussed. Then the definition of the $T$-DARMA$(p,q,m,n,l)$ model is given, while the excitation of the $T$-DARMA process is introduced by the $B$-DARMA-E or the $M$-DARMA daily precipitation models which were constructed in the previous chapters.

Section 3 concentrates on the construction of the particular models, $T$-DARMA$(0,1,1,0,1)$, $T$-DARMA$(0,1,1,1,1)$ which are shown to be suitable for modeling the daily precipitation-streamflow data in Indiana. First, the statistical properties between the daily precipitation and the daily streamflows are discussed. The use of the $T$-DARMA process makes it possible to establish statistical relationships between the precipitation input and the streamflow output. Furthermore, the estimation of parameters in the model is achieved by relating the statistical properties between these two series. After the parameters are estimated, the models are used to generate daily streamflows by inputting the daily precipitations.
Then the residual series, formed by subtracting the generated daily flows from their counterpart of observed daily flows, are studied. Further, the simultaneous plots of the generated series and the observed data are compared.

Finally the promising future study avenues of the daily precipitation-streamflow process, based on the newly built precipitation models in this study, are explored.

V.2. Daily precipitation-streamflow processes

V.2.1. Stochastic transfer processes

Whether the process from precipitation to streamflow is stochastic or deterministic has been widely discussed over the past decade. There is a strong link between the stochastic and the deterministic processes for the hydrological time series (Klemes, 1978). Since this study deals with discrete daily hydrological time series, it aims at the stochastic modeling with possible conceptual interpretations.

The pairs of observations \((Q_k, R_k)\) are available at equispaced time intervals, where \(Q_k\) and \(R_k\) are, respectively, the \(k\)-th observation of the streamflow and the precipitation in the same watershed. It is assumed that the relationship between \(Q\) and \(R\) is approximately linear. Suppose an impulse response function, \(H(B)\), with linear weights, \(h_0, h_1, h_2, \ldots\), is taken such that
\[ Q_k = H(B)R_k \]  \hspace{1cm} (5-1)

Then a general linear transfer model, (5-1), is formed, where \( B \) is a backward operator. For a discrete system a difference equation can be used (Box and Jenkins, 1976) as follows:

\[ \delta(B)Q_k = \omega(B)R_k \]  \hspace{1cm} (5-2)

where,

\[ \delta(B) = 1 - \delta_1 B - \delta_2 B^2 - \cdots - \delta_m B^m \]  \hspace{1cm} (5-2a)
\[ \omega(B) = 1 - \omega_1 B - \omega_2 B^2 - \cdots - \omega_n B^n \]  \hspace{1cm} (5-2b)

A stationary process is assumed so that the linear impulse response function, \( H(B) \), is obtained as

\[ H(B) = \omega(B) / \delta(B) \]  \hspace{1cm} (5-3)

This transfer process may be described by a Box–Jenkins model with a white noise \( N_k \), as

\[ Q_k = \delta^{-1}(B)\omega(B)R_k + N_k \]  \hspace{1cm} (5-4)

For hydrological time series with relatively long time intervals (years or months), the Box–Jenkins models work reasonably well as reviewed in chapter 1. For the daily flow series this model has failed to describe the phenomena as evidenced by the literature. Thus a new process should be built for the modeling of the daily precipitation–streamflow series.

The stochastic transfer process was conceptually interpreted as a linear reservoir model in a hydrological catchment by Klemes (1975, 1978) and O'Conner (1978). It was
pointed out that nonlinearities, like unsteady flow effects, tend to weaken as the time interval over which the streamflow is lumped increases and the catchment can be considered linear. Therefore, the use of the linear time series models is justified and appropriate. In the next subsection the general form of the Transfer Discrete Autoregressive Moving Average (T-DARMA) process is defined and discussed.

7.2.2. The T-DARMA process

A transfer process from the precipitation to the streamflow can only be effective if the properties of the input, i.e. the precipitation series, are well known in advance.

In the previous chapters two daily precipitation models, B-DARMA-E and M-DARMA, were built and were shown to be very suitable to the data in Indiana. Furthermore, their covariance structures are well defined. Therefore they are quite suitable as inputs in the precipitation-streamflow transfer process. Physically there are losses in the transfer process from the precipitations to the streamflows. The loss term may include different components such as the evapotranspiration, infiltration and deep percolation to the groundwater. Therefore, a Discrete Autoregressive Moving Average (T-DARMA(p,q,m,n,1)) model is formulated as follows:
\[ \phi(B)Q_k = \theta(B)R_{k-1} - L_j \] (5-5)

where

- \( Q_k \): daily streamflow in the \( k \)-th day,
- \( R_{k-1} \): daily precipitation in the \( (k-1) \)-th day,
- \( L_j \): daily loss in the \( j \)-th season,
- \( \phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_mB^m \),
- \( \theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_n B^n \),

\( p \): order of the autoregressive precipitation process,
\( q \): order of the moving-average precipitation process,
\( m \): order of the autoregressive transfer process,
\( n \): order of the moving average transfer process.

In this research \( R_{k-1} \) is taken to be \( R_k \) since the watersheds studied are comparatively small. Furthermore, the roots of \( \phi(B) = 0 \) are assumed to be outside the unit circle as required for stability. Since a stationary process is being considered, the loss term \( L_j \) is assumed to be a constant during the \( j \)-th season, while the input \( \{R_k\} \) is a B-DARMA-E(p,q) or M-DARMA(p,q,r) process, where \( p, q, \) and \( r \) are the same as defined in chapters 3 and 4 for the daily precipitation models. In the following sections the particular T-DARMA models illustrate the model building by using the daily precipitation-streamflow series.

V.3. Particular T-DARMA models

In this section two particular models, T-DARMA(0,1,1,0,1) and T-DARMA(0,1,1,1,1), are discussed and used to produce
daily streamflows by inputting the daily precipitations which were modeled by binary and multi-state DARMA models.

V.3.1. Statistical properties

The daily precipitation time series in Indiana were suitably modeled by binary DARMA(0,1) or multi-state DARMA(1,0) models. For the study of the T-DARMA modeling, the discussions in this section assume that a B-DARMA-E(0,1) daily precipitation model is used as the precipitation input. Since the input is governed by a stochastic autoregressive moving average process, the conceptual transfer model, T-DARMA, is a stochastic model. The autocorrelation structures of the T-DARMA(0,1,1,0,1) and the T-DARMA(0,1,1,1,1) models are discussed in the following based on the known autocorrelation function of the input precipitation series. The autocorrelation properties of each model, which are very valuable in the stochastic transfer models, are formulated respectively. The relation of the mean function between the daily precipitations and the corresponding daily streamflows is also presented. These statistical properties are then used for the estimation of parameters of the models.

V.3.1.1. The T-DARMA(0,1,1,0,1) model

It is seen from Figure (6-1) that the empirical autocorrelation functions of the daily streamflow time
series in each season have a high value in the first lag and then show a geometrical decay towards the tail. Therefore a T-DARMA(0,1,1,0,1) model whose covariance structure behaves in a similar fashion, seems to be a suitable model to try and is defined as:

\[(1-\mu B)Q_k = R_k - L_j\]  \hspace{1cm} (5-6)

where \( \mu \) is the parameter of the first order of the autoregressive component, while \( Q_k, R_k, \) and \( L_j \) are the same as in equation (5-5). The parameter, \( \mu \), reflects the storage characteristics of the watershed, i.e. a single \( \mu \) in the T-DARMA(0,1,1,0,1) means that the catchment may be modeled by a single reservoir in a series-reservoir formation. Let EV be evaporation, SP be seepage, PT be precipitation, QQ be streamflow, DN be detention water, then a conceptual graph could be sketched as follows:

On the other hand, \( m \) parameters for AR component of the transfer process in equation (5-5) mean that the watershed
has to be separated into m-layers before the seepage. This was similarly interpreted by Fagrum (1980) and Klemes (1978) except that their works were based on the effective rainfall. On the other hand the term \( L_j \) represents the loss during the process from the falling precipitation in the \( k \)-th day to the surface streamflow in the \( k \)-th day. The loss term obviously includes many factors such as the evaporation and the deep percolation. Stationary models are assumed for separate nonoverlapping seasons. The loss term \( L_j \) is regarded as a constant in each season.

Since stationarity is assumed for the T-DARMA(0,1,1,0,1) and \( L_j \) is constant during \( J \)-th season, then equation (5-6) can be expressed as

\[
Q_k = \sum_{i=0}^{\infty} (\mu B^i) \cdot R_k - \sum_{i=0}^{\infty} (\mu^i) \cdot L_j
\]

Equation (5-6a) is analogous to the expression of the unit hydrograph (Linsley et al., 1972). Since \( L_j \) is a constant during the season, the autocorrelation function of \( \{Q_k\} \) can be based on that of \( \{R_k\} \) which is known and derived in the previous chapters. Let \( R'_k = R_k - \bar{R}_j \), where \( \bar{R}_j \) is the mean value of the daily precipitations in the \( j \)-th season. Define CRQ and CRR as the autocorrelation functions of \( \{Q_k\} \) and \( \{R_k\} \), respectively. Therefore,

\[
CRQ(1) = E \left( (R'_k) + \mu R'_{k-1} + \mu^2 R'_{k-2} + \ldots \right) \\
= E \left( R'_k \right) + E \left( \mu R'_{k-1} \right) + E \left( \mu^2 R'_{k-2} \right) + \ldots
\]

\[
= \mu \cdot CR(0) + \mu \cdot CR(0) + \mu^2 \cdot CR(0) + \ldots
\]

\[
+ \mu \cdot CR(1) + 2 \mu^2 \cdot CR(0) + 2 \mu^3 \cdot CR(0) + \ldots
\]

\[
+ \mu \cdot CR(2) + 2 \mu^2 \cdot CR(1) + 2 \mu^3 \cdot CR(1) + \ldots
\]

\[\ldots \]
\[
= \{\mu \text{CRR}(0) + (1+\mu^2)\text{CRR}(1) + \mu(1+\mu^2)\text{CRR}(2) + \mu^2(1+\mu^2)\text{CRR}(3)+\ldots\} / (1-\mu^2)
\]

where \(\text{CRR}(0)=1\),

\[
\text{CRR}(2) = \{ (R'_0 + \mu R'_1 + \mu^2 R'_2 + \ldots) (R'_2 + \mu R'_1 + \mu^2 R'_0 + \ldots) \}
= \mu^2 \text{CRR}(0) + \mu^4 \text{CRR}(0) + \mu^6 \text{CRR}(0) + \ldots
+ \mu \text{CRR}(1) + 2\mu^3 \text{CRR}(1) + 2\mu^5 \text{CRR}(1) + \ldots
+ \text{CRR}(2) + 2\mu^4 \text{CRR}(2) + 2\mu^6 \text{CRR}(2) + \ldots
+ \mu \text{CRR}(3) + 2\mu^5 \text{CRR}(3) + 2\mu^7 \text{CRR}(3) + \ldots
+ \ldots
\]

\[
= \{\mu^2 \text{CRR}(0) + \mu(1+\mu^2)\text{CRR}(1) + (1+\mu^4)\text{CRR}(2) + \mu(1+\mu^4)\text{CRR}(3)+\ldots\} / (1-\mu^2)
\]

by induction it follows that if \(k \geq 2\),

\[
\text{CRR}(k) = \{\mu^k \text{CRR}(0) + \mu^{k-1} (1+\mu^2)\text{CRR}(1) + \ldots
+ \mu^2(1+\mu^2)\text{CRR}(k-1) + (1+\mu^2)^2\text{CRR}(k) + \mu(1+\mu^2)^2\text{CRR}(k+1) + \ldots\} / (1-\mu^2)
\] (5-7)

