AGGREGATION SIMULATION MODEL OF FLOW AND RAINFALL SERIES

NURUL ADZURA ISMAIL

UNIVERSITI TEKNOLOGI MALAYSIA
AGGREGATION SIMULATION MODEL OF FLOW AND RAINFALL SERIES

NURUL ADZURA ISMAIL

A thesis submitted in fulfilment of the requirements for the award of the degree of Master of Engineering (Civil)

Faculty of Civil Engineering
Universiti Teknologi Malaysia

MAY 2005
To my beloved soul mate
ACNOWLEDGEMENTS

Alhamdullilah, with His mercy and Blessings, this study was finally completed. It gives the author immense pleasure to express her heartiest gratitude to her supervisor Associate Professor Dr. Sobri Harun for his constant guidance, inspiration, suggestion and support in performing and completing the research study. The author also grateful to co-supervisor Associate Professor Dr. Zulkifli Yusop for his assistance and co-operation in carrying out the study in one way or another.

Special thanks are due to Ministry of Science, Technology and Innovations for the financial support. Special thanks are also due to the Department of Irrigation and Drainage (DID) for providing necessary references and data.

The understanding, patience, moral support and co-operation from Iman are deeply acknowledged. Special thank due to my parents, family members, Yap Sze Jeat and my fellow postgraduate students for helping in this study.

May Allah reward and bless all of them. Finally the author is expressing her sincere gratitude to Allah once again who made the study to complete.
ABSTRACT

Synthetic hydrology series is useful for evaluating the consequences of water supply management decisions and reservoir design. The main objective of this study is to identify and confirm the best model in flow and rainfall simulation. The study covers the application of aggregation and disaggregation methods for flow and rainfall stochastic simulation. In general, the application of various periodic models for the flow simulation was mostly successful. The application of disaggregation models was found to yield sufficient performance and competitive to the periodic models. It has been proven that the transformation does not always guarantee improvement in the candidate models performance. The Periodic Autoregressive of Order One (PAR (1)) model is the best performer for the monthly and annual flow simulation using periodic models for both untransformed and transformed series. The Valencia and Schaake (VLSH) model is the robust model from disaggregation group for the monthly and annual flow simulation. Simulation for monthly and annual rainfall series shows that the VLSH model is the best performer to produce sufficient results for both untransformed and transformed series. The results from this study are based on investigation from graphs and frequency analysis. The outcome of study has potential to assist the water engineers and consultant in making decisions for the operation of the water resources systems. It is suggested that the rainfall simulation should be applied in water resources planning because observed flow series are subjected to disturbance due to development.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>DECLARATION</td>
<td>ii</td>
<td></td>
</tr>
<tr>
<td>DEDICATION</td>
<td>iii</td>
<td></td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>iv</td>
<td></td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>v</td>
<td></td>
</tr>
<tr>
<td>ABSTRAK</td>
<td>vi</td>
<td></td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>vii</td>
<td></td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>xii</td>
<td></td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xvii</td>
<td></td>
</tr>
<tr>
<td>LIST OF SYMBOLS</td>
<td>xx</td>
<td></td>
</tr>
<tr>
<td>LIST OF APPENDICES</td>
<td>xxiii</td>
<td></td>
</tr>
</tbody>
</table>

## 1 INTRODUCTION

1.1 Background  
1.2 Statement of Problems  
1.3 Objectives  
1.4 Scope  
1.5 The Importance of Study  
1.6 Research Hypothesis
## 2 LITERATURE REVIEW

2.1 Introduction

2.2 Importance of Synthetic Series of Flow and Rainfall

2.3 Statistical Methods in Hydrology
   2.3.1 Time Series Analysis
   2.3.2 Frequency and Hybrid Analysis
   2.3.3 Stochastic Hydrology
   2.3.4 Simulation Methods

2.4 Review on Flow Simulation

2.5 Review on Rainfall Simulation

2.6 Summary of Literature Review

## 3 METHODOLOGY

3.1 Introduction

3.2 Hydrological Data

3.3 Periodic Models (Aggregation Models)
   3.3.1 Periodic Autoregressive (PAR) Model
   3.3.2 Periodic Autoregressive Moving Average (PARMA) Model

3.4 Disaggregation Models
   3.4.1 Valencia – Schaake
   3.4.2 Mejia-Rouselle
   3.4.3 Lane
   3.4.4 SPIGOT

3.5 Development of Flow Simulation Model
   3.5.1 Statistical Analysis of Data
   3.5.2 Fitting a Stochastic Model
   3.5.3 Synthetic Flow Data Generation
3.6 Development of Rainfall Simulation Model
3.7 Evaluation the Performance of Simulation Models

4 CASE STUDY

4.1 Introduction
4.2 River Basins
  4.2.1 Negeri Sembilan
    4.2.1.1 Sg. Muar Basins
    4.2.1.2 Sg. Triang Basins
    4.2.1.3 Sg. Linggi Basins
  4.2.2 Melaka
  4.2.3 Selangor
    4.2.3.1 Sg. Bernam Basin
    4.2.3.2 Sg. Selangor Basin
    4.2.3.3 Sg. Klang Basin
    4.2.3.4 Sg. Langat Basin
  4.2.4 Johor
    4.2.4.1 Sg. Muar Basin
    4.2.4.2 Sg. Johor Basin
  4.3 Land Use Changes
    4.3.1 Negeri Sembilan
    4.3.2 Melaka
    4.3.3 Selangor
    4.3.4 Johor
  4.4 Climatic Characteristics
    4.4.1 Impacts of Climate on the Water Resources
5 RESULTS AND DISCUSSIONS

