DISSERTATION

DISAGGREGATION OF PRECIPITATION RECORDS

Submitted by

Luis Guillermo Cadavid

Department of Civil Engineering

In partial fulfillment of the requirements

for the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 1991

COLORADO STATE UNIVERSITY

JANUARY 17,1990

WE HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER OUR SUPERVISION BY LUIS GUILLERMO CADAVID ENTITLED DISAGGREGATION OF SHORT TERM PRECIPITATION RECORDS BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

Committee on Graduate Work

Adviser

Department Head

COLORADO STATE UNIVERSITY

ABSTRACT OF DISSERTATION

DISAGGREGATION OF PRECIPITATION RECORDS

This investigation is related to temporal disaggregation of precipitation records. The objective is to formulate algorithms to disaggregate precipitation defined at a given time scale into precipitation of smaller time scales, assuming that a certain mechanism or stochastic process originates the precipitation process. The disaggregation algorithm should preserve the additivity property and the sample statistical properties at several aggregation levels.

Disaggregation algorithms were developed for two models which belong to the class of continuous time point processes: Poisson White Noise (PWN) and Neyman-Scott White Noise (NSWN). Precipitation arrivals are controlled by a counting process and storm activity is represented by instantaneous amounts of precipitation (White Noise terms). Algorithms were tested using simulated samples and data collected at four precipitation stations in Colorado.

The PWN model is the easiest and formulation of the disaggregation model was successful. The algorithm is based on the distribution of the number of arrivals (N) conditional on the total precipitation in the time interval (Y), the distribution of the White Noise terms conditional on N and Y, and the distribution of the arrival times conditional on N. Its application to disaggregate precipitation is limited due to its lack of serial correlation.

However, PWN disaggregation model performs well on PWN simulated samples.

The NSWN is more complex. Required distributions are the same as for the PWN model. Formulation of a disaggregation algorithm was based on theoretical and empirical results. A procedure for model parameters estimation based on weighted least squares was implemented. This procedure reduces the number of estimation failures as compared to method of moments. NSWN disaggregation model performed well on simulated and recorded samples given that parameters used are similar to those controlling the process at the disaggregation scale.

The main shortcoming is the incompatibility of parameter estimates at different aggregation levels. This renders the disaggregation model of limited application. Examination of variation of parameter estimates with the aggregation scale suggests the existence of a region where estimated values appear to be compatible. Finally, it is shown that the use of information at a nearby precipitation station with similar precipitation regime may improve parameter values to use in disaggregation.

Luis Guillermo Cadavid
Civil Engineering Department
Colorado State University
Fort Collins, CO 80523
Summer 1991

ACKNOWLEDGMENTS

The author wishes to express his deepest gratitude to his advisor and professor, Dr. Jose D. Salas, Professor of Civil Engineering, for his constant guidance, encouragement and support, not only in the development of this research, but in the academic and professional life of the author. His financial support is also acknowledged.

The author is also extremely grateful to his professor, Dr. Duane C. Boes, Professor and Head of the Statistics Department, for the hours he spent with the author in pursue of this research, for his cooperation and accessibility, and for the excellent courses he taught to the author during his graduate studies.

In addition, special gratitude is extended to the other members of my committee, Dr. Vujica Yevjevich, Emeritus Professor of Civil Engineering, Dr. Darrel Fontaine, Associate Professor of Civil Engineering, and Dr. Roger E. Smith, from the U. S. Agricultural Research Service, for their support, cooperation and suggestions in my graduate research.

To my wife, Olga, and my son, Felipe, for their love, care, understanding, constant support, encouragement and dedication in the achievement of this study and through all my graduate work. Special acknowledgments for her help in computer work, for her drawing of figures, for her help in typing this report.

Thanks to my family for their encouragement during my studies. To my father, Hernando, who taught me to love engineering. To my brother, Andres, for his constant help and cooperation in my country. To my brother, Eduardo, for his constant friendship. To Juan, for being also my friend.

To my wife's family, for their constant support: Doña Amparo, Nenita, Arturo, Francisco, Margarita and Jorge. My deepest thanks for their cooperation

The partial financial support from the Universidad Nacional de Colombia, Facultad de Minas, Medellin, is gratefully acknowledged. For the cooperation received from Dr. Ricardo Smith, Maria V. Velez, Lilian Posada, Dr. Oscar Mesa and Felipe Ospina, members of this institution.

Special appreciation is extended to my friends Jorge, Ana and Pablo Restrepo, for their constant friendship, care and company during the development of my graduate studies at Colorado State University.

To my graduate student fellows, Luis Pinzon, Fernando Pons, Xianou Guo, Nieng-Jin Liou and Ding Ching Wang, for their friendship and cooperation.

The Colorado State University Computer Center is duly acknowledged for providing time in the Cyber 205 supercomputer. Special thanks to John Cooley for his cooperation.

This study was funded through the Agricultural Experimental Station Project COLO 357 "Characterizing the Properties of Precipitation for Agricultural Areas in Colorado". This support is gratefully acknowledged. In addition, the technical support obtained

through the NSF Grant INT-8715102 on "U.S.-Italy Cooperative Research on Prediction of Floods and Precipitation Events" is also acknowledged.

A Olga, mi esposa, quien ha sido mi luz
A Felipe, mi hijo, ese nuevo candil que apenas comienza a iluminar
A los candiles que han de venir
A Hernando, mi padre, por darme luz
A Ruth, mi madre, cuya luz nos dejó temprano
A Kiko y Amparo, quienes dieron luz a Olga

To Olga, my wife, who has been my light
To Felipe, my son, that new candle just starting to light up
To the candles to come
To Hernando, my father, who gave me light
To Ruth, my mother, whose light left us so early
To Kiko and Amparo, who gave light to Olga

TABLE OF CONTENTS

Chapter		Page
1	INTRODUCTION	1
	1.1 General	1
	1.2 Precipitation disaggregation problem	4
	1.3 Objectives	5
	1.4 Report organization	7
2	LITERATURE REVIEW	8
	2.1 Introduction	8
	2.2 Precipitation modeling	8
	2.3 Precipitation disaggregation models	17
	2.4 Point processes theory	20
	2.5 Selection of models for disaggregation process formulation	21
3	ANALYSIS OF PRECIPITATION RECORDS	24
	3.1 Introduction	24
	3.2 Data description and selection of gauging stations	24
	3.3 Estimation of sample properties	27
	3.4 Computation of aggregated samples	31
	3.5 Results for monthly statistics	32
	3.6 Effect of the aggregation operation on sample statistics	49
	3.7 Clustering of precipitation data	64
4	POISSON WHITE NOISE MODEL FOR PRECIPITATION	69
	4.1 Introduction	69
	4.2 Model description	69
	4.3 Properties of the PWN precipitation process	72
	4.4 Exponential case	74
	4.5 Gamma case	76
	4.6 Parameter estimation by method of moments	78
	4.7 Results for estimated parameters	79
5	POISSON WHITE NOISE DISAGGREGATION MODEL FOR PRECIPITATION	89
	5.1 Introduction	89
	5.2 Disaggregation model structure	89
	5.3 Distribution of N(T) conditional on Y	90
	5.4 Distribution of U_i, conditional on Y, N(T) and U_{i-1}	92
	5.5 Distribution of T_n conditional on N(T)	98

Chapter		Page
5	5.6 Disaggregation model algorithm	99
	5.7 Testing the PWND model	101
	5.8 Application of the PWND model to historical precipitation samples	105
6	NEYMAN-SCOTT WHITE NOISE MODEL FOR PRECIPITATION	118
	6.1 Introduction	118
	6.2 Model description	118
	6.3 Probability generating functional and probability generating function	125
	6.4 Properties of the NS point process	132
	6.5 Marginal probability distribution function for $N(t_1)$	135
	6.6 Factorial moments for the NS process	143
	6.7 Properties of the NSWN precipitation process	144
	6.8 Parameter estimation by method of moments	149
	6.9 Results for parameters estimated by method of moments	151
	6.10 Estimation of parameters using Weighted Least Squares method	160
	6.11 Results for parameters estimated using Weighted Least Squares method	164
	6.12 Comparison of estimation methods	170
	6.13 Extrapolation of the marginal distribution for N(T)	180
7	NEYMAN-SCOTT WHITE NOISE DISAGGREGATION MODEL FOR PRECIPITATION	188
	7.1 Introduction	188
	7.2 Distribution of N(T) conditional on Y	189
	7.3 Distribution of U_i, conditional on Y, N(T) and U_{i-1}	190
	7.4 Distribution of T_n conditional on N(T)	191
	7.5 Simulation and sampling of arrival times	200
	7.6 Disaggregation model algorithm	202
	7.7 Testing the NSWND model	204
	7.8 Application of the PWND model to historical precipitation samples	212
	7.9 Variation of parameter estimates with the aggregation scale	223
	7.10 Improvement of parameter estimates using information at another station	229
8	CONCLUSIONS AND RECOMMENDATIONS	237
	8.1 Summary	237
	8.2 Conclusions	240
	8.3 Recommendations	245
	REFERENCES	249

LIST OF TABLES

Table		Page
3.1	Main characteristics for the precipitation recording stations used in the study	26
3.2	Monthly statistics for precipitation at Denver Wsfo Ap station	33
3.3	Monthly statistics for precipitation at Greenland 9 SE station	34
3.4	Monthly statistics for precipitation at Idaho Springs station	35
3.5	Monthly statistics for precipitation at Ward station	36
4.1	Estimated parameters for the PWN model for Denver Wsfo Ap station, for different temporal aggregation levels	80
4.2	Estimated parameters for the PWN model for Greenland 9 SE station, for different temporal aggregation levels	81
4.3	Estimated parameters for the PWN model for Idaho Springs station, for different temporal aggregation levels	82
4.4	Estimated parameters for the PWN model for Ward station, for different temporal aggregation levels	83
5.1	Summary of results for parameters estimated from PWN simulated precipitation series, for simulation 1 (Exponential case) and different aggregation levels	102
5.2	Summary of results for parameters estimated from PWN simulated precipitation series, for simulation 2 (Gamma case) and different aggregation levels	103
5.3	Comparison of statistics for PWN simulated and disaggregated series, for simulation 1 (Exponential Case) and simulation 2 (Gamma case)	104
5.4	Description of historical samples for application of the PWND model	105
5.5	Comparison of statistics for historical and PWN disaggregated hourly series, for Denver Wsfo Ap station	107
5.6	Comparison of statistics for historical and PWN disaggregated daily series, for Denver Wsfo Ap station	108
5.7	Comparison of statistics for historical and PWN disaggregated hourly series, for Greenland 9 SE station	109
5.8	Comparison of statistics for historical and PWN disaggregated daily series, for Greenland 9 SE station	110

Table		Page
5.9	Comparison of statistics for historical and PWN disaggregated 5 min. series, for Idaho Springs station	111
5.10	Comparison of statistics for historical and PWN disaggregated 5 min. series, for Ward station	112
6.1	Number of terms $m_{n,i}$ for each power λ^i, to evaluate derivatives of order n for g(z) at z=0, from n=3 to n=20	139
6.2	Sequences $\{\ell_k$, k=1,..,i+1$\}$ and coefficients $C_{n,i,j}$ to evaluate derivatives of order n for g(z) at z=0, for n from 3 to 11	140
6.3	Summary of results for parameters estimated from NSWN simulated precipitation series, using method of moments, for simulation 1, for different aggregation levels	152
6.4	Estimated parameters for the NSWN model for Denver Wsfo Ap station, for different temporal aggregation levels, using method of moments	156
6.5	Estimated parameters for the NSWN model for Greenland 9 SE station, for different temporal aggregation levels, using method of moments	157
6.6	Estimated parameters for the NSWN model for Idaho Springs station, for different temporal aggregation levels, using method of moments	158
6.7	Estimated parameters for the NSWN model for Ward station, for different temporal aggregation levels, using method of moments	159
6.8	Summary of results for parameters estimated from NSWN simulated precipitation series using WLS method, for simulation 1, for different aggregation levels	165
6.9	Estimated parameters for the NSWN model for Denver Wsfo Ap station, for different temporal aggregation levels, using WLS method	166
6.10	Estimated parameters for the NSWN model for Greenland 9 SE station, for different temporal aggregation levels, using WLS method	167
6.11	Estimated parameters for the NSWN model for Idaho Springs station, for different temporal aggregation levels, using WLS method	168
6.12	Estimated parameters for the NSWN model for Ward station, for different temporal aggregation levels, using WLS method	169
6.13	Bias, Root Mean Square Error (RMSE) and number of failures for parameters estimated for NSWN simulation 1, for two different temporal scales	171
7.1	Summary of statistics estimated from NSWN simulation 1 disaggregated hourly precipitation series, for different choices of parameters to use in the disaggregation	211

Table		Page
7.2	Description of historical samples for application of the NSWND model.....	212
7.3	Comparison of statistics for historical and NSWN disaggregated hourly series, for Denver Wsfo Ap station.....	213
7.4	Comparison of statistics for historical and NSWN disaggregated hourly series, for Greenland 9 SE station.....	214
7.5	Comparison of statistics for historical and NSWN disaggregated hourly series, for Idaho Springs station.....	215
7.6	Comparison of statistics for historical and NSWN disaggregated hourly series, for Ward station.....	216
7.7	Description of Historical samples for application of the NSWND model using information at another station..	230
7.8	Parameter values for application of the NSWND model using information at another station.....	230
7.9	Comparison of statistics for historical and NSWN disaggregated hourly series, for some months of the precipitation recording stations, when disaggregation is performed using information at a nearby station.....	231

LIST OF FIGURES

Figure		Page
3.1	Location for the precipitation recording stations used in the study	26
3.2	Monthly statistics for Denver Wsfo Ap station, for T=60 min	37
3.3	Monthly statistics for Greenland 9 SE station, for T=60 min	38
3.4	Monthly statistics for Denver Wsfo Ap station, for T_a=1440 min	39
3.5	Monthly statistics for Greenland 9 SE station, for T_a=1440 min	40
3.6	Monthly statistics for Idaho Springs station, for T=5 min	41
3.7	Monthly statistics for Ward station, for T=5 min	42
3.8	Monthly statistics for Idaho Springs station, for T_a= 60 min	43
3.9	Monthly statistics for Ward station, for T_a=60 min	44
3.10	Sample correlograms for months 03 and 04, for Denver Wsfo Ap station	45
3.11	Sample correlograms for months 09 and 10, for Greenland 9 SE station	46
3.12	Sample correlograms for months 03 and 04, for Idaho Springs station	47
3.13	Sample correlograms for months 09 and 10, for Ward station	48
3.14	Ratio of statistics for two aggregation levels for Denver Wsfo Ap station	56
3.15	Ratio of statistics for two aggregation levels for Greenland 9 SE station	57
3.16	Ratio of statistics for two aggregation levels for Idaho Springs station	58
3.17	Ratio of statistics for two aggregation levels for Ward station	59
3.18	Recorded hourly precipitation at Denver Wsfo Ap station, between 1949/05/05/01 and 1949/06/04/24	65
3.19	Recorded hourly precipitation at Greenland 9 SE station, between 1969/10/01/01 and 1969/10/31/24	66
3.20	Recorded 5 min. precipitation at Idaho Springs station, between 1984/02/12/00 and 1984/02/14/13	67

Figure		Page
3.21	Recorded 5 min. precipitation at Ward station, between 1984/04/20/00 and 1984/04/22/13.....	68
4.1	Schematic representation of the Poisson White Noise precipitation process.....	70
4.2	Historical and PWN computed probability distribution functions for the amount of precipitation, for Denver Wsfo Ap station, for months 06 and 07 and for T_a= 1440 min.....	86
4.3	Historical and PWN computed probability distribution functions for the amount of precipitation, for Greenland 9 SE station, for months 05 and 08 and for T_a= month.....	87
4.4	Historical and PWN computed probability distribution functions for the amount of precipitation, for Idaho Springs station, for months 02 and 07 and for T_a= 60 min.....	88
5.1	Sample probability distribution functions for hourly historical and PWN disaggregated precipitation series, for Denver Wsfo Ap station, for months 06 and 07.....	113
5.2	Sample probability distribution functions for daily historical and PWN disaggregated precipitation series, for Greenland 9 SE station, for months 05 and 08.....	114
5.3	Sample probability distribution functions for 5 min. historical and PWN disaggregated precipitation series, for Idaho Springs station, for months 02 and 07.....	115
6.1	Schematic representation of the Neyman-Scott White Noise precipitation process.....	121
6.2	Theoretical and estimated correlograms for NSWN simulated process, for simulation 1, for T=60 min.....	153
6.3	Theoretical and estimated correlograms for NSWN simulated process, for simulation 1, for T=1440 min...	154
6.4	Estimation results for NSWN process simulation 1, for T= 60 min.....	172
6.5	Estimation results for NSWN process simulation 1, for T_a= 1440 min.....	173
6.6	Historical and fitted correlograms for NSWN simulation 2, for T=60 min., for months 11 and 12.....	175
6.7	Role of changing weighting method on fitted correlograms (T_a=1440 min., simulation 2).....	176
6.8	Role of weighting methods and number of terms in the objective function on fitted correlograms (T=5 min., simulation 3).....	176
6.9	Historical and NSWN fitted correlograms (WLS), for Denver Wsfo Ap station, for months 03 and 04, for T_a=1440 min.....	177

Figure		Page
6.10	Historical and NSWN fitted correlograms (WLS), for Idaho Springs station, for months 01 and 02, for T=5 min.....	178
6.11	Historical and NSWN fitted correlograms (moment), for Greenland 9 SE station, for months 01 and 02, for T=60 min.....	179
6.12	Example of computed, empirically extrapolated and simulated distributions for the N(T) in the NS process, for λ=0.00009475 1/min., p=0.02273, α=0.03257 1/min. and T=60 min.....	184
6.13	Historical and NSWN computed probability distribution functions for the amount of precipitation, for Denver Wsfo Ap station, for months 03 and 08 and for T_a= 1440 min.....	186
6.14	Historical and NSWN computed probability distribution functions for the amount of precipitation, for Idaho Springs station, for months 02 and 08 and for T_a= 60 min.....	187
7.1	Example of the cdf for the first arrival time, conditional on one arrival in the interval, for p=0.730, α=0.000356 1/min. and T=1440 min.....	197
7.2	Example of the cdf for the first arrival time, conditional on two arrivals in the interval, for λ=0.000086 1/min, p=0.730, α=0.000356 1/min. and T=1440 min.....	198
7.3	Example of the cdf for the second arrival time, conditional on two arrivals in the interval, for λ=0.000086 1/min, p=0.730, α=0.000356 1/min., T=1440 min. and t_1=400 min.....	199
7.4	Theoretical and estimated correlograms for NSWN disaggregated series, for simulation 1, using parameters estimated at T_a=1440 min.....	205
7.5	Theoretical and estimated correlograms for NSWN disaggregated precipitation series, for simulation 1, using parameters estimated at T=60 min.....	206
7.6	Theoretical and estimated correlograms for NSWN disaggregated precipitation series, for simulation 1, using population parameters (Sampling by splitting and rescaling).....	208
7.7	Theoretical and estimated correlograms for NSWN disaggregated precipitation series, for simulation 1, using population parameters (Uniform sampling).....	209
7.8	Correlograms and sample probability distribution function for hourly historic and NSWN disaggregated precipitation series, for Denver Wsfo Ap station, for months 03 and 08 (T_d=60 min.).....	218

Figure		Page
7.9	Correlograms and sample probability distribution function for hourly historic and NSWN disaggregated precipitation series, for Greenland 9 SE station, for months 07 and 08 (T_d=60 min.)	219
7.10	Correlograms and sample probability distribution function for 5 min. historic and NSWN disaggregated precipitation series, for Idaho Springs station, for months 02 and 08 (T_d=5 min.)	220
7.11	Correlograms and sample probability distribution function for 5 min. historic and NSWN disaggregated precipitation series, for Ward station, for months 11 and 12 (T_d=5 min.)	221
7.12	Variation of estimated parameter values with the aggregation scale for NSWN simulation 1 (Method of moments)	224
7.13	Variation of estimated parameter values with the aggregation scale for NSWN simulation 1 (WLS method)	225
7.14	Variation of estimated parameter values for the NSWN model with the aggregation scale, for Denver Wsfo Ap station (WLS method)	226
7.15	Variation of estimated parameter values for the NSWN model with the aggregation scale, for Greenland 9 SE station (WLS method)	227
7.16	Correlograms and sample probability distribution function for hourly historic and NSWN disaggregated precipitation series, for Denver Wsfo Ap station, for month 06, obtained using information at Greenland 9 SE station (T_d=60 min.)	232
7.17	Correlograms and sample probability distribution function for hourly historic and NSWN disaggregated precipitation series, for Greenland 9 SE station, for month 06, obtained using information at Denver Wsfo Ap station (T_d=60 min.)	233
7.18	Correlograms and sample probability distribution function for hourly historic and NSWN disaggregated precipitation series, for Idaho Springs station, for month 08, obtained using information at Ward station (T_d=60 min.)	234
7.19	Correlograms and sample probability distribution function for hourly historic and NSWN disaggregated precipitation series, for Ward station, for month 04, obtained using information at Idaho Springs station (T_d=60 min.)	235

Chapter 1

INTRODUCTION

1.1 General

It is difficult to find today, some publication in hydrology or water resources where the importance of precipitation is not stressed or at least suggested. The role that precipitation plays in the hydrological cycle as a continuous process is well established, as one of the main variables driving many other processes, like evaporation, infiltration, surface runoff and streamflow. Many authors emphasize the requirement of detailed spatial and time precipitation records in order to adequately describe, on a physical basis, many of these processes.

Examples of the use of precipitation records in different theoretical and applied problems are abundant in the literature. It has been realized that description of different processes in hydrology require precipitation records at different spatial and time resolutions, since those processes occur at different temporal and spatial scales of the precipitation process. For instance, analysis of infiltration using physically based models requires point rainfall data at a resolution of minutes, while drought related problems may require records spanning over decades, perhaps in a monthly or larger time scale, and covering entire geographical regions (Berndtsson and Niemczynowicz, 1988). Giambelluca (1987) points out how the use of

monthly precipitation values in monthly water balance models may lead to large errors in the estimation of monthly evapotranspiration and groundwater recharge.

In the field of water resources systems planning, management and operation, precipitation can not be forgotten if realistic and optimal solutions are expected. Reservoir operation problems may require an adequate forecast of precipitation as an indicator of future water availability. Also, the daily operation of irrigation systems in agriculture may require daily precipitation data as input.

When detailed design for water conveyance structures is required and no discharge records are available, design discharge hydrographs are often obtained via continuous simulation, requiring as input adequate precipitation events.

The process of precipitation as it evolves in nature is not the same as that recorded in gauging stations. Precipitation is a space-time phenomenon and its complete description requires knowledge of the intensity field at any time and at any point in space. It is customary to define the intensity field as a non negative stochastic process indexed by space and time coordinates. Its description corresponds to a mixture of theories like random fields, atmospheric sciences, thermodynamics and hydrology, among others.

The intermittence of the intensity process has posed important difficulties in the description of precipitation. If it were possible to record or describe the intensity process in continuous fashion, many important features like storm duration, storm arrival time, total yield per storm and spatial and temporal distribution of intensity could be obtained directly.

However, the recorded precipitation process is usually composed of cumulative amounts of precipitation over disjoint intervals of time. This aggregation process lumps together many of the important characteristics previously mentioned. In most practical cases, it is common to find precipitation records available at hourly or daily scales, from which values at larger scales can be obtained by the simple operation of aggregation. Unfortunately, precipitation records for scales smaller than the day or the hour are scarce and it is not easy nor simple to apply the inverse operation of disaggregation.

Even if continuous traces are available at one site, it may be difficult to translate these into computer data (digitize, in computer terminology). In regard to spatial variation, radar imagery is helpful in some instances. However, recording and storing a large amount of echoes in the Plan Position Indicator might become a formidable task, as cumbersome as that posed by continuous recording. This constraint further imposes the use of discrete or aggregated precipitation series.

As stated previously, the study of several hydrologic processes require precipitation data at a level of aggregation shorter than the resolution provided by the available sample. Unfortunately, precipitation data is usually, at most, in the form of hourly values. Besides, high resolution data is unwieldy to handle and store. One of the techniques called to solve, at least partially, some of the problems described above is precipitation modeling in general and precipitation disaggregation in the absence of short sampling interval precipitation records.

Moss and Lins (1988) describe an expanded program aimed to assess the hydrologic implications of climatic changes. The program is divided in three parts, one of them being the research element. Within the research component, model development plays a very important role, for general atmospheric circulation models, basin or catchment hydrologic models and stochastic models. Within the stochastic class of models, areal and temporal disaggregation models are viewed as important tools to link general circulation models to basin or catchment hydrologic models.

1.2 Precipitation disaggregation problem

Disaggregation techniques are almost as old as the first models for streamflow generation and they appear due to the necessity to obtain hydrologic traces with a better resolution than those originally recorded or simulated. The objective of any disaggregation scheme is to provide such resolution while preserving statistical properties at more than one aggregation level (Salas et al., 1985). An additional condition has been imposed, in some cases, on disaggregation models, although this is not explicitly stated in the definitions given in the literature: Added disaggregated values must reproduce the original values. Disaggregation techniques provide additional information in the sense that they show probable distributions of the process within the original aggregation scale.

A similar working definition is adopted here for precipitation disaggregation. Given a recorded historical precipitation trace and an assumed underlying stochastic process governing precipitation formation, the objective in the disaggregation operation is to obtain precipitation data at a finer time resolution, in such a way that

properties of the underlying process, sample trace, in terms of recorded amounts, and observed statistics at several temporal levels are all preserved. All these are desirable properties for the disaggregation process.

Examination of streamflow disaggregation models, for example the Valencia-Schaake model (Salas et al. 1985), shows a similar working definition. The basic assumption is that all variables follow a multinormal distribution at different aggregation levels. If correlations are preserved normality is preserved.

Streamflow disaggregation models are difficult, if not impossible, to apply to short sampling interval precipitation records. First, they do not incorporate intermittence of the process. Second, they are not built to preserve the type of correlation found in short sampling interval precipitation records.

The focus in this research is on temporal precipitation disaggregation. One could equally think of spatial rainfall disaggregation, i. e., how to break an average value of precipitation over a large area into values corresponding to smaller areal portions. However, the problem of spatial disaggregation is not considered part of this research.

1.3 Objectives

The overall objective of the present research is to develop improved techniques for disaggregation of precipitation samples into higher resolution series, in which the intermittence property is present, such as daily, hourly, or fraction of the hour series.

In order to accomplish this overall objective, four intermediate objectives are formulated:

1. A review of the actual state of the art in precipitation modeling, and the position of that state in relation to the formulation of precipitation disaggregation schemes. The accomplishment of this objective will allow selection of models for the disaggregation problem.
2. Development of precipitation disaggregation models which could be used for different aggregation and disaggregation scales, under the presence of different mechanisms generating precipitation and under the presence of the two most common forms of precipitation in the U. S.: rain and snow. These models should be able to reproduce important characteristics observed at the aggregation and disaggregation scales.
3. Test and assessment of the performance and applicability of the developed models using precipitation data recorded in stations located in the state of Colorado, for different temporal measurement scales, and simulated series drawn from the selected models.
4. Presentation of conclusions and recommendations to guide future research in the field of precipitation disaggregation.

Analysis of developed models is based on moments and corresponding sample estimates up to the third order, including correlograms, for the amount of precipitation in a given time interval. The probability distribution function for this random variable is also analyzed.

1.4 Report organization

Chapter 2 is dedicated to review the state of the art in precipitation modeling and precipitation disaggregation. This review allowed the selection of White Noise point process models as a first approach to the problem of precipitation disaggregation.

Chapter 3 deals with examination of precipitation records for gauging stations in the state of Colorado, in particular the ones to be used in testing the models. Main seasonal statistics are estimated and characteristics of the precipitation process are examined, such as periodicity and clustering. The role of the aggregation scale on estimated statistics is also investigated.

Chapters 4 and 6 are dedicated to the review of the basic models selected to formulate disaggregation procedures. Models are presented, their properties are derived and estimation of parameters is developed. Parameters are estimated for simulated traces and for the recording precipitation stations used in this study.

In chapters 5 and 7 disaggregation models are developed based on results presented in the chapters 4 and 6. Required distributions are derived and disaggregation algorithms are outlined. Models are tested, both on simulated and recorded precipitation traces.

Due to space limitations, many derivations are not included in the text. The reader is referred to a longer version of the report for this research (Cadavid et al, 1991). Also, a large amount of computer code was developed as part of this research. No computer program listings or manuals are presented in this report. These materials can be requested from the Hydrology and Water Resources Program, Civil Engineering Department, at Colorado State University.

Chapter 2

LITERATURE REVIEW

2.1 Introduction

This chapter is devoted to review the current state of the art in precipitation modeling. It is segmented into four parts. The first one covers the broad area of precipitation modeling. In the second section, the topic of precipitation disaggregation is surveyed. Third, but not less important, some publications and articles on recent trends in the theory of stochastic point processes are listed. Finally, the chapter is closed with the selection of models for the development of this research.

This chapter fulfills the first objective described in Chapter 1. Besides, it is expected that it will serve future research in the area of precipitation modeling, in collection and review of existing literature. In this regard, key publications are pointed out.

2.2 Precipitation modeling

After scanning the available literature, several trends emerge in regard to precipitation modeling (Waymire and Gupta, 1981a), although in some instances they are not well differentiated and in other cases a mixture of them is used. Approaches based on the application of thermodynamics and fluid mechanics to precipitation generating mechanisms in the atmosphere form the first category. Although this approach just begins, some encouraging results have been

obtained. Georgakakos and Bras (1984a, 1984b) use cloud physics to formulate a physically based precipitation model, for which they report satisfactory results.

Precipitation modeling has been enhanced by the observation of regularities present in different precipitation formations, such as cyclonic thunderstorms, air mass thuderstorms and frontal thunderstorms. An example of this type of application is given by Amorocho and Wu (1977). The observed regularities are: subsynoptic areas, large mesoscale areas (LMSA), small mesoscale areas (SMSA), clusters of convective cells and convective cells (Corotis, 1976; Amorocho and Wu, 1977; Waymire and Gupta, 1981a). The enumeration order corresponds to areal extension, in such a way that subsynoptic areas cover larger areas than LMSA, and so on. Besides, the larger the area the lower the mean precipitation intensity and the longer the time they last. For instance, LMSA's are of the order of 1000 to 10000 km^2 lasting for several hours, while convective cells cover around 5 km^2 and last from 1 to about 30 minutes. For convective cells, developing, mature and dissipating stages are important concepts used in precipitation modeling in regard to time and space distribution of intensity. Downward and upward currents of moist air formed during these three stages are responsible for clustering of cells in time and space (regeneration of cells), which at the same time is the mechanism responsible for the large serial correlation observed in short sampling interval precipitation records. In many cases, the lower intensity enclosing smaller formations is called "background noise or background precipitation" (Garcia-Bartual and Marco, 1987). All these regularities are more or less present in all

formations and their analysis has been facilitated by the use of satellite and radar imagery.

Space-time or multidimensional modeling of precipitation is accomplished when some of the aforementioned regularities are reproduced using the theory of random fields. Examples are found in Bras and Rodriguez-Iturbe (1976), Waymire et al. (1984), Bell (1987), Rodriguez-Iturbe and Eagleson (1987), Eagleson et al. (1987), Sivapalan and Wood (1987), Islam et al. (1988), Jacobs et al. (1988), among others.

At this point, special mention is made of the articles by Waymire and Gupta (1981a, 1981b, 1981c), where the state of the art in stochastic modeling of precipitation up to 1980 is reviewed, important mathematical concepts and tools related to the theory of point processes are summarized (probability generating functional), and examples of application of these tools to some hydrologic processes are presented. This series is highly recommended in regard to applications of the theory of point processes in hydrology.

As previously stated, many of the processes used to describe the stochastic nature of precipitation belong to the theory of point processes. In this regard, Foufoula-Georgiou and Lettenmaier (1987) point out two alternatives for modeling precipitation using stochastic processes. In the first approach, the sequence of dry and rainy periods is described using discrete point process theory. In the second approach, an underlying continuous time process governing precipitation formation is assumed and estimation is performed based on observations of the integrated process over certain time intervals. Examples are: Cox processes (doubly stochastic), Neyman-Scott cluster processes, Bartlett-Lewis cluster processes, renewal processes, Markov

processes and Poisson processes. Models with Poisson arrivals have been used in hydrology for quite some decades now (Todorovic and Yevjevich, 1969). Neyman-Scott processes, although introduced to rainfall modeling since the 60's (Le Cam, 1961), did not develop popularity until the middle 70's (Kavvas and Delleur, 1975, 1981). The Neyman-Scott process was initially formulated to describe the distribution of galaxies in space (Neyman and Scott, 1952). The Bartlett-Lewis process was introduced as an alternative to the Neyman-Scott Rectangular Pulse Model by Rodriguez-Iturbe et al. (1987a).

Within the general class of multidimensional models for precipitation, a special set has developed most of the attention, called temporal precipitation models. In these models, spatial variability is not considered and time evolution of the intensity process at one point is studied. The models cited in the literature are: Poisson White Noise Model (PWN), Poisson Rectangular Pulse model (PRP), Neyman-Scott White Noise model (NSWN), Neyman-Scott Rectangular Pulse model (NSRP), and Bartlett-Lewis Rectangular Pulse model (BLRP) (Rodriguez-Iturbe et al. ,1984; Rodriguez-Iturbe 1986; Rodriguez-Iturbe et al., 1987). An excellent review of the NSWN model is presented by Foufoula-Georgiou and Guttorp (1986).

Temporal precipitation models exhibit two components. The first component, also known as counting process, models the arrival in time of precipitation events and dry periods duration. The second one describes internal characteristics, like precipitation amounts, intensities and storm durations. For the first component, Poisson processes, Neyman-Scott and Bartlett-Lewis clustering processes are used. For modeling internal characteristics either instantaneous amounts or volumes of precipitation, known as white noise terms, or

random intensities constant over a random durations, known as rectangular pulses, are used. Rodriguez-Iturbe et al. (1987a) add noise or jitter to the rectangular pulse. For a presentation and review of white noise models and rectangular pulse models the reader is referred to Rodriguez-Iturbe et al. (1984), Rodriguez-Iturbe (1986), Foufoula-Georgiou and Guttorp (1986), Rodriguez-Iturbe et al. (1987a) and Rodriguez-Iturbe et al. (1987b).

It must be noted that the term White Noise as used in this research is somewhat different from the term used in time series analysis. Here, White Noise terms belong to a sequence of identically and independently distributed random variables, following a common distribution, not necessarily the Normal distribution.

Neyman-Scott and Bartlett-Lewis processes were introduced in order to simulate cell clustering as described before. Arrival times of cluster centers follow a Poisson process. Around each cluster center, a random cluster size (number of cluster members) and cluster members locations are generated, representing cell activity. For example, in the Neyman-Scott process, cluster size is assumed to be geometric (Kavvas and Delleur 1975, 1981: Ramirez and Bras, 1982, 1985) or Poisson (Rodriguez-Iturbe et al., 1984), while cluster members are located to the right of the the cluster center according to an exponential distribution. On the other hand, the Bartlett-Lewis process uses a finite renewal chain to simulate cluster characteristics.

The Poissonian models, PWN and PRP, have been recognized as poor due to their inability to properly describe the correlation structure present in precipitation records. For example, PWN model lacks correlation, while the PRP model generates a Markovian correlation by

means of a finite probability of the same pulse covering two different disjoint time intervals.