The autocorrelation function of the B-DARMA-E daily precipitation process has a cut-off after the first lag as discussed in chapters 2 and 3. For the data in Indiana the B-DARMA-E was shown to be suitable for the daily precipitation time series. Based on the daily precipitation input as a B-DARMA-E process, the autocorrelation function of \(\{Q_k\}\) can be simplified as

\[
\text{CRR}(k) = \{\mu^k + \mu^{k-1} (1+\mu^2)\text{CRR}(1)\} / (1-\mu^2) .
\] (5-8)

The mean function of the streamflows can be expressed by using the equation (5-6) as follows:

\[
E\{(1-\mu B)Q_{k-1}\} = E(R_k - L_j)
\] (5-9)

Since \(E\{Q_{k-1}\} = E\{Q_k\} \) and \(L_j \) is a constant, then
\[ E(Q_k) = \{E(R_k) - L_j\} / (1 - \mu) \]  

(5-9a)

These statistical properties will be used for the estimation of the parameters in the T-DARMA(0,1,1,0,1) model.

V.3.1.2. The T-DARMA(0,1,1,1,1) model

Following the definition in equation (5-5), the T-DARMA(0,1,1,1,1) model is defined as follows:

\[ (1 - \mu B)Q_k = (1 - \nu B)R_k - L_j, \]  

(5-10)

where \( \nu R_{k-1} \) represents the fraction of the precipitation falling in the previous day that contributes to the present daily streamflow, while a single autoregressive parameter \( \mu \) means that the watershed can be conceptually represented by an one-layer storage and \( L_j \) is the loss term. As defined before, a conceptual graph can be sketched as follows:

\[ \text{Solving for the daily flow, } Q_k, \text{ equation (5-10) becomes} \]

\[ \text{Diagram showing flow and storage processes with symbols and arrows indicating inputs and outputs.} \]
\[ Q_k = \{R_{k-\nu}R_{k-\nu-1}L_j\} / (1-\mu B) \]
\[ = R_k + (\mu-\nu)R_{k-1} + \mu(\mu-\nu)R_{k-2} \]
\[ + \mu^2(\mu-\nu)R_{k-3} + \mu^3(\mu-\nu)R_{k-4} + \ldots + L_j (1-\mu)^{-1} \]

Let \( \psi = \mu-\nu \), then

\[ Q_k = R_k + \psi R_{k-1} + \psi \mu R_{k-2} + \psi \mu^2 R_{k-3} \]
\[ + \psi \mu^2 R_{k-4} + \psi \mu^3 R_{k-5} + \ldots + L_j (1-\mu)^{-1} \]

This expression is similar to that used in the unit hydrograph (Linsley et al., 1972)

By assuming that the input is a B-DARMA-E daily precipitation model, the autocorrelation function of the T-DARMA(0,1,1,1,1) model can be derived in the same way as shown in the case of the T-DARMA(0,1,1,0,1) process. Since the autocorrelation function of the B-DARMA-E model has a cut-off after the first lag, defining \( R_0 \) as before and assuming that \( \text{CRR}(0) = 1 \), and \( \text{CRR}(k) = 0 \), if \( k \geq 2 \), then,

\[ \text{CRQ}(1) = E \{ (R'_{k-1} + \psi R'_{k-2} + \psi \mu R'_{k-3} + \psi \mu^2 R'_{k-4} + \psi \mu^3 R'_{k-5} + \ldots) \} \]
\[ = \{(\psi + (\psi^2)/(1-\mu^2))\} \text{CRR}(0) + \{(\psi + (\psi^2)/(1-\mu^2))\} \text{CRR}(1) \]

Similarly substituting \( \text{CRR}(0) = 1 \),

\[ \text{CRQ}(3) = (\psi \mu^2 + (\psi^2 \mu^2)/(1-\mu^2)) + \{(\psi^2 \mu^3)/(1-\mu^2)\} \text{CRR}(1) \]

By induction, for \( k \geq 1 \),

\[ \text{CRQ}(k) = \{\psi \mu^k + (\psi^2 \mu^k)/(1-\mu^2)\} \]
\[ + \{\psi \mu^k + (\psi^2 \mu^k)/(1-\mu^2)\} \text{CRR}(1) \]

(5-12)
The relationship between the mean functions of \( \{Q_k\} \) and \( \{R_k\} \) can be obtained by taking the expectation of equation (5-10),

\[
E\{(1-\mu B)Q_k\} = E\{(1-\nu B)R_k - L_j\}
\]

(5-13)

Since \( E\{Q_{k-1}\} = E\{Q_k\} \), and \( E\{L_j\} = L_j \), and \( E\{R_k\} = E\{R_{k-1}\} \).

\[
E\{Q\} = \{(1-\nu_1)E(R) - L_j\}/(1-\mu)
\]

(5-13a)

In the next section these statistical properties will be used to estimate the parameters in the T-DARMA(0,1,1,1,1) model.

V.3.2. Estimation of parameters

The process from the observed precipitation to the observed streamflow includes many uncertain losses such as evaporation and deep percolation. Therefore, the effective rainfall concept used by previous authors as the input to the precipitation-streamflow model is not realistic. The losses may be studied through the stochastic process using the statistical relationships constructed in the last section. The relationships between the autocorrelation functions of precipitations and streamflows are used to estimate the parameters of the autoregressive moving average component of the T-DARMA model. Then the mean functions of both series are related to obtain the parameter of the uncertain loss.
It is desirable to preserve the autocorrelation structure in the daily streamflow sequence using the autocorrelation structure of the corresponding daily precipitation sequence. Suppose that the B-DARMA-E daily precipitation process is used. Equations (5-8) and (5-12) give the autocorrelation relationships between \( \{R_k\} \) and \( \{Q_k\} \), respectively, for the T-DARMA(0,1,1,0,1) and the T-DARMA(0,1,1,1,1) models. These equations are nonlinear functions of the transfer model parameters, while the CRR(1) appearing in these equations is the known value of the first autocorrelation coefficient in the precipitation series. Therefore, the parameters can be estimated through the nonlinear least squares method by fitting the theoretical autocorrelation function to the empirical autocorrelations of \( \{Q_k\} \). The estimation program discussed in chapter 3 is applicable and is used in this study. After the parameters of the autoregressive moving average components of the transfer model are obtained, equations (5-9) and (5-13a) are applied to compute the loss factor \( L_j \), respectively, in the T-DARMA(0,1,1,0,1) and T-DARMA(0,1,1,1,1) models. Table 5.1 gives the estimated parameters of these two models for the streamflow station 3275006 in the Whitewater river near Alpine and the corresponding precipitation station 12-1229 at Cambridge, Indiana. Furthermore, the compared plots of the estimated and the theoretical autocorrelation functions are given in Figures (5-2) and (5-3). It is noted that the conversion of the streamflow unit, CFS, to the precipitation unit, INCH, for this watershed is .0009436 based on the drainage area of
square miles and the assumed Thiessen Polygon weight of 1.

V.3.3. The diagnostic checking by the residual series

With the estimated parameters, both models are used to generate daily streamflows by inputting the observed precipitations. Then the residual series are formed by subtracting the generated streamflows from the observed daily streamflow series. It is desirable to have a white noise in the residual series in order to verify the model. The autocorrelation functions of both residual series, respectively, from the T-DARMA(0,1,1,0,1) and the T-DARMA(0,1,1,1,1) models are computed and plotted in Figures (5-4) and (5-5). It can be seen from the figures that the residual autocorrelations are very close to a white noise. Therefore the models are acceptable.

Furthermore, the mean values of the generated series for different seasons are computed. Table 5.2. lists statistics of the generated series of the T-DARMA(0,1,1,0,1) and the T-DARMA(0,1,1,1,1) as well as those of the observed series in each season. The results show that the statistics of different seasons are reasonably preserved. Meanwhile the autocorrelations of the generated series are plotted for the comparisons with those of the observed series. Figures (5-8) and (5-9) show that the second order moment is well preserved in the generated series.
On the other hand, the plots of the generated daily streamflows by both models and their counterparts are simultaneously plotted along with the observed precipitation series in each year. The different seasons are generated separately, i.e. 90 days in one season, and then are combined into a 360-day series. Figures (5-6) and (5-7) provide the generated daily streamflows and their corresponding observations as well as the precipitation inputs. From these comparisons, it is concluded that the models are reasonably acceptable.

V.4. Summary

The study in this chapter is devoted to the construction of a model for the daily precipitation-streamflow process. Since the process from the falling precipitation to the streamflow involves many uncertain factors, it is best to study this process by the stochastic method. On the other hand some deterministic properties of the process are physically well understood and can be determined through statistical analyses (Eagleson, 1978a, 1978b). The Transfer Discrete Autoregressive Moving Average (T-DARMA(p,q,m,n,1)) model built in this chapter is a stochastic as well as a conceptual model based on the above understanding.

The discrete autoregressive moving average components are used to model the random process from the k-th daily precipitation to the k-th daily streamflow. In the process
the discrete autoregressive component represents the characteristics of the watershed, i.e. the order \( m \) means that the watershed may be separated into \( m \)-layer storages; while the discrete moving average component represents the characteristics of the system input, i.e. the order \( n \) means that the input series of previous \( n \) days conceptually contribute or extract water to or from the system depending on the negative or positive parameters. The parameters of the discrete autoregressive moving average process are estimated through the autocorrelation functions of the streamflow series and of their corresponding precipitation series. Then the loss which is known to be indispensible (Eagleson, 1978a) is estimated from the mean functions of the two series. Therefore, the T-DARMA\((p,q,m,n,1)\) model not only considers the stochastic properties of the process, but also takes into account the water balance, which was described by Eagleson (1978a, 1978b).

Two particular models, T-DARMA\((0,1,1,0,1)\) and T-DARMA\((0,1,1,1,1)\), are used to model the daily precipitation-streamflow data in Indiana. The model construction from the estimation to the checking of residual series provides a promising way for model building. Especially, the relationship of the autocorrelations between the input and the output series gives a strong foundation for the stochastic modeling from the precipitation to the streamflow since the preservation of these relations preserves the persistence properties of the process. Furthermore, the
relationship of the mean functions between the two series connects the idea of water balance to a stochastic process.