5.1 Applications of Stochastic Models 68

5.2 Discussion on Simulated Monthly Flow Series 69

5.2.1 Seasonal Periodic Models 71

5.2.1.1 Negeri Sembilan 72

5.2.1.2 Melaka 78

5.2.1.3 Selangor 81

5.2.1.4 Johor 83

5.2.2 Disaggregation Models 87

5.2.2.1 Negeri Sembilan 88

5.2.2.2 Melaka 92

5.2.2.3 Selangor 94

5.2.2.4 Johor 976

5.2.3 Annual Mean and Standard deviation 100

5.2.3.1 Negeri Sembilan 101

5.2.3.2 Melaka 106

5.2.3.3 Selangor 108

5.2.3.4 Johor 111

5.3 Summary of Discussions on Generated Flow Simulation 115

5.4 Generation of Synthetic Rainfall Sequences 126

5.4.1 Discussion on Simulated Monthly Rainfall Series 127

5.4.1.1 Ladang Union Station 127

5.4.1.2 Ladang Hopeful Station 130

5.4.1.3 Ladang Bernam Station 131

5.4.1.4 Ladang Paya Lalang Station 132
5.4.2 Discussion on Simulated Annual Rainfall Series

5.4.2.1 Ladang Union Station 133
5.4.2.2 Ladang Hopeful Station 135
5.4.2.3 Ladang Bernam Station 136
5.4.2.4 Ladang Paya Lalang Station 137