In the family of cluster models correlation is reproduced by a finite probability of cluster members, belonging to the same cluster center, to fall in disjoint time intervals.

One of the advantages of temporal precipitation models is the physical meaning attached to the parameters of the models. In regard to cluster models, Foufoula-Georgiou and Guttorp (1986) show that selection of the distribution of the cluster size, geometric or Poisson, may result in different estimates for some properties of the process, questioning therefore the physical interpretation of parameters.

In the continuous point process formulation precipitation intensity is described first as a function of time. From this, the cumulative rainfall or average intensity process for non overlapping time intervals is obtained and its second order properties are derived. The estimation of parameters is usually done by method of moments, equating population moments to their sample counterpart values.

In doubly stochastic processes (Cox processes) some of the parameters controlling occurrences are considered samples from a different unobservable stochastic process. For instance, Smith and Karr (1985) use a Poisson arrival model with the rate being sampled from a two state Markov process. Also, Entekhabi et al. (1989) modified the NSRP model by considering the parameter of the exponential distribution controlling pulse duration as a random variable.

A lot of research effort has been invested in temporal precipitation modeling in the last years, although there are still controversies and unsolved questions about the applicability of these models. One of the most covered topics in the literature is the compatibility of models at different temporal aggregation levels or scales (Rodriguez-Iturbe et al., 1984; Foufoula-Georgiou and Guttorp, 1986; Rodriguez-Iturbe et al., 1987b; Obeysekera et al., 1987; Liu and Salas, 1988). This compatibility refers to the fact that parameters estimated at different temporal scales are different and to the fact that correlation structure is not preserved at scales different from that in which parameters were estimated. This issue has led to the problem of temporal scale and self-similarity for the precipitation processes (Rodriguez-Iturbe, 1986; Kedem and Chiu, 1987), which has been investigated trough concepts like scale of fluctuation and variance ratio (Obeysekera et al., 1987).

A second topic analyzed in the literature, in regard to temporal precipitation models, is estimation of parameters by moments and maximum likelihood methods (Smith and Karr, 1985; Foufoula-Georgiou and Guttorp, 1986; Obeysekera et al., 1987; Smith, 1987; Liu and Salas, 1988). However, important problems are still encountered. Method of moments poses problems of existence and numerical convergence. Maximum likelihood is practically impossible in most of the cases. Guttorp (1986) and Foufoula-Georgiou and Guttorp (1986) present an innovative estimation procedure, based on binary series formed from occurrence times in the precipitation sample, ignoring amounts of precipitation in the interval. With these sample series, they formulate moment and maximum likelihood estimation techniques. A different approach is used by Kavvas and Delleur (1975, 1981) and

Ramirez and Bras (1982, 1985). They fit the theoretical spectrum of counts and the log-survivor function of the Neyman-Scott process to the respective sample estimates.

Foufoula-Georgiou and Guttorp (1986) present an excellent review of the Neyman Scott White Noise model for precipitation. One of their conclusions, that the model is scale dependent or time scale inconsistent, is not new and was previously detected by other authors (Rodriguez-Iturbe et al., 1984). They recommend that any NSWN model fitted to precipitation records at a given aggregation level should not be extrapolated to other scales. Their stronger conclusion is that the NSWN model does not provide an adequate representation of precipitation generating mechanisms, and therefore no physical meaning can be attached to model parameters. Summarizing, they consider the model descriptive and time dependent, with no predictive ability.

The last approach in precipitation modeling summarized here uses tools developed for time series analysis. In this approach, it is assumed that means and covariances are sufficient to describe temporal and spatial variation of the process. A first order Markov chain is the most used type of model. However, this is valid only for Gaussian fields and poor performance has been reported (Georgakakos and Bras, 1984a and 1984b; Foufoula-Georgiou and Lettenmaier, 1987) when these models are applied to short sampling interval precipitation modeling. An example of this approach is found in Raudkivi and Lawgon (1974). Chang et al. (1984) used discrete ARMA models (DARMA) to describe daily precipitation in Indiana.

As stated before, not all authors follow a continuous description of the precipitation intensity field in order to model precipitation. For example, many start directly with the discrete process specified

at given aggregation scales. There are advocates for both types of approaches. In the later approach, works by Roldan and Woolhiser (1982), Woolhiser and Roldan (1982), Guttorp (1986), Foufoula-Georgiou and Lettenmaier (1987) and Smith (1987) are cited. Alternating Renewal, Binary, Markov, Markov Renewal and Markov Bernoulli processes are used in this theoretical setting. The main point sustained by some of these authors is the inadequacy of the conversion of a continuous process into a discrete process, which may result in misleading inferences in regard to precipitation clustering.

Without regard to which approach is used, some authors model internal characteristics of the rainfall process, others work on external properties, while others consider both. An innovative idea, in regard to internal characteristics, is introduced by Garcia-Bartual and Marco (1987). They consider a temporal rainfall model in which intensity within the storm follows a Gamma function with its maximum value and total storm yield treated as random variables. These type of exponential decaying laws are not usually included in temporal rainfall models, but they often appear in spatial representations. An additional example of the investigation of internal characteristics is given by Nguyen and Rousselle (1981), where the time distribution of hourly precipitation is investigated.

Waymire and Gupta (1981a) attribute the lack of unified approach to model precipitation to two main reasons. The first is given by the nature of rainfall itself, which is not unique. Despite the aforementioned observed regularities, precipitation in different regions of the globe or at different locations in a large geographic area behaves differently. A similar consideration is valid for different time scales. The second one stems from the diversity of

mathematical tools used in rainfall modeling. This is the reason for which they attempt to form a common and general theoretical background (Waymire and Gupta, 1981b).

A last point is emphasized in regard to stochastic modeling of precipitation. Complete description of any stochastic processes requires specification of all its finite dimensional joint distribution functions. However, except for very few cases in the reviewed literature, this important requirement is not met due to the difficulties that it poses.

2.3 Precipitation disaggregation models

The first feature found in the literature about precipitation disaggregation models is that they are scarce. The opposite is true for streamflow modeling, where a lot of attention has been given to this topic and adequate models have been formulated.

The first short sampling interval rainfall disaggregation scheme presented here was introduced in the literature by Woolhiser and Osborn (1985a). The cumulative rainfall process for a given storm is described by means of ten rescaled incremental dimensionless rainfall depths. The process is made dimensionless after dividing by the total storm depth and rescaled once the dimensionless increment in each interval is divided by the total dimensionless depth between the beginning of the current interval and the end of the storm. Although time is made also dimensionless, this is masked in by working with ten intervals. It is noted that rescaled increments for two consecutive intervals are not independent, which is accounted for by means of linear dependence between the expected value of the next rescaled increment and the same quantity in the current interval. The marginal

distribution for the rescaled increment in the the first interval is assumed to be Beta, and so are the distributions for the remaining increments conditional on the increment value in the previous interval.

As described, the model has a total of 26 parameters. The number of parameters is reduced after some of them are regressed on the index ordering the intervals. Parameters are estimated using maximum likelihood. The model is applied and tested using precipitation data for a gauging station in Arizona. Tests are performed on the basis of the likelihood ratio test and the Akaike Information Criteria. Dependence of model parameters on amount and duration of storms, reproduction of frequencies for observed intensities and preservation of internal correlations are investigated. The general conclusion is that model structure provides an acceptable approximation for summer thunderstorm rainfall in the area where it is tested.

In a subsequent publication (Woolhiser and Osborn, 1985b) seasonal and regional effects on the model are investigated. Tests are performed by examining the impact of model parameters on peak runoff frequency distributions. The conclusion, conditional on the data used, is that the model requires seasonal and regional differentiation of parameters.

A more complete disaggregation scheme is later discussed by Woolhiser and Econopouly (1986) and completed by Hershenhorn and Woolhiser (1987). In this case, beginning with daily amounts of rain, a model is proposed to obtain within the day magnitudes for the number of storms, amount, duration and arrival time for each storm. They

report how simulated sequences compare well with observed values. Storms obtained are further disaggregated by using the previously described model.

Although the described models constitute an important attempt to set up a procedure to disaggregate rainfall, they are not completely satisfactory. In general, the models are heavy, they lack simplicity in structure, estimation and application. Many of the working hypotheses and assumptions do not have a physical appealing meaning. Apparently, many transformations on the original data are required to obtain good results. Another shortcoming is the lack of flexibility in the number of intervals. If this number were increased, the number of parameters would increase by a factor of three.

A second disaggregation model was developed by Ormsbee (1989). It is completely empirical in the sense that rainfall is disaggregated by assuming linear similarity between within hour rainfall distribution and the hourly counterpart. Comparison is performed in terms of runoff quantities obtained by means of continuous simulation, for both recorded and disaggregated traces of rainfall. Although the model is simple and does not require large amounts of data or computation time, it lacks, as the previous case, physical basis.

As a last reference, Giambelluca and Oki (1987) used a 10-state seasonal first order Markov chain to disaggregate monthly rainfall values into daily values. Dimensionless values of daily precipitation were obtained simulating, with the Markov chain, the transition from the current state to the following day state. Once the process is in a given state, the dimensionless amount of rain in that state is obtained by sampling from seasonal Weibull distributions previously

fitted to the dimensionless data. Monthly amounts of precipitation are preserved after daily simulated dimensionless amounts are multiplied by the corresponding monthly value.

2.4 Point processes theory

This section is not tutorial neither complete in reference to publications related to the theory of point processes. It is intended to point out publications which were found useful in the development of this research. In this sense, it is based on the personal experience of the author.

The first recommended text (Taylor and Karlin, 1984) is basic and introductory for the general theory of stochastic processes. It presents in a clear form the basics of Poisson and renewal processes.

For a more formal presentation of the theory of Poisson processes, the reader is referred to Parzen (1964). Many variations of this process are presented and specially important to the Poisson Rectangular Pulse model is the extension of the filtered Poisson process to stochastic response functions.

The introduction of cluster processes can be reviewed in the articles by Neyman and Scott (1952), Neyman and Scott (1958), Le Cam (1961) and Moyal (1962). As mentioned before, Le Cam (1961) was the first to apply cluster processes to the study of precipitation. However, the reader must be warned that these articles are highly technical, directed mostly to the statistical community.

An important tool, developed parallel to the theory of cluster processes, is the probability generating functional for a random measure, as an extension of the probability generating function for an integer valued random variable. In regard to this, the reader is

referred to Moyal (1962), Vere-Jones (1968), Westcott (1972) and Daley and Vere-Jones (1972). As before, the publications listed above are highly technical. In regard to the theory of point process and the probability generating functional applied to stochastic processes in hydrology, the reader is encouraged to review and follow the series of three articles published by Waymire and Gupta (1981a, 1981b and 1981c). Kavvas and Delleur (1975) and Ramirez and Bras (1982) also present examples of application of the probability generating functional to precipitation modeling. In other areas of geophysics, Vere-Jones (1970) describes modeling of earthquake occurrences using clustering processes. Finally, concepts about interarrival times specification of point processes are found in Lawrance (1972).

Although most of the publications referred above are highly technical, the text by Cox and Isham (1980) was very helpful in understanding many of the concepts and derivations related to the theory of point processes. It is highly recommended to review this text after the basics of the theory of stochastic processes have been covered (Taylor and Karlin, 1984; Parzen, 1964).

2.5 Selection of models for disaggregation process formulation

Given the actual state of the art in precipitation modeling, some of the approaches related above could be used in order to formulate temporal disaggregation schemes. Approaches considering spatial variation are eliminated at once since spatial disaggregation is not one of the objectives of this research.

The approach based on the theory of point processes could be extremely useful, specially in the part dealing with temporal rainfall modeling. However two important shortcomings emerge, which have been

discussed previously. The first one is given by the incompatibility of temporal rainfall models at different aggregation scales. If disaggregation models were to be built from temporal rainfall models, doubts would arise about the basic requirements, the preservation of the underlying stochastic process and its parameters. The second difficulty stems directly from the formulation of the models. They do not specify all possible finite joint probability distribution functions, either for the intensity or the aggregated process. Since disaggregation involves addition up to recorded values, herculean tasks could be faced by following this approach. Temporal rainfall models have not been formulated with the purpose of disaggregation.

The approach based on time series analysis would be worthy to try. Traditional approaches, like autoregressive moving average models, have to be abandoned and more sophisticated tools such as DARMA or Alternating Renewal Markov models, should be tried.

Finally, the rainfall disaggregation models describe in this chapter do not show feasibility of further improvement.

Despite the aforementioned shortcomings for the point process approach and specially for temporal rainfall models, it was decided to adopt four models as the bases for the formulation of precipitation disaggregation models in this research: Poisson White Noise, Poisson Rectangular Pulse, Neyman-Scott White Noise and Neyman-Scott Rectangular Pulse. Unfortunately, during the development of this research, it was impossible to derive some of the required distributions for Rectangular Pulse type models and they were abandoned.

The followed procedure started with the simplest model, Poisson White Noise precipitation process. From there, the investigation proceeded to the Neyman-Scott White Noise model. After reviewing Chapters 4 to 6, the reader will be aware of the degree of difficulty attained in this step.

The use of approaches based on discrete point processes and based on time series analysis were left as material for further research. They could have been also selected. However, it is precisely the continuous formulation of temporal precipitation models what makes them so attractive to work the problem of precipitation disaggregation.

Chapter 3

ANALYSIS OF PRECIPITATION RECORDS

3.1 Introduction

This chapter describes the data used in the present study as well as the criteria used to select four gauging stations from larger sets of available information. Methods of estimation of seasonal statistics are presented, along with the corresponding results at different temporal aggregation levels. Some analysis is done in regard to the behavior of estimated statistics at different aggregation levels.

3.2 Data description and selection of gauging stations

Two large sets of precipitation data were available to select the gauging stations used in the development of this research. In the first set, composed of 39 precipitation recording stations located in northeastern Colorado, the basic information is given as hourly amounts of precipitation, recorded to a precision of 0.01 inches (in.), with recording periods ranging from 3 to 36 years. Part of this set was previously used by Obeysekera et al. (1986) and Liu and Salas (1988). The second set is composed of amounts of precipitation recorded every 5 minutes (min.) to a precision of 0.01 in., for 22 stations located also in northeastern Colorado (Liu and Salas, 1988). For this set of stations, the recording period is 5 years.

Due to the objectives stated for this study and the detail required in the analysis of precipitation information, it was impossible to work with 61 gauging stations. Therefore, a criteria was adopted to select four of them, two from the hourly data set and two from the 5 min. data set. The criteria is based on the type of records found in precipitation samples: zero or positive values, precipitation traces smaller than 0.01 in. and missing values. Unfortunately, in the original samples, precipitation traces smaller than 0.01 in. and missing values were recorded using the same negative value. Therefore, it was impossible to differentiate these and necessary to catalog all of them as missing data. Consequently, the criteria adopted was to select stations, within each set, with the longest recorded period, with the minimum number of missing data and with the maximum number of positive recorded values. Table 3.1 gives some characteristics for the selected precipitation stations and Figure 3.1 shows the location of these stations in Colorado. Note how for 5 min. stations the amount of missing data, approximately 30%, is relatively more important than for hourly data, where this amount does not exceed 3.7%. Spatial location was not considered in the selection of the stations. The temporal measurement scale, denoted by T in Table 3.1, is defined as the period of time over which cumulative or average amounts of precipitation are recorded. It is always given in minutes. In the sequel, the adjective temporal will be omitted, since this research does not deal with spatial aggregation or averaging, and precipitation will be understood as cumulative amounts over a given interval.

Table 3.1 Main characteristics for the precipitation recording stations used in the study.

Code	Name	Latitude North	Longitude West	Elevation (ft.)	Period of Record
52220	Denver Wsfo Ap	39°45′	104°52′	5280	1948-1983
53579	Greenland 9 SE	39°06′	104°44′	7480	1948-1983
ISGC2	Idaho Springs	39°41′	105°30′	3456	1983-1987
WRDC2	Ward	40°02′	105°32′	3005	1983-1987

Name	T (min.)	Sample Size	Positive Values	Number of Zero Values	Missing Values
Denver Wsfo Ap	60	315576	13213	297153	5210
Greenland 9 SE	60	315576	10799	293099	11678
Idaho Springs	5	526176	6377	366726	153073
Ward	5	526176	7562	358684	159930

T: Measurement scale

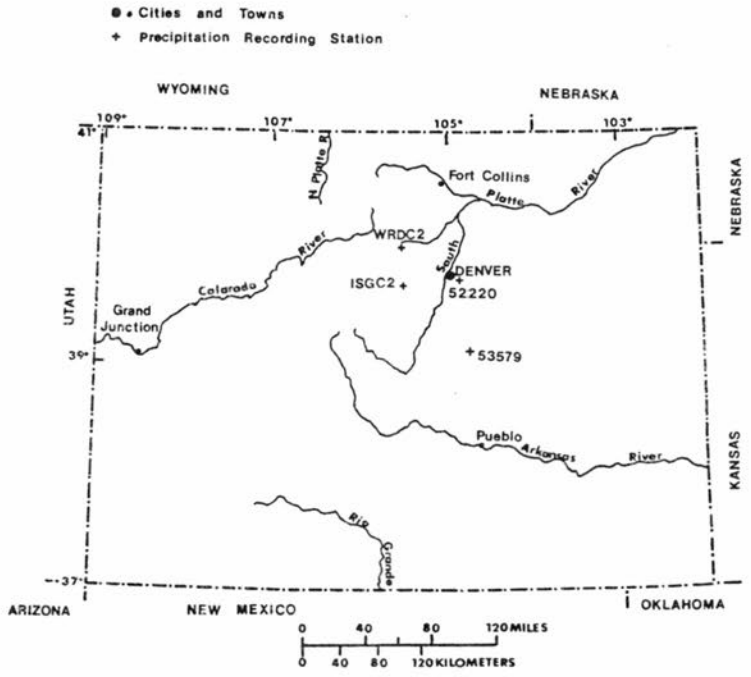


Figure 3.1 Location for the precipitation recording stations used in the study.

3.3 Estimation of sample properties

So far, features such as censored and missing data have been detected in the precipitation samples. Besides of these properties, many authors have identified periodicity, both diurnal and seasonal, as important characteristics of precipitation samples (Obeysekera et al. 1987). Diurnal periodicity appears for stations in Colorado during summer months, where the frequency of afternoon or evening thunderstorms of convective type is larger than for the rest of the day. On the other hand, for winter months, precipitation events of the frontal mass type tend to be evenly distributed along the day. Seasonal periodicity appears due to the presence of different mechanisms controlling precipitation formation during different seasons of the year.

The traditional approach used to take into account seasonal periodicity is by dividing the year in seasons and considering that the process is stationary and homogeneous within each season. Some authors suggest to account for diurnal periodicity by using models which parameters change within the day. However, this suggestion has not been developed yet, and it will not be examined in this research either.

For processing precipitation data two of the four characteristics outlined above were considered: seasonal periodicity and missing values. Diurnal periodicity was neglected and precipitation amounts were treated as if they were recorded continuously in the interval $[0,+\infty)$. In order to account for seasonal periodicity, the year was partitioned in 12 seasons, which coincide with the 12 calendar months. The treatment of missing data in the estimation of statistics will be clarified in the following paragraphs.

The aggregation scale, denoted by T_a, is the period of time over which cumulative amounts of precipitation are computed. In this work, the aggregation scale is always an integer multiple of the measurement scale, indicating that an integer number of recorded precipitation values have been added to obtain the aggregated value.

The monthly statistics estimated for each station were: mean, standard deviation, skewness coefficient, kurtosis coefficient and correlogram. The expressions used are based on Salas et al. (1985). Assume that for a given measurement or aggregation scale, the sample is represented as $y_{i,\tau,k}$, i=1,..,N, τ=1,..,NS and k=1,..,NT(τ), where N is the total number of years in the sample, NS is the number of seasons in the year and NT(τ) is the number of periods of length T or T_a in season τ. Also, let MT(τ) be the sample size available for estimation of statistics in season τ, computed as:

$$MT(\tau) = \sum_{i=1}^{N} \sum_{k=1}^{NT(\tau)} I_{[0,\infty)}(y_{i,\tau,k}) \qquad (3.1)$$

where $I_A(y)$ is the indicator function and it takes the value of one if y is contained in the set A. Equation (3.1) assumes that missing data are coded as negative values, in which cases the indicator function takes the value of zero.

Mean, $\bar{Y}_\tau$, standard deviation, S_τ, skewness, $\hat{g}_\tau$, and kurtosis, k_τ, for season τ are estimated using:

$$\bar{Y}_\tau = \frac{1}{MT(\tau)} \sum_{i=1}^{N} \sum_{k=1}^{NT(\tau)} y_{i,\tau,k} \qquad (3.2)$$

$$S_\tau = \left[\frac{1}{MT(\tau)-1} \sum_{i=1}^{N} \sum_{k=1}^{NT(\tau)} \left(y_{i,\tau,k} - \bar{Y}_\tau \right)^2 \right]^{1/2} \qquad (3.3)$$

$$\hat{g}_\tau = \frac{MT(\tau)}{[MT(\tau)-1][MT(\tau)-2]S_\tau^3} \sum_{i=1}^{N} \sum_{k=1}^{NT(\tau)} \left(y_{i,\tau,k} - \bar{Y}_\tau \right)^3 \tag{3.4}$$

$$k_\tau = \frac{MT(\tau)^2}{[MT(\tau)-1][MT(\tau)-2][MT(\tau)-3]S_\tau^4} \sum_{i=1}^{N} \sum_{k=1}^{NT(\tau)} \left(y_{i,\tau,k} - \bar{Y}_\tau \right)^4 \tag{3.5}$$

Although it is not explicitly stated, in eqs. (3.2) to (3.5) missing values are skipped in the summations.

The sample autocorrelation coefficient of lag k for season τ is obtained according to the following set of expressions:

$$MT'(\tau) = \sum_{i=1}^{N} \sum_{j=1}^{NT(\tau)-k} I_{[0,\infty)}(y_{i,\tau,j})\, I_{[0,\infty)}(y_{i,\tau,j+k}) \tag{3.6}$$

$$r_\tau(k) = \frac{\sum_{i=1}^{N} \sum_{j=1}^{NT(\tau)-k} \left(y_{i,\tau,j} - \bar{Y}_\tau^{(1)} \right)\left(y_{i,\tau,j+k} - \bar{Y}_\tau^{(2)} \right)}{MT'(\tau)\, S_\tau^{(1)}\, S_\tau^{(2)}} \tag{3.7}$$

$$\bar{Y}_\tau^{(1)} = \frac{1}{MT'(\tau)} \sum_{i=1}^{N} \sum_{j=1}^{NT(\tau)-k} y_{i,\tau,j} \tag{3.8}$$

$$\bar{Y}_\tau^{(2)} = \frac{1}{MT'(\tau)} \sum_{i=1}^{N} \sum_{j=k}^{NT(\tau)} y_{i,\tau,j} \tag{3.9}$$

$$S_\tau^{(1)} = \left[\frac{1}{MT'(\tau)-1} \sum_{i=1}^{N} \sum_{j=1}^{NT(\tau)-k} \left(y_{i,\tau,j} - \bar{Y}_\tau^{(1)} \right)^2 \right]^{1/2} \tag{3.10}$$

$$S_\tau^{(2)} = \left[\frac{1}{MT'(\tau)-1} \sum_{i=1}^{N} \sum_{j=k}^{NT(\tau)} \left(y_{i,\tau,j} - \bar{Y}_\tau^{(2)} \right)^2 \right]^{1/2} \tag{3.11}$$

Equation (3.7) has been written in such a way that cross products between the last k values in the season for year i and the first k values in the same season for year i+1 are not included, so that introduction of false correlation is avoided. Therefore, in order to

preserve the boundaries for $r_\tau(k)$ ($\pm$ 1.0), means and standard deviations are computed using the first or last NT(τ)-k values in the samples within each season. As a final observation, the lag to use in eq. (3.7) should not exceed the number of sample values in a given month NT(τ).

As shown in eq. (3.6), for correlation computations, any missing value in one of the series causes the elimination of the corresponding term in the other series. Those values are also eliminated for computation of means and standard deviations.

In addition to the statistics listed above, sample probability distribution functions (pdf) for the amount of precipitation in a given time interval were also estimated for each month. The estimation procedure selected is described in Salas et al. (1987), and referred to as frequency analysis for grouped data. Although censorship of precipitation samples was not taken into account, the pdf for the amount of precipitation in a given time interval is still composed of a discrete part, p_o^τ, and a continuous part, $g(y)^\tau$, and can be written for month τ as:

$$f_Y^\tau(y) = p_o^\tau \, I_{\{0\}}(y) + g(y)^\tau \, I_{(0,\infty)}(y) \qquad (3.12)$$

In the first indicator in (3.12), the set corresponds to a singleton, while for the second one, it corresponds to the positive real numbers, excluding zero.

The estimator for the probability of zero precipitation in any time interval for month τ is

$$\hat{p}_o^\tau = \frac{1}{MT(\tau)} \sum_{i=1}^{N} \sum_{k=1}^{NT(\tau)} I_{\{0\}}(y_{i,\tau,k}) \qquad (3.13)$$

The number of class intervals, n_c, for estimation of $g(y)^\tau$ is computed as:

$$NP(\tau) = \sum_{i=1}^{N} \sum_{k=1}^{NT(\tau)} I_{(0,\infty)}(y_{i,\tau,k}) \tag{3.14}$$

$$n_c = 1.5 + 3.322 \, \mathrm{Log}[NP(\tau)] \tag{3.15}$$

where $NP(\tau)$ is the number of positive precipitation values available for month τ.

The maximum and minimum values, y_{max} and y_{min}, among those positive in the sample, are computed next. The class interval length, Δy, is given by:

$$\Delta y = (y_{max} - y_{min})/(n_c - 1) \tag{3.16}$$

Class marks, y_n, $n=1,..,n_c$, are computed according to:

$$y_1 = y_{min} + \Delta y/2 \tag{3.17}$$

$$y_n = y_{n-1} + \Delta y, \text{ for } n=2,..,n_c \tag{3.18}$$

Finally, an estimate of $g(y_n)^{\tau}$ is computed as

$$\hat{g}(y_n)^{\tau} = \frac{1}{MT(\tau)\,\Delta y} \sum_{i=1}^{N} \sum_{k=1}^{NT(\tau)} I_{(y_n - \Delta y/2,\ y_n + \Delta y/2)}(y_{i,\tau,k}) \tag{3.19}$$

As before, missing values are not considered in the computations.

3.4 Computation of aggregated samples

The derivation of samples for a given aggregation scale T_a, from the sample given at a measurement scale T, is accomplished by adding recorded precipitation values for an integer number of periods of length T. In the sequel, this integer number of periods will be known as scale ratio, will be denoted by R and it is simply

$$R = T_a / T \tag{3.20}$$

The aggregated sample, denoted as $x_{i,\tau,j}$, $i=1,..,N$, $\tau=1,..,NS$ and $j=1,..,NT(\tau)/R$, is computed from the recorded sample using

$$x_{i,\tau,j} = \sum_{k=(j-1)R+1}^{jR} y_{i,\tau,k} \qquad (3.21)$$

In principle, missing values are skipped in the summation in eq. (3.21). Since this study is related to precipitation modeling, any attempt to fill in missing data, previously to fitting any model, is worthless. It was decided to declare an aggregated value as missing only when all of the R values in that period were missing. Otherwise, the aggregated value was made equal to the summation of the non missing values.

Seasonal statistics for the aggregated series are obtained using the procedure described in Section 3.3, replacing $NT(\tau)$ by $NT(\tau)/R$ and $y_{i,\tau,k}$ by $x_{i,\tau,k}$.

3.5 Results for monthly statistics

For the four gauging stations selected in this study, monthly statistics were estimated for a total of ten samples. Four of these samples correspond to the recorded process and six to aggregated series. For hourly samples, daily and monthly samples were derived, while for 5 min. samples, hourly series were obtained. Some results are given in Tables 3.2 to 3.5 and in Figures 3.2 to 3.13. Additional plots for correlograms and sample probability distribution functions are given along the text as required (Cadavid et al., 1991). For the monthly cases, the aggregation scale is the product of the number of days in the month times the number of minutes in one day. Also, for these cases, a different notation, $\hat{r}^{*}_{\tau}(k)$, has been adopted for the sample autocorrelation function, since these coefficients now

Table 3.2 Monthly statistics for precipitation at Denver Wsfo Ap station.

T_a = 60 min.

r	$\bar{Y}_r$ (in.)	S_r (in.)	$\hat{g}_r$	k_r	$r_r(1)$	$r_r(2)$	$\hat{p}_o^r$
1	0.0006801	0.0049772	12.11	205.62	0.71691	0.57100	0.9665131
2	0.0009086	0.0061519	11.15	175.27	0.68910	0.50174	0.9611395
3	0.0017654	0.0099400	10.12	149.03	0.75612	0.57768	0.9355223
4	0.0024694	0.0155213	12.25	224.28	0.69156	0.51496	0.9359616
5	0.0034961	0.0228363	21.81	960.54	0.47548	0.33986	0.9294167
6	0.0024015	0.0218696	19.64	591.11	0.44566	0.26083	0.9604574
7	0.0024328	0.0291912	23.68	768.34	0.20808	0.07550	0.9688556
8	0.0019180	0.0255126	29.22	1204.05	0.22711	0.08915	0.9733677
9	0.0015879	0.0167423	34.67	1990.73	0.29277	0.22636	0.9669833
10	0.0012549	0.0102260	18.67	644.42	0.59933	0.42638	0.9653524
11	0.0011448	0.0074182	10.04	134.50	0.72793	0.56389	0.9580052
12	0.0007904	0.0060122	13.45	254.63	0.75781	0.62583	0.9661213

T_a = 1440 min.

r	$\bar{Y}_r$ (in.)	S_r (in.)	$\hat{g}_r$	k_r	$r_r(1)$	$r_r(2)$	$\hat{p}_o^r$
1	0.0163226	0.0649583	7.40	77.53	0.13211	0.01616	0.8101382
2	0.0218061	0.0774986	5.93	49.84	0.08530	-0.03051	0.8030612
3	0.0423687	0.1380392	9.13	140.71	0.15227	-0.04510	0.7161290
4	0.0592381	0.2117069	8.41	102.80	0.05288	-0.03717	0.7123810
5	0.0838618	0.2654400	5.89	49.42	0.18217	0.01565	0.6599078
6	0.0576095	0.2112194	7.71	81.95	0.17241	0.02770	0.7038095
7	0.0583871	0.1849830	5.11	35.00	0.11189	0.05657	0.7059908
8	0.0460125	0.1671292	6.23	50.21	0.22378	0.02164	0.7204301
9	0.0380741	0.1419492	5.93	44.33	0.19281	0.02931	0.7972222
10	0.0301165	0.1197811	6.52	58.76	0.21452	-0.01881	0.8422939
11	0.0274630	0.0927824	4.85	31.24	0.18084	-0.04898	0.8166667
12	0.0189606	0.0894527	13.29	252.89	0.08562	-0.01674	0.8252688

T_a = Month

r	$\bar{Y}_r$ (in.)	S_r (in.)	$\hat{g}_r$	k_r	$r_r^*(1)$	$r_r^*(2)$	$\hat{p}_o^r$
1	0.5060000	0.3790173	0.904	2.878	-0.00869	0.14110	0.0
2	0.6157143	0.4367412	0.545	2.745	0.25003	-0.08628	0.0
3	1.3134286	0.8419618	1.883	8.584	-0.12137	0.01647	0.0
4	1.7771429	1.0148068	0.737	3.248	-0.03522	-0.02657	0.0
5	2.5997143	1.7250737	0.608	3.519	0.30198	0.28168	0.0
6	1.7282857	1.3159530	0.722	2.679	0.01442	-0.03789	0.0
7	1.8100000	1.2566154	1.690	7.067	0.11701	0.03277	0.0
8	1.4263889	1.2045513	1.855	7.748	-0.07728	0.05448	0.0
9	1.1422222	1.0664524	1.340	5.030	-0.02473	0.13738	0.0
10	0.9336111	0.9066678	1.746	6.619	0.02527	0.03813	0.0
11	0.8238889	0.5279409	1.442	5.955	-0.18474	-0.16777	0.0
12	0.5877778	0.5962411	2.387	9.532	0.07146	0.06738	0.0

Table 3.3 Monthly statistics for precipitation at Greenland 9 SE station

T = 60 min.

T_a = Month

τ	$\bar{Y}_\tau$ (in.)	S_τ (in.)	$\hat{g}_\tau$	k_τ	$r_\tau(1)$	$r_\tau(2)$	$\hat{p}_o^\tau$
1	0.0004928	0.0044552	14.64	273.15	0.33363	0.27145	0.9775606
2	0.0007259	0.0054339	12.31	192.91	0.40728	0.33988	0.9666269
3	0.0015395	0.0099860	9.55	111.68	0.45365	0.42059	0.9506144
4	0.0018894	0.0125486	10.60	152.05	0.47697	0.36627	0.9506627
5	0.0030465	0.0200147	14.42	375.71	0.35303	0.25007	0.9457752
6	0.0027236	0.0246846	20.59	643.61	0.29861	0.16409	0.9637932
7	0.0037136	0.0360284	26.78	1165.90	0.23827	0.07237	0.9549006
8	0.0033891	0.0344460	21.02	619.05	0.30741	0.08667	0.9636665
9	0.0013185	0.0144451	26.59	1186.13	0.28296	0.14386	0.9755950
10	0.0008649	0.0078730	13.38	237.73	0.46125	0.36387	0.9776705
11	0.0009577	0.0078314	12.98	238.79	0.45557	0.37100	0.9701135
12	0.0006582	0.0060834	14.41	260.91	0.40307	0.38676	0.9753730

T_a = 1440 min.

τ	$\bar{Y}_\tau$ (in.)	S_τ (in.)	$\hat{g}_\tau$	k_τ	$r_\tau(1)$	$r_\tau(2)$	$\hat{p}_o^\tau$
1	0.0117821	0.0411307	5.24	38.48	0.16647	-0.00790	0.8707295
2	0.0173953	0.0542382	4.32	25.08	0.06205	-0.01618	0.8396323
3	0.0367867	0.1214637	7.62	83.03	0.19350	-0.02297	0.7516159
4	0.0446609	0.1446564	8.57	128.35	0.14886	-0.03833	0.7507163
5	0.0725092	0.1973372	5.55	50.84	0.20096	-0.01673	0.7047970
6	0.0651877	0.1919705	5.02	36.10	0.17734	0.06068	0.7490119
7	0.0887137	0.2392607	6.23	61.10	0.08858	0.02712	0.6411992
8	0.0811439	0.2389898	5.81	48.63	0.07041	0.08758	0.6955720
9	0.0313902	0.1095939	5.93	51.37	0.19753	0.05199	0.8341323
10	0.0205835	0.0872723	6.87	63.22	0.22611	0.01867	0.8850987
11	0.0256286	0.0860681	6.04	50.62	0.22821	0.00680	0.8561905
12	0.0157348	0.0753473	11.95	206.03	0.11814	-0.01567	0.8682796

τ	$\bar{Y}_\tau$ (in.)	S_τ (in.)	$\hat{g}_\tau$	k_τ	$r_\tau^*(1)$	$r_\tau^*(2)$	$\hat{p}_o^\tau$
1	0.3645714	0.2134499	0.117	2.159	0.21843	0.18979	0.0
2	0.4900000	0.2522371	0.831	4.280	0.12694	0.33532	0.0
3	1.1382857	0.9897608	2.240	8.177	-0.09511	-0.19019	0.0
4	1.3360000	1.0023977	0.716	2.607	0.16796	-0.14947	0.0
5	2.2457143	1.4521126	0.936	3.753	0.30838	0.30940	0.0
6	1.8848571	1.2380354	0.901	3.778	0.06357	-0.03071	0.02857
7	2.6979412	1.2714291	0.465	2.938	-0.02839	-0.22582	0.0
8	2.4433333	1.5453063	1.006	4.440	-0.09509	-0.03478	0.0
9	0.9354286	0.7166029	0.890	3.256	-0.10763	-0.01482	0.0
10	0.6369444	0.7045721	2.431	11.337	0.16115	0.06723	0.02778
11	0.6848571	0.6563732	1.528	4.805	0.19474	-0.04714	0.02857
12	0.4877778	0.5287295	2.459	10.655	0.01980	0.09003	0.05556

Table 3.4 Monthly statistics for precipitation at Idaho Springs station

T = 5 min.