From the generation point of view, the autocorrelation function of the residual series shows that the daily streamflow series generated by inputting the daily precipitation are quite satisfactory. The compared plots of the generated and the corresponding observed series are very satisfactory. Moreover, the comparison of mean values of generated series with those of observed series seems to conclude that the T-DARMA(0,1,1,1,1) is better than the T-DARMA(0,1,1,0,1).
Table 5.1. Estimated parameters of transfer models
Station: 12-1229 and 3275000

<table>
<thead>
<tr>
<th>Season</th>
<th>T-DARMA $(0,1,1,0,1)$</th>
<th>T-DARMA $(0,1,1,1,1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\mu$</td>
<td>$L_j$</td>
</tr>
<tr>
<td>1</td>
<td>.89889</td>
<td>.02103</td>
</tr>
<tr>
<td>2</td>
<td>.98991</td>
<td>.07467</td>
</tr>
<tr>
<td>3</td>
<td>.89392</td>
<td>.08503</td>
</tr>
<tr>
<td>4</td>
<td>.79881</td>
<td>.04305</td>
</tr>
</tbody>
</table>
Table 5.2. Mean values of observed and generated series

Stations: 12-1229 and 3275000

<table>
<thead>
<tr>
<th>Season</th>
<th>Observed</th>
<th>T-DARMA ((0,1,1,0,1))</th>
<th>T-DARMA ((0,1,1,1,1))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>329.4</td>
<td>972.4</td>
<td>930.5</td>
</tr>
<tr>
<td>2</td>
<td>717.3</td>
<td>704.7</td>
<td>723.2</td>
</tr>
<tr>
<td>3</td>
<td>263.2</td>
<td>283.3</td>
<td>279.3</td>
</tr>
<tr>
<td>4</td>
<td>318.3</td>
<td>343.1</td>
<td>332.2</td>
</tr>
</tbody>
</table>
Figure (5-1) Empirical autocorrelations of daily streamflows; station: 3275000
Figure (G-2) Empirical autocorrelation functions, station: 3275000
Figure(5-3)  Empirical and T-DARMA(0,1,1,1,1) model autocorrelation functions: station: 3275000
Figure (3-4) Autocorrelations of the residual series based on T-DARMA(0,1,1,0,1) models; station: 3275000
Figure (5-5) Autocorrelations of the residual series based on T-DARMA(0,1.1.1,1.1) models; station: 3275000
Figure (5-6) Generated daily flows by the T-DARMA(0,1,1,0,1) model and its observed counterparts; station: 3275000
Figure (C-7) Generated daily flows by the T-DARMA(0,1,1,1,1) model and its observed counterparts; station 3275000
Figure (5.8) Autocorrelations of T-DARMA(0, 1, 1, 0, 1) generated and observed series, station: 3275000
Figure (5-9) Autocorrelations of T-DARMA(0,1,1,1,1) generated and observed series; station: 3275000
VI.1. The B-DARMA-E modeling

The binary discrete autoregressive moving average process (B-DARMA) as a model for the wet-dry precipitation sequence was reported to be reasonably good by Buishand (1978). Nevertheless, the DARMA(1,1) was forcibly used for all the wet-dry sequences which could vary from one climate to another. The modeling results in this study show that DARMA(1,1) is not always the most suitable model for the data in Indiana. Therefore, it is more appropriate to identify the suitable model from a family of DARMA processes, which were constructed in the previous chapters of this thesis, in order to fit the process to different climatological regions.

The magnitude of the precipitation in the B-DARMA process is expressed by an exponential distribution which is shown to be suitable for fitting the nonzero precipitation sequence. Thus the B-DARMA-E model is formulated. The magnitudes of precipitations in each season form a nonzero precipitation series which are verified to be independent of their occurrence time intervals. Since the B-DARMA process can replicate the persistencies of wet and dry spells, it is reasonable to combine the B-DARMA process with the exponential distribution to form the B-DARMA-E model. The
results of fitting this model to the precipitation data in Indiana are satisfactory.

The model identification by means of the autocorrelation function is convenient for the choice of a suitable combination of the autoregressive and the moving average components. This is done by visual comparison of the empirical and the theoretical autocorrelation plots. This method also provides a wide selection of possible models before a final model is selected so that a truly good model will not be neglected.

The initial values of parameters, using the estimator in equation (3-2) suggested by Jacobs and Lewis (1978b), is shown to be capable of smoothing the geometrically decreasing tail of the serial correlations, and is efficient in the obtainment of the iterated results by the nonlinear least squares method.

The autocorrelation function is a very useful tool to investigate a stochastic process. Klemes (1978) in a discussion of the transfer rainfall-runoff modeling emphasized that the relation between the two processes can be obtained by equating the respective autocorrelation functions. This procedure is also used in this paper for the T-DARMA process. The results of the illustrative example in Chapter 3 show that the second moment property of the precipitation is well preserved by the B-DARMA-E model.
Run lengths of hydrological time series have been discussed in two different ways: one concentrates on the physical study (Saldarriaga and Yevjevich, 1970; Yevjevich, 1972; Sen, 1979); the other is derived for the purpose of mathematical discussions on modeling (Jacobs and Lewis, 1978a, 1978b; Buishand, 1973). This research starts with the mathematical modeling of the stochastic process and ends up with the applications of the constructed models to hydrology. Thus the run lengths defined in this paper aim at both mathematical and physical achievement. The definitions of run lengths in chapter 2 are used for the diagnostic checking of the B-DARMA-E model building. The probability distribution of the run length of 0 is equivalent to that of the dry period, which, for example, provides information to decide whether an irrigation is needed in a season. On the other hand, the probability distribution of the run length of 1 is equivalent to that of the wet period, which, for example, gives information for deciding the planting day in a farming area.

From the illustrative example on the B-DARMA-E modeling in chapter 3, the three-step procedure of identification, estimation, and diagnostic checking is efficient in fitting the model to the daily precipitation time series. Meanwhile the criterion of the best model selection by the minimum sum of squared errors of run length distributions provides a validation of model’s suitability in preserving the
persistence of wet and dry spells.

III.2. The \textit{M-DARMA} process

For the study of daily precipitation series, all previous works as reviewed in chapter 1 are inclined towards the modeling of the occurrence processes as wet-dry sequences because they were shown to be easily modeled by the Markov chain or other discrete models which can preserve the sequence persistence. The \textit{B-DARMA-E} model built in this work includes the submodel \textit{B-DARMA}, which is used to model the wet-dry sequence. The attached exponential model is able to give the magnitude of the precipitation in the discrete modeling, while the family of the \textit{B-DARMA} processes can provide more flexibility in the model selection. A significant improvement could be achieved if the discrete process can carry the discrete precipitation quantities in the model and also can picture the different state persistances in a single process. The \textit{M-DARMA} model based on the above idea is given in chapter 4 and provides very promising results.

For the discussion of the \textit{M-DARMA} process, the autocorrelation function is used to check the second order properties of the observed precipitations in a transformed discrete series. By the comparisons of autocorrelation plots in Figures (4-1a) to (4-1d), it is concluded that the
3-state discrete daily precipitation series used in this research is reasonably adequate.

By means of the newly defined multi-state run lengths for diagnostic checking, the three steps of model building used in the B-DARMA-E process are still applicable and equally efficient. The identification and the estimation procedures through the autocorrelations are similar to those in the B-DARMA-E model.

Conceptually the infinite-state series means the real-valued continuous observations. However it is extremely difficult to discuss the run length, if the states in a discrete series are too many. In fact, it is impossible to obtain the run length physically in an infinite-state series since it is no longer discrete. Therefore, finding suitable finite states for the discrete daily precipitation time series is critical in the construction of the M-DARMA model. In chapter 4, a three-state discrete series is shown to be adequate and their multi-state run lengths are defined. By statistical derivations, the probability distributions of 3-state run lengths are constructed in chapter 2 and are used for checking the M-DARMA modeling.

Furthermore, the studies of floods and droughts have caused many discrepancies due to different definitions (Changnon, 1980). Physically a flood should not only include the high peak, but should also depend on how large
the discharge volume will be and how long the high water stage will last, i.e. the flood volume and duration are equally important. So does the drought, i.e. the severity of a drought should include its lowest volume and its duration besides the occurrence time. Thus it is more suitable to study floods and droughts through a model which can provide the persistancies in every state. The modeling results in chapter 4 convinces us that the M-DARMA model is able to meet the above needs.

III.3. The T-DARMA process

The transfer system from the precipitation to the streamflow not only physically includes the concept of water balance where the law of conservation of mass has to hold, but also depends on many uncertain factors, such as direct evaporation, evapotranspiration, infiltration, and groundwater storage. In reality, it is known that each factor above may contribute to or extract water from the system. Nevertheless it is not certain how it goes and how much it does. Thus it is believed that a stochastic system should be used in order to describe this kind of phenomenon.

Based on the precipitation models, B-DARMA-E and M-DARMA, whose statistical properties were previously constructed, the autocorrelation relationship of the input daily precipitation and the output daily streamflow was studied
and used to construct the T-DARMA model. On the other hand, water balance is considered by the relationship of mean functions, which is done by taking the expectation of both sides of equation (5-5). Therefore, the T-DARMA process conceptually introduces the physical law of conservation to a stochastic model. These statistical connections between the two hydrological time series strongly show that the T-DARMA model, which is based on the above relationship, is an adequate process to describe the daily precipitation-streamflow phenomenon. The T-DARMA model is a stochastic process since the stochastic precipitation is used as the input.

In order to preserve the autocorrelation property of a daily streamflow series through the input of the daily precipitation series, equations (5-3) and (5-12) are used to estimate the parameters of the autoregressive moving average process based on the known autocorrelations of the input series. These nonlinear equations of model parameters can be solved by the nonlinear least squares method. In order to keep the water balance in the process, equations (5-9a) and (5-13a) are used to estimate the loss term in the model. Therefore, the transfer process through the first two moments relates the precipitation to the streamflow.
CHAPTER VII
CONCLUSIONS AND RECOMMENDATIONS

VII.1. Conclusions

The following conclusions are arranged by the order of the three processes, B-DARMA-E, M-DARMA, and T-DARMA built in the previous chapters.

(1) The newly developed stochastic process, B-DARMA-E, consisting of a submodel, Binary Discrete Autoregressive Moving Average (B-DARMA) for the wet and dry days sequence, and an independent Exponential distribution for the nonzero precipitation sequence, adequately fits the daily precipitation time series in Indiana.

(2) The three-step procedure of model building developed for the B-DARMA, is simple and effective in modeling the daily precipitation time series. The identification step based on the autocorrelation function is quite suitable since the second moment is easily obtained and is well understood by engineers. Moreover, the estimation step can serve to check the identification as well as to compute the parameters of the identified models.

(3) The diagnostic checking step based on the discrete Kolmogorov-Smirnov method for testing the goodness-of-fit of the probability distributions of the run lengths of 0 and 1, which govern the characteristics of the B-DARMA process is
statistically efficient. Furthermore, the criterion for the selection of the best model is to choose the one with the smallest error between the estimated and the theoretical run length probability distributions.