5.5 Summary of Discussion on the Generated Rainfall Simulation 138

6 CONCLUSION AND RECOMMENDATION

6.1 Conclusion 140
6.2 Recommendation 143

REFERENCES 144

APPENDICES

Appendix A 158
Appendix B 190
Appendix C 212
Appendix D 233
Appendix E 242
Appendix F 248
# LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE NO.</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>List of selected flow gauging stations</td>
<td>58</td>
</tr>
<tr>
<td>4.2</td>
<td>List of selected rainfall stations</td>
<td>59</td>
</tr>
<tr>
<td>5.1</td>
<td>Simplified term of the candidate models</td>
<td>69</td>
</tr>
<tr>
<td>5.2</td>
<td>Transformation type selected for the Sg. Pedas</td>
<td>70</td>
</tr>
<tr>
<td>5.3</td>
<td>Transformation type selected for the Sg. Pedas using SPIGOT model</td>
<td>71</td>
</tr>
<tr>
<td>5.4</td>
<td>The best performance in the preservation of monthly statistical</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td>characteristics for Sg. Pedas flow series</td>
<td></td>
</tr>
<tr>
<td>5.5</td>
<td>The best performance in the preservation of monthly statistical</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td>characteristics for Sg. Muar flow series</td>
<td></td>
</tr>
<tr>
<td>5.6</td>
<td>The best performance in the preservation of monthly statistical</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td>characteristics for Sg. Gemencheh flow series</td>
<td></td>
</tr>
<tr>
<td>5.7</td>
<td>The best performance in the preservation of monthly statistical</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>characteristics for Sg. Melaka flow series</td>
<td></td>
</tr>
<tr>
<td>5.8</td>
<td>The best performance in the preservation of monthly statistical</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>characteristics for Sg. Kesang flow series</td>
<td></td>
</tr>
<tr>
<td>5.9</td>
<td>The best performance in the preservation of monthly statistical</td>
<td>82</td>
</tr>
<tr>
<td></td>
<td>characteristics for Sg. Selangor flow series</td>
<td></td>
</tr>
<tr>
<td>5.10</td>
<td>The best performance in the preservation of monthly statistical</td>
<td>82</td>
</tr>
<tr>
<td></td>
<td>characteristics for Sg. Bernam flow series</td>
<td></td>
</tr>
<tr>
<td>5.11</td>
<td>The best performance in the preservation of monthly statistical</td>
<td>86</td>
</tr>
<tr>
<td></td>
<td>characteristics for Sg. Segamat flow series</td>
<td></td>
</tr>
<tr>
<td>Section</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>5.12</td>
<td>The best performance in the preservation of monthly statistical characteristics for Sg. Muar flow series</td>
<td></td>
</tr>
<tr>
<td>5.13</td>
<td>The best performance in the preservation of monthly statistical characteristics for Sg. Lenggor flow series</td>
<td></td>
</tr>
<tr>
<td>5.14</td>
<td>The best performance in the preservation of monthly statistical characteristics for Sg. Sembrong flow series</td>
<td></td>
</tr>
<tr>
<td>5.15</td>
<td>The best performance in the preservation of monthly statistical characteristics for Sg. Pedas flow series</td>
<td></td>
</tr>
<tr>
<td>5.16</td>
<td>The best performance in the preservation of monthly statistical characteristics for Sg. Muar flow series</td>
<td></td>
</tr>
<tr>
<td>5.17</td>
<td>The best performance in the preservation of monthly statistical characteristics for Sg. Gemencheh flow series</td>
<td></td>
</tr>
<tr>
<td>5.18</td>
<td>The best performance in the preservation of monthly statistical characteristics for Sg. Melaka flow series</td>
<td></td>
</tr>
<tr>
<td>5.19</td>
<td>The best performance in the preservation of monthly statistical characteristics for Sg. Kesang flow series</td>
<td></td>
</tr>
<tr>
<td>5.20</td>
<td>The best performance in the preservation of monthly statistical characteristics for Sg. Selangor flow series</td>
<td></td>
</tr>
<tr>
<td>5.21</td>
<td>The best performance in the preservation of monthly statistical characteristics for Sg. Bernam flow series</td>
<td></td>
</tr>
<tr>
<td>5.22</td>
<td>The best performance in the preservation of monthly statistical characteristics for Sg. Segamat flow series</td>
<td></td>
</tr>
<tr>
<td>5.23</td>
<td>The best performance in the preservation of monthly statistical characteristics for Sg. Muar flow series</td>
<td></td>
</tr>
<tr>
<td>5.24</td>
<td>The best performance in the preservation of monthly statistical characteristics for Sg. Lenggor flow series</td>
<td></td>
</tr>
<tr>
<td>5.25</td>
<td>The best performance in the preservation of monthly statistical characteristics for Sg. Sembrong flow series</td>
<td></td>
</tr>
<tr>
<td>5.26</td>
<td>The best performance in the preservation of annual mean and standard deviation for Sg. Pedas flow series</td>
<td></td>
</tr>
<tr>
<td>5.27</td>
<td>The best performance in the preservation of annual mean and standard deviation for Sg. Muar flow series</td>
<td></td>
</tr>
<tr>
<td>Section</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>5.28</td>
<td>The best performance in the preservation of annual mean and standard deviation for Sg. Gemencheh flow series</td>
<td>105</td>
</tr>
<tr>
<td>5.29</td>
<td>The best performance in the preservation of annual mean and standard deviation for Sg. Melaka flow series</td>
<td>108</td>
</tr>
<tr>
<td>5.30</td>
<td>The best performance in the preservation of annual mean and standard deviation for Sg. Kesang flow series</td>
<td>108</td>
</tr>
<tr>
<td>5.31</td>
<td>The best performance in the preservation of annual mean and standard deviation for Sg. Selangor flow series</td>
<td>110</td>
</tr>
<tr>
<td>5.32</td>
<td>The best performance in the preservation of annual mean and standard deviation for Sg. Bernam flow series</td>
<td>110</td>
</tr>
<tr>
<td>5.33</td>
<td>The best performance in the preservation of annual mean and standard deviation for Sg. Segamat flow series</td>
<td>114</td>
</tr>
<tr>
<td>5.34</td>
<td>The best performance in the preservation of annual mean and standard deviation for Sg. Muar flow series</td>
<td>114</td>
</tr>
<tr>
<td>5.35</td>
<td>The best performance in the preservation of annual mean and standard deviation for Sg. Lenggor flow series</td>
<td>114</td>
</tr>
<tr>
<td>5.36</td>
<td>The best performance in the preservation of annual mean and standard deviation for Sg. Sembrong flow series</td>
<td>115</td>
</tr>
<tr>
<td>5.37</td>
<td>Basic Catchment Characteristics</td>
<td>118</td>
</tr>
<tr>
<td>5.38</td>
<td>The best performance in the preservation of monthly mean standard deviation</td>
<td>119</td>
</tr>
<tr>
<td>5.39</td>
<td>The best performance in the preservation of monthly standard deviation</td>
<td>120</td>
</tr>
<tr>
<td>5.40</td>
<td>The best performance in the preservation of monthly lag-one season to season correlation coefficient</td>
<td>121</td>
</tr>
<tr>
<td>5.41</td>
<td>The best performance in the preservation of annual mean standard deviation</td>
<td>122</td>
</tr>
<tr>
<td>5.42</td>
<td>The best performance in the preservation of annual standard deviation</td>
<td>123</td>
</tr>
<tr>
<td>5.43</td>
<td>Frequency analysis for mean monthly (untransformed and transformed flow series)</td>
<td>124</td>
</tr>
<tr>
<td>5.44</td>
<td>Frequency analysis for monthly standard deviation (untransformed and transformed flow series)</td>
<td>124</td>
</tr>
</tbody>
</table>
5.45 Frequency analysis for annual mean (untransformed and transformed flow series) 125
5.46 Frequency analysis for annual standard deviation (untransformed and transformed flow series) 125
5.47 Transformation for the disaggregation models (Ladang Union) transformed series 126
5.48 The best performance in the preservation of monthly statistical characteristics for Ladang Union Station 129
5.49 The best performance in the preservation of monthly statistical characteristics for Ladang Hopeful Station 130
5.50 The best performance in the preservation of monthly statistical characteristics for Ladang Bernam Station 132
5.51 The best performance in the preservation of monthly statistical characteristics for Ladang Paya Lalang Station 133
5.52 The best performance in the preservation of annual statistical properties for Ladang Union Station 135
5.53 The best performance in the preservation of annual statistical properties for Ladang Hopeful Station 136
5.54 The best performance in the preservation of annual statistical properties for Ladang Bernam Station 137
5.55 The best performance in the preservation of annual statistical properties for Ladang Paya Lalang Station 138
5.56 Basic rainfall station characteristics 139
5.57 The best performance in the preservation of monthly mean 139
5.58 The best performance in the preservation of monthly standard deviation 139
5.59 The best performance in the preservation of monthly lag-one season to season correlation coefficient 139
5.60 The best performance in the preservation of annual mean 139
5.61 The best performance in the preservation of annual standard deviation 140
5.62 Frequency analysis for mean monthly and monthly standard deviation (untransformed and transformed rainfall series) 140