τ	$\bar{Y}_\tau$ (in.)	S_τ (in.)	$\hat{g}_\tau$	k_τ	$r_\tau(1)$	$r_\tau(2)$	$\hat{p}_o^\tau$
1	0.0002547	0.0017737	10.75	230.70	0.25269	0.28805	0.9763262
2	0.0004874	0.0025347	5.95	43.28	0.58621	0.67938	0.9595122
3	0.0001207	0.0011477	10.67	147.29	0.18445	0.19485	0.9884593
4	0.0001601	0.0013659	10.52	162.91	0.39180	0.37793	0.9850894
5	0.0000822	0.0010367	15.88	338.30	0.33658	0.27216	0.9928228
6	0.0001633	0.0019787	19.91	561.77	0.49601	0.41441	0.9894793
7	0.0002498	0.0033656	27.24	1034.59	0.65573	0.47055	0.9876537
8	0.0003329	0.0040560	28.57	1263.63	0.71317	0.55197	0.9838920
9	0.0002406	0.0106349	108.54	9784.89	0.38566	0.20153	0.9919039
10	0.0000463	0.0010781	61.40	5890.69	0.26773	0.19383	0.9964561
11	0.0003904	0.0021816	6.78	64.27	0.45688	0.51587	0.9651364
12	0.0002762	0.0016558	6.14	45.55	0.04178	0.09343	0.9725571

T_a = 60 min.

τ	$\bar{Y}_\tau$ (in.)	S_τ (in.)	$\hat{g}_\tau$	k_τ	$r_\tau(1)$	$r_\tau(2)$	$\hat{p}_o^\tau$
1	0.0029778	0.0117532	16.54	413.55	0.68238	0.46113	0.8161616
2	0.0057225	0.0243752	6.78	56.66	0.91200	0.84436	0.8483771
3	0.0014121	0.0064867	8.99	113.32	0.56301	0.33660	0.9095806
4	0.0018825	0.0103056	8.08	80.41	0.71492	0.55227	0.9337776
5	0.0009501	0.0061687	10.91	151.83	0.47899	0.37325	0.9519339
6	0.0018804	0.0135363	11.95	198.54	0.38086	0.16050	0.9566295
7	0.0029266	0.0242398	18.76	487.92	0.20179	0.07064	0.9458863
8	0.0038999	0.0313368	20.25	609.54	0.37593	0.17952	0.9434683
9	0.0028188	0.0572481	42.27	1960.46	0.02574	0.00407	0.9688958
10	0.0005410	0.0053664	18.23	420.30	0.14537	0.06379	0.9771767
11	0.0046043	0.0184925	6.05	48.21	0.82507	0.74022	0.8841348
12	0.0032535	0.0094797	4.24	28.41	0.86863	0.78902	0.8563171

Table 3.5 Monthly statistics for precipitation at Ward station
T = 5 min.

τ	$\bar{Y}_\tau$ (in.)	S_τ (in.)	$\hat{g}_\tau$	k_τ	$r_\tau(1)$	$r_\tau(2)$	$\hat{p}_o^\tau$
1	0.0000385	0.0006341	17.11	314.64	0.06578	0.08122	0.9962441
2	0.0004028	0.0020769	6.26	74.20	0.24581	0.34073	0.9613895
3	0.0001135	0.0010923	10.18	118.60	0.17474	0.23815	0.9889755
4	0.0004434	0.0026457	8.30	100.11	0.66053	0.67333	0.9653969
5	0.0005960	0.0044361	35.18	2516.50	0.51198	0.50135	0.9581723
6	0.0003732	0.0071452	89.74	7890.23	0.41144	0.43298	0.9802281
7	0.0004041	0.0045374	26.37	994.17	0.71801	0.50466	0.9801361
8	0.0003497	0.0040986	29.15	1452.18	0.60646	0.43882	0.9825144
9	0.0001740	0.0021773	56.97	5920.69	0.33908	0.34074	0.9865223
10	0.0002562	0.0027001	62.18	6629.07	0.34337	0.20862	0.9786616
11	0.0002516	0.0016431	7.09	62.44	0.32006	0.39202	0.9759543
12	0.0000583	0.0007704	13.42	188.50	0.05710	0.13137	0.9942347

T_a = 60 min.

τ	$\bar{Y}_\tau$ (in.)	S_τ (in.)	$\hat{g}_\tau$	k_τ	$r_\tau(1)$	$r_\tau(2)$	$\hat{p}_o^\tau$
1	0.0004526	0.0031995	15.35	381.89	0.60738	0.34696	0.9671533
2	0.0047368	0.0154022	5.08	36.39	0.83079	0.73155	0.8404344
3	0.0013295	0.0067778	7.19	66.87	0.65276	0.49714	0.9405303
4	0.0052206	0.0254079	8.27	94.18	0.74068	0.59026	0.9074394
5	0.0069025	0.0336450	12.29	232.29	0.36646	0.21879	0.8760991
6	0.0042637	0.0405869	30.37	1182.87	0.37341	0.05208	0.9271396
7	0.0045426	0.0306026	14.25	270.17	0.21736	0.13811	0.9203840
8	0.0040410	0.0273778	11.84	178.04	0.26094	0.20547	0.9266996
9	0.0020361	0.0142049	16.71	389.55	0.25090	0.16666	0.9372303
10	0.0029912	0.0167944	12.13	215.78	0.42226	0.29327	0.9140351
11	0.0029719	0.0126769	6.83	62.00	0.83306	0.67408	0.8941219
12	0.0006888	0.0043308	9.02	106.12	0.70860	0.58794	0.9623548

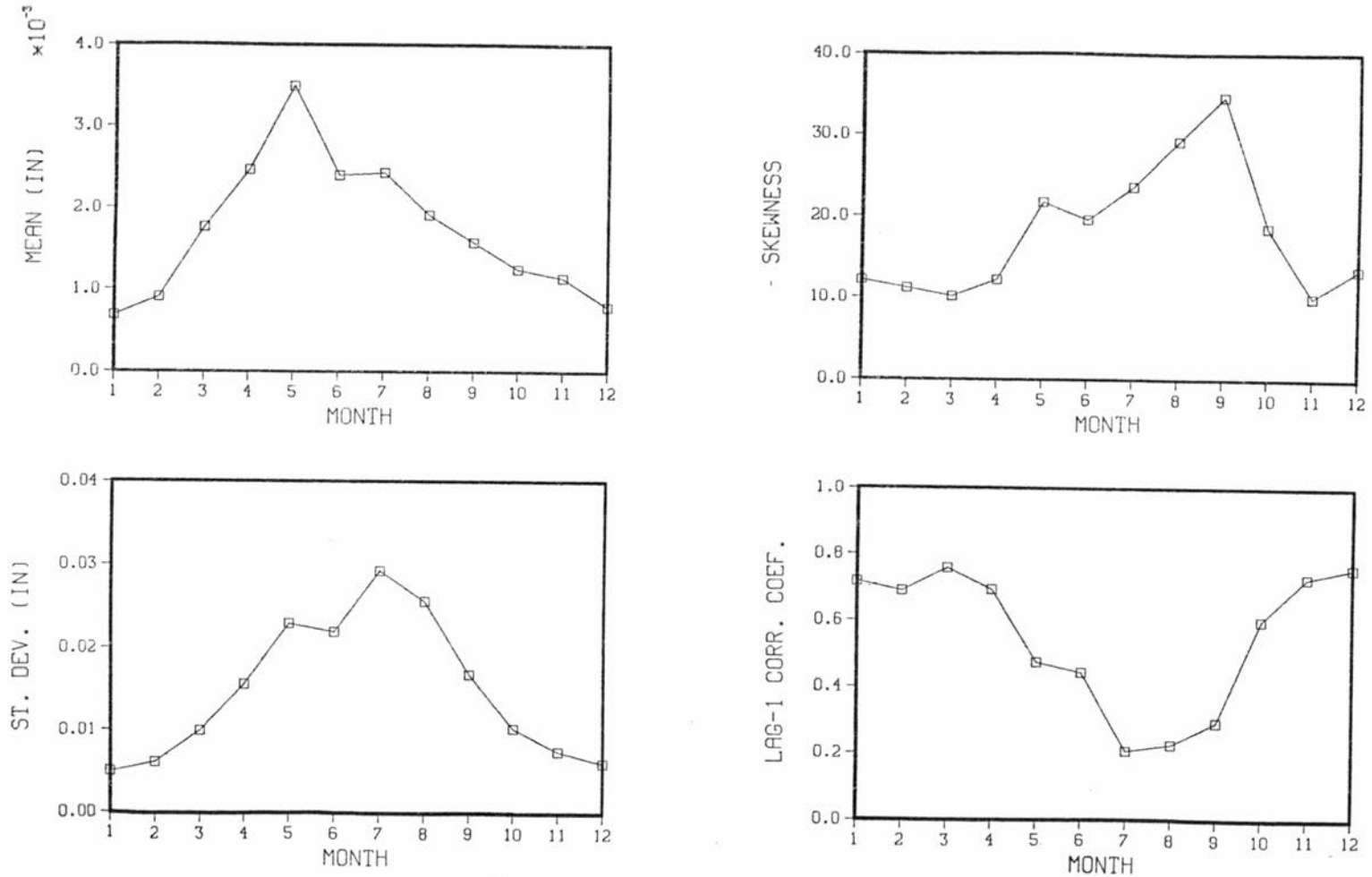


Figure 3.2 Monthly statistics for Denver Wsfo Ap station, for T=60 min..

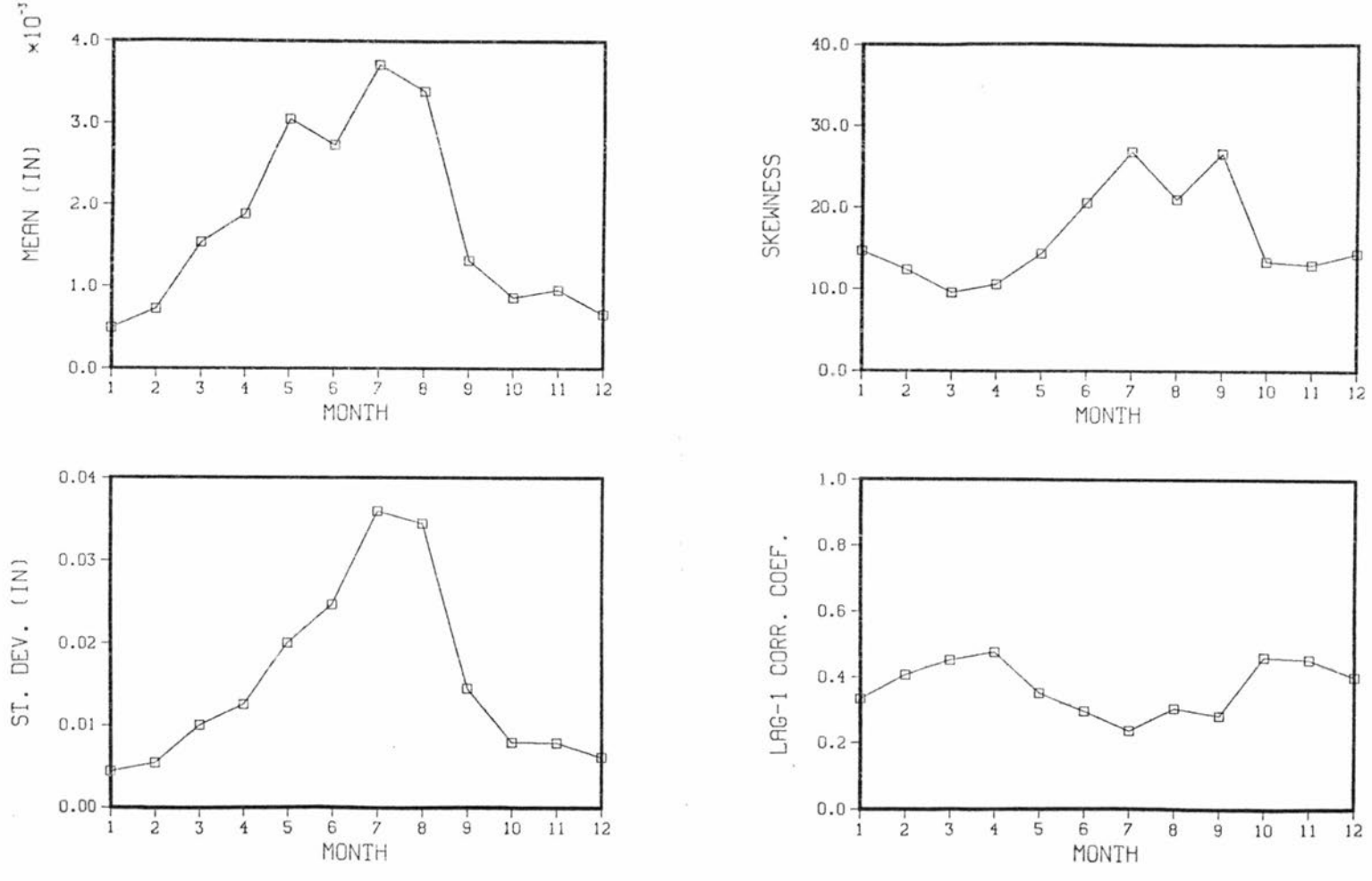


Figure 3.3 Monthly statistics for Greenland 9 SE station, for T=60 min..

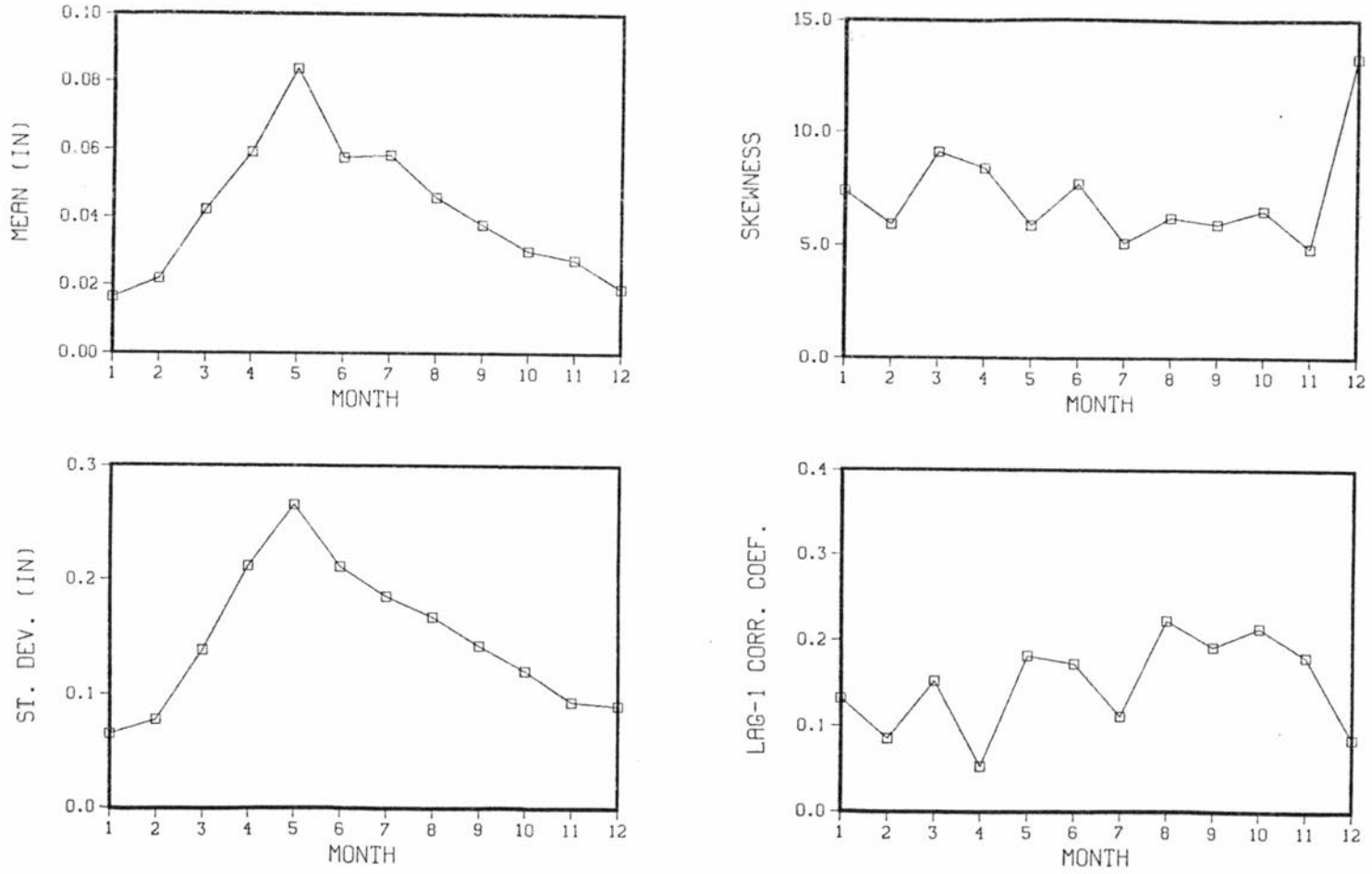


Figure 3.4 Monthly statistics for Denver Wsfo Ap station, for T_a=1440 min..

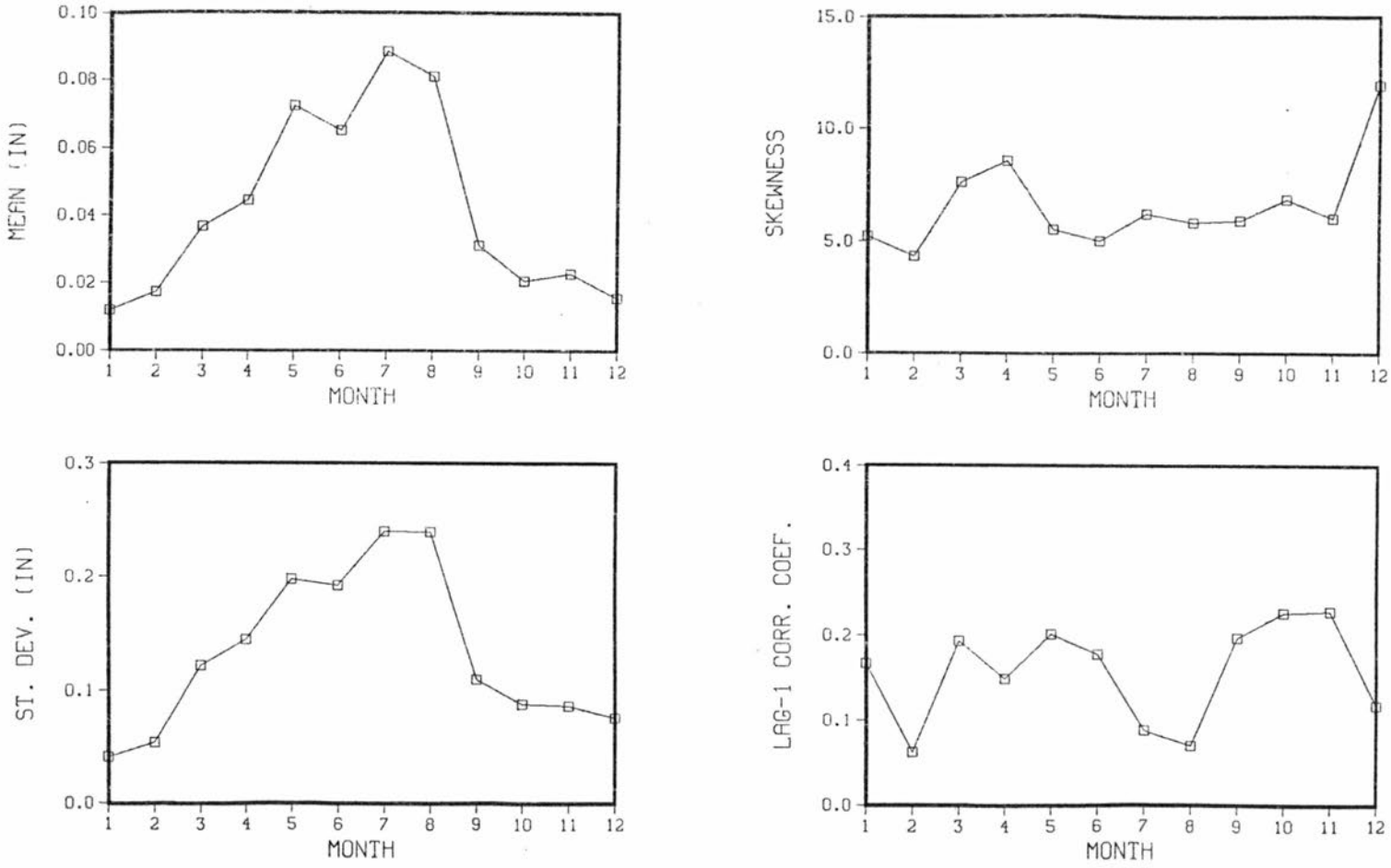


Figure 3.5 Monthly statistics for Greenland 9 SE station, for T_a=1440 min..

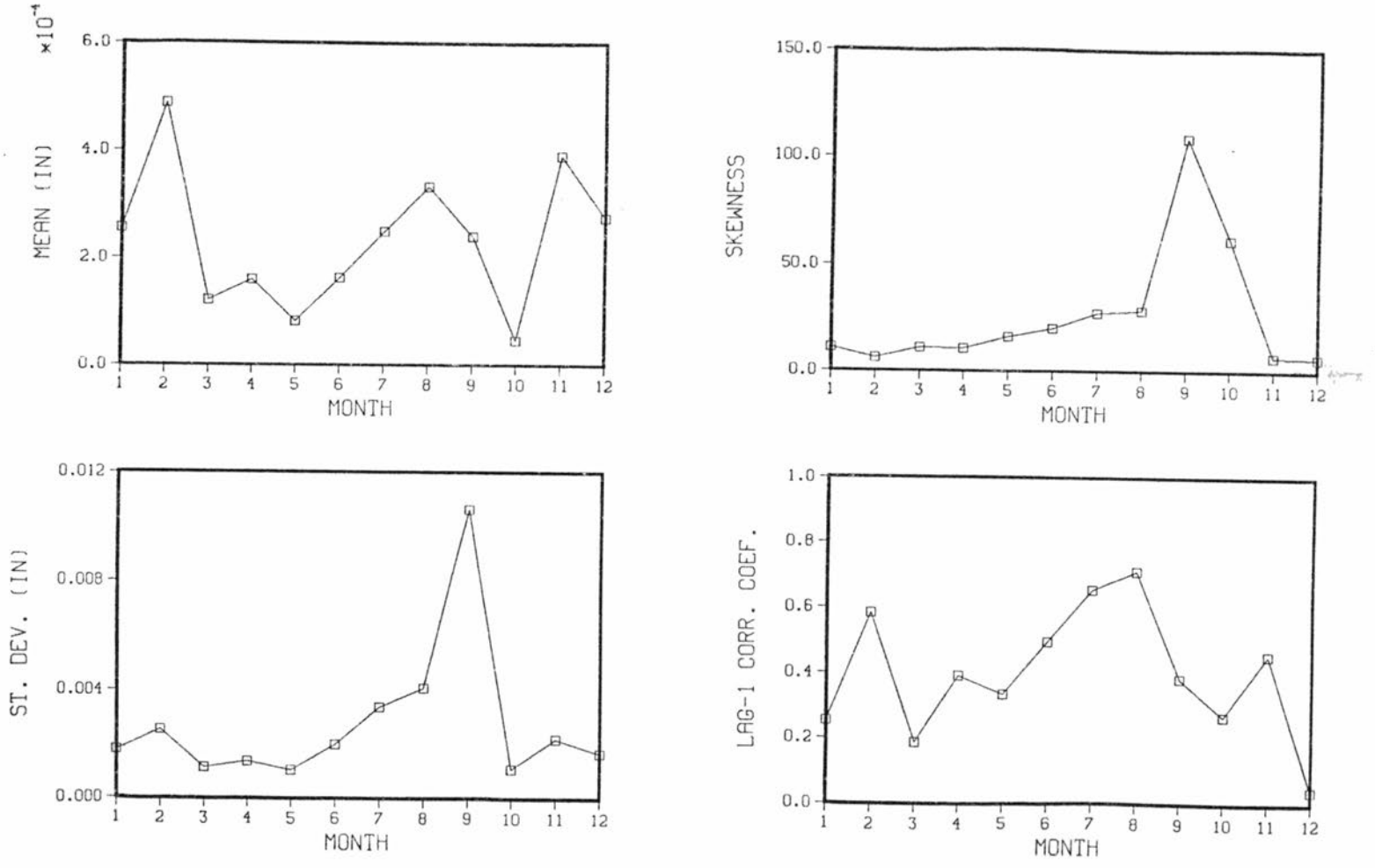


Figure 3.6 Monthly statistics for Idaho Springs station, for T=5 min..

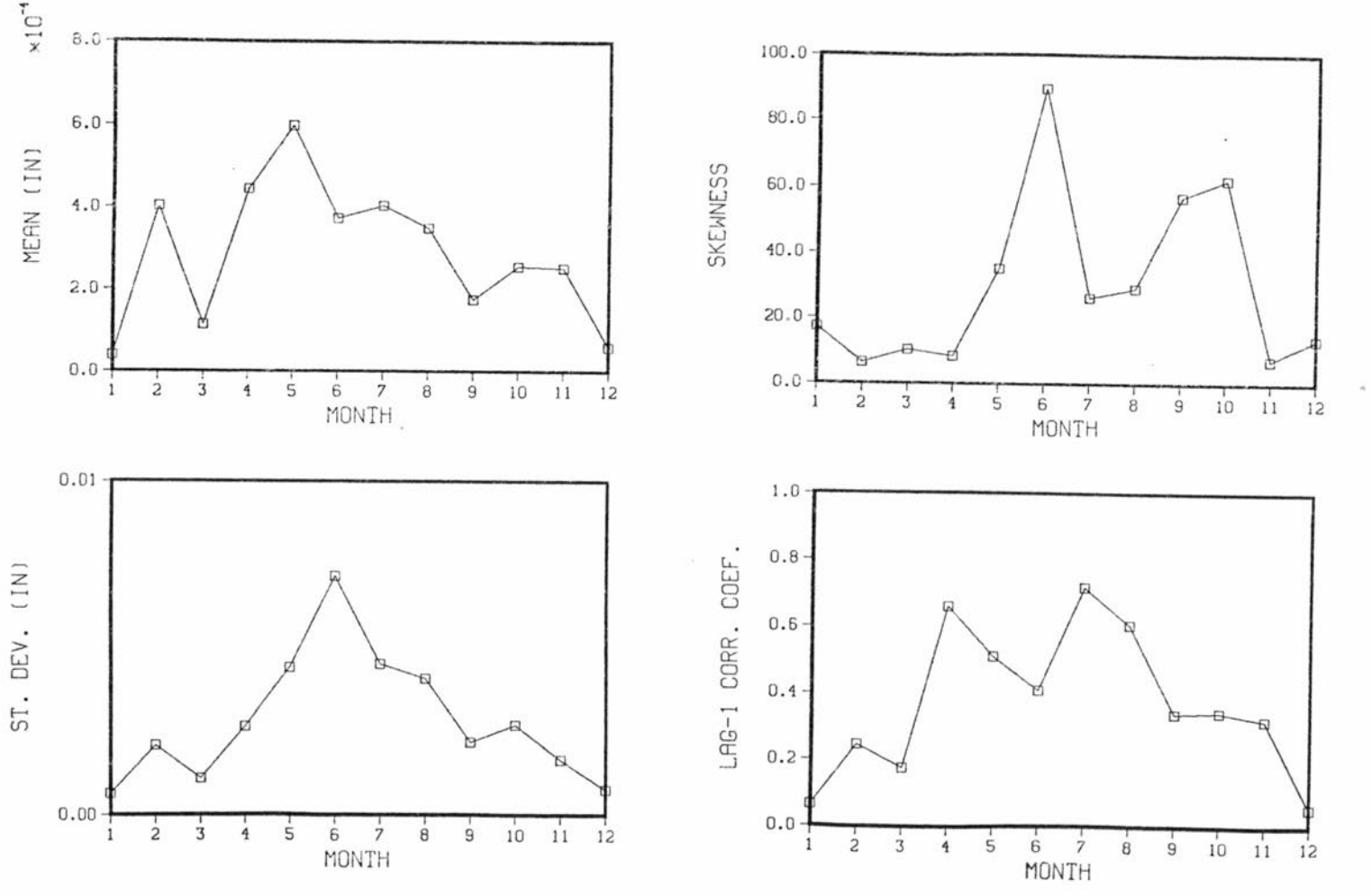


Figure 3.7 Monthly statistics for Ward station, for T=5 min..

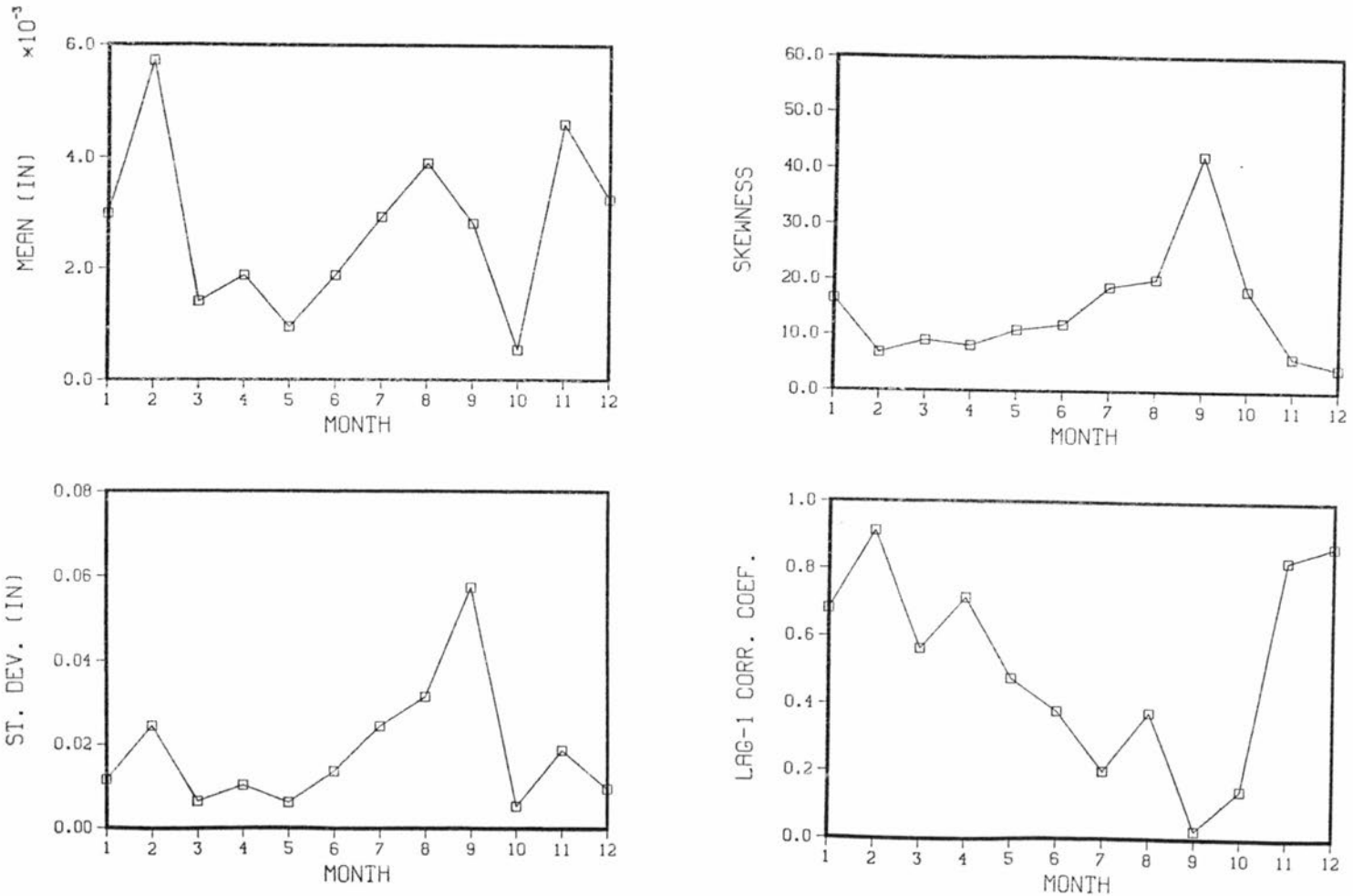


Figure 3.8 Monthly statistics for Idaho Springs station, for T_a = 60 min..