(4) The persistence of the wet and the dry spells can effectively be preserved by the models by preserving the run lengths of 0 and 1. The behaviors of floods and droughts can be interpreted through the probability distributions of the run lengths of 0 and 1 respectively since the latter are statistically defined.

(5) A Multi-state Discrete Autoregressive Moving Average (M-DARMA) model is developed for the modeling of the daily precipitation time series which can be transformed into a multi-state discrete sequence, where the transformation is based on the preservation of the second moment property of the observed series. A 3-state daily precipitation series was found to be a reasonably acceptable multi-state discrete series and was used to illustrate the modeling of the M-DARMA process.

(6) The three-step procedure including the identification, the estimation, and the diagnostic checking, developed for the M-DARMA model building is demonstrated to be effective by an illustrative example. The identification step uses the autocorrelation function as a main tool. The estimation scheme by the nonlinear least squares method
through the autocorrelation function not only makes the fine estimation of the parameters but is also used as a further check of the suitability of the candidate models identified in the identification step.

(7) The concept of the multi-state run length is established for the purpose of diagnostic checking in the M-DARMA model building. This diagnostic checking scheme using the multi-state run lengths is a suitable and sensitive method for testing the model. Moreover, the selection of the best model based on the best preservation of the probability distributions of the multi-state run lengths is a reasonable method since it is conceptually suitable to preserve the persistance of the observed precipitation time series which are described by the run lengths.

(8) One of the advantages of the multi-state runs is that it brings the wet and the dry periods into a more understandable form, i.e. the highest-state run represents floods while the lowest-state run represents droughts. Since the probability distributions of the different run lengths can be obtained by the conditional probability expressions given in chapter 2, these distributions are useful in applications such as irrigation or reservoir operation where the informations about floods and droughts are important.
(9) The Transfer Discrete Autoregressive Moving Average (T-DARMA(p,q,m,n,1)) model built in this study is a stochastic as well as a conceptual model, while the model parameters are estimated through the autocorrelation functions of the streamflow series.

(10) Two particular models, T-DARMA(0,1,1,0,1) and T-DARMA(0,1,1,1,1), are used to model the daily precipitation-streamflow data in Indiana. The model construction from the estimation to the checking of residual series provides a promising way for model building. Especially, the relationship of the autocorrelations between the input and the output series gives a strong foundation for the stochastic modeling from the precipitation to the streamflow since the preservation of these relations preserves the persistence properties of the process. Furthermore, the relationship of the mean functions between the two series connects the water balance to a stochastic process.

VII.2. Recommendations

(1) The exponential distribution in a B-DARMA-E model can be improved through modeling the process in an independent wet spell instead of forming the nonzero precipitation sequence, which has to rely on the assumption that precipitation occurrences are independent of their occurrence time intervals.
(2) In the M-DARMA model the precipitation magnitude can be distinguished better by increasing the state number, which will however complicate the run length structure. Therefore, further studies are needed to decide the most suitable discrete state number.

(3) The particular T-DARMA models constructed in this research provide very promising results. Nevertheless the general modeling of the T-DARMA(p,q,m,n,1), whose general statistical structure is no longer as simple as the particular cases given in this study, is waiting to be pursued.
BIBLIOGRAPHY


Court, A. "Hydrology Precipitation Research, 1975-1976", Reviews of Geophysics and Space Physics,
Vol. 17, No. 6, 1979.


Richardson, C. W. "Stochastic models of Daily Precipitation, Temperature, and Solar


1981.


APPENDIX A

COVARIANCE STRUCTURES OF PARTICULAR DARMA MODELS

In this appendix, the second order moments are given for the DARMA models \((p,0)\), \((0,q)\), \((1,1)\), \((1,2)\), and \((2,1)\).

A.1. Correlation functions for the DARMA\((p,0)\)

From the definition given in equation (2-4), \(\{Y_n\}\) is a sequence of discrete random variables with a specified marginal distribution, \(\pi\), i.e.

\[
P(Y_n=k) = \pi_k, \quad k=0,1,2,... \quad (A-1)
\]

It is assumed that \(\{X_n\}\) starts with \(X_0\) which has the same probability distribution as \(\{Y_n\}\) and that

\[
P(Y_n=1) = 1-P(Y_n=0) = \rho, \quad \text{for } 0 \leq \rho < 1 \quad (A-2)
\]

Let \(\{m_2\}\) be a sequence of independent random variables taking values in \(\{1,2,\ldots,p\}\) with

\[
P\{m_2=i\} = \alpha_i, \quad i=1,2,\ldots,p \quad (A-3)
\]

where,

\[
\alpha_1 + \alpha_2 + \ldots + \alpha_p = 1 \quad (A-3a)
\]

Define

\[
\hat{\rho} = E(X_n) = E(Y_n) \quad (A-4)
\]

By equations (2-4) and (2-4a) the probability that

\(X_{n+k} = X_{n+k-j}\) is \(\rho \alpha_j, \quad j=1,2,\ldots,p\).

With

\[
X'_n = X_n - \hat{\rho} \quad (A-5)
\]

the correlation function of \(\{X_n\}\) can be derived from
equation (2-4) as follows:

\[ \text{Corr}(i) = E(X'_{n}X'_{n-1}) \]
\[ = E(X'_{n-1}V_{n}X'_{n-m2}) + E(X'_{n-1}(1-V_{n})Y_{n}X'_{n-m2}) \]

Since \( Y_{n} \) and \( X_{n-1} \) are independent, it follows that

\[ \text{Corr}(i) = E(X'_{n-1}V_{n}X'_{n-m2}) \] (A-7)

Thus equation (A-7) yields the result that

\[ \text{Corr}(i) = \rho \alpha_{1} \text{Corr}(0) + \rho \alpha_{2} \text{Corr}(1) + \ldots + \rho \alpha_{p} \text{Corr}(p-1) \] (A-8)

In the same way the following expressions can be derived

\[ \text{Corr}(2) = \rho \alpha_{1} \text{Corr}(1) + \rho \alpha_{2} \text{Corr}(0) + \ldots + \rho \alpha_{p} \text{Corr}(p-2) \]

\[ \ldots \]

\[ \text{Corr}(p) = \rho \alpha_{1} \text{Corr}(p-1) + \rho \alpha_{2} \text{Corr}(p-2) + \ldots + \rho \alpha_{p} \text{Corr}(0) \]

and for \( k > p \)

\[ \text{Corr}(k) = \rho \alpha_{1} \text{Corr}(p+k-1) + \rho \alpha_{2} \text{Corr}(p+k-2) + \ldots + \rho \alpha_{p} \text{Corr}(k) \] (A-9)

A.2. Correlation functions for DARMA(0,q)

Let

\[ P(m_{2}=i) = \beta_{i}, \quad i=0,1,\ldots,q-1. \] (A-10)

where

\[ \beta_{0} \beta_{1} \ldots + \beta_{q-1} = 1 \] (A-10a)

By the definition in equation (2-6) the covariance function of \( \{X_{n}\} \) can be derived as follows:

for \( i \leq j \leq q-1 \),
\[
\text{Cov}(X_{n+j}, X_n) = \text{Cov}(Y_{n+j-m_3}, Y_{n-m_3})
\]
\[
= \sum_{k=j}^{q-1} \text{Var}(Y) \sum_{m_3=k}^{q-1-j} P(m_3=k) P(m_3=j-k)
\]
\[
= \sum_{k=0}^{q-1} P(m_3=k) P(m_3=j+k)
\]
\[
= \sum_{k=j}^{q-1} \beta_k \beta_{k-j}
\]
\[
\text{Corr}(j) = \frac{\text{Cov}(j)}{\sqrt{\text{Var}(Y_j) \cdot \text{Var}(Y_{n-j})}}
\]
\[
\text{Corr}(j) = 0
\]

where \(m_3\) take values in \(\{0, 1, 2, \ldots, q-1\}\).

Therefore, for \(1 \leq j < q-1\),
\[
\text{Corr}(j) = \sum_{k=j}^{q-1} \beta_k \beta_{k-j}
\]

and for \(j \geq q\),
\[
\text{Corr}(j) = 0
\]

A.3. Covariance functions for DARMA(1,1)

Since \(\{X_n\}, \{A_n\}, \{Y_n\}\) are stationary with the same mean, the covariance function can be derived from equations (2-15) and (2-16) follows:
\[
\text{Cov}(X_n, X_{n-k}) = E(X_nX_{n-k}) - E(X_n)E(X_{n-k})
\]
\[
= \beta^2 E(Y_nY_{n-k}) - E(Y_n)E(Y_{n-k})
\]
\[
+ \beta(1-\beta) [E(Y_nA_{n-1-k}) - E(Y_n)E(A_{n-1-k})]
\]
\[
+ \beta(1-\beta) [E(A_{n-1}Y_{n-k}) - E(A_{n-1})E(Y_{n-k})]
\]
\[
+ (1-\beta)^2 [E(A_{n-1}A_{n-1-k}) - E(A_{n-1})E(A_{n-1-k})]
\]
\[
(A-14)
\]

It is assumed that \(\{Y_n\}\) is a sequence of independent random variables with an identical distribution, i.e.
\[
E(Y_nY_{n-k}) = E(Y_n)E(Y_{n-k})
\]
\[
(A-15)
\]
From the independence between \( \{A_{n-1-k}\} \) and \( \{Y_n\} \) it follows that

\[
E(Y_n A_{n-1-k}) = E(Y_n) E(A_{n-1-k}) \tag{A-16}
\]

Furthermore, by the same reasons the following expression can be derived.

\[
E(A_{n-k} Y_{n-k}) - E(A_{n-k}) E(Y_{n-k}) = (1 - \rho) \{ E(Y_{n-k})^2 - E(Y_{n-k}) \}^2 \]
\[
= (1 - \rho) \text{Var}(Y) \tag{A-17}
\]

and for \( i=1, 2, \ldots, k-1 \),

\[
E(A_{n-i} Y_{n-k}) - E(A_{n-i}) E(Y_{n-k}) = \rho \{ E(A_{n-i} Y_{n-k}) - E(A_{n-i}) E(Y_{n-k}) \} \tag{A-18}
\]

By induction method the third term in equation (A-14) can be simplified as follows:

\[
E(A_{n-1} Y_{n-k}) - E(A_{n-1}) E(Y_{n-k}) = (1 - \rho) \rho^{k-1} \text{Var}(Y) \tag{A-19}
\]

and

\[
E(A_{n-1} A_{n-1-k}) - E(A_{n-1}) E(A_{n-1-k}) = \rho^k \text{Var}(Y) \tag{A-20}
\]

Replacing above expressions in equation (A-14), the covariance function is obtained as

\[
\text{Cov}(k) = 8 (1 - \beta) (1 - \rho) \rho^{k-1} \text{Var}(Y) + (1 - \beta)^2 \rho^k \text{Var}(Y)
\]
\[
= (1 - \beta) (1 - \beta - 2 \rho \beta) \rho^{k-1} \text{Var}(Y) \tag{A-21}
\]