5.63 Frequency analysis for annual mean and annual standard deviation (untransformed and transformed rainfall series) 140

6.1 Rank of the best performance for the seasonal (monthly) flow simulation models 142

6.2 Rank of the best performance for the annual flow simulation models 142

6.3 Rank of the best performance for the rainfall simulation models 143
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE NO.</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>Sg. Pedas flow series (untransformed flows)</td>
<td>70</td>
</tr>
<tr>
<td>5.2</td>
<td>Sg. Pedas flow series (transformed flows)</td>
<td>71</td>
</tr>
<tr>
<td>5.3</td>
<td>Monthly mean for PAR models for Sg. Pedas (untransformed flows)</td>
<td>73</td>
</tr>
<tr>
<td>5.4</td>
<td>Monthly mean for PARMA models for Sg. Pedas (untransformed flows)</td>
<td>73</td>
</tr>
<tr>
<td>5.5</td>
<td>Monthly mean for PAR models for Sg. Pedas (transformed flows)</td>
<td>74</td>
</tr>
<tr>
<td>5.6</td>
<td>Monthly mean for PARMA models for Sg. Pedas (transformed flows)</td>
<td>74</td>
</tr>
<tr>
<td>5.7</td>
<td>Monthly standard deviation for PAR models for Sg. Pedas (untransformed flows)</td>
<td>74</td>
</tr>
<tr>
<td>5.8</td>
<td>Monthly standard deviation for PARMA models for Sg. Pedas (untransformed flows)</td>
<td>74</td>
</tr>
<tr>
<td>5.9</td>
<td>Monthly standard deviation for PAR models for Sg. Pedas (transformed flows)</td>
<td>75</td>
</tr>
<tr>
<td>5.10</td>
<td>Monthly standard deviation for PARMA models for Sg. Pedas (transformed flows)</td>
<td>75</td>
</tr>
<tr>
<td>5.11</td>
<td>Monthly Lag-1 season to season correlation for PAR models for Sg. Pedas (untransformed flows)</td>
<td>75</td>
</tr>
<tr>
<td>5.12</td>
<td>Monthly Lag-1 season to season correlation for PARMA models for Sg. Pedas (untransformed flows)</td>
<td>75</td>
</tr>
<tr>
<td>Section</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>5.13</td>
<td>Monthly lag-one season to season correlation for PAR models for Sg. Pedas (transformed flows)</td>
<td></td>
</tr>
<tr>
<td>5.14</td>
<td>Monthly lag-one season to season correlation for PARMA models for Sg. Pedas (transformed flows)</td>
<td></td>
</tr>
<tr>
<td>5.15</td>
<td>Monthly mean for disaggregation models for Sg. Pedas (untransformed flows)</td>
<td></td>
</tr>
<tr>
<td>5.16</td>
<td>Monthly mean for disaggregation models for Sg. Pedas (transformed flows)</td>
<td></td>
</tr>
<tr>
<td>5.17</td>
<td>Monthly std. deviation for disaggregation models for Sg. Pedas (untransformed flows)</td>
<td></td>
</tr>
<tr>
<td>5.18</td>
<td>Monthly std. deviation for disaggregation models for Sg. Pedas (transformed flows)</td>
<td></td>
</tr>
<tr>
<td>5.19</td>
<td>Monthly lag-one season to season correlation coefficient for disaggregation models for Sg. Pedas (untransformed flows)</td>
<td></td>
</tr>
<tr>
<td>5.20</td>
<td>Monthly lag-one season to season correlation coefficient for disaggregation models for Sg. Pedas (transformed flows)</td>
<td></td>
</tr>
<tr>
<td>5.21</td>
<td>Annual mean for Sg. Pedas (Untransformed)</td>
<td></td>
</tr>
<tr>
<td>5.22</td>
<td>Annual mean for Sg. Pedas (Transformed)</td>
<td></td>
</tr>
<tr>
<td>5.23</td>
<td>Annual standard deviation for Sg. Pedas (Untransformed)</td>
<td></td>
</tr>
<tr>
<td>5.24</td>
<td>Annual standard deviation for Sg. Pedas (Transformed)</td>
<td></td>
</tr>
<tr>
<td>5.25</td>
<td>Monthly mean for Ladang Union Station (Untransformed)</td>
<td></td>
</tr>
<tr>
<td>5.26</td>
<td>Monthly mean for Ladang Union Station (Transformed)</td>
<td></td>
</tr>
<tr>
<td>5.27</td>
<td>Monthly standard deviation for Ladang Union Station (Untransformed)</td>
<td></td>
</tr>
<tr>
<td>5.28</td>
<td>Monthly standard deviation for Ladang Union Station (Transformed)</td>
<td></td>
</tr>
<tr>
<td>5.29</td>
<td>Monthly lag one season to season correlation for Ladang Union Station (Untransformed)</td>
<td></td>
</tr>
<tr>
<td>5.30</td>
<td>Monthly lag one season to season correlation for Ladang Union Station (Transformed)</td>
<td></td>
</tr>
<tr>
<td>5.31</td>
<td>Annual mean for Ladang Union Station (Untransformed)</td>
<td></td>
</tr>
<tr>
<td>5.32</td>
<td>Annual mean for Ladang Union Station (Transformed)</td>
<td></td>
</tr>
</tbody>
</table>
5.33 Annual standard deviation for Ladang Union Station (Untransformed) 134
5.34 Annual standard deviation for Ladang Union Station (Transformed) 135
LIST OF SYMBOLS