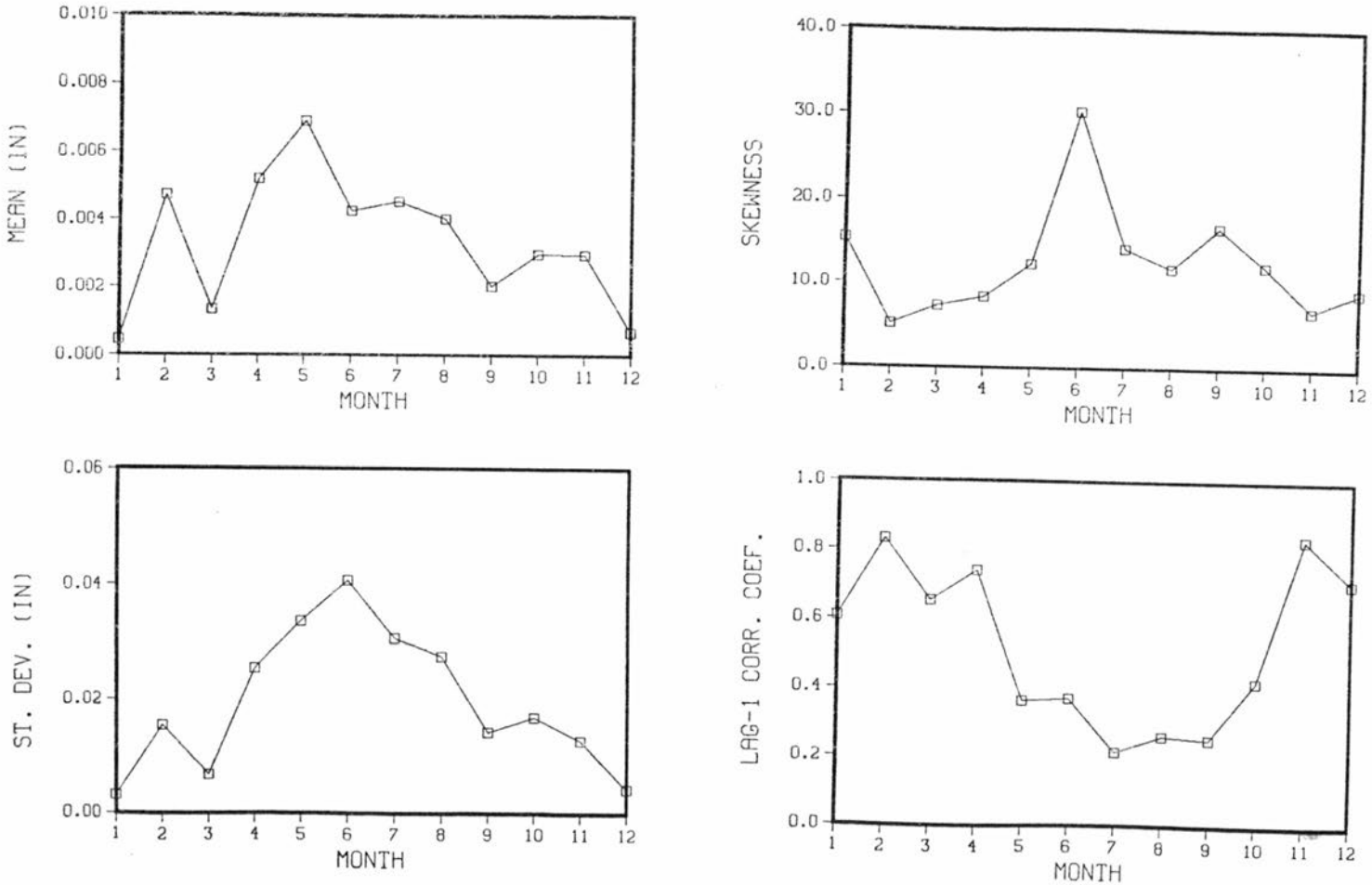


Figure 3.9 Monthly statistics for Ward station, for T_a=60 min..

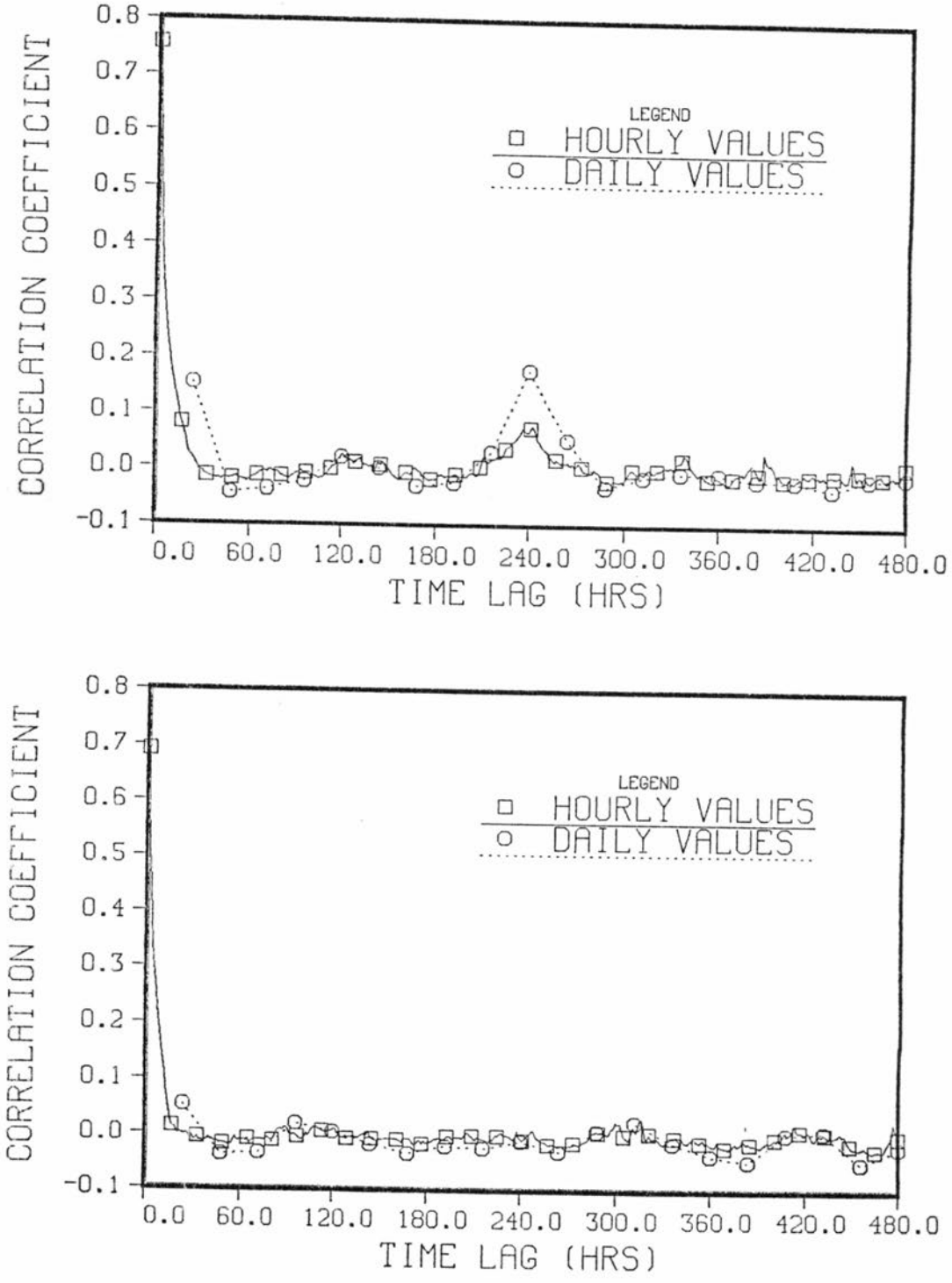


Figure 3.10 Sample correlograms for months 03 and 04, for Denver Wsfo Ap station.

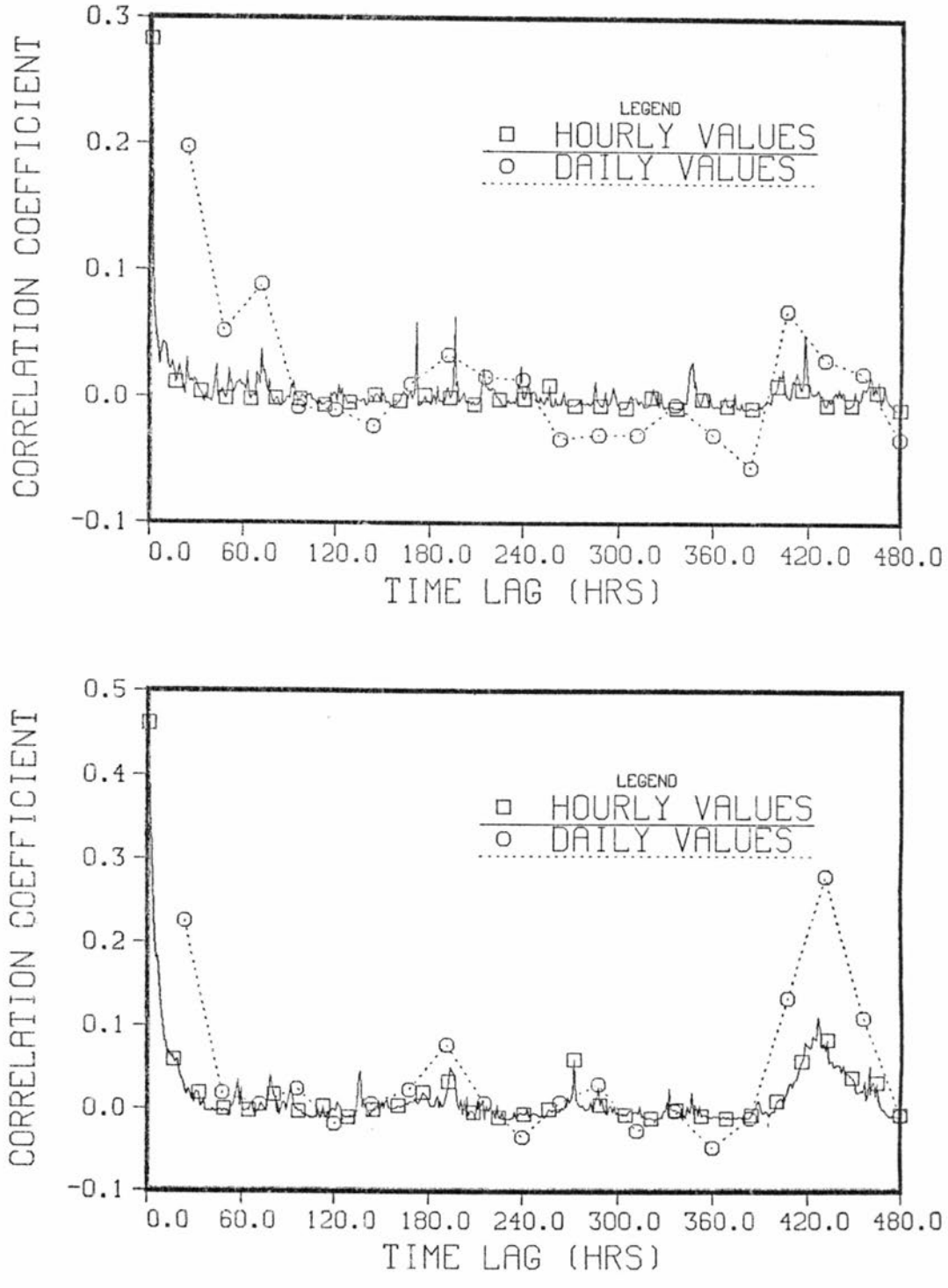
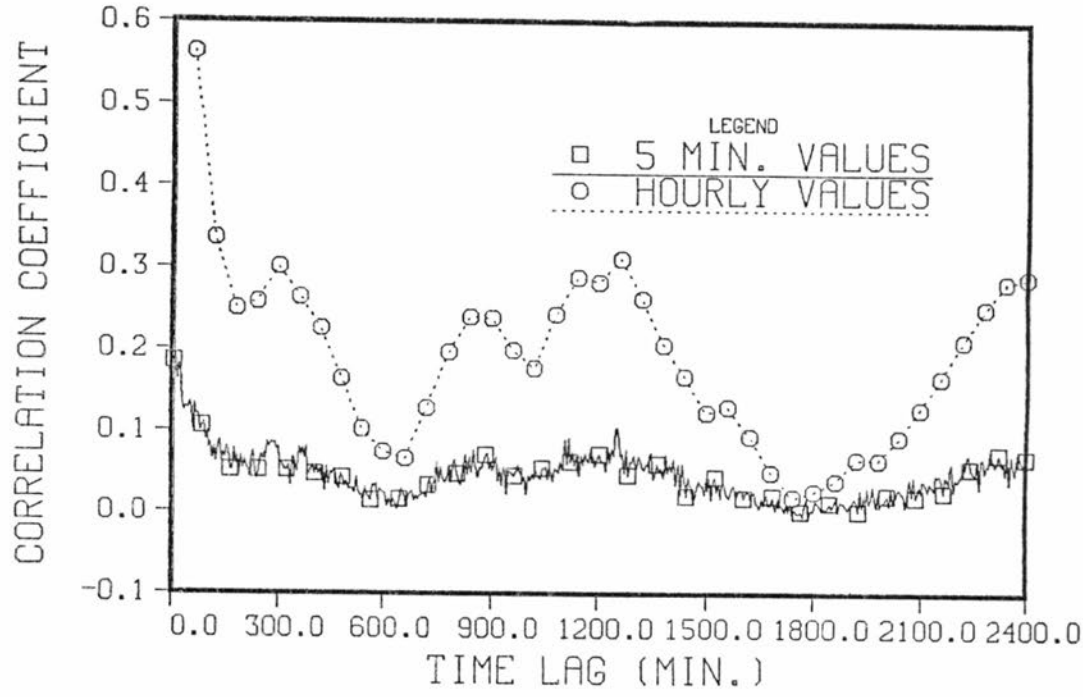


Figure 3.11 Sample correlograms for months 09 and 10, for Greenland 9 SE station.



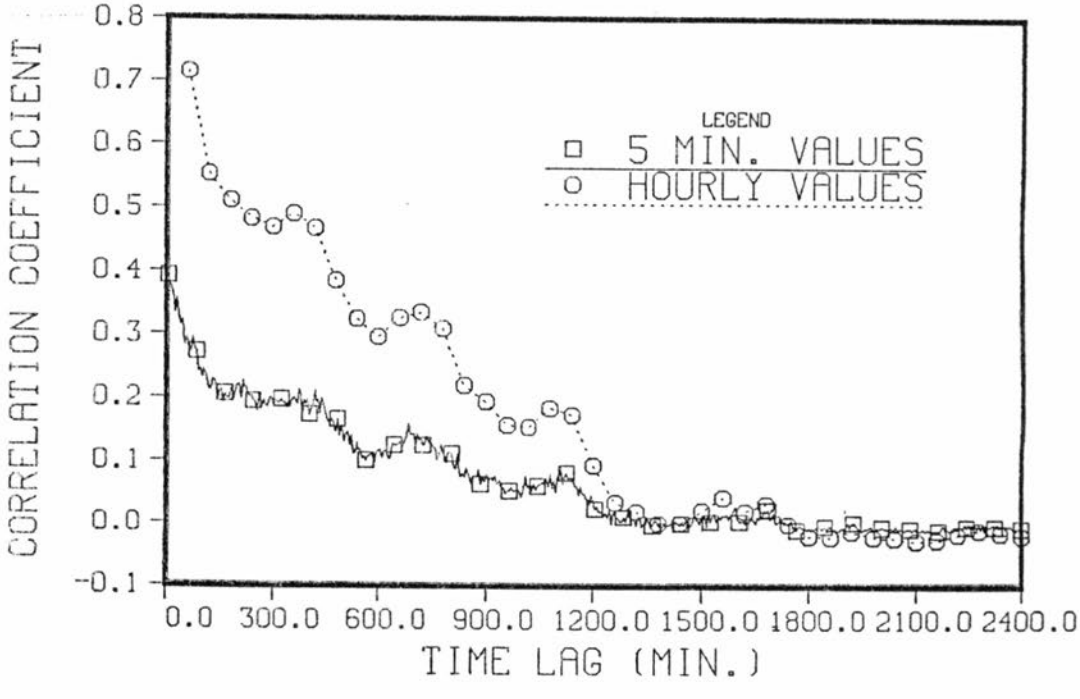


Figure 3.12 Sample correlograms for months 03 and 04, for Idaho Springs station.

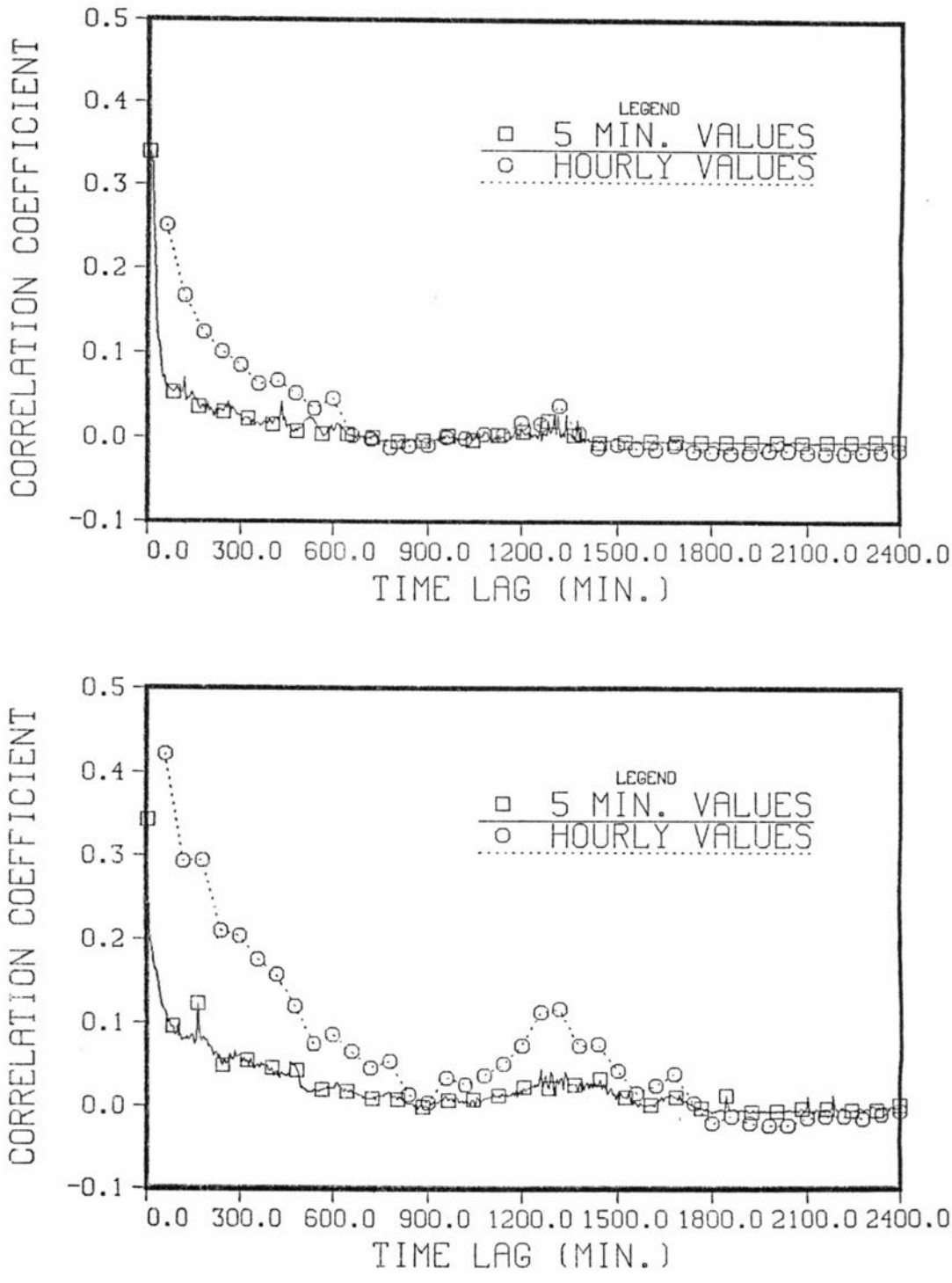


Figure 3.13 Sample correlograms for months 09 and 10, for Ward station.

express linear dependence between consecutive seasonal values and not between values within the season, as is the case for smaller aggregation scales.

The information presented in tables and graphs confirms the periodic behavior for the adopted monthly partition and for the analyzed precipitation recording stations, although this periodicity is stronger at the measurement level and tends to disappear on certain statistics at larger aggregation scales, like for skewness and correlation coefficients going from hourly to daily scales. In most of the cases, summer or transition summer months exhibit smaller correlation than other months, presenting also larger amounts of precipitation and larger variability. This behavior is expected for summer convective storms in Colorado. A large departure from this behavior is observed for Idaho Springs and Ward stations, for T= 5 min., for which correlation tends to increase for summer months. Also observe how in some cases estimated lag-1 correlation coefficient is larger for T=5 min. than for T_a=60 min.. This point will be discussed in a larger extent in the next section. Finally, note how larger sample variability for computed statistics is suggested in samples for 5 min. and for the corresponding derived hourly samples.

For correlograms presented in Figures 3.10 to 3.13 a common characteristic is highlighted. All of them tend to be of the exponential type.

3.6 Effect of the aggregation operation on sample statistics

In the previous section, the question of the behavior of sample statistics with changes in the aggregation level was posed. The analysis presented try to answer this question, both at the estimation

and population level. Derivations for the results given in this section are lengthy and therefore a detailed presentation is not possible.

Assume that the sample series at the measurement scale is denoted as $y_{i,k}$, i=1,..,m, k=1,..,n, where n is the number of intervals of length T in the aggregation period and m is the number of aggregation periods. The aggregated series x_i, i=1,..,m, is obtained as

$$x_i = \sum_{k=1}^{n} y_{i,k} \quad , \quad i=1,\ldots,m \tag{3.22}$$

The subscript for the season has been eliminated, indicating that values are observed sequentially for the indices i and k, i.e., $y_{i+1,1}$ was taken after $y_{i,n}$. In the following presentation no missing data are considered in the series and biased estimators are used in computing statistics.

Statistics under consideration are means, variances and correlation functions. The general question is how these statistics are related for two different aggregation levels.

Using moment estimators, mean and variance for the original series $\bar{Y}$ and S_Y^2, are given by:

$$\bar{Y} = \frac{1}{mn} \sum_{i=1}^{m} \sum_{k=1}^{n} y_{i,k} \tag{3.23}$$

$$S_Y^2 = \frac{1}{mn} \left(\sum_{i=1}^{m} \sum_{k=1}^{n} y_{i,k}^2 - mn\, \bar{Y}^2 \right) \tag{3.24}$$

To obtain the covariance of lag j, it is necessary to consider if j is a multiple or not of the aggregation period length n. With this in mind, let:

$$e = \text{int}(j/n) \tag{3.25}$$

$$d = j/n - e \tag{3.26}$$

In words, e represents the closest integer less than or equal to j/n and d is the fractional part of j/n. Based on the values taken by e and d, the covariance of lag j for the original series, denoted by $\hat{\gamma}_Y(j)$, is obtained using one of the following expressions

$$\hat{\gamma}_Y(j) = \frac{1}{mn-j} \sum_{i=1}^{m-e} \sum_{k=1}^{n} \left(y_{i,k} - \bar{Y} \right) \left(y_{i+e,k} - \bar{Y} \right)$$

for $d=0$ and $e \geq 0$ (3.27)

$$\hat{\gamma}_Y(j) = \frac{1}{mn-j} \left[\sum_{i=1}^{m-e} \sum_{k=1}^{(e+1)n-j} \left(y_{i,k} - \bar{Y} \right) \left(y_{i+e,k+j-en} - \bar{Y} \right) + \sum_{i=1}^{m-e-1} \sum_{k=(e+1)n-j+1}^{n} \left(y_{i,k} - \bar{Y} \right) \left(y_{i+e+1,k-(e+1)n-j} - \bar{Y} \right) \right]$$

for $d \neq 0$ and $e \geq 0$ (3.28)

From eq. (3.27), S_Y^2 is obtained after making $e=0$.

Using the transformation in (3.22), the mean for the aggregated series, $\bar{X}$, is:

$$\bar{X} = n \bar{Y} \tag{3.29}$$

In order to obtain the variance, S_X^2, one proceeds from the basic estimator

$$S_X^2 = \frac{1}{m} \sum_{i=1}^{m} \left(x_i - \bar{X} \right)^2 = \frac{1}{m} \sum_{i=1}^{m} \left(\sum_{k=1}^{n} y_{i,k} - n \bar{Y} \right)^2$$

and the final result for S_X^2 is:

$$S_X^2 = n S_Y^2 + \frac{2}{m} \left[\sum_{k=1}^{n-1} (mn-k) \hat{\gamma}_Y(k) - \sum_{k=1}^{n-1} \sum_{i=1}^{m-1} \sum_{\ell=n+1-k}^{n} \left(y_{i+1,\ell-n+k} - \bar{Y} \right) \left(y_{i,\ell} - \bar{Y} \right) \right] \tag{3.30}$$

If the correlation coefficient of lag j for the original series, $r_Y(j)$, is estimated using

$$r_Y(j) = \hat{\gamma}_Y(j)/S_Y^2 \tag{3.31}$$

and the standardized original series, $z_{i,k}$, is defined as

$$z_{i,k} = (y_{i,k} - \bar{Y})/S_X \tag{3.32}$$

then S_X^2 can be written in terms of the sample autocorrelation function for the original series in the following way

$$S_X^2 = S_Y^2\left\{n + \frac{2}{m}\left[\sum_{k=1}^{n-1}(mn-k)r_Y(k) - \sum_{k=1}^{n-1}\sum_{i=1}^{m-1}\sum_{\ell=n+1-k}^{n} z_{i+1,\ell-n+k}\, z_{i,\ell}\right]\right\} \tag{3.33}$$

The estimated covariance of lag δ for the aggregated series, $\hat{\gamma}_X(\delta)$, is defined as

$$\hat{\gamma}_X(\delta) = \frac{1}{m-\delta}\sum_{i=1}^{m-\delta}\left(x_i - \bar{X}\right)\left(x_{i+\delta} - \bar{X}\right)$$

or after using eq. (3.22)

$$\hat{\gamma}_X(\delta) = \frac{1}{m-\delta}\sum_{i=1}^{m-\delta}\left[\sum_{k=1}^{n}\left(y_{i,k} - \bar{Y}\right)\sum_{k=1}^{n}\left(y_{i+\delta,k} - \bar{Y}\right)\right]$$

Letting

$$\theta = n\,\delta \tag{3.34}$$

$$\omega_1 = n(\delta-1)+1 \tag{3.35}$$

$$\omega_2 = n(\delta+1)-1 \tag{3.36}$$

the former equation becomes

$$\hat{\gamma}_X(\delta) = \frac{1}{m-\delta}\left[\sum_{i=\omega_1}^{\theta-1}\sum_{j=1}^{m-\delta}\sum_{k=1}^{i-\omega_1+1}\left(y_{j,k+n-i+\omega_1-1} - \bar{Y}\right)\left(y_{j+\delta,k} - \bar{Y}\right) + (mn-\theta)\,\hat{\gamma}_Y(\theta) + \sum_{i=\theta+1}^{\omega_2}\sum_{j=1}^{m-\delta}\sum_{k=1}^{n-i+\theta}\left(y_{j,k-1} - \bar{Y}\right)\left(y_{j+\delta,k+i-\theta} - \bar{Y}\right)\right]$$

In the above equation, θ is a lag for the covariance in the original series. The remaining terms stand for cross products, for lags i, $i=\omega_1,..,\theta-1$ and $i=\theta+1,..,\omega_2$.

After additional covariance terms for the original series are introduced, the final result for the covariance of lag δ for the aggregated series is

$$\hat{\gamma}_X(\delta) = \frac{1}{m-\delta}\left[\sum_{i=n\delta-n+1}^{n\delta+n-1}(mn-i)\,\hat{\gamma}_Y(i) - \sum_{i=n\delta-n+1}^{n\delta-1}\sum_{j=1}^{m-\delta+1}\sum_{k=1}^{n\delta-i}\left(y_{j,k}-\bar{Y}\right)\left(y_{j+\delta-1,k+i-n\delta+n}-\bar{Y}\right) - \sum_{i=n\delta+1}^{n\delta+n-1}\sum_{j=1}^{m-\delta-1}\sum_{k=n\delta+n-i+1}^{n}\left(y_{j,k}-\bar{Y}\right)\left(y_{j+\delta+1,k+i-n\delta-n}-\bar{Y}\right)\right] \quad (3.37)$$

Equation (3.37) transform into (3.30) for $\delta=0$.

Using eq. (3.32) and defining the sample autocorrelation function for the aggregated series, $r_X(\delta)$, in the same way as in (3.31), eq. (3.37) is modified and presented as

$$r_X(\delta) = \frac{1}{m-\delta}\,\frac{S_X^2}{S_Y^2}\left[\sum_{i=n\delta-n+1}^{n\delta+n-1}(mn-i)\,r_Y(i) - \sum_{i=n\delta-n+1}^{n\delta-1}\sum_{j=1}^{m-\delta+1}\sum_{k=1}^{n\delta-i} z_{j,k}\,z_{j+\delta-1,k+i-n\delta+n} - \sum_{i=n\delta+1}^{n\delta+n-1}\sum_{j=1}^{m-\delta-1}\sum_{k=n\delta+n-i+1}^{n} z_{j,k}\,z_{j+\delta+1,k+i-n\delta-n}\right] \quad (3.38)$$

In the following paragraphs, similar results will be given for the population case. Assume that $Y_{i,j}$ represents an stationary stochastic process, with j=1,..,n, where n is the length of the

aggregation period, and i is an element of the positive integers. Mean, variance and covariance function will be denoted as μ_Y, σ_Y^2 and $\gamma_Y(.)$, respectively for the original process. The objective here is to derive mean, variance and covariance function for the aggregated process, obtained as

$$X_i = \sum_{j=1}^{n} Y_{i,j} \tag{3.39}$$

The mean μ_X is given by the simple relation

$$\mu_X = n\,\mu_Y \tag{3.40}$$

The variance of the aggregated process σ_X^2 is easy to obtain using expected value theory. An intermediate result is

$$\sigma_X^2 = n\,\sigma_Y^2 + 2 \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} \gamma_Y(j-k)$$

which becomes

$$\sigma_X^2 = n\,\sigma_Y^2 + 2 \sum_{k=1}^{n-1} (n-k)\,\gamma_Y(k) \tag{3.41}$$

or

$$\sigma_X^2 = \sigma_Y^2 \left[n + 2 \sum_{k=1}^{n-1} (n-k)\,\rho_Y(k) \right] \tag{3.42}$$

where $\rho_Y(.)$ is the population autocorrelation function for the original process.

A similar procedure is followed in order to obtain the covariance of lag δ for the aggregated process denoted by $\gamma_X(\delta)$. The first important result is

$$\gamma_X(\delta) = \sum_{k=1}^{n} \sum_{j=1}^{n} \gamma_Y(n\delta+j-k)$$

which, after some manipulation, becomes

$$\gamma_X(\delta) = n\,\gamma_Y(n\delta) + \sum_{j=1}^{n-1} (n-j) \left[\gamma_Y(n\delta+j) + \gamma_Y(n\delta-j) \right] \tag{3.43}$$

In terms of autocorrelation functions, eq. (3.43) becomes

$$\rho_X(\delta) = \frac{\sigma_Y^2}{\sigma_X^2}\left[n\,\rho_Y(n\delta) + \sum_{j=1}^{n-1}(n-j)\left[\rho_Y(n\delta+j) + \rho_Y(n\delta-j)\right]\right] \quad (3.44)$$

No analytical derivations are obtained here for the probability distribution function for the amount of precipitation in a given time interval. However, it is important to mention that the pdf at a larger aggregation scale is the convolution of the pdf at the measurement scale. Also, as the aggregation scale increases, the probability of zero precipitation must decrease.

At this point, it is possible to present some analysis on the behavior of sample statistics at the measurement and aggregation scales. Figures 3.14 to 3.17 show the ratio between some statistics estimated at the aggregation level scale and the same statistics at the measurement level. No plots of this type are provided for monthly and daily aggregation scales, although results were similar.

The mean is the statistic with the best behavior, since in all cases the ratio is very close to the aggregation period length, number of days in the month for monthly and daily aggregation scales, 24 for daily and hourly samples and 12 for hourly and 5 min. samples. This behavior is shown in eqs. (3.29) and (3.40). Small discrepancies are due to the fact that the number of periods with missing data in the aggregated series is not an integer multiple of the number of missing periods in the measured series. For example, as stated before, if five missing hourly values are present in a given day, the value for that day is not declared missing, but made equal to the sum of the remaining 19 values.

In the analysis of standard deviations and correlation functions, eqs. (3.33), (3.38), (3.42) and (3.44) show that the behavior of these

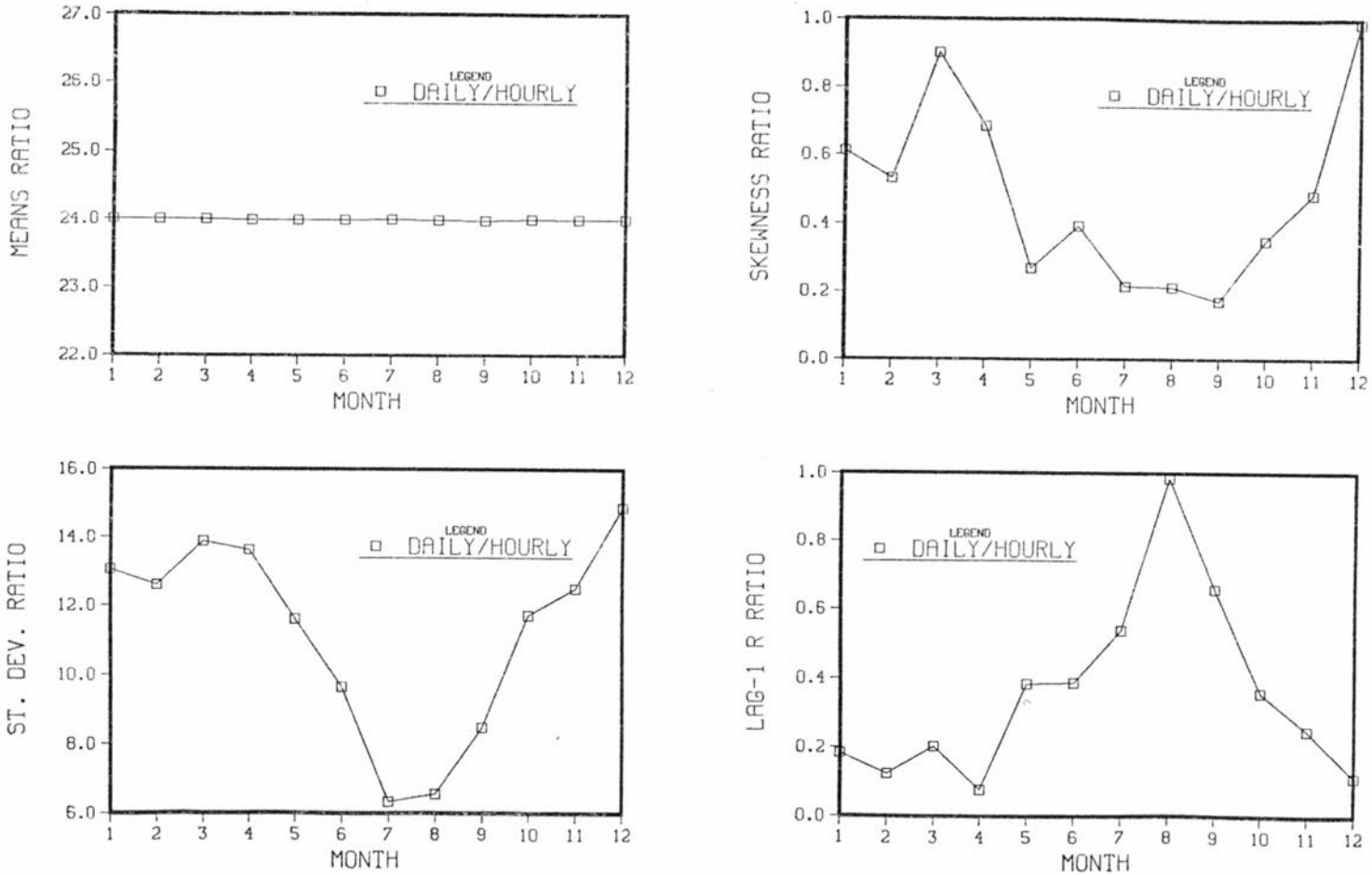


Figure 3.14 Ratio of statistics for two aggregation levels for Denver Wsfo Ap station.