A.4. Covariance functions of the DARMA(1,2) model

Based on the definition in equation (2-18), the covariance structure of the DARMA(1,2) process can be derived as follows:
\[ \text{Cov}(k) = \text{Cov}(X_n, X_{n-k}) = E(X_nX_{n-k}) - E(X_n)E(X_{n-k}) \]
\[ = \gamma_1^2 \{ E(Y_nY_{n-k}) - E(Y_n)E(Y_{n-k}) \} \]
\[ + \gamma_1 \gamma_2 \{ E(Y_nY_{n-k-1}) - E(Y_n)E(Y_{n-k-1}) \} \]
\[ + \gamma_1 (1 - \gamma_1 - \gamma_2) \{ E(Y_nA_{n-k-2})E(Y_n)E(A_{n-k-2}) \} \]
\[ + \gamma_1 \gamma_2 \{ E(Y_{n-1}Y_{n-k}) - E(Y_{n-1})E(Y_{n-k}) \} \]
\[ + \gamma_2^2 \{ E(Y_{n-1}Y_{n-k-1}) - E(Y_{n-1})E(Y_{n-k-1}) \} \]
\[ + \gamma_2 (1 - \gamma_1 - \gamma_2) \{ E(Y_{n-1}A_{n-k-2})E(Y_{n-1})E(A_{n-k-2}) \} \]
\[ + \gamma_1 (1 - \gamma_1 - \gamma_2) \{ E(A_{n-2}Y_{n-k}) - E(A_{n-2})E(Y_{n-k}) \} \]
\[ + \gamma_2 (1 - \gamma_1 - \gamma_2) \{ E(A_{n-2}Y_{n-k-1}) - E(A_{n-2})E(Y_{n-k-1}) \} \]
\[ + (1 - \gamma_1 - \gamma_2)^2 \{ E(A_{n-2}A_{n-k-2}) - E(A_{n-2})E(A_{n-k-2}) \} \]
\[ (A-22) \]

In equation (A-22), the first six terms are simply equal to 0, while the last three terms can be obtained in the same way as in A.3. They are given as follows:

\[ E(A_{n-2}Y_{n-k}) - E(A_{n-2})E(Y_{n-k}) = (1 - \rho) \rho ^{k-2} \text{Var}(Y) \]  \[ (A-23) \]

\[ E(A_{n-2}Y_{n-k-1}) - E(A_{n-2})E(Y_{n-k-1}) = (1 - \rho) \rho ^{k-1} \text{Var}(Y) \]  \[ (A-23a) \]

\[ E(A_{n-2}A_{n-k-2}) - E(A_{n-2})E(A_{n-k-2}) = \rho ^k \text{Var}(Y) \]  \[ (A-23b) \]

Therefore, for \( k \geq 2, \)

\[ \text{Cov}(k) = (1 - \gamma_1 - \gamma_2) \rho ^{k-2} \{ (1 - \gamma_1 - 2\gamma_2) \rho ^2 - (\gamma_1 - \gamma_2) \rho + \gamma_1 \} \text{Var}(Y) \]  \[ (A-24) \]

A.5. Covariance function of DARMA(2,1)

Let

\[ Y_n \text{ with probability } \beta, \]
\[ X_n = \]
\[ A_{n-1} \text{ with probability } 1 - \beta, \]  \[ (A-25) \]
where

\[ Y_n \text{ with probability } 1-\rho, \]
\[ A_n = A_{n-1} \text{ with probability } \rho \alpha, \]
\[ A_{n-2} \text{ with probability } \rho (1-\alpha). \]

When the sequence \( \{Y_n\} \) is identically and independently \( \pi \)-distributed, it follows that \( \{A_n\} \) becomes a DAR(2) process (Jacobs and Lewis, 1973c). Thus \( \{X_n\} \) has the same marginal distribution and mean as \( \{Y_n\} \), i.e.

\[ P(X_n=k) = \beta P(Y_n=k) + (1-\beta) P(A_{n-1}=k) \]
\[ = \pi_k \quad (A-26) \]

Furthermore, the covariance function of the DARMA(2,1) model can be derived as follows:

\[ \text{Cov}(k) = \text{Cov}(X_n, X_{n-k}) = \text{E}(X_n X_{n-k}) - \text{E}(X_n) \text{E}(X_{n-k}) \]
\[ = \sigma^2 \{ \text{E}(Y_n Y_{n-k}) - \text{E}(Y_n) \text{E}(Y_{n-k}) \} \]
\[ + \beta(1-\beta) \{ \text{E}(Y_n A_{n-k-1}) - \text{E}(Y_n) \text{E}(A_{n-k-1}) \} \]
\[ + \beta(1-\beta) \{ \text{E}(A_{n-1} Y_{n-k}) - \text{E}(A_{n-1}) \text{E}(Y_{n-k}) \} \]
\[ + (1-\beta)^2 \{ \text{E}(A_{n-1} A_{n-k-1}) - \text{E}(A_{n-1}) \text{E}(A_{n-k-1}) \} \quad (A-27) \]

From the independence between \( Y_n \) and \( Y_{n-k} \) and between \( Y_n \) and \( A_{n-k-1} \), the first two terms in equation (A-27) are equal to 0. Using the properties of the DAR(2) process, the covariance function of DARMA(2,1) is obtained from the last two terms as follows:

If \( k=1 \),

\[ \text{Cov}(1) = \beta(1-\beta)(1-\rho) \text{Var}(Y) + (1-\beta)^2 \{ \rho \alpha \text{Var}(Y) / (1-\rho(1-\alpha)) \} \quad (A-28) \]

If \( k=2 \),

\[ \text{Cov}(2) = \beta(1-\beta) \rho \alpha \text{Var}(Y) + (1-\beta)^2 \rho \alpha (1-\rho) / (\alpha + 2\rho - 1) \text{Var}(Y) / (1-\rho(1-\alpha)) \quad (A-28a) \]

If \( k>2 \),
\[ \text{Cov}(k) = \beta(1-\beta)\text{CCV}(k-1) + (1-\beta)^2\text{CV}(k) \] (A-29)

where \( \text{CCV}(k-1) \) is the \((k-1)\) lag cross covariance between \( \{Y_{n-k}\} \) and \( \{A_{n-1}\} \) by assuming \( \text{CCV}(0) \) is the variance of \( \{Y_n\} \), i.e.

\[ \text{CCV}(k-1) = \rho(1-\rho)\text{CCV}(k-2) + (1-\rho)\text{CCV}(k-3), k=3,4,... \] (A-29a)

On the other hand \( \text{CV}(k) \) is the \(k\)-th autocovariance function of \( \{A_n\} \), i.e.

\[ \text{CV}(k) = \rho(1-\rho)\text{CV}(k-1) + (1-\rho)\text{CV}(k-2), \quad k \geq 3 \] (A-29b)

Therefore, the covariance function of the DARMA(2,1) process is obtained as shown in equation (A-29).
APPENDIX B

BINARY AND MULTI-STATE RUN LENGTHS

B.1. DARMA(0,1) models

The definition of the binary run length was given in Section II.4. This appendix intends to derive the series \( \{a_n\} \), \( \{b_n\} \), and \( \{c_n\} \) which are used in the expressions of run lengths for binary and multi-state DARMA(0,1) models in chapter 2. For the binary DARMA process, \( P(X_0 = 0) = \pi_0 = 1 - \pi_1 \). Therefore,

\[
a_0 = P(X_0 = 1) = \pi_1
\]

\[
a_1 = P(X_0 = 1, X_1 = 0) = P(X_0 = 1, Y_1 = 0) + P(X_0 = 1, Y_0 = 0)(1 - \beta)
\]

\[
= a_0 \pi_0 + (1 - \beta)\{P(Y_0 = 1, Y_0 = 0) + P(Y_1 = 1, Y_0 = 0)(1 - \beta)\}
\]

\[
= \beta \pi_0 a_0 + (1 - \beta)^2 \pi_0 \pi_1 \quad (B-2)
\]

\[
a_2 = P(X_0 = 1, X_1 = 0, X_2 = 0)
\]

\[
= P(X_0 = 1, X_1 = 0, Y_2 = 0) + P(X_0 = 1, X_1 = 0, Y_1 = 0)(1 - \beta)
\]

\[
= \beta \pi_0 a_1 + (1 - \beta)\{P(X_0 = 1, Y_1 = 0) + P(X_0 = 1, Y_0 = 0, Y_1 = 0)\}
\]

\[
= \beta \pi_0 a_1 + \beta (1 - \beta) \pi_0 a_0 + (1 - \beta)^2 \pi_0 + (1 - \beta)^2 \pi_0 \pi_1
\]

\[
= \beta \pi_0 a_1 + \beta (1 - \beta) \pi_0 a_0 + (1 - \beta)^2 \pi_0^2 \pi_1 \quad (B-3)
\]

\[
a_3 = \beta \pi_0 a_2 + \beta (1 - \beta) \pi_0 a_1 + \beta (1 - \beta)^2 \pi_0 a_0 + (1 - \beta)^3 \pi_0^2 \pi_1
\]

\[
= \beta \pi_0 a_2 + \beta (1 - \beta) \pi_0 a_1 + \beta (1 - \beta)^2 \pi_0^2 \pi_1 \quad (B-4)
\]

and by induction,

\[
a_n = \beta \pi_0 a_{n-1} + \beta (1 - \beta) \pi_0 a_{n-2} + \ldots
\]

\[
+ \beta (1 - \beta)^n \pi_0 a_0 + (1 - \beta)^n \pi_0 \pi_1 \quad (B-5)
\]
Similarly, replacing \( P(X_k=1)=\pi_1 \) by \( P(X_k=0)=\pi_0 \) and \( P(X_k=0)=\pi_0 \) by \( P(X_k=1)=\pi_1 \), the sequence \( \{b_n\} \) is obtained as follows:

\[
\begin{align*}
  b_0 &= P(X_0=0)=\pi_0 \\
  b_1 &= P(X_0=0,X_1=1)=\beta \pi_1 b_0 + (1-\beta)^2 \pi_1 \pi_0 \\
  b_2 &= \beta \pi_1 b_1 + \beta (1-\beta) \pi_1 b_0 + (1-\beta)^3 \pi_1^2 \pi_0 \\
  b_n &= \beta \pi_1 b_{n-1} + \beta (1-\beta) \pi_1 b_{n-2} + \ldots \\
  &\quad + \beta (1-\beta)^{n-1} \pi_1 b_{n-2} + \ldots \\
  &\quad + \beta (1-\beta)^{n+1} \pi_1^{n-1} b_0 + (1-\beta)^n \pi_1^n \pi_0
\end{align*}
\]

and by induction,

\[
\begin{align*}
  b_n &= \beta \pi_1 b_{n-1} + \beta (1-\beta) \pi_1 b_{n-2} + \ldots \\
  &\quad + \beta (1-\beta)^{n-1} \pi_1 b_{n-2} + \ldots \\
  &\quad + \beta (1-\beta)^{n+1} \pi_1^{n-1} b_0 + (1-\beta)^n \pi_1^n \pi_0
\end{align*}
\]