GENERAL

mm  Millimeter
km  Kilometer
m   Meter
km² Kilometer square
km³ Kilometer cube
%   Percent
°C  Degree Celsius
N   Sample size
X   Original observed series
a, b transformation coefficients
τ   season
ν   year
µ   mean
σ   standard deviation
ω   Number of season in the year
Y   monthly flow or rainfall
Yτ  Column matrix containing the seasonal values
Yτ-1 A column matrix of the previous matrix of the previous substation
ε   The current value from a completely random series (stochastic term)
εν,τ Matrix of an independent random variable at year ν and season τ
Qt  Matrix of the annual/seasonal flow or rainfall value of year
χ²  chi-square test density function
PERIODIC MODELS (AGGREGATION MODELS)

PAR  Periodic autoregressive model
PARMA  Periodic autoregressive and moving average model
A, B  Parameter matrices
Y  Monthly flow
$Y_{v,\tau}$  Matrix of monthly flow at year $v$ and season $\tau$
$r_{k}(\varepsilon)$  Correlogram of the residuals
L  Maximum lag considered
$\sigma^2$  Maximum likelihood estimate of the residual variance
c  Lower bound of three parameter log normal distribution
$p$  is the term of periodic autoregressive parameter
$q$  Order of autoregressive parameter of MA model
G  Backward shift operator of PARMA models
$\rho_{k,\tau}$  Autocorrelation function
$\phi_{1,\tau}, \ldots, \phi_{p,\tau}$  Seasonal autoregressive parameters
$\theta_{1,\tau}, \ldots, \theta_{q,\tau}$  Seasonal moving average parameters
$\theta_{\tau}(G)$  Periodic polynomials

DISAGGREGATION MODELS

A, B and C  Parameter matrices
$Q_v$  Annual series vector
$Q_{v,\tau}$  Generated annual flow/rainfall vector.
m  Rank of residuals matrix
$M_0$  Lag-zero correlation matrix or population moment
$M^1$  Lag-one correlation matrix or population moment
$M^{-1}$  Inverse matrix of population moment
$M^T$  Transpose matrix of population moment
$Q$  Single value and a column vector which contains an annual value
$Q_t$  Matrix of the annual rainfall/flow value of year
$p$  Order of autoregressive parameter of AR model
$q$  Order of autoregressive parameter of MA model
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVD</td>
<td>Singular Value Decomposition</td>
</tr>
<tr>
<td>VLSH</td>
<td>Valencia-Schaake disaggregation model</td>
</tr>
<tr>
<td>MJRS</td>
<td>Mejia-Rouselle disaggregation model</td>
</tr>
<tr>
<td>LANE</td>
<td>Lane disaggregation model</td>
</tr>
<tr>
<td>SPIGOT</td>
<td>Stedinger disaggregation model</td>
</tr>
</tbody>
</table>
**LIST OF APPENDICES**

<table>
<thead>
<tr>
<th>APPENDIX</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.1</td>
<td>Monthly skewness for Periodic models for untransformed and transformed Sg. Pedas flow series.</td>
<td>159</td>
</tr>
<tr>
<td>A.2</td>
<td>Graphical plots for Periodic models for untransformed and transformed Sg. Muar flow series.</td>
<td>160</td>
</tr>
<tr>
<td>A.3</td>
<td>Graphical plots for Periodic models for untransformed and transformed Sg. Gemencheh flow series.</td>
<td>164</td>
</tr>
<tr>
<td>A.4</td>
<td>Graphical plots for Periodic models for untransformed and transformed Sg. Melaka flow series.</td>
<td>166</td>
</tr>
<tr>
<td>A.5</td>
<td>Graphical plots for Periodic models for untransformed and transformed Sg. Kesang flow series.</td>
<td>170</td>
</tr>
<tr>
<td>A.6</td>
<td>Graphical plots for Periodic models for untransformed and transformed Sg. Selangor flow series.</td>
<td>174</td>
</tr>
<tr>
<td>A.7</td>
<td>Graphical plots for Periodic models for untransformed and transformed Sg. Bernam flow series.</td>
<td>178</td>
</tr>
<tr>
<td>A.8</td>
<td>Graphical plots for Periodic models for untransformed and transformed Sg. Segamat flow series.</td>
<td>182</td>
</tr>
<tr>
<td>A.9</td>
<td>Graphical plots for Periodic models for untransformed and transformed Sg. Muar, Johor flow series.</td>
<td>184</td>
</tr>
<tr>
<td>A.10</td>
<td>Graphical plots for Periodic models for untransformed and transformed Sg. Lenggor flow series.</td>
<td>186</td>
</tr>
<tr>
<td>A.11</td>
<td>Graphical plots for Periodic models for untransformed and transformed Sg. Sembrong flow series.</td>
<td>188</td>
</tr>
</tbody>
</table>
APPENDIX B

B.1 Monthly skewness for Disaggregation models for untransformed and transformed Sg. Pedas flow series. 191
B.2 Graphical plots for disaggregation models for untransformed and transformed Sg. Muar flow series. 192
B.3 Graphical plots for disaggregation models for untransformed and transformed Sg. Gemencheh flow series. 194
B.4 Graphical plots for disaggregation models for untransformed and transformed Sg. Melaka flow series. 196
B.5 Graphical plots for disaggregation models for untransformed and transformed Sg. Kesang flow series. 198
B.6 Graphical plots for disaggregation models for untransformed and transformed Sg. Selangor flow series. 200
B.7 Graphical plots for disaggregation models for untransformed and transformed Sg. Bernam flow series. 202
B.8 Graphical plots for disaggregation models for untransformed and transformed Sg. Segamat flow series. 204
B.9 Graphical plots for disaggregation models for untransformed and transformed Sg. Muar flow series. 206
B.10 Graphical plots for disaggregation models for untransformed and transformed Sg. Lenggor flow series. 208
B.11 Graphical plots for disaggregation models for untransformed and transformed Sg. Sembrong flow series. 210
APPENDIX C