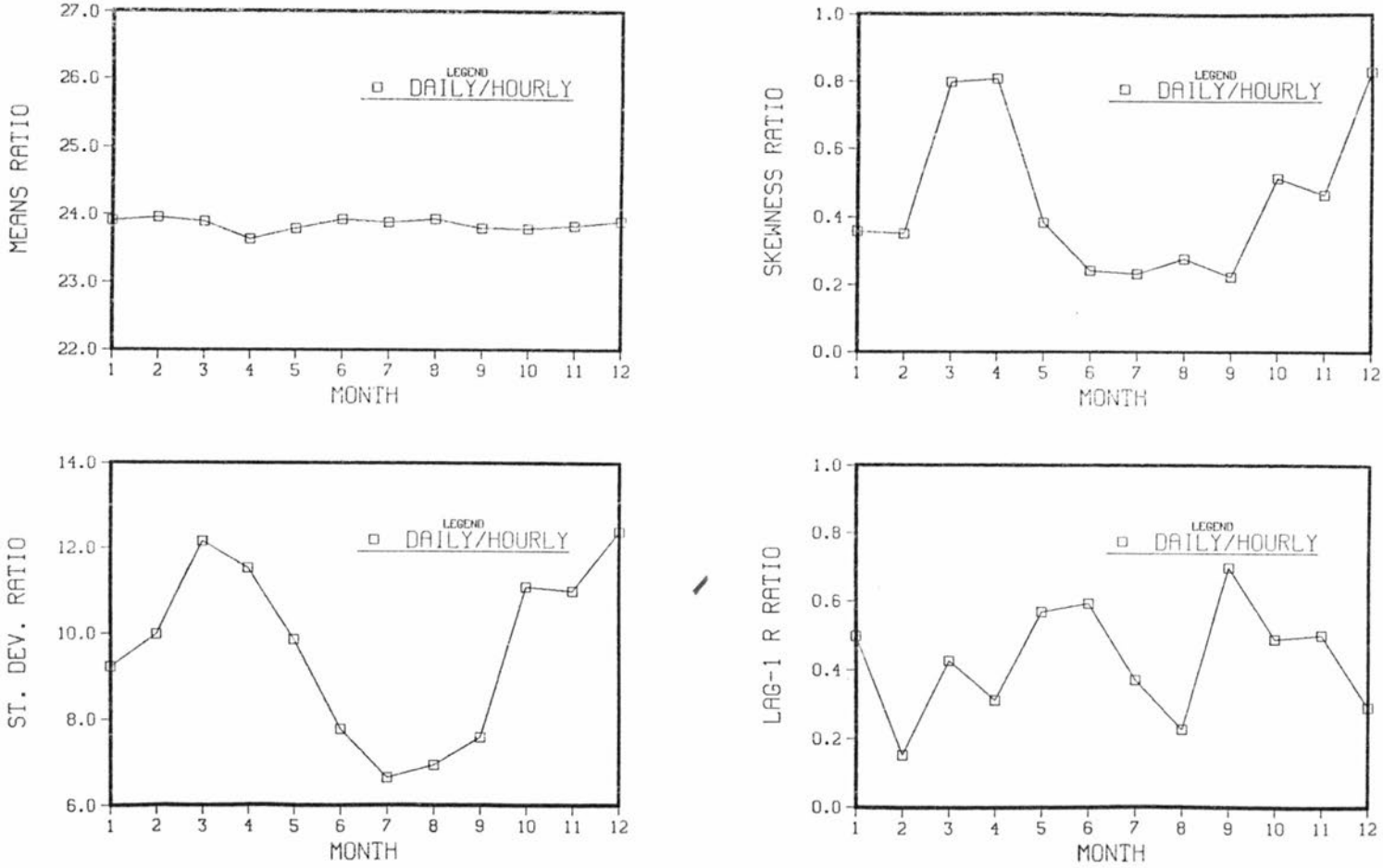


Figure 3.15 Ratio of statistics for two aggregation levels for Greenland 9 SE station.

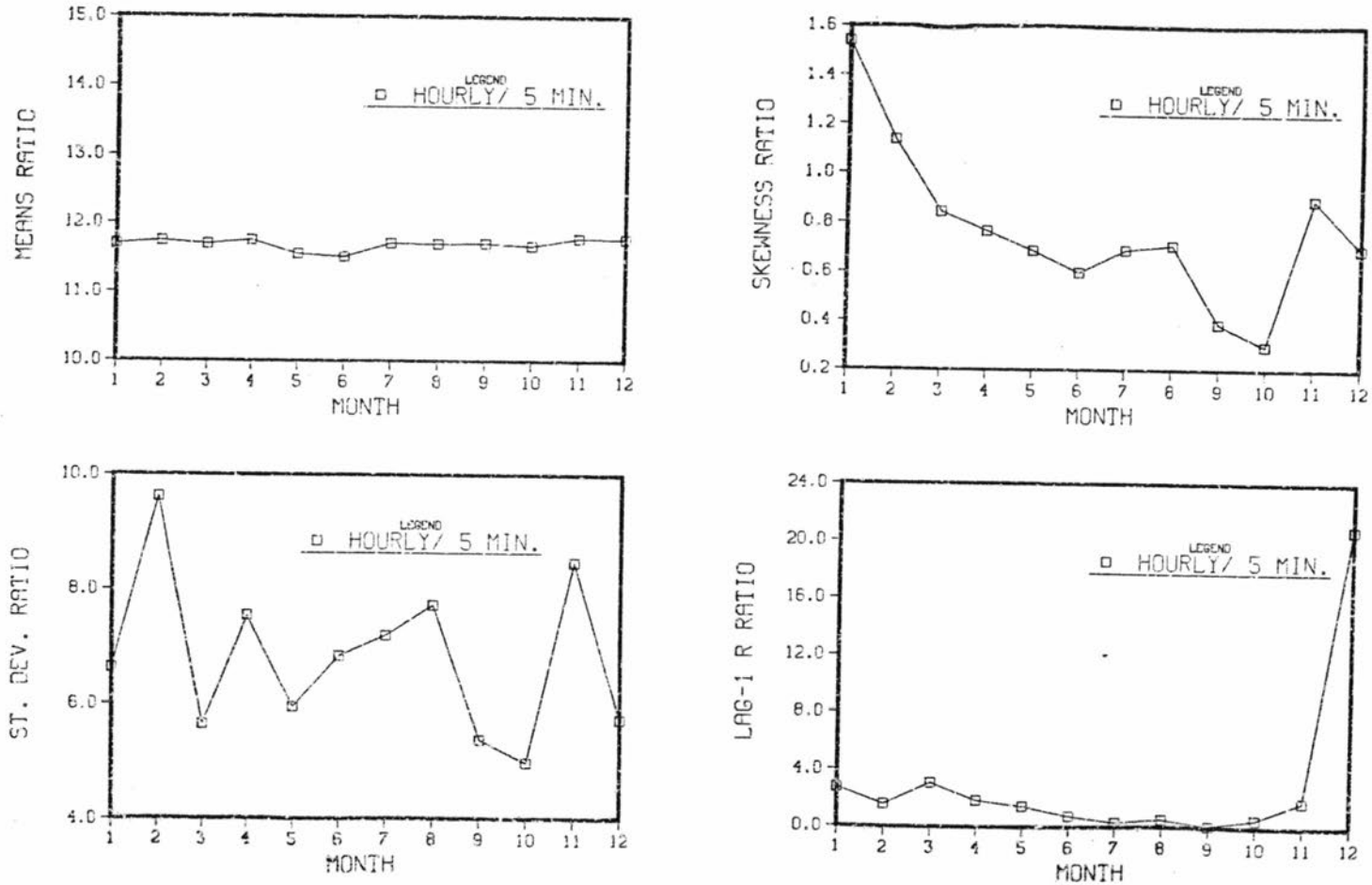


Figure 3.16 Ratio of statistics for two aggregation levels for Idaho Springs station.

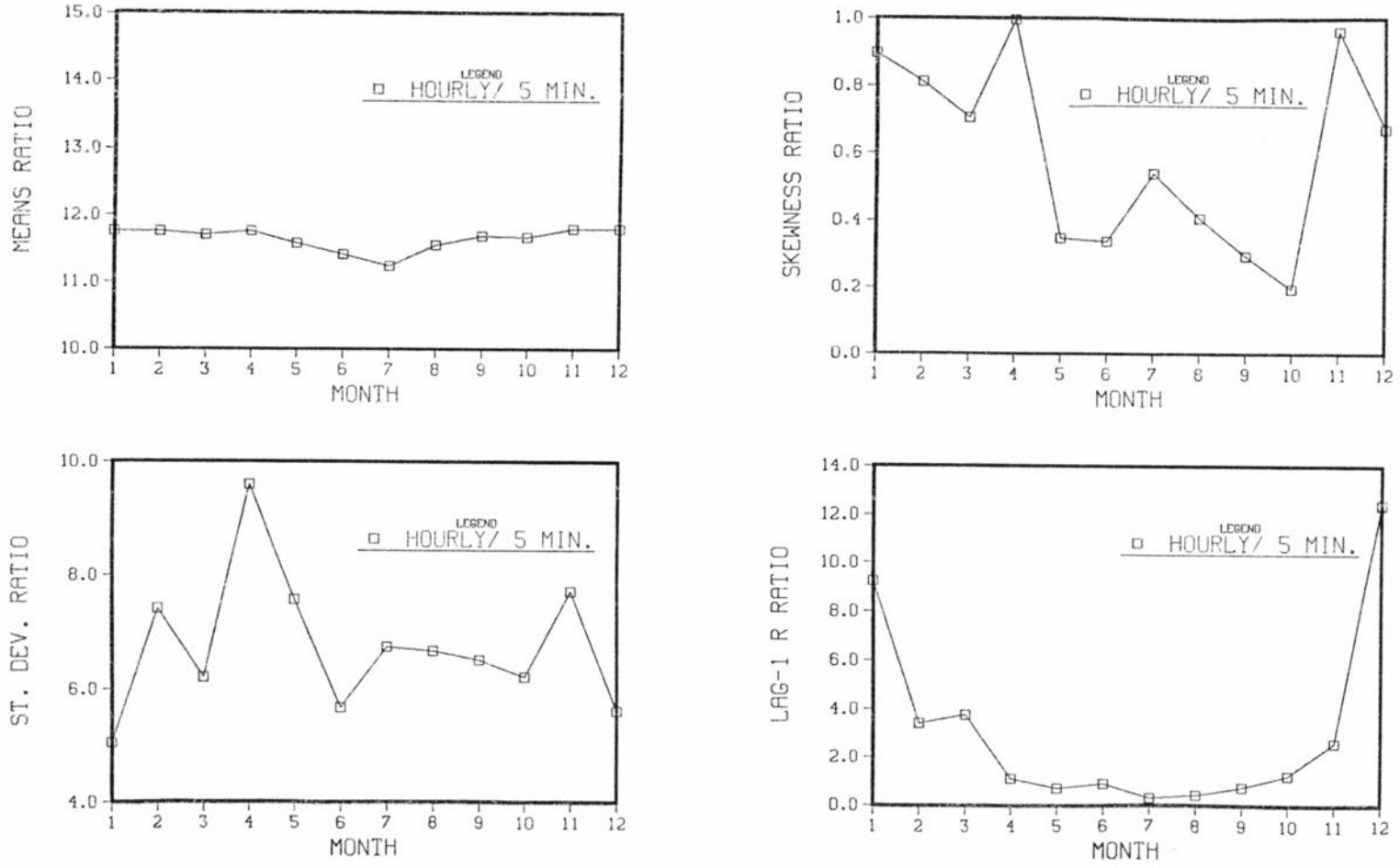


Figure 3.17 Ratio of statistics for two aggregation levels for Ward station.

is highly dependent on the second order properties of the original series or process, as expected, in the form of linear combinations. Furthermore, it is not possible to establish a one to one relationship between the correlation for the aggregated series at a given lag and the same lag in the measured series. From eqs. (3.33) and (3.42), it is most probable to obtain an increase in the variance when the aggregation period increases, given the positive correlation usually present in short sampling interval precipitation data. For all the cases analyzed here, as shown in Figures 3.14 to 3.17, the result was an increase in the standard deviation and consequently in the variance.

The behavior of the sample autocorrelation function is even more difficult to predict than the behavior of the population counterpart, due to the presence of the three summations of cross products in eq. (3.38). Also observe, in eqs. (3.38) and (3.44), the ratio of variances at the two temporal scales. If the ratio is small, as was the case, one should expect a decrease in the population and sample autocorrelation functions when a larger aggregation scale is used. In fact, analysis of empirical data and common sense show that the larger the measurement time interval the less correlated precipitation data should be. Tables 3.2 to 3.5 and Figures 3.10 to 3.17 show that in fact a decrease in correlation was obtained from hourly to daily data, but for some months an increase was obtained from 5 min. to hourly samples. Equations (3.38) and (3.44) do not inhibit this increase. Islam et. al. (1988) report an increase in correlation from the hourly to the daily level, due to diurnal periodicity.

Figures 3.16 and 3.17 show that increase in correlation appears in months where snow precipitation is dominant or at least present.

This type of storm, frontal mass, tends to last longer, with low precipitation intensity, as opposite to summer thunderstorms, with short duration and high intensity. If in the average, storms last longer than 1 hr., an increase in correlation from 5 min. to hourly values is highly probable, since at the hourly level positive values will be paired with positive values. In the summer months case, if the average storm lasts less than 1 hr., at the hourly level positive values will be paired with zero values, and a decrease in correlation will be obtained.

To further illustrate the preceeding point, assume 5 min. records are available with each positive value followed by a zero value and the zero value followed by another positive value, and so on. Using eq (3.27), the lag-1 correlation coefficient for this sample can be shown to be

$$r_Y(1) = -\bar{Y}^2/S_Y^2 \tag{3.45}$$

When an hourly series is derived from the former, all values in the new sample will be positive, and the resulting lag-1 correlation coefficient $r_X(1)$ will be, almost surely, larger than the one given in eq. (3.45).

Although the above hypotheses require further testing, they were accepted as adequate explanations for the present case. This yields the conclusion that correlation in precipitation records is highly controlled by the mean storm duration and by the way values are distributed in time, i.e., the so called precipitation clustering.

Another important issue which affects computation of statistics is the inclusion or not of zero values in the estimation procedure. It was not possible to find strong mathematical or analytical

arguments in favor of one or the other method. If zero values were eliminated systematically from the samples, the distribution of the amount of precipitation would be truncated, generating a shift to the right of the raw moments, starting with the mean. Variance could be larger or smaller. Correlations are more difficult to analyze, since these involve the joint distribution of the amount of precipitation in two disjoint intervals. However, elimination of zero values would decrease appreciably sample sizes available for correlation estimation.

In order to clarify the preceeding point, a simulation for an AR(1) process was performed, where the number of zeros in the data was increased from 0 to 90% in increments of 10%. Zero values were uniformly located within the sample. Results from the simulation show that as the number of zeros increases, mean, variance and autocorrelation functions decrease, while skewness and kurtosis increase.

Derivation of properties for the precipitation models used in this study do not differentiate between zero and positive states of the process. In view of this and the preceeding arguments, it was decided to keep zero precipitation values in the samples, as the most logical way to estimate statistics and model parameters.

A final word on estimated statistics. Skewness, kurtosis and probability of zero values behaved as expected, i.e., the tendency in them was to decrease with an increase in the measurement scale.

An additional point, considered before, is the comparison of correlation functions at two different levels of aggregation, for example T and T_a. In order to solve this, assume the existence of an underlying stochastic process, $\{X(t), -\infty \leq t \leq +\infty\}$, defining evolution

of precipitation intensity with time. The amount of precipitation in the kth interval of length T, Y_k, is

$$Y_k = \int_{(k-1)T}^{kT} X(t)dt \tag{3.46}$$

and the average amount of precipitation in the same interval, Z_k, is

$$Z_k = Y_k/T \tag{3.47}$$

The process X(t) is not observed. Either Y_k or Z_k are recorded. Note that Z_k can be interpreted as an estimate of X(t) located at the time t=kT - T/2.

It can be shown (Parzen, 1964) that the covariance for the cumulative amount of precipitation, $\gamma_Y(.)$, is related to the covariance of the intensity process, $\gamma_X(\tau)$, $-\infty \leq \tau \leq +\infty$, as

$$\gamma_Y(\delta) = \int_0^T \int_{\delta T}^{(\delta+1)T} \gamma_X(s-r)\, ds\, dr \tag{3.48}$$

The covariance for the average process, $\gamma_Z(.)$, is

$$\gamma_Z(\delta) = \gamma_Y(\delta)/T^2 \tag{3.49}$$

Equations (3.48) and (3.49) show that population autocorrelation functions for the aggregated and the averaged process, $\rho_Y(.)$ and $\rho_Z(.)$, are the same,

$$\rho_Z(\delta) = \gamma_Z(\delta)/\gamma_Z(0) = \gamma_Y(\delta)/\gamma_Y(0) \tag{3.50}$$

The argument is as follows. In the samples, $y_{i,k}$ for the measurement scale T, and x_i for the aggregated series T_a, two different correlograms were estimated. If j and δ denote lags, then $r_Y(j)$ plotted against T(j-.5) and $r_X(\delta)$ plotted against $T_a(\delta-.5)$ are both estimates of the continuous correlogram $\rho_X(\tau)$, $-\infty \leq \tau \leq +\infty$, for

the intensity process. This allows comparison of correlograms estimated at two different aggregation levels. In fact, this is the way correlograms presented in Figures 3.10 to 3.13 were plotted.

3.7 Clustering of precipitation data

Clustering for precipitation records could be defined as the process by which precipitation events arrive at a point in space following a non uniform rate in time. Once a storm has arrived, the probability of other storms arriving close to this is higher than if the state were dry at that time.

As argued before, clustering is responsible for the correlation structure present in precipitation records (Waymire and Gupta, 1981a, 1981b, 1981c). Figures 3.18 to 3.21 present examples of precipitation events recorded in the gauging stations used in this study, where the process of clustering is easily observed.

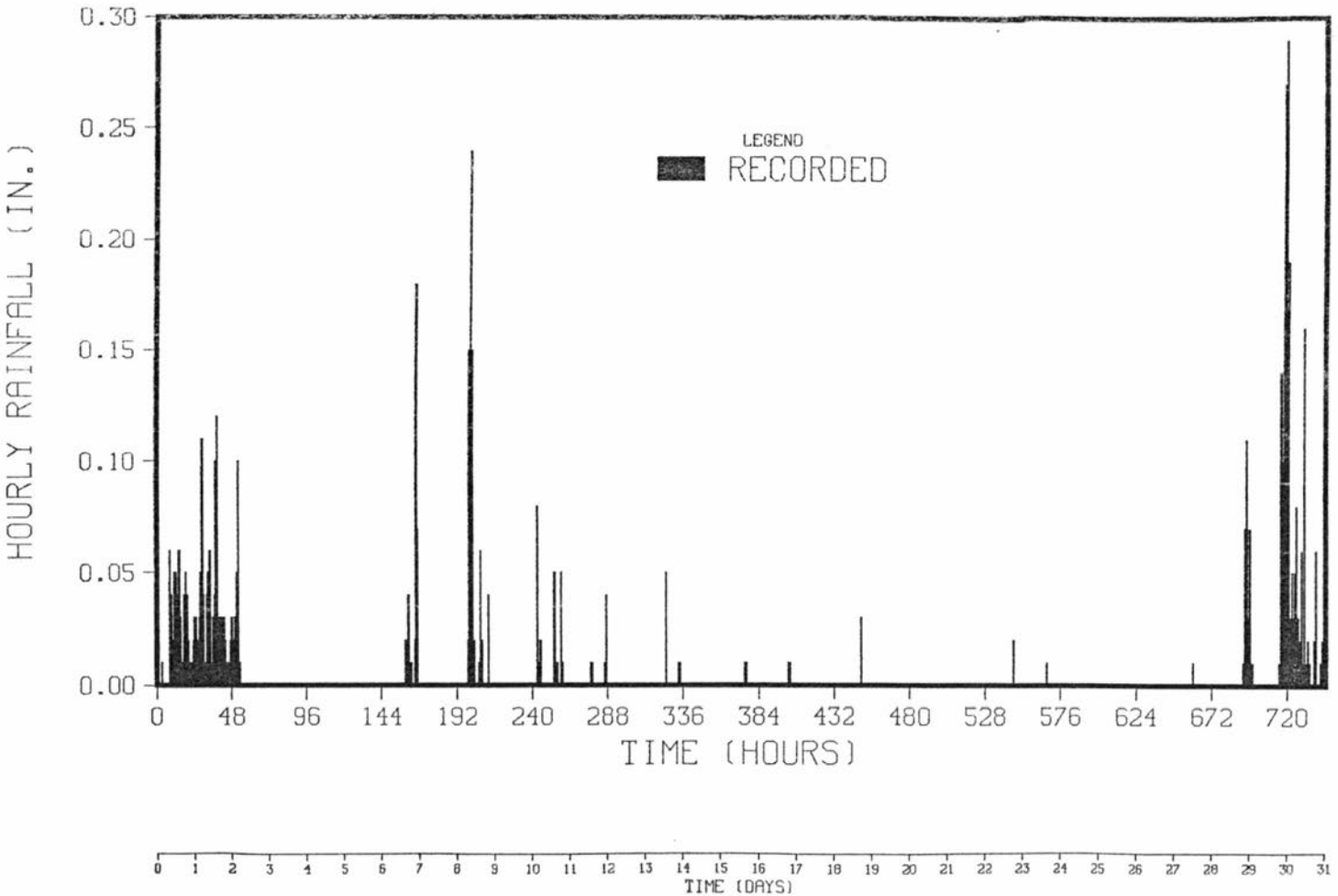


Figure 3.18 Recorded hourly precipitation at Denver Wsfo Ap station, between 1949/05/05/01 and 1949/06/04/24.

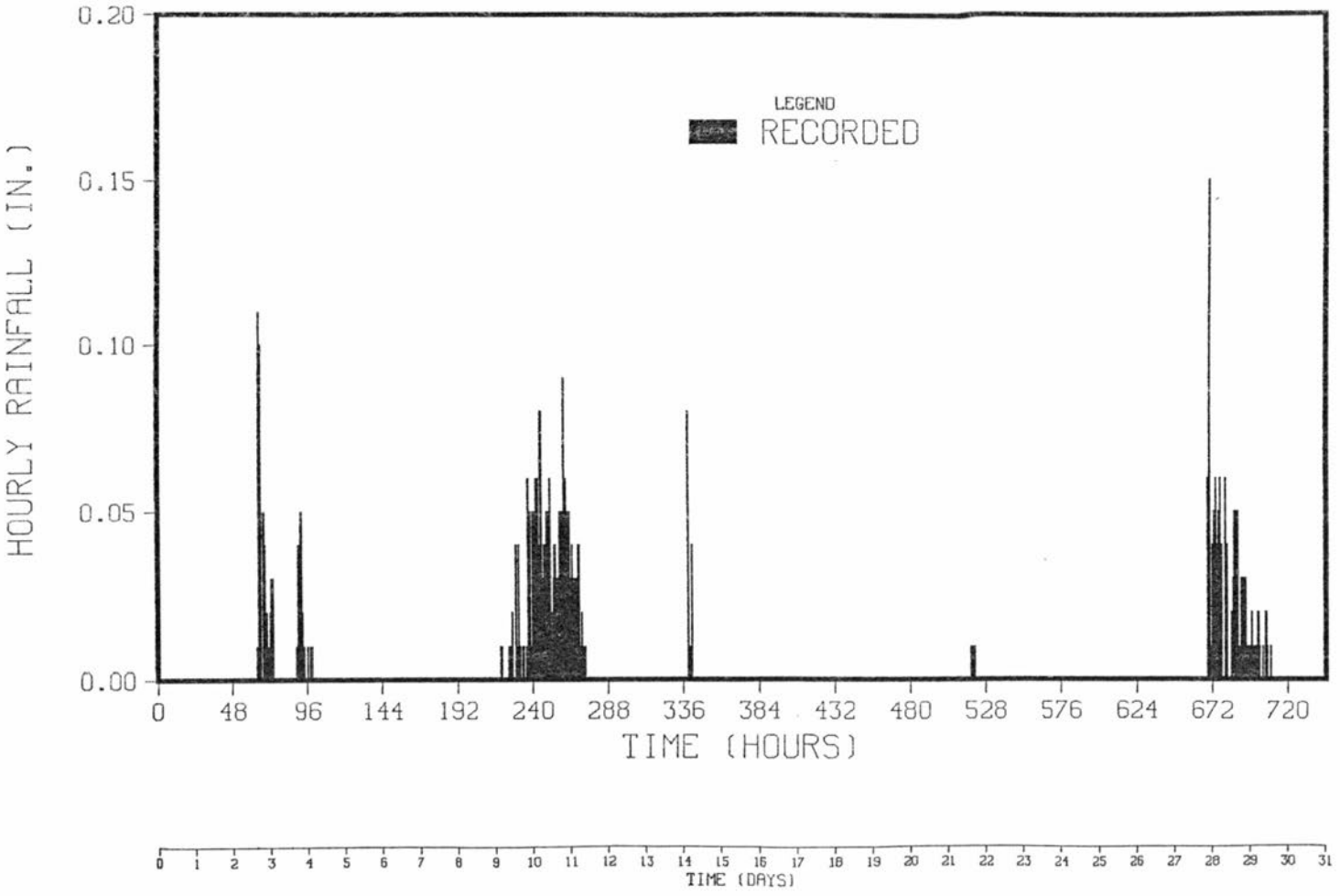


Figure 3.19 Recorded hourly precipitation at Greenland 9 SE station, between 1969/10/01/01 and 1969/10/31/24.

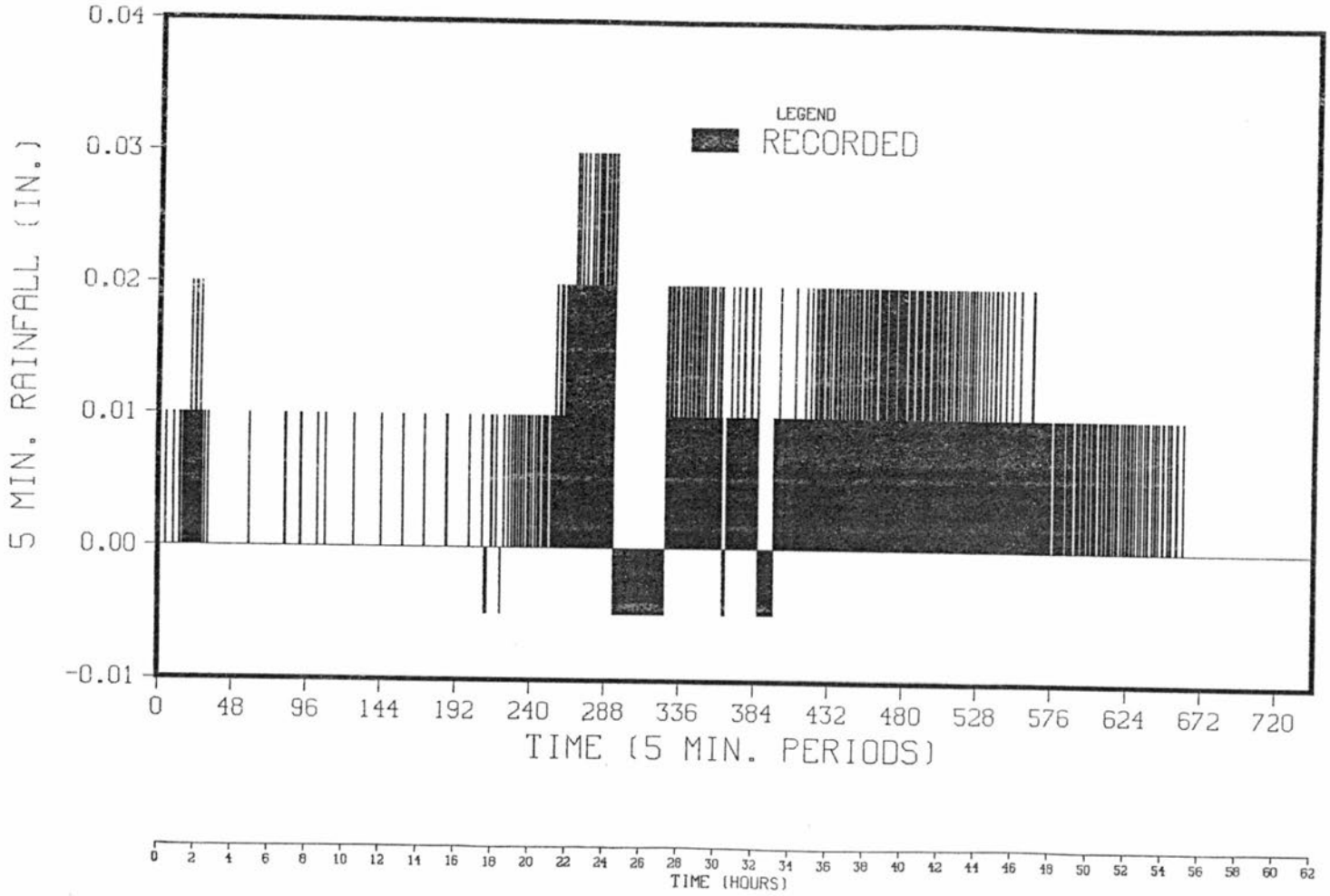


Figure 3.20 Recorded 5 min. precipitation at Idaho Springs station, between 1984/02/12/00 and 1984/02/14/13 (Negative values are missing).

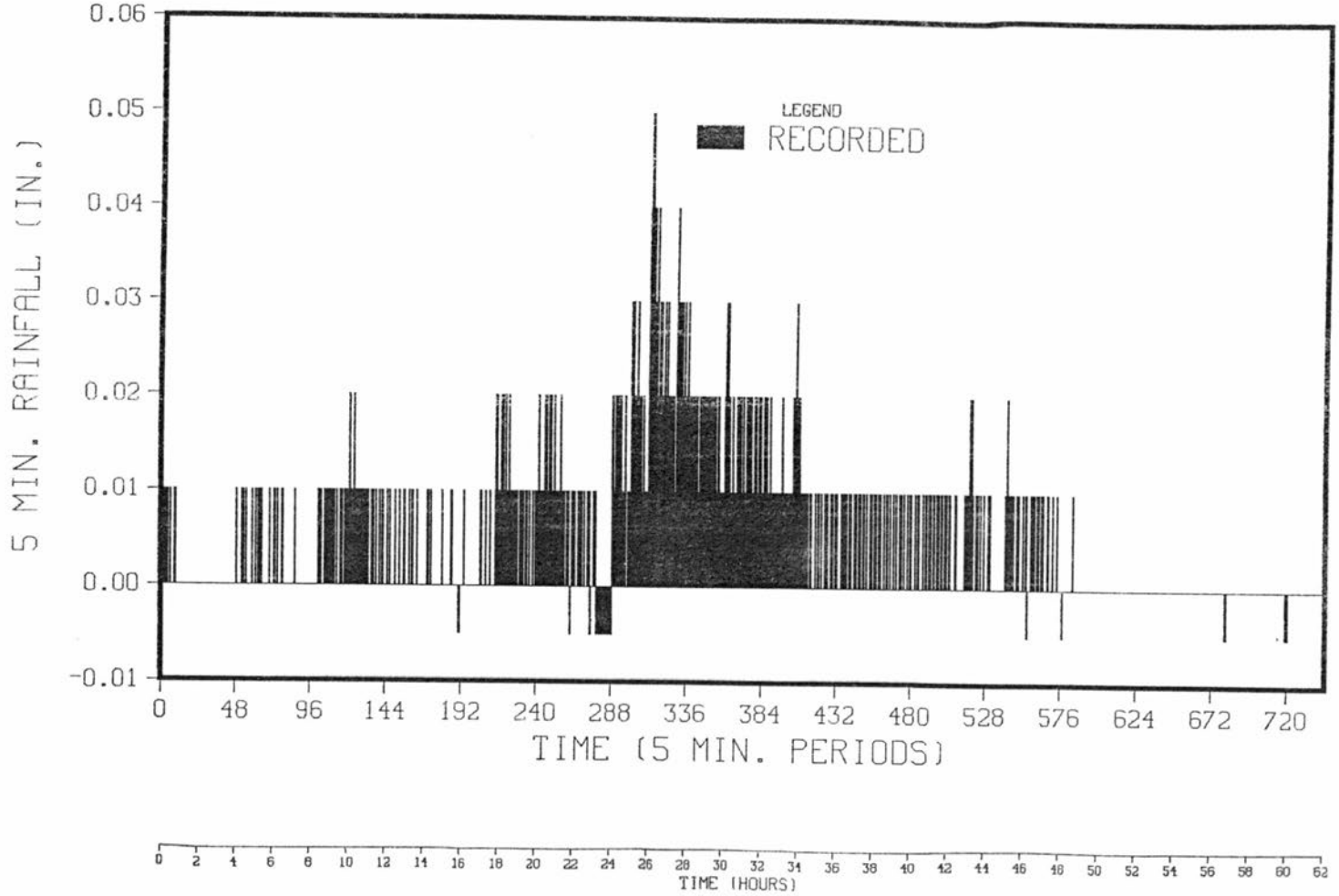


Figure 3.21 Recorded 5 min. precipitation at Ward station, between 1984/04/20/00 and 1984/04/22/13 (Negative values are missing).

Chapter 4

POISSON WHITE NOISE MODEL FOR PRECIPITATION

4.1 Introduction

In this chapter, a brief review of the Poisson White Noise (PWN) model for precipitation is presented. The model is described and derivation of its main properties is outlined. Two particular cases are presented, when the White Noise terms are exponential and gamma distributed, respectively. Method of moments is used to estimate model parameters for the four precipitation recording stations used in the study.

4.2 Model description

Figure 4.1 gives an schematic representation of the PWN model for precipitation. Following other authors (Waymire and Gupta, 1981a; Rodriguez-Iturbe et al., 1984), the process is described as follows: Precipitation events arrive in time according to an homogeneous Poisson process of rate λ, the arrival times being denoted by T_k. A second random variable, U_k, is associated to each arrival, giving instantaneous amount or volume of precipitation. The sequence $\{U_k, k=1,2..\}$ is formed by independent random variables and they are assumed independent of the arrival times. They follow the same distribution denoted by $f_U(u)$. Therefore, in the PWN process for precipitation the arrival component is Poissonian and internal properties of precipitation are modeled using White Noise terms.

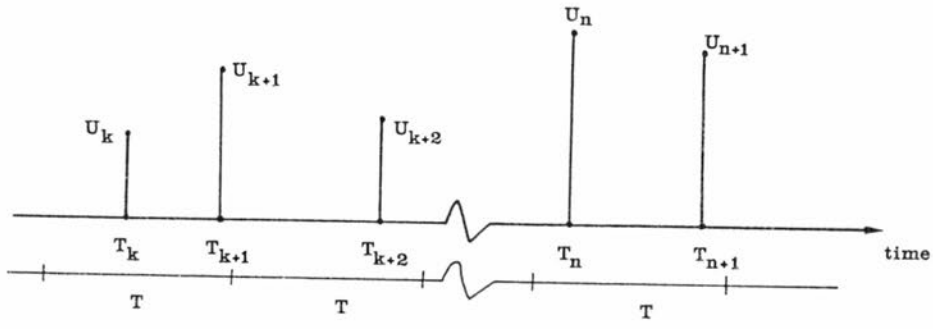


Figure 4.1 Schematic representation of the Poisson White Noise precipitation process.

It must be emphasized that the term White Noise as used here is not limited to the Normal distribution.