In the multi-state DARMA models a finite number of states are assumed. Thus \( \{c_n\} \) for the multi-state run length in the DARMA(0,1) process can be derived by replacing \( P(X_k=1)=\pi_1 \) by \( P(X_k\neq i)=1-\pi_1 \) and \( P(X_k=0)=\pi_0 \) by \( P(X_k=i)=\pi_i \) in the derivation of \( \{a_n\} \). Doing these replacements,

\[
\begin{align*}
  c_0 &= P(X_0=i)=1-\pi_i \\
  c_1 &= P(X_0\neq i,X_1=i)=\beta \pi_i c_0 + (1-\beta)^2 \pi_i (1-\pi_i) \\
  c_2 &= \beta \pi_i c_1 + \beta (1-\beta) \pi_i c_0 + (1-\beta)^3 \pi_i^2 (1-\pi_i) \\
  c_3 &= \beta \pi_i c_2 + \beta (1-\beta) \pi_i c_1 + \beta (1-\beta)^2 \pi_i^3 c_0 + (1-\beta)^4 \pi_i^3 (1-\pi_i)
\end{align*}
\]

and by induction,

\[
\begin{align*}
  c_n &= \beta \pi_i c_{n-1} + \beta (1-\beta) \pi_i c_{n-2} + \ldots \\
  &\quad + \beta (1-\beta)^{n-1} \pi_i c_{n-2} + \ldots \\
  &\quad + \beta (1-\beta)^{n+1} \pi_i^{n-1} c_0 + (1-\beta)^n \pi_i^n (1-\pi_i)
\end{align*}
\]

E.2. DARMA(1,1) models

This appendix is used to derive the transition matrices which are defined in chapter 2 for the binary and the multi-
state run lengths in the DARMA(1,1) models. First, following equations (2-15), (2-15a), (2-29), and (2-29a), the transition matrices, \( H_0 \) and \( H_1 \), for the binary DARMA(1,1) are derived as follows:

\[
H_0(0,0) = P(X_{n+1} = 0, A_{n+1} = 0 | A_n = 0) \\
= \beta \rho P(Y_{n+1} = 0, A_n = 0 | A_n = 0) \\
+ \beta (1-\rho) P(A_n = 0, Y_{n+1} = 0 | A_n = 0) \\
+ \rho (1-\beta) P(A_n = 0 | A_n = 0) \\
+ (1-\rho) (1-\beta) P(A_n = 0, Y_{n+1} = 0 | A_n = 0) \\
= \beta \rho \pi_0 \beta (1-\rho) \pi_0 + \rho (1-\beta) + (1-\rho) (1-\beta) \pi_0 \\
= \rho (1-\beta) + (1-\rho) (1-\beta) \pi_0 + \beta \rho \pi_0 \\
= \rho (1-\beta) + (1-\rho) (1-\beta) \pi_0 \\
= \rho (1-\beta) + (1-\rho) (1-\beta) \pi_0 
\]  

(B-16)

\[
H_0(0,1) = P(X_{n+1} = 0, A_{n+1} = 1 | A_n = 0) \\
= (1-\beta) (1-\rho) P(Y_{n+1} = 1 | A_n = 0) \\
= (1-\beta) (1-\rho) \pi_1 
\]  

(B-17)

\[
H_0(1,0) = P(X_{n+1} = 0, A_{n+1} = 0 | A_n = 1) \\
= (1-\rho) \beta P(Y_{n+1} = 0) = \beta (1-\rho) \pi_0 
\]  

(B-18)

\[
H_0(1,1) = P(X_{n+1} = 0, A_{n+1} = 1 | A_n = 1) \\
= \beta \rho P(Y_{n+1} = 0) \\
= \beta \rho \pi_0 
\]  

(B-19)

Similarly

\[
H_1(0,0) = P(X_{n+1} = 1, A_{n+1} = 0 | A_n = 0) \\
= \beta \rho P(Y_{n+1} = 1) \\
= \beta \rho \pi_1 
\]  

(B-20)
\[ H_0(0, 0) = P(X_{n+1} = 0, A_{n+1} = 0 | A_n = 0) \]
\[ = \beta \rho P(Y_{n+1} = 0) \]
\[ = \beta \rho \pi_0 \]
\[ = \rho (1 - \beta) \pi_0 \tag{B-24} \]

\[ H_1(1, 0) = P(X_{n+1} = 1, A_{n+1} = 0 | A_n = 1) \]
\[ = (1 - \beta)(1 - \rho) \pi_0 \]
\[ = (1 - \beta)(1 - \rho) \pi_1 \]

\[ H_1(1, 1) = P(X_{n+1} = 1, A_{n+1} = 1 | A_n = 1) \]
\[ = \beta \rho (1 - \rho) \pi_1 \]
\[ + \beta (1 - \rho) P(Y_{n+1} = 1 | A_n = 1) \]
\[ + (1 - \beta) P(A_n = 1 | A_n = 1) \]
\[ + (1 - \rho)(1 - \beta) P(A_n = 1, Y_{n+1} = 1 | A_n = 1) \]
\[ = \beta \rho \pi_1 + \beta (1 - \rho) \pi_1 + \rho (1 - \beta) + (1 - \rho)(1 - \beta) \pi_1 \]
\[ = \rho (1 - \beta) + (1 - \rho)(1 - \beta) \pi_1 \tag{B-23} \]

For the derivation of the transition matrices in the multi-state DARMA(1,1) model, the 3 states are assumed, i.e. \( \pi_0 + \pi_1 + \pi_2 = 1 \) in equation (2-42). By equations (2-15), (2-15a), (2-29), and (2-29a), the transition matrices, \( H_0 \), \( H_1 \), and \( H_2 \), are obtained as follows:

\[ H_0(0, 0) = P(X_{n+1} = 0, A_{n+1} = 0 | A_n = 0) \]
\[ = \beta \rho P(Y_{n+1} = 0, A_{n+1} = 0 | A_n = 0) + \beta (1 - \rho) P(Y_{n+1} = 0) \]
\[ + (1 - \beta) P(A_n = 0 | A_n = 0) + (1 - \beta)(1 - \rho) P(Y_n = 0) \]
\[ = \rho (1 - \beta) + (1 - \rho)(1 - \beta) \pi_0 \tag{B-24} \]
\[ \text{H}_0(0,1) = P(X_{n+1} = 0, A_{n+1} = 1 | A_n = 0) \]
\[ = (1-\beta)(1-\rho)P(Y_{n+1} = 1) \]
\[ = (1-\beta)(1-\rho)\pi_1 \]  \hspace{1cm} (B-25)

\[ \text{H}_0(0,2) = P(X_{n+1} = 0, A_{n+1} = 2 | A_n = 0) \]
\[ = (1-\beta)(1-\rho)P(Y_{n+1} = 2) \]
\[ = (1-\beta)(1-\rho)\pi_2 \]  \hspace{1cm} (B-26)

\[ \text{H}_0(1,0) = P(X_{n+1} = 0, A_{n+1} = 0 | A_n = 1) \]
\[ = \beta(1-\rho)P(Y_{n+1} = 0) = \beta(1-\rho)\pi_0 \]  \hspace{1cm} (B-27)

\[ \text{H}_0(1,1) = P(X_{n+1} = 0, A_{n+1} = 1 | A_n = 1) \]
\[ = \beta\rho P(Y_{n+1} = 0) = \beta\rho \pi_0 \]  \hspace{1cm} (B-28)

\[ \text{H}_0(1,2) = P(X_{n+1} = 0, A_{n+1} = 2 | A_n = 1) \]
\[ = \beta\rho P(Y_{n+1} = 0, A_n = 2 | A_n = 1) \]
\[ + \beta(1-\rho)P(Y_{n+1} = 0, Y_{n+1} = 2) \]
\[ + (1-\beta)\rho P(A_n = 0, A_n = 2 | A_n = 1) \]
\[ = \beta (1-\rho)P(Y_{n+1} = 0) = \beta(1-\rho)\pi_0 \]  \hspace{1cm} (B-29)

\[ \text{H}_0(2,0) = P(X_{n+1} = 0, A_{n+1} = 0 | A_n = 0) \]
\[ = \beta(1-\rho)P(Y_{n+1} = 0) = \beta(1-\rho)\pi_0 \]  \hspace{1cm} (B-30)

\[ \text{H}_0(2,1) = P(X_{n+1} = 0, A_{n+1} = 1 | A_n = 2) = 0 \]  \hspace{1cm} (B-31)

\[ \text{H}_0(2,2) = P(X_{n+1} = 0, A_{n+1} = 2 | A_n = 2) \]
\[ = \beta\rho P(Y_{n+1} = 0) = \beta\rho \pi_0 \]  \hspace{1cm} (B-32)

In a similar way
\begin{align}
H_1(0, 0) &= \mathbb{P}(X_{n+1} = 1, A_{n+1} = 0 | A_n = 0) \\
          &= \beta r \pi_1 \tag{B-33} \\
H_1(0, 1) &= \mathbb{P}(X_{n+1} = 1, A_{n+1} = 1 | A_n = 0) \\
          &= \beta (1-r) \pi_1 \tag{B-34} \\
H_1(0, 2) &= \mathbb{P}(X_{n+1} = 1, A_{n+1} = 2 | A_n = 0) = 0 \tag{B-35} \\
H_1(1, 0) &= \mathbb{P}(X_{n+1} = 1, A_{n+1} = 0 | A_n = 1) \\
          &= (1-\beta) (1-r) \pi_0 \tag{B-36} \\
H_1(1, 1) &= \mathbb{P}(X_{n+1} = 1, A_{n+1} = 1 | A_n = 1) \\
          &= \beta (1-\beta) + (1-\beta)(1-\beta) \pi_1 \tag{B-37} \\
H_1(1, 2) &= \mathbb{P}(X_{n+1} = 1, A_{n+1} = 2 | A_n = 1) \\
          &= (1-\beta) (1-r) \pi_2 \tag{B-38} \\
H_1(2, 0) &= \mathbb{P}(X_{n+1} = 1, A_{n+1} = 0 | A_n = 2) = 0 \tag{B-39} \\
H_1(2, 1) &= \mathbb{P}(X_{n+1} = 1, A_{n+1} = 1 | A_n = 2) \\
          &= \beta (1-\beta) \pi_1 \tag{B-40} \\
H_1(2, 2) &= \mathbb{P}(X_{n+1} = 1, A_{n+1} = 2 | A_n = 2) \\
          &= \beta r \pi_1 \tag{B-41} \\
H_2(0, 0) &= \mathbb{P}(X_{n+1} = 2, A_{n+1} = 0 | A_n = 0) \\
          &= \beta r \pi_2 \tag{B-42} \\
H_2(0, 1) &= \mathbb{P}(X_{n+1} = 2, A_{n+1} = 1 | A_n = 0) = 0 \tag{B-43} \\
H_2(0, 2) &= \mathbb{P}(X_{n+1} = 2, A_{n+1} = 2 | A_n = 0) \\
          &= \beta (1-r) \pi_2 \tag{3-44}
\end{align}
\[ H_2(1,0) = P(X_{n+1} = 2, A_{n+1} = 0 | A_n = 1) = 0 \]  \hfill (B-45)

\[ H_2(1,1) = P(X_{n+1} = 2, A_{n+1} = 1 | A_n = 1) \]
\[ = \beta \rho \pi_2 \]  \hfill (B-46)

\[ H_2(1,2) = P(X_{n+1} = 2, A_{n+1} = 2 | A_n = 1) \]
\[ = \beta (1 - \rho) \pi_2 \]  \hfill (B-47)

\[ H_2(2,0) = P(X_{n+1} = 2, A_{n+1} = 0 | A_n = 2) \]
\[ = (1 - \beta) (1 - \rho) \pi_0 \]  \hfill (B-48)

\[ H_2(2,1) = P(X_{n+1} = 2, A_{n+1} = 1 | A_n = 2) \]
\[ = (1 - \beta) (1 - \rho) \pi_1 \]  \hfill (B-49)

\[ H_2(2,2) = P(X_{n+1} = 2, A_{n+1} = 2 | A_n = 2) \]
\[ = \rho (1 - \beta) (1 - \rho (1 - \beta)) \pi_2 \]  \hfill (B-50)