C.1  Box-Whisker plots for Sg. Muar  213  
C.2  Box-Whisker plots for Sg. Gemencheh  215  
C.3  Box-Whisker plots for Sg. Melaka  217  
C.4  Box-Whisker plots for Sg. Kesang  219  
C.5  Box-Whisker plots for Sg. Selangor  221  
C.6  Box-Whisker plots for Sg. Bernam  223  
C.7  Box-Whisker plots for Sg. Segamat  225  
C.8  Box-Whisker plots for Sg. Muar, Johor  227  
C.9  Box-Whisker plots for Sg. Lenggor  229  
C.10 Box-Whisker plots for Sg. Sembrong  230  

APPENDIX D

D.1  Monthly skewness for Ladang Union  234  
D.2  Seasonal Rainfall Series Analysis (Ladang Hopeful)  235  
D.3  Seasonal Rainfall Series Analysis (Ladang Bernam)  237  
D.4  Seasonal Rainfall Series Analysis (Ladang Paya Lalang)  239  

APPENDIX E

E.1  Annual Rainfall Series Analysis (Ladang Hopeful)  242  
E.2  Annual Rainfall Series Analysis (Ladang Bernam)  244  
E.3  Annual Rainfall Series Analysis (Ladang Paya Lalang)  246  

APPENDIX F

F.1  Parameter Estimation PAR (1)model  248
CHAPTER 1

INTRODUCTION

1.1 Background

Hydrological data such as flows and rainfall are the basic information used for the design of water resources systems. Hence, the studies are needed regarding the relatively accuracy of the data required for various types of water resources planning and management. Therefore, modeling rainfall and flow series at a useful time and specified scale for different applications has been important problem in hydrology for the last 30 years (Sanso and Guenni, 1999). This situation needs one to select the most appropriate time interval of hydrological data for the design purposes because the characteristics (mostly statistical) behaviors between time interval is different to each other. Based on the literature review done by previous researchers, the overall statistical characteristics (such as standard deviation, skewness coefficient and lag-one correlation coefficient) decrease as the time interval increases. Long sequences of daily rainfall or flows are required increasingly, not only for hydrological purposes but also to provide inputs for models of crop growth, landfills, tailing dams, land disposal of liquid waste and other environmentally-sensitive projects (Buras, 1975). Rainfall and flows are generally measured at the daily time scale and this forms the basis for monthly and annual rainfall and flows series. The need for hourly data for hydrological applications, especially in flood studies, suggests the use of disaggregation model (Koutsoyannis and Onof, 2001). Meanwhile, observations taken in minutes or hours will exhibit temporal dependence will tend decrease and to be
very small or non-existent of the annual scale. Beside of that, as the time interval is longer, the underlying time series becomes simpler to analyze and to model; conversely, as the sample time series is smaller, the amount of information contained in the sample is longer but the characteristics of the series become more complex and the corresponding statistical modeling are more difficult.

Lane (1980) suggested simulating the very short time period because at the short time interval weather persistence and season has an effect. Beside, Salas et al. (1980) proposed the aggregation model regarding to the basic form of the original on lower level time series. In fact, the need to preserve annual and seasonal time series properties inspired the development of simulation models (Salas, 1989; Grygier and Stedinger, 1991, Shah et al., 1996). The monthly and annual hydrological data have been used for the short and medium term planning and operation of water resource systems (Maheepala and Perera, 1996).

Despite the resolution of the time interval, the data quality and accuracy should also be taken into account. The data uncertainties and randomness that is one of the factors that stems from difficulties in estimating future demands for water developments. Shah et al. (1996) stated, there may be some situations where rainfall exhibits reasonable spatial uniformity (e.g. frontal storms over basins with gently varying topography), these tend to be the exception rather than the rule, particularly where a rainfall regime is dominated by convective storms or is subjected to pronounce orographic effects. This situation shows that by applying stochastic approach can currently provide the only effective route towards a hydrological description of rainfall in the absence of satisfactory mathematical and physical representatives of the laws governing its complexity (Stedinger and Taylor, 1982). The stochastic models of daily rainfall with annually varying parameters usually do not preserve the variance of monthly and annual precipitation (Buishand, 1977; Zucchini and Guttorp, 1991; Woolhiser et al., 1993; Boughton, 1999). This underestimation may be due to real long-term trends in rainfall, changes in the data collection techniques or in rain gauge exposure, model inadequacies, and/or the existence of large-scale atmospheric circulation patterns that do not exhibit annual
periodicities (Woolhiser, 1992). Hydrology and its process must also be affected by human activities and various disruptions in nature.

Many hydrologists use the forecasting technique to design and implementations the water resources systems. The main objective of forecasting is to use the time series model fitted to a data set to obtain the most accurate estimate or prediction of the future unknown series. However, forecasting itself unable to test whether or not a class of time series models statistically preserves important historical statistics of the data sets to which the set of models is fitted (Box and Jenkins, 1976). One would quite naturally, like to employ models, which can account for the key statistical characteristics of hydrological time series. However, when a mathematical model can be employed for determining exactly of a system, the model is said to be deterministic. Deterministic models are designed and used for identifying and evaluating system performance in its uncertain environmental and it is a point of implicit stochastic models (Troch et al., 1993).