The Poisson process is a popular model in the literature on stochastic process, mainly because most of its properties are simply derived, but it has also been shown to be, in some cases, an adequate model for the occurrence of events in space and time. In particular, the model described above as PWN is presented by some authors as Compound Poisson process (Parzen, 1964; Taylor and Karlin, 1984; Waymire and Gupta, 1981a). The simplicity in deriving the properties of a Poisson process stems from the fact that the random variables representing the number of occurrences in disjoint intervals are independent and identically distributed, following a Poisson distribution with intensity λ. Parameters in the model are λ and the parameters indexing the distribution $f_U(u)$.

Let N(0,t) represent the Poisson counting process, i.e., the number of precipitation bursts in the interval from 0 to t. The precipitation intensity at time t, X(t), is defined by means of

$$X(t)\,dt = U(t)\,dN(t),\quad t\geq 0 \tag{4.1}$$

where dN(t) represents the increments of the counting process and is computed as

$$dN(t) = N(0,t+dt) - N(0,t) \tag{4.2}$$

This increment is one if an arrival takes place in dt. Otherwise is zero.

The instantaneous precipitation intensity process is not observed. Cumulative amounts of precipitation over time intervals of length T (measurement scale) are recorded. Therefore, properties of the process must be estimated from samples composed of these amounts.

In order to obtain an expression for the cumulative amount of precipitation in the kth interval of length T, Y_k, one starts with the cumulative amount of precipitation in the interval (0,t), denoted as $N_U(t)$, and given by

$$N_U(t) = \int_0^t X(s)ds \tag{4.3}$$

Using eq. (4.1) and given that the PWN process has no multiplicities (Section 6.2), eq. (4.3) becomes

$$N_U(t) = \sum_{i=0}^{N(0,t)} U_i \tag{4.4}$$

with $U_0=0$. Equation (4.4) is the definition of a Compound Poisson process. From this, Y_k is obtained as:

$$Y_k = N_U(kT) - N_U(kT-T) \tag{4.5}$$

The following expressions are equivalent to (4.5):

$$Y_k = \sum_{i=0}^{N(0,kT)} U_i - \sum_{i=0}^{N(0,kT-T)} U_i\ ,\quad Y_k = \sum_{i=N(0,kT-T)}^{N(0,kT)} U_i,\quad \text{and}$$

Note that an arbitrary Y_k can be expressed as

$$\sum_{i=0}^{N(T)} U_i \tag{4.6}$$

where N(T) now denotes the number of occurrences in an interval of length T and it has a Poisson distribution with parameter λT, i.e.,

$$f_{N(T)}(n) = \exp(-\lambda T)\ (\lambda T)^n / n! \tag{4.7}$$

Following a similar argument in the opposite direction, one can show that Y_k is also given by

$$Y_k = \int_0^T X(t)dt \tag{4.8}$$

Properties for the cumulative amounts of precipitation are obtained from (4.6). An alternate way to proceed is to use eqs. (4.1) and (4.8), after properties for the precipitation intensity process are derived. However, the last approach will not be used here.

4.3 Properties of the PWN precipitation process

Proceeding from eq. (4.6), the mean for Y_k, μ_Y, is computed as

$$\mu_Y = E\left[\sum_{i=0}^{N(T)} U_i \right] = \sum_{n=0}^{\infty} E\left[\sum_{i=0}^{N(T)} U_i \,\middle|\, N(T)=n \right] f_{N(T)}(n)$$

Since the White Noise terms are independent among themselves and independent from the counting process, μ_Y finally becomes

$$\mu_Y = \lambda\ T\ E[U] \tag{4.9}$$

where E[U] is the mean of the White Noise distribution $f_U(u)$.

Using a similar approach for the second raw moment for Y_k, the variance for the amount of precipitation in the interval of length T, σ_Y^2, is shown to be

$$\sigma_Y^2 = \lambda\ T \left(\sigma_U^2 + E^2[U] \right) \tag{4.10}$$

where σ_U^2 is the variance of the White Noise terms

The covariance of lag δ between Y_k and $Y_{k+\delta}$, denoted by $\gamma_Y(\delta)$, is given by

$$\gamma_Y(\delta) = E\left[\sum_{i=0}^{N(T)} U_i \sum_{j=0}^{N(T)} U_j \right] - \mu_Y^2$$

The first summation refers to the interval k and the second to the interval $k+\delta$. However, in the first summation, all random variables are independent from those appearing in the second summation, since they refer to different time intervals, and therefore

$$\gamma_Y(\delta) = 0 \tag{4.11}$$

and consequently

$$\rho_Y(\delta) = 0 \tag{4.12}$$

where $\rho_Y(\delta)$ represents the population autocorrelation function for the cumulative amounts of precipitation. A stronger proof of the result in eq. (4.11) is obtained using the definition for Y_k given in (4.5) and the second order properties for $N_U(t)$, which are very similar to those corresponding to Y_k:

$$E\left[N_U(t) \right] = \lambda\, t\, E[U]$$

$$Var\left(N_U(t) \right) = \lambda\, t \left(\sigma_U^2 + E^2[U] \right)$$

$$Cov\left(N_U(t_1)\, N_U(t_2) \right) = \lambda\, Min\left(t_1, t_2 \right) \left(\sigma_U^2 + E^2[U] \right)$$

Therefore, any Compound Poisson process exhibits correlation different from zero as long as time intervals overlap.

From eq. (4.6), Y_k takes the value of zero only when N(T) is zero. Besides, if the positive part is conditioned on N(T), the marginal distribution for Y_k can be written as

$$f_Y(y) = f_{N(T)}(0) I_{\{0\}}(y) + \left(\sum_{n=1}^{\infty} f_{Y|N(T)=n}(y) f_{N(T)}(n) \right) I_{(0,\infty)}(y) \tag{4.13}$$

The distribution of the amount of precipitation, conditional on a given number of arrivals in the interval, is the nth convolution of the distribution of the White Noise terms

$$f_{Y|N(T)=n}(y) = f_U^{(n)}(y)$$

with

$$f_U^{(1)}(y) = f_U(y) \text{ and } f_U^{(n)}(y) = \int_0^\infty f_U^{(n-1)}(y-z) f_U(z) dz$$

and therefore $f_Y(y)$ becomes

$$f_Y(y) = f_{N(T)}(n) I_{\{0\}}(y) + \left[\sum_{n=1}^{\infty} f_U^{(n)}(y) \; f_{N(T)}(n) \right] I_{(0,\infty)}(y) \qquad (4.14)$$

or, after using the Poisson distribution,

$$f_Y(y) = \exp(-\lambda T) I_{\{0\}}(y) + \left[\sum_{n=1}^{\infty} f_U^{(n)}(y) \frac{\exp(-\lambda T)(\lambda T)^n}{n!} \right] I_{(0,\infty)}(y) \qquad (4.15)$$

Different forms for eqs. (4.13) to (4.15) are presented by Waymire and Gupta (1981a). Following the same technique outlined above, the moment generating function for Y_k, $m_Y(x)$, can be shown to be of the form

$$m_Y(x) = f_{N(T)}(0) + \sum_{n=1}^{\infty} m_{S_n}(x) \; f_{N(T)}(n) \qquad (4.16)$$

where $m_{S_n}(x)$ is the moment generating function for the sum of n White Noise terms, $S_n = \sum_{i=1}^{n} U_i$. Again, using the Poisson distribution for the number of occurrences in the interval of length T, eq. (4.16) becomes

$$m_Y(x) = \exp(-\lambda T) + \sum_{n=1}^{\infty} m_{S_n}(x) \frac{\exp(-\lambda T)(\lambda T)^n}{n!} \qquad (4.17)$$

4.4 Exponential case

Developments presented in the preceeding section did not require any assumptions on the distribution of the White Noise terms. However, this type of assumption is required if parameter estimation

or more complicated operations, like computation of distributions, are desired.

In this section, results are presented for the case when the White Noise terms are independently and identically distributed according to an exponential distribution with parameter μ ($\mu>0$)

$$f_U(u) = \mu \exp(-\mu u) I_{(0,\infty)}(u) \tag{4.18}$$

Equation (4.18) does not allow negative values for the random variable U. This constraint is required as long as the process under analysis is positive valued.

The parameter in (4.18), μ, has dimensions 1/length, in this case 1/in.. Mean and variance for the exponential distribution (Mood et al.,1974) are

$$E[U] = 1/\mu$$

$$\sigma_U^2 = 1/\mu^2$$

From eqs. (4.9) and (4.10), mean and variance for the amount of precipitation in a time interval of length T are:

$$\mu_Y = \lambda T/\mu \tag{4.19}$$

$$\sigma_Y^2 = 2\lambda T/\mu^2 \tag{4.20}$$

Given that the random variables U_i, i=1,..,n, are iid exp(μ), their sum, S_n, is gamma distributed with shape parameter n and scale parameter μ (Mood et al., 1974). Using this result in (4.15), the distribution for Y_k becomes

$$f_Y(y)=\exp(-\lambda T) I_{\{0\}}(y) + \frac{\exp(-\mu y-\lambda T)}{y} \left[\sum_{n=1}^{\infty} \frac{(\lambda T \mu y)^n}{(n!)^2} \right] I_{(0,\infty)}(y) \tag{4.21}$$

Similarly, from eq. (4.16), the moment generating function for Y_k is given by:

$$m_Y(x) = \exp[\lambda T x/(\mu - x)] \tag{4.22}$$

Although the moment generating function has a nice closed solution, the case is not the same for the probability distribution function.

4.5 Gamma case

Assume the White Noise terms in the PWN model are independently and identically distribute following a Gamma distribution with scale parameter μ and shape parameter r ($\mu>0$, $r>0$). The distribution of U is of the form

$$f_U(u) = \frac{\mu^r}{\Gamma(r)} u^{r-1} \exp(-\mu u)\, I_{(0,\infty)}(u) \tag{4.23}$$

Again, μ has dimensions 1/length and r is dimensionless. Mean and variance for the distribution are (Mood et al., 1974)

$$E[U] = r/\mu$$

$$\sigma_U^2 = r/\mu^2$$

The gamma case collapses into the exponential case when r=1.

Taking the two preceeding equations to (4.9) and (4.10), mean and variance for Y_k for the gamma case become

$$\mu_Y = \lambda T r/\mu \tag{4.24}$$

$$\sigma_Y^2 = \lambda T r\,(r+1)/\mu^2 \tag{4.25}$$

Given that the random variables U_i, i=1,..,n, are iid Gamma(μ,r), their sum, S_n, is also gamma distributed with scale parameter μ and

shape parameter nr. Under this condition, the probability distribution function for Y_k takes the form

$$f_Y(y)=\exp(-\lambda T)I_{\{0\}}(y)+\frac{\exp(-\mu y-\lambda T)}{y}\left\{\sum_{n=1}^{\infty}\frac{[\lambda T(\mu y)^r]^n}{\Gamma(nr)\ n!}\right\}I_{(0,\infty)}(y) \qquad (4.26)$$

and the moment generating function for Y_k, from eq. (4.17), becomes

$$m_Y(x) = \exp\left\{\lambda T\left[\left(\frac{\mu}{\mu-x}\right)^r - 1\right]\right\} \qquad (4.27)$$

When the White Noise terms in the PWN model are considered as Gamma distributed, three parameters, λ, μ and r, must be estimated in order to define the model. If method of moments were to be used, as it will, mean and variance would not be sufficient to perform the estimation. An additional property is required and the one selected here is the skewness coefficient for the amount of precipitation in a time interval of length T, denoted by g_Y, and defined as

$$g_Y = \left(E\left[Y_k^3\right] - 3\,\mu_Y\,E\left[Y_k^2\right] + 2\,\mu_Y^3\right) / \sigma_Y^3$$

The second raw moment for Y_k is obtained from eqs. (4.24) and (4.25)

$$E\left[Y_k^2\right] = \lambda Tr\,(\lambda Tr+r+1) / \mu^2$$

The third raw moment is computed from the moment generating function for Y_k, eq. (4.27), using (Mood et al., 1974)

$$E\left[Y_k^3\right] = \left.\frac{d^3 m_Y(x)}{dx}\right|_{x=0}$$

which, after performing differentiations and evaluations, becomes

$$E\left[Y_k^3\right] = \lambda Tr\left[(\lambda Tr+r+1)^2 + (\lambda Tr+r+1) + \lambda Tr^2\right] / \mu^3$$

Replacing expressions for the raw moments in the definition for the the skewness, the final result for g_Y is

$$g_Y = (r+2)\,[\lambda Tr(r+1)]^{-1/2} \qquad (4.28)$$

A nice result, obtained from the preceeding derivation, shows that the third centered moment for Y_k, for the gamma case, is given by

$$E\left[\left(Y_k - \mu_Y\right)^3\right] = \lambda Tr\,(r+1)(r+2) / \mu^3$$

Using mathematical induction, it is possible to show that the following relationship holds for the ith centered moment for Y_k

$$E\left[\left(Y_k - \mu_Y \right)^i\right] = \lambda T r \prod_{j=1}^{i-1} (r+i)/\mu^i, \quad i \geq 2 \tag{4.29}$$

4.6 Parameter estimation by method of moments

Method of moments is used here to estimate parameters for the PWN model, for both the exponential and the gamma cases. In this method, sample estimated statistics are equated to the population counterparts and model parameters are obtained by solving the resulting set of equations. Equations (4.21) and (4.26) suggest the use of maximum likelihood estimators, since amounts of precipitation in disjoint intervals are independent. However, the infinite summations in the continuous part of the distribution makes this estimation technique cumbersome to develop and apply.

In Chapter 3, statistics for the four precipitation gauging stations were estimated, for different aggregation scales for each sample, the smallest of them being the measurement scale. To account for seasonal periodicity, samples were divided in calendar months. The month was denoted by the index τ. In order to keep this partition in the estimation of parameters, the same index is attached to the parameters. The reader must remember that mean, standard deviation and skewness were denoted respectively by $\overline{Y}_\tau$, S_τ and $\hat{g}_\tau$, while, in general, measurement scale was denoted by T.

For the exponential case, eqs. (4.19) and (4.20) provide moment estimators for the model, given by

$$\hat{\mu}_\tau = 2\overline{Y}_\tau / S_\tau^2 \tag{4.30}$$

$$\hat{\lambda}_\tau = \bar{Y}_\tau \hat{\mu}_\tau \; / \; T \tag{4.31}$$

Similar results are obtained for the gamma case using eqs. (4.24), (4.25) and (4.28)

$$\phi_\tau = \bar{Y}_\tau \hat{g}_\tau \; / \; S_\tau \tag{4.32}$$

$$\hat{r}_\tau = (2 - \phi_\tau) \; / \; (\phi_\tau - 1) \tag{4.33}$$

$$\hat{\lambda}_\tau = \left(\frac{\bar{Y}_\tau}{S_\tau} \right) \frac{\hat{r}_\tau + 1}{T \, \hat{r}_\tau} \tag{4.34}$$

$$\hat{\mu}_\tau = \lambda_\tau T \, \hat{r}_\tau \; / \; \bar{Y}_\tau \tag{4.35}$$

In eq. (4.32), ϕ_τ is an intermediate variable for estimation. For the estimated parameters to have physical meaning, $\hat{r}_\tau$ must be positive. From eq. (4.33) this happens when ϕ_τ is between 1.0 and 2.0.

4.7 Results for estimated parameters

Tables 4.1 to 4.4 give estimated parameters for the PWN precipitation model, using the method of moments, for both the exponential and the gamma case, for different temporal aggregation levels. Statistics used in the estimation process are listed in Tables 3.2 to 3.5.

The first feature observed in Tables 4.1 to 4.4 is the presence of negative parameter estimated values for some months in the gamma case, when the constraint in ϕ_τ is violated. Although these values are listed for sake of completeness, the corresponding models can not be used. The fact that the PWN gamma model may yield negative

Table 4.1 Estimated parameters for the PWN model for Denver Wsfo Ap station, for different temporal aggregation levels.

T= 60 min.

	Exponential model		Gamma model			
τ	$\hat{\mu}_\tau$ (1/in.)	$\hat{\lambda}_\tau$ (1/min.) x 10^{-3}	$\hat{r}_\tau$	$\hat{\mu}_\tau$ (1/in.)	$\hat{\lambda}_\tau$ (1/min.) x 10^{-3}	ϕ_τ
1	54.90761	0.622377	0.52854	41.96426	0.899959	1.654218
2	48.01580	0.727119	0.54489	37.08950	1.030779	1.647296
3	35.73554	1.051458	0.25292	22.38687	2.604371	1.798135
4	20.50053	0.843733	0.05298	10.79328	8.385223	1.949689
5	13.40794	0.781258	-0.57238	2.86672	-0.291832	3.338551
6	10.04224	0.401940	-0.13556	4.34042	-1.281463	2.156829
7	5.70995	0.231519	0.02736	2.93310	4.346263	1.973365
8	5.89344	0.188393	-0.16414	2.46303	-0.479663	2.196380
9	11.32981	0.299843	-0.56292	2.47602	-0.116401	3.287913
10	24.00090	0.501978	-0.22520	9.29794	-0.863524	2.290656
11	41.60662	0.793854	0.81768	37.81382	0.882356	1.550151
12	43.73308	0.576110	0.30140	28.45721	1.243764	1.768400

T_a= 1440 min.

	Exponential model		Gamma model			
τ	$\hat{\mu}_\tau$ (1/in.)	$\hat{\lambda}_\tau$ (1/min.) x 10^{-3}	$\hat{r}_\tau$	$\hat{\mu}_\tau$ (1/in.)	$\hat{\lambda}_\tau$ (1/min.) x 10^{-3}	ϕ_τ
1	7.736598	0.087695	0.16494	4.506323	0.309693	1.858415
2	7.261398	0.109960	0.49621	5.432290	0.165780	1.668355
3	4.447032	0.130843	-0.44512	1.233772	-0.081556	2.802209
4	2.643389	0.108742	-0.26112	0.976563	-0.153842	2.353414
5	2.380463	0.138631	0.16066	1.381459	0.500749	1.861575
6	2.582596	0.103320	-0.09393	1.169996	-0.498284	2.103677
7	3.412583	0.138368	0.63277	2.785984	0.178519	1.612455
8	3.294588	0.105272	0.39843	2.303632	0.184743	1.715085
9	3.779146	0.099921	0.69368	3.200333	0.121984	1.590430
10	4.198149	0.087801	0.56641	3.288005	0.121407	1.638403
11	6.380383	0.121683	1.29677	7.327145	0.107759	1.435393
12	4.739092	0.062400	-0.44969	1.303965	-0.038174	2.817184

T_a= Month

		Exponential model			Gamma model		
τ	T_a (min.)	$\hat{\mu}_\tau$ (1/in.)	$\hat{\lambda}_\tau$ (1/min.) x 10^{-3}	$\hat{r}_\tau$	$\hat{\mu}_\tau$ (1/in.)	$\hat{\lambda}_\tau$ (1/min.) x 10^{-3}	ϕ_τ
1	44640	7.044699	0.079852	3.82567	16.99770	0.050362	1.20723
2	40320	6.455961	0.098587	-5.32749	-13.96900	0.040040	0.76892
3	44640	3.705541	0.109027	-0.48365	0.95667	-0.058193	2.93670
4	43200	3.451323	0.141979	2.43154	5.92167	0.100184	1.29141
5	44640	1.747190	0.101751	-12.9844	-10.46953	0.046957	0.91656
6	43200	1.996019	0.079854	-20.2754	-19.23704	0.037957	0.94812
7	44640	2.292470	0.092951	-0.30288	0.79905	-0.106962	2.43449
8	44640	1.966153	0.062824	-0.16413	0.82172	-0.159971	2.19636
9	43200	2.008619	0.053108	1.29797	2.30787	0.047012	1.43517
10	44640	2.271431	0.047505	0.25315	1.42323	0.117579	1.79799
11	44640	5.911912	0.109112	-0.20006	2.36458	-0.218136	2.25010
12	44640	3.306734	0.043540	-0.26036	1.22288	-0.061842	2.35203

Table 4.2 Estimated parameters for the PWN model for Greenland 9 SE station, for different temporal aggregation levels.

T= 60 min.

r	Exponential model $\hat{\mu}_r$ (1/in.)	Exponential model $\hat{\lambda}_r$ (1/min.) x 10^{-3}	Gamma model $\hat{r}_r$	Gamma model $\hat{\mu}_r$ (1/in.)	Gamma model $\hat{\lambda}_r$ (1/min.) x 10^{-3}	ϕ_r
1	49.65537	0.407836	0.61506	40.09809	0.535462	1.619173
2	49.16810	0.594852	0.55294	38.17764	0.835322	1.643938
3	30.87639	0.792236	1.11993	32.72786	0.749818	1.471714
4	23.99735	0.755676	0.67690	20.12058	0.936027	1.596338
5	15.21013	0.772294	-0.16333	6.36293	-1.978053	2.195215
6	8.93966	0.405801	-0.21396	3.51343	-0.745372	2.272213
7	5.72183	0.354143	-0.43194	1.62517	-0.232871	2.760384
8	5.71264	0.322678	-0.06348	2.67498	-2.379960	2.067790
9	12.63772	0.277714	-0.29917	4.42844	-0.325280	2.426881
10	27.90714	0.402281	1.13068	29.73059	0.379034	1.469333
11	31.23062	0.498492	0.70250	26.58502	0.604047	1.587372
12	35.57092	0.390213	0.78761	31.79349	0.442825	1.559405

T_a= 1440 min.

r	Exponential model $\hat{\mu}_r$ (1/in.)	Exponential model $\hat{\lambda}_r$ (1/min.) x 10^{-3}	Gamma model $\hat{r}_r$	Gamma model $\hat{\mu}_r$ (1/in.)	Gamma model $\hat{\lambda}_r$ (1/min.) x 10^{-3}	ϕ_r
1	13.92901	0.113967	0.99902	13.92222	0.114023	1.500243
2	11.82636	0.142863	1.59882	15.36732	0.116109	1.384789
3	4.98687	0.127396	-0.23514	1.90713	-0.207192	2.307430
4	4.26857	0.132387	-0.39288	1.29575	-0.102283	2.647142
5	3.72396	0.187514	-0.03819	1.79086	-2.360834	2.039713
6	3.53775	0.160151	0.42198	2.51531	0.269836	1.703243
7	3.09940	0.190944	-0.23629	1.18352	-0.308562	2.309402
8	2.84137	0.160110	0.02707	1.45914	3.037529	1.973644
9	5.22699	0.113941	0.43294	3.74498	0.188561	1.697865
10	5.40501	0.077259	0.61238	4.35747	0.101711	1.620199
11	6.16346	0.097710	0.66126	5.11955	0.122737	1.601953
12	5.54314	0.060569	-0.33100	1.85417	-0.061203	2.494780

T_a= Month

r	T_a (min.)	Exponential model $\hat{\mu}_r$ (1/in.)	Exponential model $\hat{\lambda}_r$ (1/min.) x 10^{-3}	Gamma model $\hat{r}_r$	Gamma model $\hat{\mu}_r$ (1/in.)	Gamma model $\hat{\lambda}_r$ (1/min.) x 10^{-3}	ϕ_r
1	44640	16.00371	0.130701	-2.24995	-10.002	0.036305	0.199972
2	40320	15.40310	0.187190	0.62937	12.549	0.242306	1.613732
3	44640	2.32392	0.059258	-0.36572	0.737	-0.051382	2.576591
4	43200	2.65923	0.082239	-25.4197	-32.469	0.039502	0.959049
5	44640	2.13002	0.107155	1.23330	2.379	0.097020	1.447768
6	43200	2.45947	0.107309	1.69499	3.314	0.085309	1.371058
7	44640	3.33794	0.201737	-79.5268	-131.058	0.099600	0.987265
8	44640	2.04637	0.112006	0.69241	1.732	0.136884	1.590873
9	43200	3.64320	0.078887	5.18524	11.267	0.047050	1.161675
10	44640	2.56614	0.036614	-0.16486	1.072	-0.092733	2.197417
11	44640	3.17928	0.048775	0.68168	2.673	0.060164	1.594643
12	44640	3.48967	0.038131	-0.21192	1.375	-0.070894	2.268915

Table 4.3 Estimated parameters for the PWN model for Idaho Springs station, for different temporal aggregation levels.

T= 5 min.

	Exponential model		Gamma model			
τ	$\hat{\mu}_\tau$ (1/in.)	$\hat{\lambda}_\tau$ (1/min.) x 10^{-3}	$\hat{r}_\tau$	$\hat{\mu}_\tau$ (1/in.)	$\hat{\lambda}_\tau$ (1/min.) x 10^{-3}	ϕ_τ
1	161.9192	8.248169	0.84061	149.015	9.030121	1.543296
2	151.7268	14.79033	5.89838	523.335	8.648928	1.144961
3	183.2654	4.424027	7.19410	750.847	2.519489	1.122039
4	171.6261	5.495474	3.29916	368.924	3.580593	1.232603
5	152.9662	2.514765	2.85652	294.959	1.697562	1.259301
6	83.4173	2.724409	0.55529	64.869	3.815364	1.642968
7	44.1060	2.203534	-0.02131	21.583	-50.57838	2.021783
8	40.4714	2.694583	-0.25633	15.049	-3.908651	2.344694
9	4.25460	0.204731	-0.31298	1.462	-0.224692	2.455577
10	79.6696	0.737740	-0.38912	24.334	-0.579083	2.636993
11	164.0550	12.80942	3.69096	384.788	8.139949	1.213175
12	201.4827	11.12991	41.7633	4308.037	5.698203	1.023384

T_a= 60 min.

	Exponential model		Gamma model			
τ	$\hat{\mu}_\tau$ (1/in.)	$\hat{\lambda}_\tau$ (1/min.) x 10^{-3}	$\hat{r}_\tau$	$\hat{\mu}_\tau$ (1/in.)	$\hat{\lambda}_\tau$ (1/min.) x 10^{-3}	ϕ_τ
1	43.11349	2.139722	-0.68647	6.75850	-0.488612	4.189577
2	19.26280	1.837189	0.68984	16.27550	2.250204	1.591772
3	67.11936	1.579654	0.04469	35.05956	18.46208	1.957218
4	35.45017	1.112249	1.10423	37.29761	1.059757	1.475233
5	49.93580	0.790733	0.46885	36.67401	1.238642	1.680806
6	20.52486	0.643249	0.51539	15.55163	0.945659	1.659893
7	9.96174	0.485900	-0.20941	3.93782	-0.917200	2.264879
8	7.94281	0.516269	-0.34209	2.61283	-0.496443	2.519964
9	1.72017	0.080813	-0.07536	0.79527	-0.495751	2.081505
10	37.57172	0.338771	0.19287	22.40916	1.047607	1.838311
11	26.92789	2.066401	0.97264	26.55951	2.095465	1.506934
12	72.40884	3.926369	1.19075	79.31493	3.611877	1.456464

Table 4.4 Estimated parameters for the PWN model for Ward station, for different temporal aggregation levels.

T= 5 min.

	Exponential model		Gamma model			
τ	$\hat{\mu}_\tau$ (1/in.)	$\hat{\lambda}_\tau$ (1/min.) x 10^{-3}	$\hat{r}_\tau$	$\hat{\mu}_\tau$ (1/in.)	$\hat{\lambda}_\tau$ (1/min.) x 10^{-3}	ϕ_τ
1	191.5028	1.474571	24.7854	2468.984	0.767032	1.038781
2	186.7619	15.04554	3.65843	435.009	9.579052	1.214664
3	190.2575	4.318847	16.2298	1639.052	2.292476	1.058038
4	126.6906	11.23492	1.55492	161.8424	9.230155	1.391401
5	60.5722	7.220210	-0.73164	8.1275	-1.324143	4.726362
6	14.6198	1.091225	-0.72879	1.9825	-0.203033	4.687303
7	39.2559	3.172660	-0.25826	14.5588	-4.555992	2.348185
8	41.6346	2.911925	-0.32763	13.9969	-2.987921	2.487282
9	73.4079	2.554594	-0.71855	10.3302	-0.500301	4.553059
10	70.2829	3.601293	-0.79592	7.1715	-0.461681	5.900134
11	186.3857	9.378929	10.6769	1088.201	5.128681	1.085639
12	196.4563	2.290680	62.7476	6261.807	1.163593	1.015686

T_a= 60 min.

	Exponential model		Gamma model			
τ	$\hat{\mu}_\tau$ (1/in.)	$\hat{\lambda}_\tau$ (1/min.) x 10^{-3}	$\hat{r}_\tau$	$\hat{\mu}_\tau$ (1/in.)	$\hat{\lambda}_\tau$ (1/min.) x 10^{-3}	ϕ_τ
1	88.42606	0.667027	-0.14589	37.76263	-1.952490	2.170814
2	39.93461	3.152704	0.77881	35.51809	3.600399	1.562172
3	57.88164	1.282560	1.43735	70.53883	1.087435	1.410282
4	16.17382	1.407284	0.42993	11.56372	2.340287	1.699335
5	12.19538	1.402976	-0.34236	4.01006	-1.347461	2.520598
6	5.17660	0.367858	-0.54335	1.18194	-0.154572	3.189878
7	9.70103	0.734465	-0.10277	4.35198	-3.205783	2.114552
8	10.78255	0.726205	0.33854	7.21645	1.435650	1.747081
9	20.18146	0.684858	-0.28318	7.23317	-0.866761	2.395064
10	21.21028	1.057403	-0.13821	9.13936	-3.296542	2.160380
11	36.98605	1.831981	0.66484	30.78784	2.293761	1.600660
12	73.44916	0.843196	1.30465	84.63736	0.447483	1.433905

estimated parameter values is a strong shortcoming and a disadvantage when compared to the exponential case.

Although no plots are presented for estimated parameter values, it is easy to verify that they exhibit seasonal periodicity. In some cases, periodicity observed at the measurement scale is preserved at larger aggregation scales, but in other cases the shape of the seasonal curve is different. For instance, for Denver Wsfo Ap station, the parameter in the exponential distribution, $\hat{\mu}_\tau$, tends to be smaller for summer months, for the three aggregation scales used with this station. In the same station, the rate of arrival for the exponential case, $\hat{\lambda}_\tau$, shows the same periodic behavior for hourly and monthly aggregation scales, but for T_a=1440 min. no periodic behavior is detected. On the other hand, for Greenland 9 SE station, the rate of arrival for the exponential model, $\hat{\lambda}_\tau$, tends to be smaller for summer months for the hourly aggregation scale, but larger in the same months for the remaining T_a values. Despite these discrepancies, the general trend in the periodic behavior is to decrease the rate of arrivals and to increase the mean of the White Noise distribution for summer months. This observation agrees with the periodic behavior of sample statistics described in Chapter 3.

Due to the relative large number of failures of the gamma case in estimating parameters, any comparison with the exponential case is difficult. However, in most of the cases, the shape parameter in the gamma distribution, $\hat{r}_\tau$, tend to be smaller than 1.0, indicating that a distribution similar to the exponential, but with a steeper decaying shape, would be more adequate for White Noise description. Fewer estimated r_τ values larger than 1.0 tend to suggest the use of

distributions with a single maximum. This point was not investigated further.

Tables 4.1 to 4.4 are examples of one of the main shortcomings found in the application of the PWN model to precipitation samples. Parameter estimates are different at different aggregation levels. In the exponential case the general tendency in the arrival rate, $\hat{\lambda}_\tau$, is to decrease with an increase in the aggregation level. The same behavior is observed for the parameter in the exponential distribution. As the aggregation scale increases, the model is recognizing a smaller number of arrivals and generating instantaneous bursts of precipitation with a larger mean value.

An explanation for this behavior is found in the statistics used to estimate parameters. As the aggregation scale increases, more information is lumped together and less information is conveyed by the statistics included in the estimation process.

As a last point allowing judgment about the two cases of the PWN precipitation model analyzed here, Figures 4.2 to 4.4 compare sample and PWN computed sample probability distribution functions for the amount of precipitation. In using eqs. (4.21) and (4.26), infinite summations were approximated by computing terms until the last one was smaller than a given tolerance.

In the figures, in general, the exponential case tends to perform better than the gamma model, providing, in some cases, an adequate fit to the sample pdf. However, the model is not able to reproduce observed high frequencies for small values of precipitation. Also, comparison of both cases, where gamma estimation was possible, shows no appreciable benefit in using the gamma model, i.e., by introducing a third parameter in the PWN precipitation model.

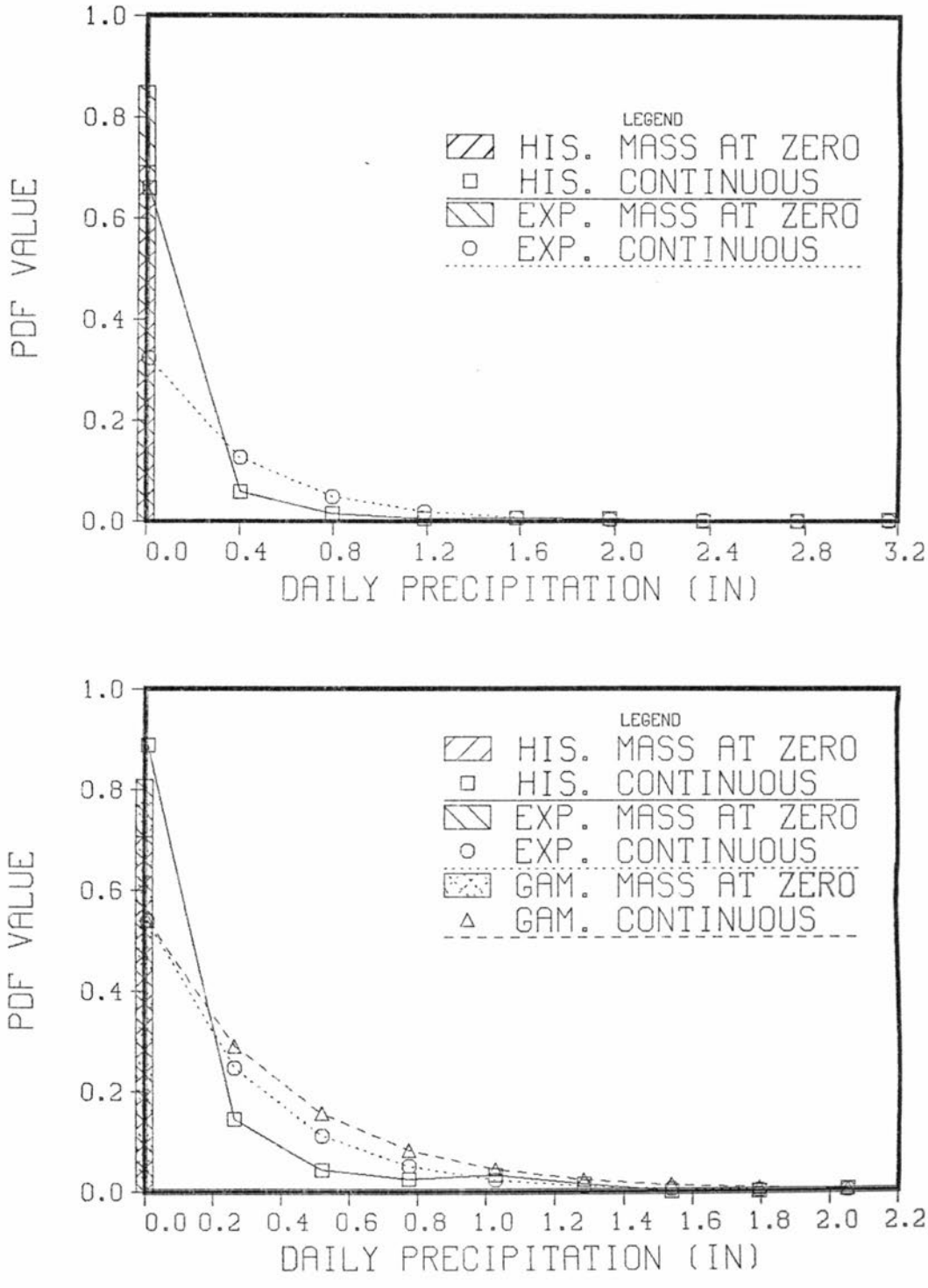
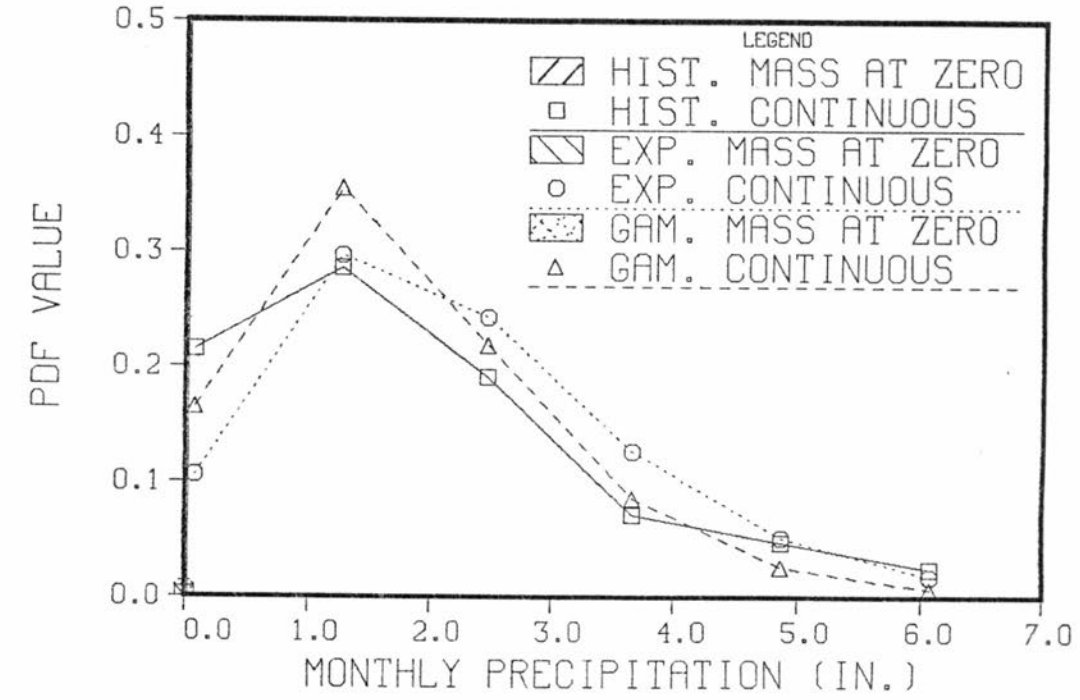


Figure 4.2 Historical and PWN computed probability distribution functions for the amount of precipitation, for Denver Wsfo Ap station, for months 06 and 07 and for T_a= 1440 min..