B.3. DARMA(1,2) model

The definition of the DARMA(1,2) model was given in equations (2-16) and (2-18a), while its bivariate Markov chain for the run length is defined in equations (2-37) and (2-38). This appendix is aimed at deriving the transition matrices for the binary and the multi-state run lengths in the DARMA(1,2) model. First \( \{X_n\} \) is assumed to take values in \( \{0, 1\} \) with probabilities, \( \pi_0 \), and \( \pi_1 \), for the binary run length. Thus the transition matrices for the binary run length, \( H_0 \) and \( H_1 \), are derived as follows:
\( H_0(0, 0) = P(X_{n+1} = 0, A_n = 0 | A_{n-1} = 0) \)
\[ = \gamma_1 \rho P(Y_{n+1} = 0, A_{n-1} = 0 | A_{n-1} = 0) \]
\[ \times \gamma_1 (1-\rho) P(Y_{n+1} = 0, Y_n = 0) \]
\[ \times \gamma_2 P(Y_n = 0) \]
\[ \times (1-\gamma_1 - \gamma_2) (1-\rho) P(Y_n = 0) \]
\[ \times \gamma_1 \rho \pi_0 + \gamma_1 (1-\rho) \pi_0^2 + \gamma_2 \rho \pi_0 \]
\[ \times \gamma_2 (1-\rho) \pi_0 + (1-\gamma_1 - \gamma_2) (1-\rho) \pi_0 \]
\[ = (1-\gamma_1 - \gamma_2) \rho \pi_0 + (1-\rho - \gamma_1 + \gamma_2) \rho \pi_0 + \gamma_1 (1-\rho) \pi_0^2 \text{ (B-51)} \]

\( H_0(0, 1) = P(X_{n+1} = 0, A_n = 1 | A_{n-1} = 0) \)
\[ = \gamma_1 (1-\rho) P(Y_{n+1} = 0, Y_n = 1) \]
\[ \times (1-\gamma_1 - \gamma_2) (1-\rho) P(Y_n = 1) \]
\[ = \gamma_1 (1-\rho) \pi_0 \pi_1 + (1-\gamma_1 - \gamma_2) (1-\rho) \pi_1 \text{ (B-52)} \]

\( H_0(1, 0) = P(X_{n+1} = 0, A_n = 0 | A_{n-1} = 1) \)
\[ = \gamma_1 (1-\rho) P(Y_{n+1} = 0, Y_n = 0) \]
\[ \times \gamma_2 (1-\rho) P(Y_n = 0) \]
\[ = \gamma_1 (1-\rho) \pi_0^2 + \gamma_2 (1-\rho) \pi_0 \text{ (B-53)} \]

\( H_0(1, 1) = P(X_{n+1} = 0, A_n = 1 | A_{n-1} = 1) \)
\[ = \gamma_1 \rho P(Y_{n+1} = 0, A_{n-1} = 1) \]
\[ \times \gamma_1 (1-\rho) P(Y_{n+1} = 0, Y_n = 1) \]
\[ \times \gamma_2 P(Y_n = 0, A_{n-1} = 1) \]
\[ = \gamma_1 \rho \pi_0 + \gamma_1 (1-\rho) \pi_0 \pi_1 + \gamma_2 \rho \pi_0 \]
\[ = \rho (\gamma_1 + \gamma_2) \pi_0 + \gamma_1 (1-\rho) \pi_0^2 \text{ (B-54)} \]

\( H_1(0, 0) = P(X_{n+1} = 1, A_n = 0 | A_{n-1} = 0) \)
\[ = \gamma_1 \rho P(Y_{n+1} = 1) + \gamma_1 (1-\rho) P(Y_{n+1} = 1, Y_n = 0) \]
\[ +\gamma_2 \rho \mathbb{P}(Y_n=1) \]
\[ = (\gamma_1 + \gamma_2) \rho \pi_1 + \gamma_1 (1-\rho) \pi_0 \pi_1 \]  \hspace{1cm} (B-55)

\[ H_1(0, 1) = \mathbb{P}(X_{n+1} = 1, A_n = 1 | A_{n-1} = 0) \]
\[ = \gamma_1 (1-\rho) \mathbb{P}(Y_{n+1} = 1, Y_n = 1) + \gamma_2 (1-\rho) \mathbb{P}(Y_n = 1) \]
\[ = \gamma_1 (1-\rho) \pi_1 \pi_1 + \gamma_2 (1-\rho) \pi_1 \]  \hspace{1cm} (B-56)

\[ H_1(1, 0) = \mathbb{P}(X_{n+1} = 1, A_n = 0 | A_{n-1} = 1) \]
\[ = \gamma_1 (1-\rho) \mathbb{P}(Y_{n+1} = 1, Y_n = 0) + (1-\gamma_1 - \gamma_2)(1-\rho) \mathbb{P}(Y_n = 0) \]
\[ = \gamma_1 (1-\rho) \pi_0 \pi_1 + (1-\gamma_1 - \gamma_2)(1-\rho) \pi_0 \]  \hspace{1cm} (B-57)

\[ H_1(1, 1) = \mathbb{P}(X_{n+1} = 1, A_n = 1 | A_{n-1} = 1) \]
\[ = \gamma_1 \rho \mathbb{P}(Y_{n+1} = 1) + \gamma_2 (1-\rho) \mathbb{P}(Y_{n+1} = 1, Y_n = 1) \]
\[ + \gamma_2 \rho \mathbb{P}(Y_n = 1) + \gamma_2 (1-\rho) \mathbb{P}(Y_n = 1) \]
\[ +=(1-\gamma_1 - \gamma_2) \rho + (1-\gamma_1 - \gamma_2)(1-\rho) \mathbb{P}(Y_n = 1) \]
\[ = \gamma_1 \rho \pi_1 + \gamma_1 (1-\rho) \pi_1 \pi_1 + \gamma_2 \rho \pi_1 + \gamma_2 (1-\rho) \pi_1 \]
\[ +=(1-\gamma_1 - \gamma_2) \rho + (1-\gamma_1 - \gamma_2)(1-\rho) \pi_1 \]
\[ = (1-\gamma_1 - \gamma_2) \rho + (1-\rho - \gamma_1 - \gamma_2 \rho) \pi_1 + \gamma_1 (1-\rho) \pi_1 \]  \hspace{1cm} (B-58)

For the multi-state run length, \( \{X_n\} \) takes values in \( \{0, 1, 2\} \) with probabilities \( \pi_0, \pi_1, \) and \( \pi_2 \). Then the transition matrices for the multi-state run length, \( H_0, H_1, \) and \( H_2 \), are obtained in the following:

\[ H_0(0, 0) = \mathbb{P}(X_{n+1} = 0, A_n = 0 | A_{n-1} = 0) \]
\[ = (1-\gamma_1 - \gamma_2) \rho + (1-\rho - \gamma_1 - \gamma_2 \rho) \pi_0 + \gamma_1 (1-\rho) \pi_0^2 \]  \hspace{1cm} (B-59)

\[ H_0(0, 1) = \mathbb{P}(X_{n+1} = 0, A_n = 1 | A_{n-1} = 0) \]
\[ = \gamma_1 (1-\rho) \pi_0 \pi_1 + (1-\gamma_1 - \gamma_2)(1-\rho) \pi_1 \]  \hspace{1cm} (B-60)
\[ H_0(0,2) = P(X_{n+1} = 0, A_n = 2 | A_{n-1} = 0) \]
\[ = \gamma_1 (1-\rho) P(Y_{n+1} = 0, Y_n = 2) + (1-\gamma_1 - \gamma_2) (1-\rho) P(Y_n = 2) \]
\[ = \gamma_1 (1-\rho) \pi_0 \pi_2 + (1-\gamma_1 - \gamma_2) (1-\rho) \pi_2 \quad (B-61) \]

\[ H_0(1,0) = P(X_{n+1} = 0, A_n = 0 | A_{n-1} = 1) \]
\[ = \gamma_1 (1-\rho) P(Y_{n+1} = 0, Y_n = 1) + \gamma_2 (1-\rho) P(Y_n = 0) \]
\[ = \gamma_1 (1-\rho) \pi_0 \pi_2 + \gamma_2 (1-\rho) \pi_0 \quad (B-62) \]

\[ H_0(1,1) = P(X_{n+1} = 0, A_n = 1 | A_{n-1} = 1) \]
\[ = \gamma_1 (1-\rho) P(Y_{n+1} = 0, Y_n = 1) + \gamma_1 \beta P(Y_{n+1} = 0) \]
\[ + \gamma_2 \rho P(Y_n = 0) \]
\[ = \rho (\gamma_1 \pi_0 + \gamma_2) \pi_0 + \gamma_1 (1-\rho) \pi_0 \pi_1 \quad (B-63) \]

\[ H_0(1,2) = P(X_{n+1} = 0, A_n = 2 | A_{n-1} = 1) \]
\[ = \gamma_1 (1-\rho) P(Y_{n+1} = 0, Y_n = 2) \]
\[ = \gamma_1 (1-\rho) \pi_0 \pi_2 \quad (B-64) \]

\[ H_0(2,0) = P(X_{n+1} = 0, A_n = 0 | A_{n-1} = 2) \]
\[ = \gamma_1 (1-\rho) P(Y_{n+1} = 0, Y_n = 0) + \gamma_2 (1-\rho) P(Y_n = 0) \]
\[ = \gamma_1 (1-\rho) \pi_0 + \gamma_2 (1-\rho) \pi_0 \quad (B-65) \]