Stochastic modelings were preceded by structural analysis of temporal and spatial stochastic process such as the analysis of errors, trend types and composition, intermittency, periodicity and stochasticity (Hipel and McLeod, 1994). By considering stochastic hydrology it becomes the light of its application to simulation and optimization in water resources planning and operation. Furthermore, synthetic hydrology stays in fact the overall science of fitting stochastic models to hydrological series and using these models for simulation purposes. The goal of simulation is to employ the fitted model to generate a set of stochastically equivalent observation series, which could possibly occur in the future. Simulation or generation of samples may be very effective tool for experimentally finding the sampling frequency distributions of testing parameters and various other estimates (Yevjevich, 1989). The synthetic flow or rainfall data generation (or in short synthetic hydrology) was later suggests as the term operational hydrology would be more appropriate. It is because, synthetic hydrology can be useful in both analytical and simulation models.
However, there is nothing about approximate methods that makes better use of the limited data and most such approximate methods have been demonstrated to be highly unreliable. There are many techniques, which can be used to adopt the limited data to simulation. By careful use of simulation models, data of poor quality can be checked, missing records completed and a considerable extension of the record can be made (Hipel et al., 1977b; Koch, 1985; Grygier and Stedinger, 1991). The most critical data for simulation are the flows and rainfall series and without both series it is impossible to carry on a study by any other reasonable adequate hydrologic technique and a simulation study. It will usually take no more time to develop the necessary data for simulation than it will require developing the estimates desired. Meanwhile, one simulation of rainfall or flows runs, it will provide an abundance of data, which can answer many hydrologic problems. As an example, if one wish to explain the effect of changing vegetal cover on the watershed, of increasing the amount of urban land use, or other possible land use changes, this is easily done with simulation (Chatfield, 1979). Using conventional methods, is would be difficult, if not impossible, to estimate the effect of such changes. In any case, if time and cost are measured against the quality and completeness of the results, simulation is far ahead of the conventional technique (Hipel and McLeod, 1994). In such case, Loucks et al. (1981) stated that, the type of condensation and storage of data in the format models and estimated parameters are more useful rather than keeping the original data in master data files as is done at present with data bank storage for final backup and for verification purposes when needed.

Based on the simulation methods, the basic disaggregation and aggregation models can be used to simulate and generate both data sequences. Disaggregation models are generally considered a very variable feature for flows and rainfall simulation (Salas et al., 1980). The earliest model such as Thomas-Fiering model (Thomas and Fiering, 1962), H-C model (Harms and Campbell, 1967) and Box-Jenkins model (Box and Jenkins, 1976) seem to currently unsuitable for a fully simulation or generating use. The first well-accepted model was presented by Valencia and Schaake (1973) by developing the Valencia-Schaake technique (VLSH). Further studies, modification and applications of disaggregation model, have been made in the past year such as Mejia and Rouselle (MJRS) model (Mejia and Rouselle,
Aggregation is a new study technique by which an assumption is made regarding the basic form of the original or lower level time series are calculated. Work also has been performed using Fourier series model (Yevjevich, 1984). In order to model adequately the seasonally varying correlation structure and to preserve the stationary statistical properties within each season, one would have to consider the families of the periodic autoregressive (PAR) model (Salas et al., 1980) or periodic autoregressive and moving average (PARMA) model (Vecchia, 1985a). The application of these models has been attractive in simulation area mainly because, the form has an intuitive type of time dependence and they are simplest models to use.

Recent development of using the alternative simulation models available led to developing many software packages. For instance well-known packages are IMSL, STATGRAPHICS, ITSM, SASS/ETS, SPSS and MATLAB. However, despite of the availability of such general-purpose programs, specialized software for simulation of hydrological time series have been attractive because of several reasons (i.e. HEC-4 (U.S ARMY Corps. of Engineer, 1971), LAST (Lane and Frevert, 1988), SPIGOT (Grygier and Stedinger, 1991) and SAMS (Salas et al., 1996)).

Based on the above-mentioned fact, this study will focus on rainfall and flows simulation based on the disaggregation procedures using VLSH model, MJRS model, LANE model and SPIGOT model. Despite of that, the aggregation models in the class of PAR and PARMA models will be used for simulation the flows sequences. Two software packages namely SAMS and SPIGOT would be used to generate the historical flows and rainfall sequences. The effectiveness of models depends on the estimation of model parameters, fitting stage and diagnostic check. The model estimated stage needs to be checked in order to verify how well it represents the historical flow and rainfall series. The evaluation of the selected models are based on the preservation of statistical characteristics such as mean, standard deviation, skewness coefficient and lag one season to season correlation coefficient. It is therefore necessary to evaluate the validity of a model before it is used for such purposes.
1.2 Statement of Problems