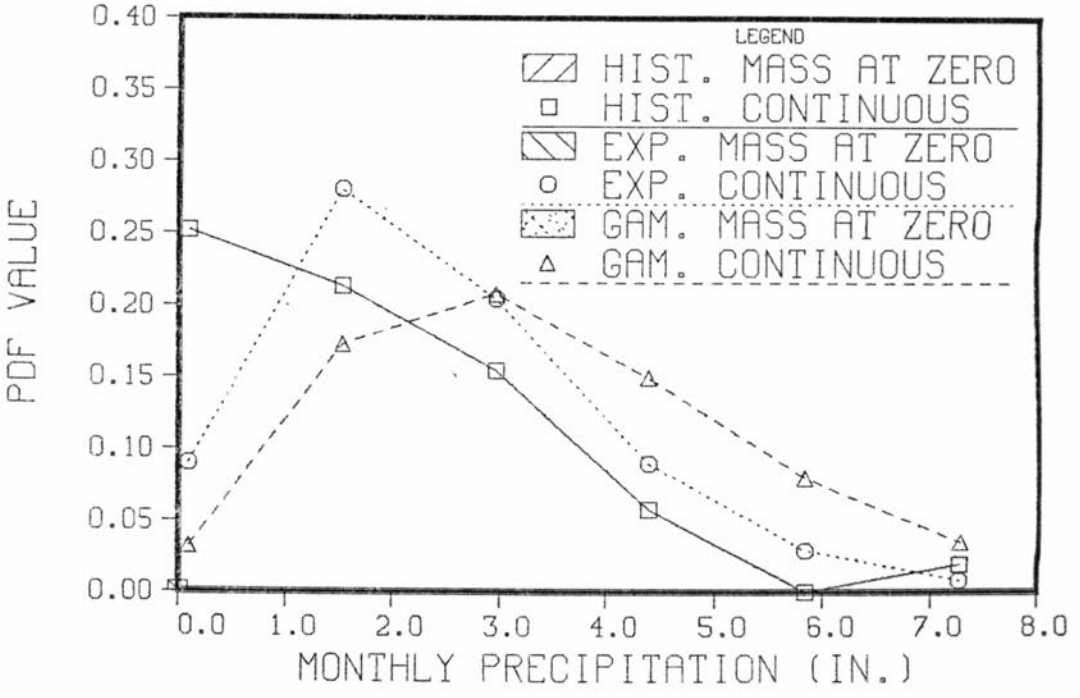


Figure 4.3 Historical and PWN computed probability distribution functions for the amount of precipitation, for Greenland 9 SE station, for months 05 and 08 and for T_a = month.

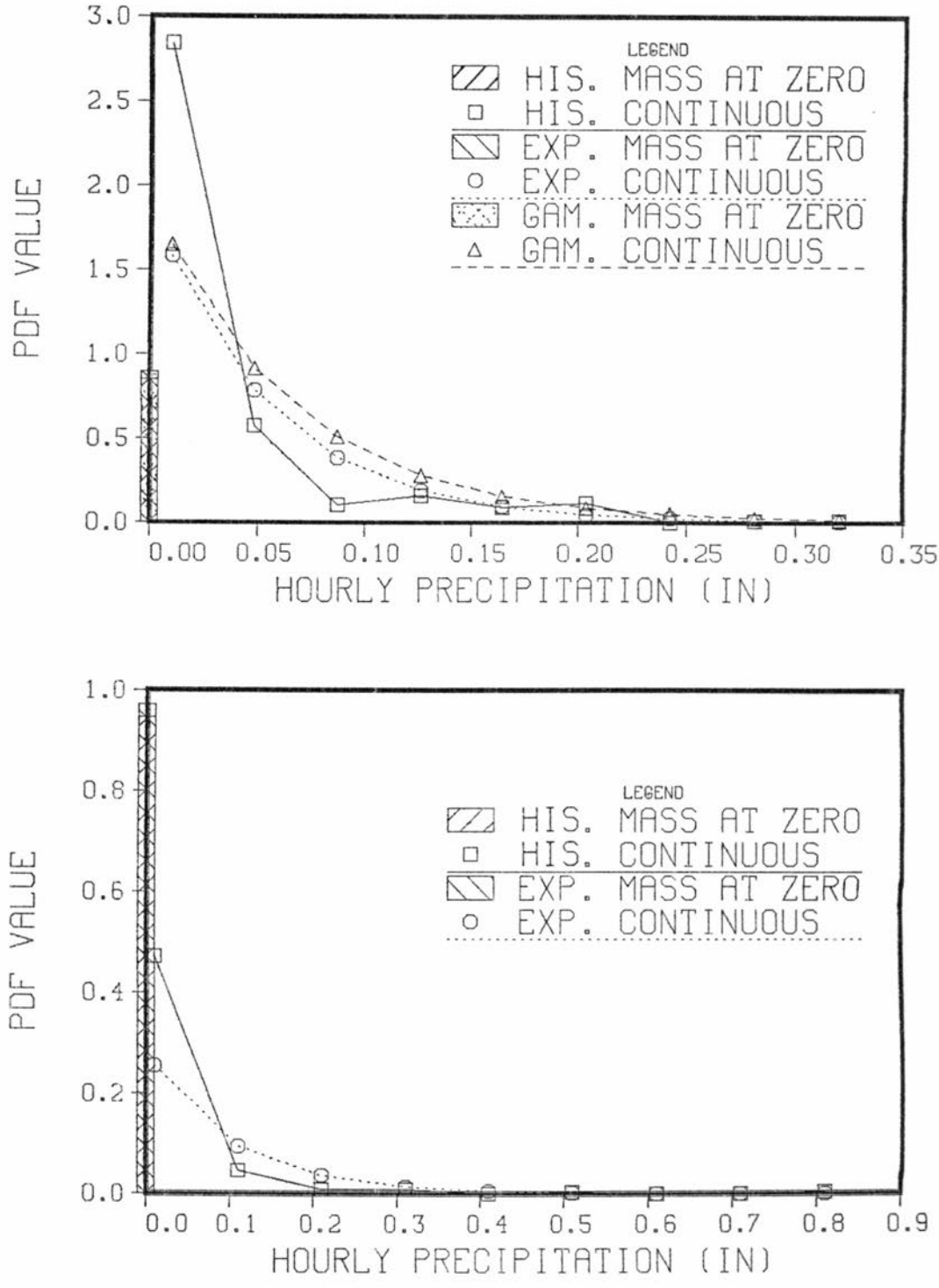


Figure 4.4 Historical and PWN computed probability distribution functions for the amount of precipitation, for Idaho Springs station, for months 02 and 07 and for T_a= 60 min..

Chapter 5

POISSON WHITE NOISE DISAGGREGATION MODEL FOR PRECIPITATION

5.1 Introduction

In this chapter, a precipitation disaggregation model based on the Poisson White Noise (PWN) model is derived. First, model structure is described, followed by the derivation of the required distributions. Next, the PWN disaggregation algorithm is described. Finally, the model is tested and applied to some months for the precipitation recording stations used in this study. The derived disaggregation model will be referred to as PWND model.

5.2 Disaggregation model structure

As stated in Chapter 1, an underlying stochastic processes controlling precipitation occurrence in time is assumed. The information available is this stochastic process and the recorded amounts of precipitation for a given aggregation scale T_a. Given a disaggregation scale T_d, of which T_a is an integer multiple, the idea is to obtain some realization of cumulative amounts of precipitation at the temporal level T_d which preserve the recorded amounts of precipitation and the underlying stochastic process.

The first step is to simulate the number of arrivals in the interval of length T_a, conditional on the recorded amount of precipitation. Given the number of arrivals, the amount of precipitation must be partitioned into White Noise values which add up

to the initial amount. Next, the arrival times for the White Noise terms are simulated. Finally, addition of the White Noise values over disjoint intervals of length T_d generates the required disaggregated precipitation amounts.

In order to perform disaggregation as described above, and given that a PWN process controls precipitation formation, the following distributions are required:

- Distribution of the number of occurrences in an interval of length T, N(T), given the recorded precipitation in that interval, Y:

  $f_{N(T)|Y}(n)$, n=1,2,...

- Distribution of the ith White Noise term, U_i, conditional on Y, conditional on N(T) and conditional on i-1 previous simulated White Noise terms $\underline{U}_{i-1}$:

  $f_{U_i|(Y,N(T),\underline{U}_{i-1})}(u_i)$, $u_i>0$, i=1,..,n

- Joint distribution of the arrival times in the interval of length T, $\underline{T}_n$, conditional on N(T):

  $f_{\underline{T}_n|N(T)}(\underline{t}_n)$

To alleviate notation, the subscript in the temporal scale has been dropped. Also, underlined characters denote vectors and the index for each vector stands for its dimension.

5.3 Distribution of N(T) conditional on Y

The distribution of the number of occurrences in a time interval of length T, conditional on the precipitation recorded in the same interval, is computed using the definition for conditional distributions

$$f_{N(T)|Y}(n) = f^{*}_{N(T),Y}(n,y) \;/\; f^{*}_{Y}(y)\; I_{\{1,2,..\}}(n) \qquad (5.1)$$

The distribution in the numerator of eq. (5.1) is the joint distribution of the number of occurrences and the total precipitation. The star (*) indicates that distributions admit only positive values for the random variables.

Marginal distributions for the number of occurrences and the cumulative amount of precipitation in a time interval of length T were presented in Chapter 4 and denoted as $f_{N(T)}(n)$ and $f_{Y}(y)$ respectively. Given that positive precipitation is recorded in the interval, the truncated marginal distributions for N(T) and Y are:

$$f^{*}_{N(T)}(n) = f_{N(T)}(n) \;/\; (1 - f_{N(T)}(0))\; I_{\{1,2,..\}}(n) \qquad (5.2)$$

$$f^{*}_{Y}(y) = \frac{1}{1 - f_{N(T)}(0)} \sum_{n=1}^{\infty} f_{U}^{(n)}(y)\; f_{N(T)}(n)\; I_{(0,\infty)}(y) \qquad (5.3)$$

where, as stated previously, $f_{U}^{(n)}(.)$ denotes the nth convolution of the distribution of the white noise terms $f_{U}(.)$, i.e., the distribution of

$$S_n = \sum_{i=1}^{n} U_i \qquad (5.4)$$

Using again the definition for conditional distributions,

$$f^{*}_{Y|N(T)}(y) = f^{*}_{N(T),Y}(n,y) \;/\; f^{*}_{N(T)}(n)\;\; I_{(0,\infty)}(y)$$

and taking into account that (Section 4.3)

$$f^{*}_{Y|N(T)=n}(y) = f_{U}^{(n)}(y)$$

equation (5.1) is transformed into

$$f_{N(T)|Y}(n) = [\; f_{U}^{(n)}(y)\; f^{*}_{N(T)}(n)\;] \;/\; f^{*}_{Y}(y)\;\; I_{\{1,2,..\}}(n)$$

which, after using eqs. (5.2) and (5.3), gives the final general form for the required distribution

$$f_{N(T)|Y}(n) = \frac{f_U^{(n)}(y)\, f_{N(T)}(n)}{\sum_{k=1}^{\infty} f_U^{(k)}(y)\, f_{N(T)}(k)}\, I_{\{1,2,..\}}(n) \qquad (5.5)$$

When the White Noise terms are identically and independently distributed according to an exponential distribution with parameter μ ($\mu>0$) and the Poisson distribution is used for the number of occurrences in the interval of length T, eq. (5.5) simplifies to

$$f_{N(T)|Y}(n) = \frac{n(\lambda T\mu y)^n/(n!)^2}{\sum_{k=1}^{\infty} k(\lambda T\mu y)^k/(k!)^2}\, I_{\{1,2,..\}}(n) \qquad (5.6)$$

Similarly, when the Gamma distribution with shape parameter r and scale parameter μ is used for the White Noise terms, eq. (5.5) becomes

$$f_{N(T)|Y}(n) = \frac{[\lambda T(\mu y)^r]^n/[(n!)\Gamma(nr)]}{\sum_{k=1}^{\infty} [\lambda T(\mu y)^r]^k/[(k!)\Gamma(kr)]}\, I_{\{1,2,..\}}(n) \qquad (5.7)$$

5.4 Distribution of U_i, conditional on Y, N(T) and $\underline{U}_{i-1}$

The objective in this section is to derive the distribution of the ith White Noise term, U_i, given the recorded precipitation in the interval of length T, given N(T) and given that i-1 previous White Noise terms, U_j, j=1,..,i-1, have taken values u_j, j=1,..,i-1. Since N(T) has the value n and Y was recorded as y, then y represents the addition of n White Noise terms. Therefore, the following equivalent notation is adopted to indicate that n is implicit in the expression for the required distribution:

$$f_{U_i|(Y,N(T),\underline{U}_{i-1})}(u_i) = f_{U_i|(Y,\underline{U}_{i-1})}(u_i),\ i=1,..,n-1 \qquad (5.8)$$

and by definition

$$f_{U_1|(Y,\underline{U}_0)}(u_1) = f_{U_1|Y}(u_1)$$

The required distribution is computed starting from the definition for conditional distributions

$$f_{U_i|(Y,\underline{U}_{i-1})}(u_i) = f_{\underline{U}_{i-1},U_i,Y}(\underline{u}_{i-1},u_i,y) \;/\; f_{\underline{U}_{i-1},Y}(\underline{u}_{i-1},y)$$

or

$$f_{U_i|(Y,\underline{U}_{i-1})}(u_i) = f_{\underline{U}_i,Y}(\underline{u}_i,y) \;/\; f_{\underline{U}_{i-1},Y}(\underline{u}_{i-1},y) \qquad (5.9)$$

The joint distribution of the vector with i White Noise terms and the total amount of precipitation is obtained as

$$f_{\underline{U}_i,Y}(\underline{u}_i,y) = \int_0^\infty\int_0^\infty \cdots \int_0^\infty f_{\underline{U}_{n-1},Y}(\underline{u}_{n-1},y)\; du_{i+1}du_{i+2}..du_{n-1} \qquad (5.10)$$

where the integration is performed n-i-1 times, starting with u_{n-1} and ending with u_{i+1}.

The first required distribution, $f_{\underline{U}_{n-1},Y}(\underline{u}_{n-1},y)$, is obtained from $f_{\underline{U}_n}(\underline{u}_n)$, which is the joint distribution of the White Noise terms, using the transformation technique (Mood et al., 1974):

$$f_{\underline{U}_{n-1},Y}(\underline{u}_{n-1},y) = |J|\; f_{\underline{U}_n}(\underline{u}_n) \qquad (5.11)$$

where $|J|$ is the absolute value of the Jacobian of the transformation. In order to obtain $|J|$, the following transformation and inverse transformation are defined

Transformation	Inverse Transformation
$U_i = U_i,\ i=1,..,n-1$	$U_i = U_i,\ i=1,..,n-1$
$Y = \sum_{i=1}^{n-1} U_i + U_n$	$U_n = Y - \sum_{i=1}^{n-1} U_i$

(5.12)

The Jacobian of the transformation is the determinant of an $n \times n$ matrix formed with the partial derivatives of the original variables with respect to the transformed variables, or

$$j_{ik} = \frac{\partial u_i}{\partial u_k},\ 1 \le i \le n-1,\ 1 \le k \le n-1; \quad j_{nk} = \frac{\partial u_n}{\partial u_k},\ 1 \le k \le n-1; \quad j_{nn} = \frac{\partial u_n}{\partial y}$$

where j_{ik} is the generic element of the matrix. Therefore, according to (5.12), the elements of the matrix are

$$j_{ii} = 1,\ 1 \le i \le n; \quad j_{ik} = 0,\ 1 \le i \le n-1,\ 1 \le k \le n,\ i \ne k; \quad j_{nk} = -1,\ 1 \le k \le n-1$$

so that the matrix in the Jacobian becomes

$$\begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ . & . & . & \dots & . & . \\ . & . & . & \dots & . & . \\ 0 & 0 & 0 & & 1 & 0 \\ -1 & -1 & -1 & \dots & -1 & 1 \end{bmatrix}$$

Using cofactors, it is easy to show that J=1 and therefore

$$|J| = 1 \tag{5.13}$$

The next step required to expand eq. (5.11) is to obtain the joint distribution of n-1 White Noise terms. Since these are iid as $f_U(u)$

$$f_{\underline{U}_{n-1}}(\underline{u}_{n-1}) = \prod_{k=1}^{n-1} f_U(u_k)\ I_{(0,\infty)}(u_k) \tag{5.14}$$

Replacing eqs. (5.12) to (5.14) in (5.11) gives the following result

$$f_{\underline{U}_{n-1},Y}(\underline{u}_{n-1},y) = \left(\prod_{k=1}^{n-1} f_U(u_k)\ I_{(0,\infty)}(u_k) \right) f_U\left(y - \sum_{k=1}^{n-1} u_k \right) I_{(0,\infty)}\left(y - \sum_{k=1}^{n-1} u_k \right) \tag{5.15}$$

The set of inequalities defined by the indicator functions

$$0 \le u_i < \infty,\ i=1,..,n-1; \quad 0 \le y - \sum_{k=1}^{n-1} u_k < \infty$$

is equivalent to the set

$$0 \le u_1 \le y; \quad 0 \le u_i \le y - \sum_{k=1}^{i-1} u_k,\ i=2,..,n\ ; \quad 0 \le y < \infty$$

so that eq. (5.15) becomes

$$f_{\underline{U}_{n-1},Y}(\underline{u}_{n-1},y) = \left(\prod_{k=1}^{n-1} f_U(u_k)\, I_{\left(y - \sum_{j=1}^{k-1} u_j \right)}(u_k) \right) f_U\left(y - \sum_{k=1}^{n-1} u_k \right) I_{(0,\infty)}(y) \tag{5.16}$$

To simplify notation, let

$$R_k = y - \sum_{j=1}^{k-1} u_j$$

$$R_1 = y \tag{5.17}$$

Then, eq. (5.16) can be written as

$$f_{\underline{U}_{n-1},Y}(\underline{u}_{n-1},y) = \left(\prod_{k=1}^{n-1} f_U(u_k)\, I_{(0,R_k)}(u_k) \right) f_U(R_n)\, I_{(0,\infty)}(y) \tag{5.18}$$

After using eq. (5.18) in (5.10), the joint distribution of $\underline{U}_i$ and Y becomes

$$f_{\underline{U}_i,Y}(\underline{u}_i,y) = \left(\prod_{k=1}^{i} f_U(u_k)\, I_{(0,R_k)}(u_k) \right) I_{(0,\infty)}(y)$$

$$\int_0^{R_{i+1}} \int_0^{R_{i+2}} \cdots \int_0^{R_{n-1}} \left(\prod_{k=i+1}^{n-1} f_U(u_k) \right) f_U(R_n)\, du_{i+1} du_{i+2} \cdots du_{n-1} \tag{5.19}$$

The most inner integral in (5.19) is solved as

$$\int_0^{R_{n-1}} f_U(u_{n-1})\, f_U(R_n)\, du_{n-1} = \int_0^{R_{n-1}} f_U(u_{n-1})\, f_U(R_{n-1} - u_{n-1})\, du_{n-1}$$

$$= f_U^{(2)}(R_{n-1}) \tag{5.20}$$

i.e., the integral represents the distribution function for the sum of two White Noise terms evaluated at R_{n-1}.

Using the result in (5.20), the second most inner integral in (5.19) is

$$\int_0^{R_{n-2}} f_U(u_{n-2}) \int_0^{R_{n-1}} f_U(u_{n-1})\ f_U(R_n)\ du_{n-1}du_{n-2} =$$

$$\int_0^{R_{n-2}} f_U(u_{n-2})\ f_U^{(2)}(R_{n-2}- u_{n-2})\ du_{n-2} = f_U^{(3)}(R_{n-2}) \tag{5.21}$$

which is the three fold convolution for the common distribution of the White Noise evaluated at R_{n-2}.

Now, assuming that the jth most inner integral is

$$\int_0^{R_{n-j}} \int_0^{R_{n-j+1}} .. \int_0^{R_{n-1}} \left(\prod_{k=j}^{n-1} f_U(u_k) \right) f_U(R_n)\ du_{n-j}du_{n-j+1}..du_{n-1} =$$

$$f_U^{(j+1)}(R_{n-j}) \tag{5.22}$$

it can be shown that the j+1st inner integral is

$$\int_0^{R_{n-j-1}} \int_0^{R_{n-j}} .. \int_0^{R_{n-1}} \left(\prod_{k=j-1}^{n-1} f_U(u_k) \right) f_U(R_n)du_{n-j-1}du_{n-j}..du_{n-1} =$$

$$f_U^{(j+2)}(R_{n-j-1}) \tag{5.23}$$

In conclusion, it has been shown, using mathematical induction, that the multiple integral in eq. (5.19) is the n-i fold convolution of the distribution of the White Noise terms evaluated at R_{i+1}. Therefore, eq. (5.19) turns into

$$f_{\underline{U}_i,Y}(\underline{u}_i,y) = \left(\prod_{k=1}^{i} f_U(u_k) I_{(0,R_k)}(u_k) \right) f_U^{(n-i)}(R_{i+1}) I_{(0,\infty)}(y) \tag{5.24}$$

Taking eq. (5.24) to eq. (5.9), the required distribution is finally obtained as

$$f_{U_i|(Y,\underline{U}_{i-1})}(u_i) = f_U(u_i)\ \frac{f_U^{(n-i)}(R_{i+1})}{f_U^{(n-i+1)}(R_i)}\ I_{(0,R_i)}(u_i) \tag{5.25}$$

Equation (5.25) corresponds to a probability distribution function: First, it is positive for all values of u_i, since all involved terms are densities and consequently positive. Second, due to convolution properties, it integrates up to one. Equation (5.25) is also valid for i=1:

$$f_{U_1|Y}(u_1) = f_U(u_1)\, f_U^{(n-1)}(y-u_1) \,/\, f_U^{(n)}(y)\, I_{(0,y)}(u_1) \qquad (5.26)$$

When the White Noise terms follow the exponential distribution with parameter μ, S_n is gamma distributed with scale parameter μ and shape parameter n:

$$f_{S_n}(s) = f_U^{(n)}(s) = \frac{\mu^n}{(n-1)!}\, s^{n-1} \exp(-\mu s)\, I_{(0,\infty)}(s) \qquad (5.27)$$

Making the appropriate replacements in (5.25), $f_{U_i|(Y,\underline{U}_{i-1})}(u_i)$ for the exponential case becomes

$$f_{U_i|(Y,\underline{U}_{i-1})}(u_i) = (n-i)\, (R_i - u_i)^{n-i-1} / R_i^{\,n-i}\, I_{(0,R_i)}(u_i) \qquad (5.28)$$

In particular, for i=1

$$f_{U_1|(Y)}(u_1) = (n-1)\, (y - u_1)^{n-2} \,/\, y^{n-1}\, I_{(0,y)}(u_1) \qquad (5.29)$$

The cumulative distribution function (cdf) for eq. (5.28) is

$$F_{U_i|(Y,\underline{U}_{i-1})}(u_i) = \frac{R_i^{\,n-i} - \left(R_i - u_i \right)^{n-i}}{R_i^{\,n-i}}\, I_{(0,R_i]}(u_i) + I_{(R_i,\infty)}(u_i) \qquad (5.30)$$

Assume a value F_i, $0 \le F_i \le 1$, is assigned to the cdf given in (5.30). The corresponding quantile is obtained by means of

$$u_i = R_i - \left[R_i^{\,n-i} - \left(1 - F_i \right) \right]^{1/(n-i)} \qquad (5.31)$$

Equation (5.31) facilitates simulation procedures.

When the White Noise terms follow a gamma distribution with scale parameter μ and shape parameter r, S_n is also gamma with parameters μ and nr

$$f_{S_n}(s) = f_U^{(n)}(s) = \frac{\mu^{nr}}{\Gamma(nr)} s^{nr-1} \exp(-\mu s)\, I_{(0,\infty)}(s) \qquad (5.32)$$

Replacement of eq. (5.32) in (5.25) yields to

$$f_{\underline{U}_i|(Y,\underline{U}_{i-1})}(u_i) = \frac{\Gamma[(n-i+1)r]}{\Gamma(nr)\,\Gamma[(n-i)r]} u_i^{r-1} \frac{(R_i - u_i)^{[(n-i)r-1]}}{R_i^{[(n-i+1)r-1]}} I_{(0,R_i)}(u_i)$$

or

$$f_{U_i|(Y,\underline{U}_{i-1})}(u_i) = \frac{1}{B(r,(n-i)r)} u_i^{r-1} \frac{(R_i - u_i)^{[(n-i)r-1]}}{R_i^{[(n-i+1)r-1]}} I_{(0,R_i)}(u_i) \qquad (5.33)$$

where B(.,.) is the Beta function. According to this expression, U_i/R_i has a Beta distribution with parameters r and (n-i)r. For i=1, eq. (5.33) turns into

$$f_{U_1|Y}(u_1) = \frac{1}{B(r,(n-1)r)} u_1^{r-1} \frac{(y - u_1)^{[(n-1)r-1]}}{y^{(nr-1)}} I_{(0,y)}(u_1) \qquad (5.34)$$

The cumulative distribution function for the pdf given in (5.33) is the incomplete Beta function and therefore it has to be computed using numerical integration. Also, for simulation purposes, the inversion of the cdf has to be done numerically. In this sense, introduction of a third parameter in the PWND model complicates its operation.

5.5 Distribution of $\underline{T}_n$ conditional on N(T)

Due to the independence of the number of arrivals in disjoint intervals for the Poisson process, arrival times are well studied in the literature. The result presented here is for the distribution of the arrival times conditional on the number of arrivals in the interval taking a value n. In this case, the joint distribution of

the arrival times is the same as the joint distribution of n order statistics drawn from a uniform (0,T) population (Taylor and Karlin, 1984). The distribution is

$$f_{\underline{T}_n|N(T)}(\underline{t}_n) = \frac{n!}{T^n} I_{(0,t_2)}(t_1)\, I_{(t_1,t_3)}(t_2)\ldots I_{(t_{n-1},T)}(t_n) \qquad (5.35)$$

Therefore, for simulation purposes, the n arrival times are sampled from a uniform (0,T) distribution.

5.6 Disaggregation model algorithm

In order to operate the PWND model, information must be provided about aggregation and disaggregation scales, type of model (exponential or gamma), model parameters and recorded precipitation sample. Since a monthly seasonal partition was adopted for the precipitation recording stations, the model operates on the sample for a given month.

The ratio between aggregation and disaggregation scales is

$R = T_a/T_d$

and the result must be an integer number. The aggregated or recorded sample is denoted by y_i, $i=1,..,N_a$, and the disaggregated series is denoted by y_j^d, $j=1,..,RN_a$, where N_a is the total sample size, collection of all years, available for the specific month at the aggregation scale.

To disaggregate the ith recorded precipitation value y_i, the model proceeds as follows:

- If $y_i = 0$, the value is disaggregated into zero values:

  $y_j^d = 0$, $j=(R-1)i+1,..,Ri$

- If y_i is missing ($y_i<0$), make the disaggregated series, y_j^d, equal to a negative value representing missing data, for j from (R-1)i+1 up to Ri.
- If y_i is positive, the algorithm operates in the following way:
  1. Sample a value for $N(T_a)$, n, conditional on y_i, from the distributions given in eqs. (5.6) or (5.7), according to the type of model selected.
  2. If n=1, make $u_1 = y_i$. Otherwise, sample n-1 White Noise values from the distributions defined by eqs. (5.31) or (5.33) and (5.17), depending on the type of model. The last White Noise value is computed as
  $$u_n = y_i - \sum_{k=1}^{n-1} u_k$$
  3. Sample n random values uniform in the interval $(0,T_a)$ and denote them by t'_k, the prime indicating that they are not ranked.
  4. The sequence of ordered pairs $((i-1)T_a+t'_k,u_k)$, k=1,..,n, is a sample realization of the process in the interval $((i-1)T_a,iT_a)$, which preserves the PWN process and the recorded amounts of precipitation. From this sequence, disaggregated values are computed using the expression
  $$y_j^d = \sum_{k=1}^{n} u_k I_{[(j-1)T_d,jT_d)}((i-1)T_a+t'_k),\ j=(i-1)R+1,..,Ri$$

The description given here for the disaggregation procedure is rather short. Details on approximation and inversion of the distributions in (5.6), (5.7) and (5.33) are skipped for reasons of space.

5.7 Testing the PWND model

The procedures and results presented in this section are not intended as exact statistical tests of hypothesis on the PWND model. They are oriented to asses the goodness in operating the PWND model and its performance when the underlying process controlling precipitation formation is a PWN process.

In order to achieve this objective, two different simulations of a PWN precipitation process were carried out, for the exponential and gamma cases respectively, the first one corresponding to estimated parameters for month 05 of the Denver Wsfo Ap station and the second one to estimated parameters for month 12 of the Idaho Springs station, both parameter sets for T_a=60 min.. Simulations represent samples drawn from a stationary processes, since the same parameters were used for every month in the year. The number of years in each simulation was 33. In the sequel, they will be referred to as simulations 1 and 2. Parameter values used in the simulations are given in Tables 5.1 and 5.2.

Results from both simulations were hourly samples of precipitation. Daily series were obtained by aggregation of hourly values. Monthly statistics were computed using the procedures described in Chapter 3. As expected, no significant correlation was observed in the simulated samples at any of the aggregation scales.

Based on the proper monthly statistics, model parameters were estimated for two different aggregation levels, hourly and daily, using moment estimators (Section 4.6). Although this chapter is not related directly to parameter estimation, Tables 5.1 and 5.2 present some indicators for estimation results, based on samples of size 12.

Table 5.1 Summary of results for parameters estimated from PWN simulated precipitation series, for simulation 1 (Exponential case) and different aggregation levels.

Parameter: λ, Units: 1/min. x 10^{-3}, Population Value: 0.78125		
	T=60 min.	T_a=1440 min.
Mean	0.78957	0.79063
Standard Deviation	0.02837	0.03458
Maximum	0.83145	0.84653
Minimum	0.73808	0.73063
Parameter: μ, Units: 1/in., Population Value: 13.40794		
	T=60 min.	T_a=1440 min.
Mean	13.48279	13.51175
Standard Deviation	0.501162	0.814282
Maximum	14.22012	14.85350
Minimum	12.51807	12.27014

Tables 5.1 and 5.2 give surprising results, in the sense that parameter estimates appear more consistent, through different aggregation scales, than parameters estimated for the precipitation stations did. Furthermore, moment estimators seem to provide adequate values for the parameters, regardless of the aggregation scale. However, a larger sample variability is observed as the aggregation scale increases. Also, parameter estimates for the gamma model tend to be less consistent and more variable than in the exponential case. A possible explanation for the difference between results in simulated and historical series is that the former are samples from a truly PWN process, while the later are not. This explanation may appear more than evident. Surprisingly, the reader will realize in Section 6.9 that the case is not the same for the Neyman-Scott White Noise model.

In order to verify PWND model performance, two months were selected from each one of the simulated series and daily values

Table 5.2 Summary of results for parameters estimated from PWN simulated precipitation series, for simulation 2 (Gamma case) and different aggregation levels.

Parameter: λ, Units: 1/min. x 10^{-3}, Population Value: 3.61187		
	T=60 min.	T_a=1440 min.
Mean	3.17540	3.52911
Standard Deviation	0.11853	0.71693
Maximum	3.36662	5.52492
Minimum	2.98010	2.59786
Parameter: μ, Units: 1/in., Population Value: 79.31493		
	T=60 min.	T_a=1440 min.
Mean	86.43762	80.57882
Standard Deviation	4.599321	19.10057
Maximum	94.52680	125.2309
Minimum	76.84243	46.13704
Parameter: r, Population Value: 1.19075		
	T=60 min.	T_a=1440 min.
Mean	1.46497	1.32111
Standard Deviation	0.12649	0.53968
Maximum	1.70840	2.60646
Minimum	1.20237	0.44657

disaggregated into hourly values, using parameters estimated at the daily level. For simulation 1, months 04 and 07 were chosen, while for simulation 2 the model was applied to months 02 and 12. For the disaggregated series, monthly statistics were computed using the procedures presented in Chapter 3. Comparison of statistics for simulated and disaggregated hourly series is presented in Table 5.3.

Results shown in Table 5.3 allow the conclusion that the disaggregation algorithm was well designed. Even more, given that the underlying process is PWN, the disaggregation model has the ability to preserve all the statistics listed in Table 5.3. However, the lack of serial correlation, the major shortcoming in PWN type models, is also

Table 5.3 Comparison of statistics for PWN simulated and disaggregated series, for simulation 1 (Exponential Case) and simulation 2 (Gamma Case).