\[ H_0(2,1) = P(X_{n+1} = 0, A_n = 1 | A_{n-1} = 2) \]
\[ = \gamma_1 (1-\rho) P(Y_{n+1} = 0, Y_n = 1) \]
\[ = \gamma_1 (1-\rho) \pi_0 \pi_1 \quad (B-66) \]

\[ H_0(2,2) = P(X_{n+1} = 0, A_n = 2 | A_{n-1} = 2) \]
\[ = \gamma_1 (1-\rho) P(Y_{n+1} = 0, Y_n = 2) + \gamma_1 \beta P(Y_{n+1} = 0) \]
\[ + \gamma_2 \rho P(Y_n = 0) \]
\[ = \gamma_1 (1-\rho) \pi_0 \pi_2 + \gamma_1 \beta \pi_0 \pi_2 + \gamma_2 \rho \pi_0 \quad (B-67) \]
\[ H_1(0,0) = P(X_{n+1} = 1, A_n = 0 | A_{n-1} = 0) \]
\[ = \gamma_1 (1 - \rho) P(Y_{n+1} = 1, Y_n = 0) + \gamma_1 \rho P(Y_n = 1) \]
\[ + \gamma_2 \rho P(Y_n = 1) \]
\[ = \gamma_1 (1 - \rho) \pi_1 \pi_0 + \gamma_1 \rho \pi_1 + \gamma_2 \pi_1 \rho \]
\[ = \gamma_1 (1 - \rho) \pi_1 \pi_0 + \rho (\gamma_1 + \gamma_2) \pi_1 \] (B-68)

\[ H_1(0,1) = P(X_{n+1} = 1, A_n = 1 | A_{n-1} = 0) \]
\[ = \gamma_1 (1 - \rho) P(Y_{n+1} = 1, Y_n = 1) + \gamma_2 (1 - \rho) P(Y_n = 1) \]
\[ = \gamma_2 (1 - \rho) \pi_1 + \gamma_1 (1 - \rho) \pi_1 \] (B-69)

\[ H_1(0,2) = P(X_{n+1} = 1, A_n = 2 | A_{n-1} = 0) = \gamma_1 (1 - \rho) P(Y_{n+1} = 1, Y_n = 2) \]
\[ = \gamma_1 (1 - \rho) \pi_1 \pi_2 \] (B-70)

\[ H_1(1,0) = P(X_{n+1} = 1, A_n = 0 | A_{n-1} = 1) \]
\[ = \gamma_1 (1 - \rho) P(Y_{n+1} = 1, Y_n = 0) + (1 - \gamma_1 - \gamma_2) (1 - \rho) P(Y_n = 0) \]
\[ = \gamma_1 (1 - \rho) \pi_1 \pi_0 + (1 - \gamma_1 - \gamma_2) (1 - \rho) \pi_0 \] (B-71)

\[ H_1(1,1) = P(X_{n+1} = 1, A_n = 1 | A_{n-1} = 1) \]
\[ = \gamma_1 (1 - \rho) P(Y_{n+1} = 1, Y_n = 1 | A_{n-1} = 1) + \gamma_2 \rho P(Y_n = 1) \]
\[ + \gamma_2 (1 - \rho) P(Y_n = 1) + \gamma_2 \rho P(Y_n = 1) \]
\[ + (1 - \gamma_1 - \gamma_2) (1 - \rho) P(Y_n = 1) + (1 - \gamma_1 - \gamma_2) \rho \]
\[ = \gamma_1 (1 - \rho) \pi_1 + \gamma_2 \rho \pi_1 + \gamma_2 (1 - \rho) \pi_1 + \gamma_2 \rho \pi_1 \]
\[ + (1 - \gamma_1 - \gamma_2) (1 - \rho) \pi_1 + (1 - \gamma_1 - \gamma_2) \rho \]
\[ = (1 - \gamma_1 - \gamma_2) P + (1 - \rho - \gamma_1 - 2\gamma_1 \rho + \gamma_2 \rho) \pi_1 + \gamma_1 (1 - \rho) \pi_1 \] (B-72)

\[ H_1(1,2) = P(X_{n+1} = 1, A_n = 2 | A_{n-1} = 1) \]
\[ = \gamma_1 (1 - \rho) P(Y_{n+1} = 1, Y_n = 2) + (1 - \gamma_1 - \gamma_2) (1 - \rho) P(Y_n = 2) \]
\[ = \gamma_1 (1 - \rho) \pi_1 \pi_2 + (1 - \gamma_1 - \gamma_2) (1 - \rho) \pi_2 \] (B-73)
\[ H_1 (Z, 0) = P(X_{n+1} = 1, A_n = 0 \mid A_{n-1} = 2) \]
\[ = \gamma_1 (1-\rho) P(Y_{n+1} = 1, Y_n = 0) = \gamma_1 (1-\rho) \pi_1 \pi_0 \]  
\[ (B-74) \]

\[ H_2 (Z, 1) = P(X_{n+1} = 1, A_n = 1 \mid A_{n-1} = 2) \]
\[ = \gamma_1 (1-\rho) P(Y_{n+1} = 1, Y_n = 1) \]
\[ = \gamma_1 (1-\rho) \pi_1 \pi_2 + \gamma_2 (1-\rho) \pi_1 \]
\[ (B-75) \]

\[ H_2 (Z, 2) = P(X_{n+1} = 1, A_n = 2 \mid A_{n-1} = 2) \]
\[ = \gamma_1 (1-\rho) P(Y_{n+1} = 2, Y_n = 0) \]
\[ = \gamma_1 (1-\rho) \pi_1 \pi_2 + \gamma_1 \rho \pi_1 + \gamma_2 \rho \pi_1 \]
\[ (B-76) \]

\[ H_2 (0, 0) = P(X_{n+1} = 2, A_n = 0 \mid A_{n-1} = 0) \]
\[ = \gamma_1 (1-\rho) P(Y_{n+1} = 2, Y_n = 0) + \gamma_1 \rho P(Y_{n+1} = 2) + \gamma_2 \rho P(Y_n = 2) \]
\[ = \gamma_1 (1-\rho) \pi_2 \pi_0 + \rho (\gamma_1 \pi_2 + \gamma_2 \pi_2) \]
\[ (B-77) \]

\[ H_2 (0, 1) = P(X_{n+1} = 2, A_n = 1 \mid A_{n-1} = 0) \]
\[ = \gamma_1 (1-\rho) P(Y_{n+1} = 2, Y_n = 1) \]
\[ = \gamma_1 (1-\rho) \pi_2 \pi_1 \]
\[ (B-78) \]

\[ H_2 (0, 2) = P(X_{n+1} = 2, A_n = 2 \mid A_{n-1} = 0) \]
\[ = \gamma_1 (1-\rho) P(Y_{n+1} = 2, Y_n = 2) + \gamma_2 (1-\rho) P(Y_n = 2) \]
\[ = \gamma_1 (1-\rho) \pi_2 \pi_2 + \gamma_2 (1-\rho) \pi_2 \]
\[ (B-79) \]

\[ H_2 (1, 0) = P(X_{n+1} = 2, A_n = 0 \mid A_{n-1} = 1) \]
\[ = \gamma_1 (1-\rho) P(Y_{n+1} = 2, Y_n = 0) \]
\[ = \gamma_1 (1-\rho) \pi_2 \pi_0 \]
\[ (B-80) \]

\[ H_2 (1, 1) = P(X_{n+1} = 2, A_n = 1 \mid A_{n-1} = 1) \]
\[ = \gamma_1 (1-\rho) P(Y_{n+1} = 2, Y_n = 1) + \gamma_1 \rho P(Y_{n+1} = 2) + \gamma_2 \rho P(Y_n = 2) \]
\[
= \gamma_1 (1 - \rho) \pi_2 \pi_1 + \gamma_1 \rho \pi_2 + \gamma_2 \rho \pi_2 \\
= \gamma_1 (1 - \rho) \pi_2 \pi_1 + \rho (\gamma_1 + \gamma_2) \pi_2 
\]  \hspace{1cm} (B-81)

\[
H_2 (1, 2) = P(X_{n+1} = 2, A_n = 2 | A_{n-1} = 1) \\
= \gamma_1 (1 - \rho) P(Y_{n+1} = 2, Y_n = 2) + \gamma_2 (1 - \rho) P(Y_n = 2) \\
= \gamma_1 (1 - \rho) \pi_2 + \gamma_2 (1 - \rho) \pi_2 
\]  \hspace{1cm} (B-82)

\[
H_2 (2, 0) = P(X_{n+1} = 2, A_n = 0 | A_{n-1} = 2) \\
= \gamma_1 (1 - \rho) P(Y_{n+1} = 2, Y_n = 0) + (1 - \gamma_1 - \gamma_2) (1 - \rho) P(Y_n = 0) \\
= \gamma_1 (1 - \rho) \pi_0 + (1 - \gamma_1 - \gamma_2) (1 - \rho) \pi_0 
\]  \hspace{1cm} (B-83)

\[
H_2 (2, 1) = P(X_{n+1} = 2, A_n = 1 | A_{n-1} = 2) \\
= \gamma_1 (1 - \rho) P(Y_{n+1} = 2, Y_n = 1) + (1 - \gamma_1 - \gamma_2) (1 - \rho) P(Y_n = 1) \\
= \gamma_1 (1 - \rho) \pi_1 + (1 - \gamma_1 - \gamma_2) (1 - \rho) \pi_1 
\]  \hspace{1cm} (B-84)

\[
H_2 (2, 2) = P(X_{n+1} = 2, A_n = 2 | A_{n-1} = 3) \\
= \gamma_1 \rho P(Y_{n+1} = 2, Y_n = 2) + \gamma_1 \rho P(Y_n = 2) \\
= (1 - \gamma_1 - \gamma_2) \rho + (1 - \rho - \gamma_1 + 2 \gamma_1 \rho + \gamma_2 \rho) \pi_2 + \gamma_1 (1 - \rho) \pi_2 
\]  \hspace{1cm} (B-85)
APPENDIX C

ESTIMATIONS OF PARAMETERS IN THE DARMA(1,1) MODEL

The autocorrelation function of the DARMA(1,1) given in equation (2-17) is a nonlinear function of the model parameters, i.e.

$$\text{Corr}(k) = (1-\beta)(\rho+\beta-2\rho\beta)^{k-1},$$ \hspace{1cm} (C-1)

The derivative of Corr(k) with respect to $\rho$ is

$$d'(\rho) = (1-\beta)(1-2\beta)^{k-1} + (1-\beta)(\rho+\beta-2\rho\beta)(k-1)^{k-2},$$ \hspace{1cm} (C-2)

and the derivative with respect to $\beta$ is

$$d'(\beta) = -(\rho+\beta-2\rho\beta)^{k-1} + (1-\beta)(1-2\rho)^{k-1}.$$ \hspace{1cm} (C-3)

Then the computer program using the least squares method described in Section III.3.2. is required to input two subprograms FUN and PAR, where the former is to provide correlation function in (C-1) and the latter is to give the derivatives (C-2) and (C-3) in the computation of normal equations by iteration. Then the values of $\rho$ and $\beta$ are estimated.