Rainfall and flow series are essential parameters for the water resources planning and management. However, the observed rainfall data have randomness, systematic (or inconsistency) and sampling errors based on the effects of anticipate climate change and historical flow data is due to the non-homogeneity (conceived as changes in nature by humans and natural disruptions). As this problem cannot, in general, be solved analytically, a simulation approach must be adopted in which a stochastic model of rainfall or flow is used to generate a long synthetic input series to the mathematical model; the required magnitude frequency relationship can then be estimated from the derived synthetic output series. One of the major problems in water resources design is the selection of the stochastic process to model the given flow or rainfall record. This involves using the historical rainfall and flow records to estimate the model parameters of an appropriate model, which may then be used to simulate the desired length of data series. Various types of stochastic models are available for use in engineering design, such as, aggregation and disaggregation models. For such a system, if simulation is conducted which used only the historical records as inputs data and is then used as a basis for decision. It is implied that the future history of the system will repeat the same pattern, which is hardly ever likely to be the case. Worse little idea of the risks, which will be encountered in making any decision, will be obtained. To avoid this situation, statistical models have been developed which generate synthetic records of flow or rainfall that are statistically similar to observed flow and rainfall records, that can be used in simulating the behavior of water resource systems. However, the generation of flow series required the totally undisturbed observed data sequences. In fact, this situation is quite impossible due to the above-mentioned problem. Due to this need, the rainfall simulation is carried out to overcome the possibility of the weakness in flow simulation. Despite, the simulated rainfall can be transformed to flow using the simple monthly linear rainfall-flow model and the rainfall simulation itself, will supply the synthetic rainfall data to the rainfall-runoff model. The need for the long-term planning of reservoir planning, management and design required the good model for synthetic data generation. However, to show the widespread applicability, the generated data series were evaluate from the preservation of the historical statistical
properties. The study on synthetic simulation will identify the right model to preserve the historical statistical characteristics. In addition, use of the simulation techniques offers the potential benefits to solve natural processes of rainfall and flow pattern based on the statistical characteristics. In reality, the flow and rainfall processes are random and uncertain. Therefore, the stochastic time series modeling is essential to model the random component in the system.

1.3 Objectives

The objectives of this study are:

i. To propose the application of aggregation and disaggregation models in the flow simulation and the application of disaggregation model in the rainfall simulation.

ii. To investigate the performance of aggregation and disaggregation model in the flow simulation and the disaggregation model in the rainfall simulation.

iii. To identify and confirm the best model in flow and rainfall simulation.

1.4 Scope of Study

The study covers the application of aggregation and disaggregation methods for flow simulation and the disaggregation models in the rainfall simulation. The aggregation model usually called as a seasonal series model follows a periodic autoregressive (PAR) and periodic autoregressive and moving average (PARMA) models. For disaggregation model, the well-known model namely, Valencia and Schaake (VLSH), Mejia and Rouselle (MJRS), Lane (LANE) and Grygier and Stedinger (SPIGOT) models are used for flow simulation. Meanwhile, the VLSH, MJRS and LANE models are used for rainfall simulation. The data analysis methods consist of time series plots and Box-Whisker plots. The development stages are
designed to decide the families of models to be considered for fitting to a flow and rainfall series. The overall methodology to fitting models consists of identification, estimation and diagnostic checking. At the identification stage the most suitable models to fit to the data can be selected by examining various types of graphs. If a transformation is required but this fact is not discovered at the identification stage, the need for a data transformation will probably detected at the diagnostic check of model development when properties of the residuals are examined. The data will be transformed using either Box-Cox; Logarithmic or Power transformation. Efficient estimates of the model parameters can be obtained at the estimation stage by employing the method of moments. Following this, the fitted models can be checked for possible inadequacies. The diagnostic checks were employed to ensure that the selected model adequately describes the flow and rainfall series under consideration by subjecting the model to a range of statistical tests. The results of generating monthly sequences will be investigated and compared the historical mean, standard deviation, lag one correlation coefficient and skewness coefficient to identify the best model. The best model is identified based on the model ability to preserve the statistical properties. The flow and rainfall gauging stations are based on four state namely; Negeri Sembilan, Melaka, Selangor and Johor. Despite of this, river basin and catchment area under studies are: Sg. Linggi basin, Sg. Triang basin, Sg. Muar basin, Sg. Selangor basin, Sg. Bernam basin, Sg. Melaka catchment and Sg. Johor basin. Four rainfall stations are under studies which two of them in the Sg. Selangor catchment area and the others in Sg. Melaka catchment and Sg. Segamat catchment. The duration of the monthly records range from less than 20 years to more than 40 years for some stations.

1.5 The Importance of Study

The need for monthly and annual data for hydrological applications, especially in flood studies, suggests the use of aggregation and disaggregation model to use the available data information. In this way, the model would provide a continuous simulation tool for use for simulation studies and design. This study will present an
improved aggregation and disaggregation method for generation of alternative sequences of monthly and annual hydrologic data sequences. This study also proposed the significant advantage over the current models for such studies. The proposed model is therefore a valuable tool for flow and rainfall simulation studies, which abound Malaysia. Using the synthetic data then provides a broad base for development of proper water resources planning and management. The results of this study also provide a new tool for keeping the data in the form of models and estimated parameters rather than in original data.

1.6 Research Hypothesis

To achieve the goal the following hypothesis have been made;

i. The applications of the VLSH model yield a better performance than the widely used disaggregation models for flow and rainfall simulation.

ii. The modeling of periodic series is more complex than modeling the annual series because the former have the influence of the annual cycle which produces the periodic variations in some or all of the statistical characteristics of the series.

iii. The preservation of historical statistical characteristics of rainfall simulation yields a better performance than flow simulation.

iv. The transformation of rainfall and flow series to normal distribution does guarantee the best results in the rainfall and flow simulation.
REFERENCES