T_a = 1440.0 min., T_d = 60.0 min., R = 24

Month τ		Historic	Exponential Model
04			
	$\hat{Y}_\tau \times 10^{-3}$ (in.)	3.6670	3.6670
	$S_\tau \times 10^{-3}$ (in.)	23.8306	24.1452
	$\hat{g}_\tau$	10.0379	9.5573
	$r_\tau(1)$	0.00396	0.00454
	$\hat{p}_o^\tau$	0.95558	0.95641
	No of positive values	1567	1538
	Maximum value (in.)	0.65	0.61
07			
	$\hat{Y}_\tau \times 10^{-3}$ (in.)	3.4949	3.4949
	$S_\tau \times 10^{-3}$ (in.)	22.4358	21.8993
	$\hat{g}_\tau$	9.7469	10.3187
	$r_\tau(1)$	0.00408	0.00508
	$\hat{p}_o^\tau$	0.95617	0.95575
	No of positive values	1598	1613
	Maximum value (in.)	0.80	0.76

T_a = 1440.0 min., T_d = 60.0 min., R = 24

Month τ		Historic	Gamma Model
02			
	$\hat{Y}_\tau \times 10^{-3}$ (in.)	3.2322	3.2322
	$S_\tau \times 10^{-3}$ (in.)	9.5453	9.6261
	$\hat{g}_\tau$	4.1722	4.2837
	$r_\tau(1)$	0.00469	-0.00435
	$\hat{p}_o^\tau$	0.84730	0.84657
	No of positive values	5028	5052
	Maximum value (in.)	0.14	0.14
12			
	$\hat{Y}_\tau \times 10^{-3}$ (in.)	3.2063	3.2063
	$S_\tau \times 10^{-3}$ (in.)	9.4227	9.5508
	$\hat{g}_\tau$	4.1168	4.1072
	$r_\tau(1)$	0.00634	0.00503
	$\hat{p}_o^\tau$	0.84529	0.85245
	No of positive values	5640	5379
	Maximum value (in.)	0.12	0.15

present in Table 5.3. The only statistic preserved exactly is the mean, since cumulative amounts of precipitation are also preserved exactly.

5.8 Application of the PWND model to historical precipitation samples

In order to formulate final conclusions on the applicability of the PWND model to historical samples, disaggregation was performed on some months for each one of the precipitation recording stations used in this study, for different combinations of aggregation and disaggregation scales. Table 5.4 gives a summary of months, models and temporal scales used for each one of the recording stations. As before for the simulations, disaggregation was performed with parameters estimated at the aggregation scale T_a.

Table 5.4 Description of historical samples for application of the PWND model.

Station	Month	Model types	T_a (min.)	T_d (min.)	R
Denver Wsfo Ap	06	Exponential	1440	60	24
Denver Wsfo Ap	07	Exponential, Gamma	1440	60	24
Denver Wsfo Ap	11	Exponential, Gamma	1440	60	24
Denver Wsfo Ap	04	Exponential, Gamma	43200	1440	30
Denver Wsfo Ap	06	Exponential, Gamma	43200	1440	30
Denver Wsfo Ap	07	Exponential	44640	1440	31
Greenland 9 SE	05	Exponential	1440	60	24
Greenland 9 SE	07	Exponential	1440	60	24
Greenland 9 SE	09	Exponential, Gamma	1440	60	24
Greenland 9 SE	05	Exponential, Gamma	44640	1440	31
Greenland 9 SE	08	Exponential, Gamma	44640	1440	31
Greenland 9 SE	10	Exponential	44640	1440	31
Idaho Springs	02	Exponential, Gamma	60	5	12
Idaho Springs	07	Exponential	60	5	12
Idaho Springs	09	Exponential	60	5	12
Ward	04	Exponential, Gamma	60	5	12
Ward	06	Exponential	60	5	12
Ward	08	Exponential, Gamma	60	5	12

For some of the months selected to test the disaggregation model, probability distribution functions for the amount of precipitation were presented in Figures 4.2 to 4.4. Again, these provide information on how good the PWN model fits the data at the aggregation scale.

After the disaggregation model was run for the samples listed in Table 5.4, monthly statistics were computed for the resulting disaggregated series. Comparison of results is given in Tables 5.5 to 5.10 in the same form as Table 5.3 did for the simulation runs.

Examination of Tables 5.5 to 5.10 yields the conclusion that the PWND model is not suitable for disaggregating precipitation series in any of the stations or scales considered. There is no suggestion of preservation of any other statistic different from the mean. The disaggregation model tends to produce larger standard deviations, skewness coefficients and probabilities of zero precipitation than those observed in historical samples. Also, the model does not reproduce the number of positive values, yielding generally a smaller value. Consequently, the continuous part in the pdf of the amount of precipitation is not reproduced either, as exemplified in Figures 5.1 to 5.3. In regard to the maximum precipitation value observed in the samples, the PWND model produces larger values, indicating how aggregated amounts are not partitioned into amounts similar to the historical ones. Good performance of the model is not even achieved for disaggregation of monthly precipitation series into daily precipitation series. As before for the simulation results, the PWND model does not produce any serial correlation.

Table 5.5 Comparison of statistics for historical and PWN disaggregated hourly series, for Denver Wsfo Ap station. T_a = 1440.0 min., T_d=60.0 min., R = 24

	Historic	Exponential Model	Gamma Model
Month τ: 06			
$\bar{Y}_\tau \times 10^{-3}$ (in.)	2.4015	2.4004	
$S_\tau \times 10^{-3}$ (in.)	21.8696	41.4074	
$\hat{g}_\tau$	19.6415	31.2497	
$r_\tau(1)$	0.44566	-0.00337	
$\hat{p}_o^\tau$	0.96046	0.98722	
No of positive values	996	322	
Maximum value (in.)	1.11	2.06	
No of missing values	732	720	
Month τ: 07			
$\bar{Y}_\tau \times 10^{-3}$ (in.)	2.4328	2.4328	2.4328
$S_\tau \times 10^{-3}$ (in.)	29.1912	37.0328	38.5639
$\hat{g}_\tau$	23.6784	24.5129	25.7636
$r_\tau(1)$	0.20808	0.00608	-0.00235
$\hat{p}_o^\tau$	0.96886	0.98690	0.98698
No of positive values	811	341	339
Maximum value (in.)	1.59	1.73	2.05
No of missing values	744	744	744
Month τ: 11			
$\bar{Y}_\tau \times 10^{-3}$ (in.)	1.1448	1.1443	1.1443
$S_\tau \times 10^{-3}$ (in.)	7.4182	18.6346	19.3909
$\hat{g}_\tau$	10.0448	22.9707	24.9000
$r_\tau(1)$	0.72793	0.00085	-0.00349
$\hat{p}_o^\tau$	0.95801	0.99178	0.99205
No of positive values	1088	213	206
Maximum value (in.)	0.18	0.75	0.95
No of missing values	12	0	0

Table 5.6 Comparison of statistics for historical and PWN disaggregated daily series, for Denver Wsfo Ap station. T_a= Month, T_d=1440.0 min.

	Historic	Exponential Model	Gamma Model
Month τ: 04, R=24			
$\bar{Y}_\tau \times 10^{-2}$ (in.)	5.9238	5.9238	5.9238
$S_\tau \times 10^{-1}$ (in.)	2.1171	1.8643	1.8895
$\hat{g}_\tau$	8.4107	5.5635	4.1088
$r_\tau(1)$	0.05288	0.00620	-0.03856
$\hat{p}_o^\tau$	0.71238	0.80476	0.86190
No of positive values	302	205	145
Maximum value (in.)	3.25	2.47	1.67
No of missing values	30	30	30
Month τ: 06, R=30			
$\bar{Y}_\tau \times 10^{-2}$ (in.)	5.7610	5.7610	
$S_\tau \times 10^{-1}$ (in.)	2.1122	2.4589	
$\hat{g}_\tau$	7.7129	5.8060	
$r_\tau(1)$	0.17241	-0.00182	
$\hat{p}_o^\tau$	0.70381	0.90285	
No of positive values	311	102	
Maximum value (in.)	3.16	2.56	
No of missing values	30	30	
Month τ: 07, R=31			
$\bar{Y}_\tau \times 10^{-2}$ (in.)	5.8387	5.8387	
$S_\tau \times 10^{-1}$ (in.)	1.8498	2.2656	
$\hat{g}_\tau$	5.1086	6.5337	
$r_\tau(1)$	0.11189	-0.04680	
$\hat{p}_o^\tau$	0.70599	0.86912	
No of positive values	319	142	
Maximum value (in.)	2.05	3.45	
No of missing values	31	31	

Table 5.7 Comparison of statistics for historical and PWN disaggregated hourly series, for Greenland 9 SE station. T_a= 1440.0 min., T_d=60.0 min., R = 24

	Historic	Exponential Model	Gamma Model
Month τ: 05			
$\bar{Y}_\tau \times 10^{-3}$ (in.)	3.0465	3.0213	
$S_\tau \times 10^{-3}$ (in.)	20.0147	40.7665	
$\hat{g}_\tau$	14.4219	25.3294	
$r_\tau(1)$	0.35303	-0.00067	
$\hat{p}_o^\tau$	0.94578	0.98658	
No of positive values	1399	349	
Maximum value (in.)	0.85	2.54	
No of missing values	984	768	
Month τ: 07			
$\bar{Y}_\tau \times 10^{-3}$ (in.)	3.7136	3.6964	
$S_\tau \times 10^{-3}$ (in.)	36.0284	49.3419	
$\hat{g}_\tau$	26.7805	28.6486	
$r_\tau(1)$	0.23827	-0.00546	
$\hat{p}_o^\tau$	0.95490	0.98388	
No of positive values	1114	400	
Maximum value (in.)	2.24	3.07	
No of missing values	2083	1968	
Month τ: 09			
$\bar{Y}_\tau \times 10^{-3}$ (in.)	1.3185	1.3079	1.3079
$S_\tau \times 10^{-3}$ (in.)	14.4451	22.4741	22.3142
$\hat{g}_\tau$	26.5882	30.1860	28.8778
$r_\tau(1)$	0.28296	-0.00328	-0.00322
$\hat{p}_o^\tau$	0.97560	0.99249	0.99225
No of positive values	606	188	194
Maximum value (in.)	0.96	1.36	1.36
No of missing values	1089	888	888

Table 5.8 Comparison of statistics for historical and PWN disaggregated daily series, for Greenland 9 SE station. T_a= Month, T_d=1440.0 min.

	Historic	Exponential Model	Gamma Model
Month τ: 05, R=31			
$\bar{Y}_\tau \times 10^{-2}$ (in.)	7.2509	7.2509	7.2509
$S_\tau \times 10^{-1}$ (in.)	1.9734	2.6009	2.6575
$\hat{g}_\tau$	5.5512	5.2121	4.9862
$r_\tau(1)$	0.20096	0.05231	-0.00367
$\hat{p}_o^\tau$	0.70480	0.86544	0.87373
No of positive values	320	146	137
Maximum value (in.)	2.78	2.79	2.24
No of missing values	32	31	31
Month τ: 08, R=31			
$\bar{Y}_\tau \times 10^{-2}$ (in.)	8.1144	8.1144	8.1144
$S_\tau \times 10^{-1}$ (in.)	2.3899	2.6973	2.8467
$\hat{g}_\tau$	5.8129	5.1149	6.0698
$r_\tau(1)$	0.07741	0.00452	-0.01293
$\hat{p}_o^\tau$	0.69557	0.84498	0.83243
No of positive values	330	173	187
Maximum value (in.)	2.85	3.01	3.67
No of missing values	32	31	31
Month τ: 10, R=31			
$\bar{Y}_\tau \times 10^{-2}$ (in.)	2.0584	2.0547	
$S_\tau \times 10^{-1}$ (in.)	0.8727	1.2283	
$\hat{g}_\tau$	6.8695	8.1569	
$r_\tau(1)$	0.22611	0.03047	
$\hat{p}_o^\tau$	0.88510	0.94713	
No of positive values	128	59	
Maximum value (in.)	1.14	1.74	
No of missing values	2	0	

Table 5.9 Comparison of statistics for historical and PWN disaggregated 5 min. series, for Idaho Springs station. T_a = 60.0 min., T_d=5.0 min., R = 12

	Historic	Exponential Model	Gamma Model
Month τ: 02			
$\tilde{Y}_\tau \times 10^{-3}$ (in.)	0.4874	0.4769	0.4769
$S_\tau \times 10^{-3}$ (in.)	2.5347	7.1029	7.1354
$\hat{g}_\tau$	5.95432	23.9832	23.9956
$r_\tau(1)$	0.58320	-0.00256	0.00180
$\hat{p}_o^\tau$	0.95951	0.98707	0.98699
No of positive values	1069	349	351
Maximum value (in.) 0	0.03	0.32	0.32
No of missing values	13917	13332	13332
Month τ: 07			
$\tilde{Y}_\tau \times 10^{-3}$ (in.)	0.2498	0.2439	
$S_\tau \times 10^{-3}$ (in.)	3.3656	7.0335	
$\hat{g}_\tau$	27.2398	65.0148	
$r_\tau(1)$	0.65573	-0.00120	
$\hat{p}_o^\tau$	0.98765	0.99544	
No of positive values	524	198	
Maximum value (in.)	0.19	0.81	
No of missing values	2198	1176	
Month τ: 09			
$\tilde{Y}_\tau \times 10^{-3}$ (in.)	0.2406	0.2349	
$S_\tau \times 10^{-3}$ (in.)	10.6349	16.541	
$\hat{g}_\tau$	108.540	146.34	
$r_\tau(1)$	0.38566	-0.00020	
$\hat{p}_o^\tau$	0.99190	0.99741	
No of positive values	244	80	
Maximum value (in.)	1.49	2.71	
No of missing values	13062	12336	

Table 5.10 Comparison of statistics for historical and PWN disaggregated 5 min. series, for Ward station. T_a = 60.0 min., T_d=5.0 min., R = 12

	Historic	Exponential Model	Gamma Model
Month τ: 04			
$\bar{Y}_\tau \times 10^{-3}$ (in.)	0.4434	0.4350	0.4350
$S_\tau \times 10^{-3}$ (in.)	2.6457	7.2811	7.4109
$\hat{g}_\tau$	8.30228	28.9615	29.2657
$r_\tau(1)$	0.66053	0.00134	0.00608
$\hat{p}_o^\tau$	0.96540	0.99200	0.99200
No of positive values	942	222	222
Maximum value (in.)	0.07	0.44	0.44
No of missing values	15977	15456	15456
Month τ: 06			
$\bar{Y}_\tau \times 10^{-3}$ (in.)	0.3732	0.3553	
$S_\tau \times 10^{-3}$ (in.)	7.1452	10.0740	
$\hat{g}_\tau$	89.7420	74.8515	
$r_\tau(1)$	0.41144	0.00103	
$\hat{p}_o^\tau$	0.98023	0.99390	
No of positive values	586	190	
Maximum value (in.)	0.94	1.21	
No of missing values	13562	12072	
Month τ: 08			
$\bar{Y}_\tau \times 10^{-3}$ (in.)	0.3497	0.3368	0.3368
$S_\tau \times 10^{-3}$ (in.)	4.0986	7.8869	7.9770
$\hat{g}_\tau$	29.1518	41.2291	41.2024
$r_\tau(1)$	0.60646	-0.00055	-0.00163
$\hat{p}_o^\tau$	0.98251	0.99383	0.99383
No of positive values	532	195	195
Maximum value (in.)	0.31	0.58	0.58
No of missing values	14215	13044	13044

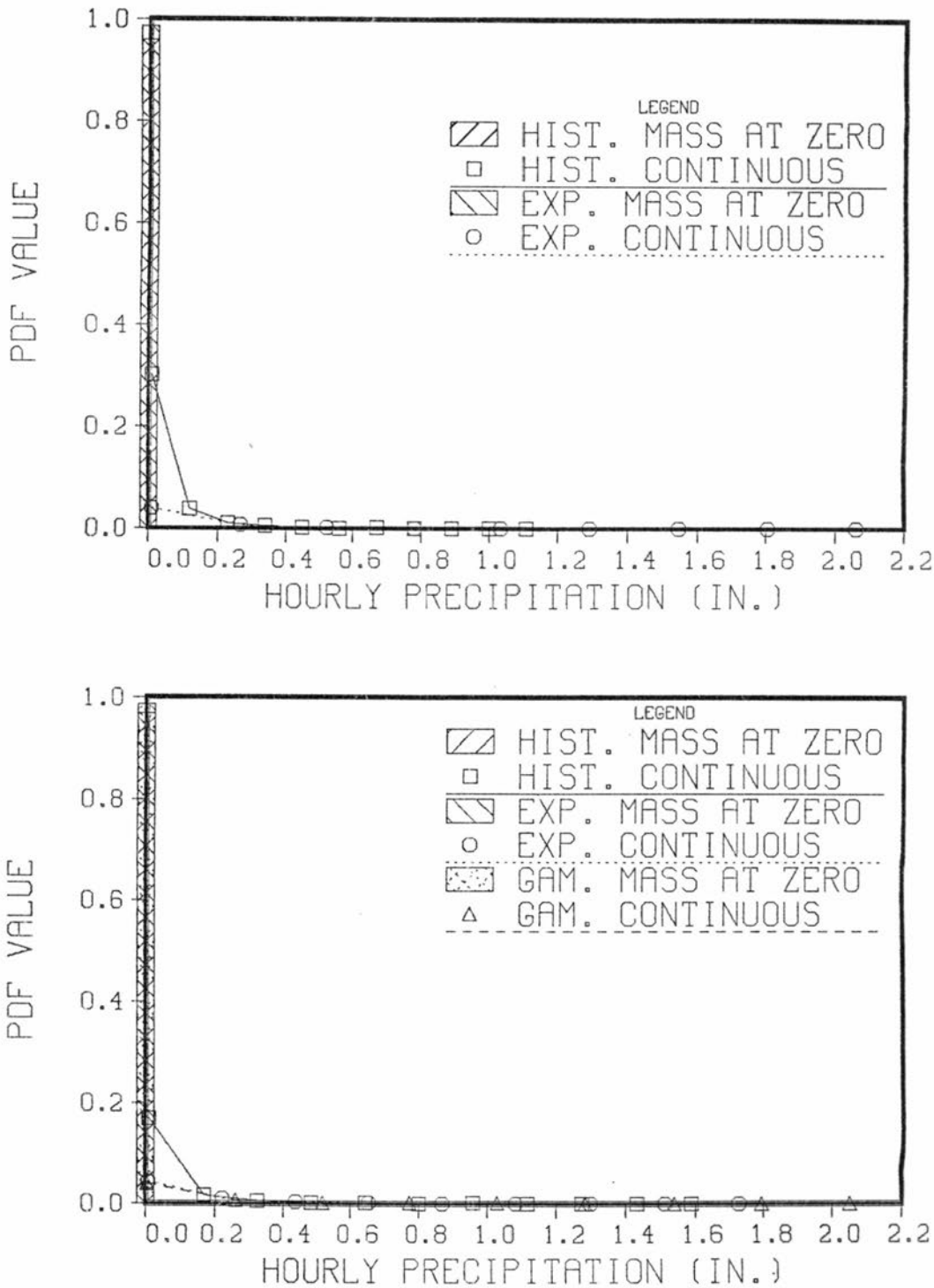
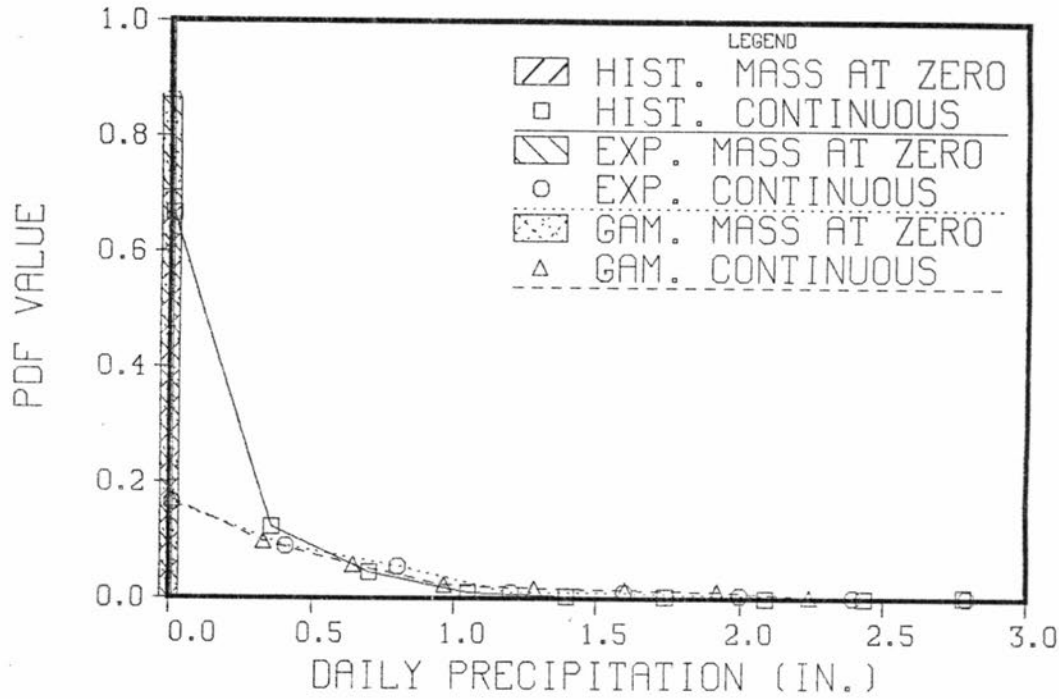


Figure 5.1 Sample probability distribution functions for hourly historical and PWN disaggregated precipitation series, for Denver Wsfo Ap station, for months 06 and 07.



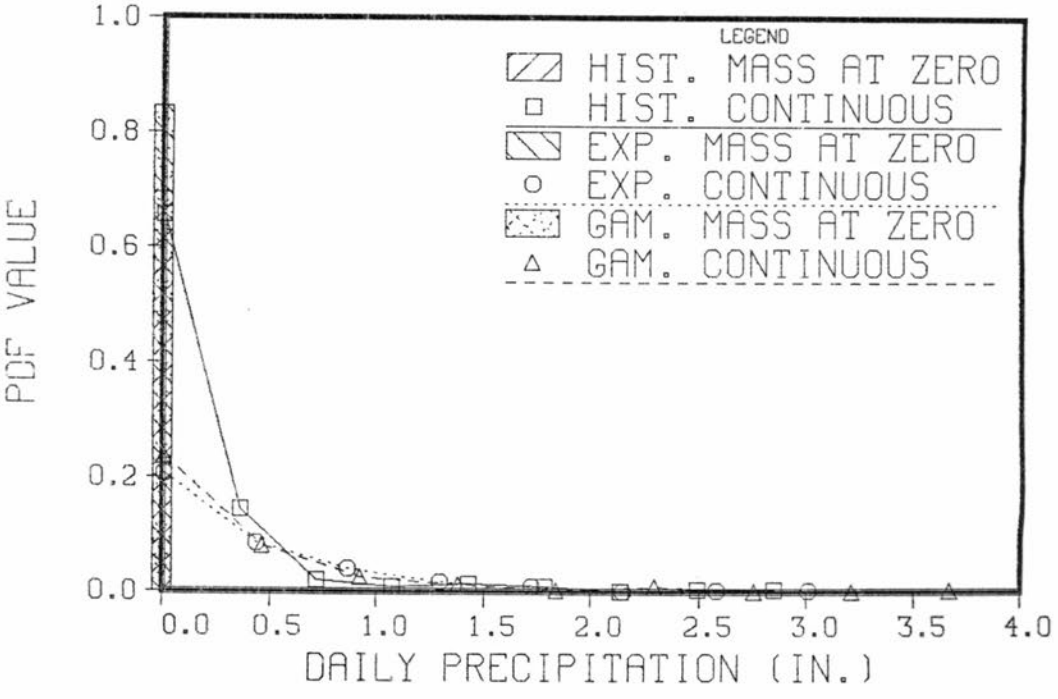


Figure 5.2 Sample probability distribution functions for daily historical and PWN disaggregated precipitation series, for Greenland 9 SE station, for months 05 and 08.

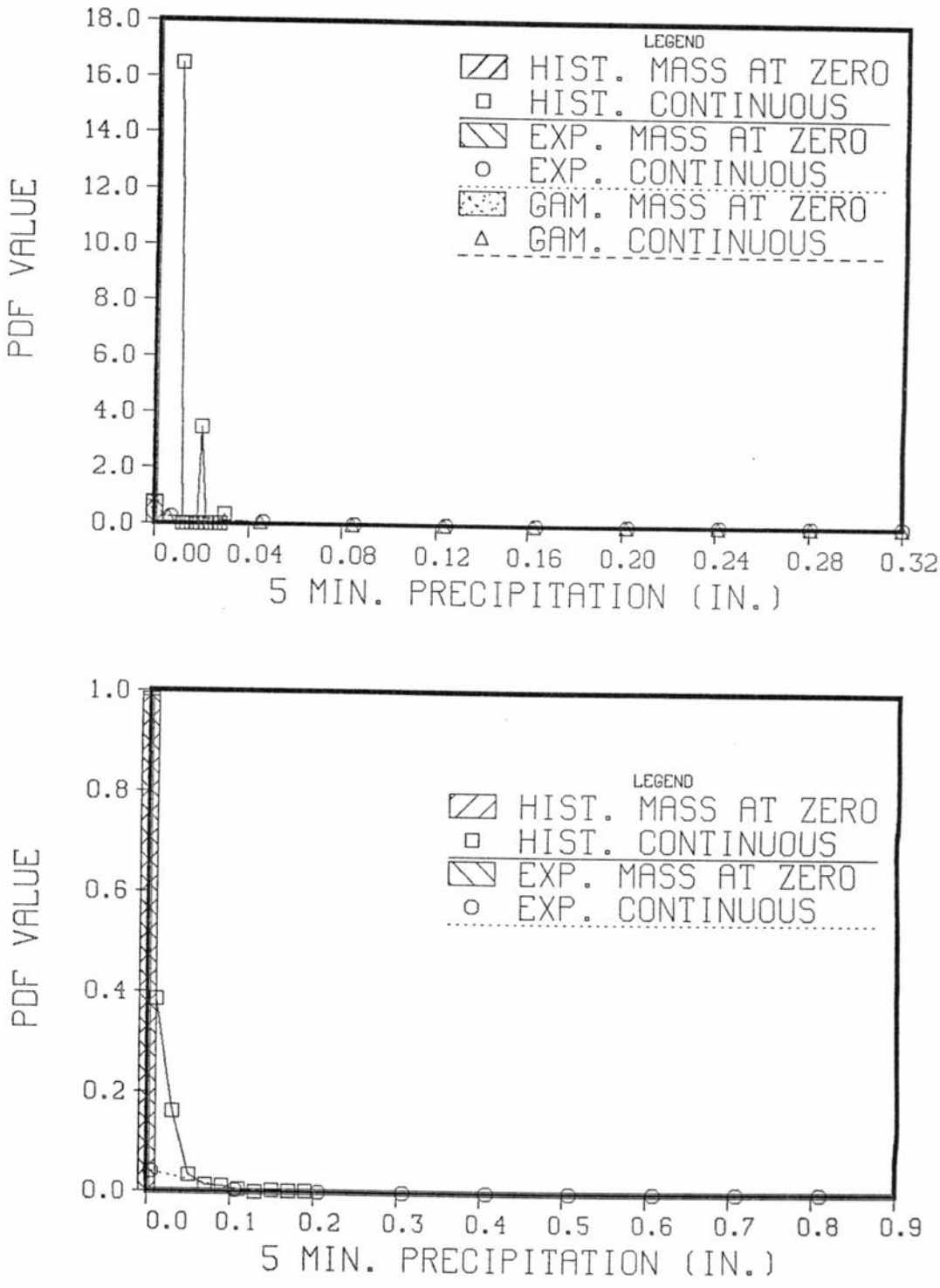


Figure 5.3 Sample probability distribution functions for 5 min. historical and PWN disaggregated precipitation series, for Idaho Springs station, for months 02 and 07.

As a last point, no benefit is obtained when the gamma model is used, in comparison to the exponential case, at least in regard to precipitation records. Furthermore, introduction of the gamma model seems to make preservation worse as compared to the exponential case.

Two reasons are found explaining the bad performance of the PWND model. First, as sustained by many authors, the PWN model does not reproduce precipitation generating mechanisms, for the range of temporal scales considered here. For example, Poisson processes do not represent adequately the arrival in time of precipitation events. They lack the property of clustering, responsible for the type of correlation observed in precipitation records. Also, the use of White Noise terms does not appear realistic in regard to storm activity.

The second reason, already discussed in Chapter 4, is the lack of consistency for parameters estimated at different aggregation scales. This violates the condition of preservation, or uniqueness, of the stochastic process controlling precipitation formation.

Parameter values estimated at the aggregation scale cause the distribution of the number of arrivals to be mostly concentrated in one or two arrivals. Therefore, when the PWND model is operated, a small number of arrivals in the interval is simulated. If the operation of the model were done with parameters similar to those obtained at the disaggregation scale, a larger number of occurrences would be generated and the model would be in a better position to reproduce correlation. However, a serious shortcoming still remains in the uniform distribution for the arrival times.

Additional research could have been attempted in order to improve parameter estimation, like maximum likelihood or use of different

statistics. However, this was not considered worth trying given the lack of correlation in the model.

The PWND model could be used to disaggregate monthly precipitation series to daily or weekly series, given that low correlation is expected at these temporal scales. However, as shown before, the model does not preserve other statistics and for this reason this option does not appear feasible.

The evidence presented renders the PWN and PWND models doubtful for their application to precipitation records. However, three points validate the development and presentation of the models. First, given that the PWN model is easy to handle, it allowed the general formulation of the disaggregation procedure to be applied further in this research. Second, some of the theoretical results obtained for the PWND model are applicable to any model having White Noise terms. Finally, it has been shown that given that the underlying process belongs to the Poissonian family, the models perform excellently. They can be applied in hydrology other areas of science, for processes which evolve in time or space with independent increments.

Chapter 6

NEYMAN-SCOTT WHITE NOISE MODEL FOR PRECIPITATION

6.1 Introduction

The outline of this chapter is, in a certain extent, similar to Chapter 4, where the PWN model for precipitation was presented. The Neyman-Scott White Noise (NSWN) model for precipitation is described and some of its main properties are presented. However, a larger degree of difficulty is found in the analytical treatment of the NSWN model, and some approximations are needed. Two different methodologies are used to estimate model parameters, both of them based on method of moments, although one of them is an extension of approaches presented by other authors.

Most of the properties related to the NSWN precipitation process are presented as results. Detailed derivations for some of them are given in Cadavid et al. (1991).

6.2 Model description

In the past, several authors have used Neyman-Scott (NS) type processes to model occurrence of precipitation in time. The first one was Le Cam (1961), followed by Kavvas and Delleur (1975). In the 80's, this type of models have been analyzed in more depth. The works by Waymire and Gupta (1981a, 1981b, 1981c), Rodriguez-Iturbe et al. (1984), Foufoula-Georgiou and Guttorp (1986) and Obeysekera et al.

(1987) are cited among others. For a more detailed account of the literature available in NS processes the reader is referred to Chapter 2.

It is important to emphasize that NS processes are used to model the arrival of precipitation in time. They belong to a larger class of processes, known as cluster processes or Moyal processes (Ramirez and Bras, 1982). Internal characteristics of precipitation, like cell activity, can be modeled in any suitable way, like for example using White Noise terms or rectangular pulses. In fact, these two approaches are the most popular in the literature. This chapter deals exclusively with Neyman-Scott White Noise (NSWN) models for precipitation.

In the evolution of the NS cluster process along time, a primary or first level process triggers the arrival of cluster centers. Given that a cluster has arrived, a second level process generates the cluster size or number of cluster members and their location with respect to the cluster center.

Most of the NS cluster processes related in the literature rest on two basic assumptions. First, they are stationary in the sense that the process of cluster centers is stationary and the clusters, relative to their center, are independent and identically distributed. The second hypothesis states that cluster centers arrive according to a Poisson process with constant intensity. For this reason they are also known as stationary Poisson cluster processes. Examples of these are the NS process and the Bartlett-Lewis process (Cox and Isham, 1980).

When the NSWN process is used as a model for precipitation, arrival of cluster centers correspond to the arrival of precipitation

generating mechanisms. The occurrence of a storm, associated with a cluster center, is the second level of the process and is formed by a given number of precipitation bursts (cluster size), which are located to the right of the cluster center according to a common distance distribution. Each precipitation burst carries a random amount or volume of instantaneous precipitation, known as White Noise term, independent of the remaining variables in the process.

The assumptions and descriptions given before summarize into the so called structural postulates of the process (Ramirez and Bras, 1982), which also describe completely the NSWN precipitation process (Figure 6.1):

1. Precipitation occurs in clusters in the time domain.
2. Cluster centers arrive in time according to a Poisson process with rate λ. This process is denoted by $N_1(t)$.
3. A cluster arriving at time T_k^* is characterized by the cluster size, denoted by $N_2(.|T_k^*)$, and by the time of occurrence of each cluster members with respect to the cluster center $T_{k,j}$, $j=1,..,N_2(.|T_k^*)$.
4. Cluster sizes are mutually independent and identically distributed and also independent from other variables in the process, following a common distribution $f_{N_2}(m)$.
5. For any cluster, the times of occurrence of events within the cluster are independent and identically distributed according to a common distribution $f_T(t)$. The same assumption applies to the White Noise terms associated with each event, denoted by $U_{k,j}$, $j=1,..,N_2(.|T_k^*)$, following a common distribution $f_U(u)$.

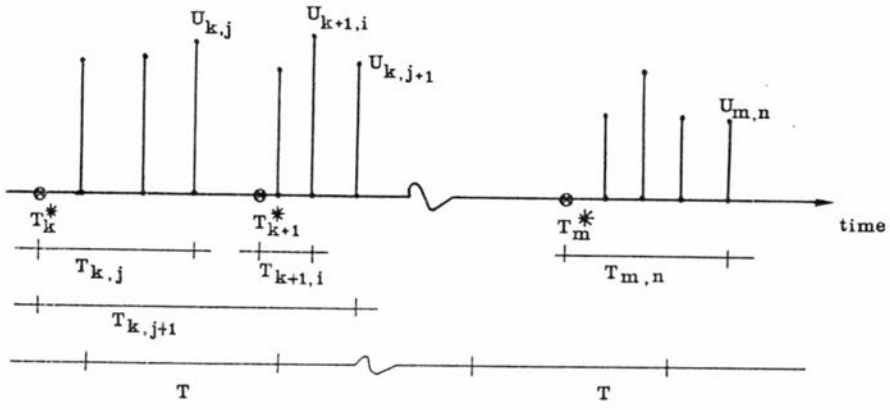


Figure 6.1 Schematic representation of the Neyman-Scott White Noise precipitation process.

In order to keep the assumption of stationarity, none of the distributions mentioned above is indexed by the arrival times of the primary process T_k^*. Also, note in Figure 6.1, the possibility of cluster overlap. Finally, as before for the PWN model, White Noise terms are not limited to follow a Normal distribution.

The instantaneous precipitation intensity, X(t), is given by

$$X(t)\ dt = U(t)\ dN(t) \tag{6.1}$$

where dN(t) represents the increments in the NS counting process. The NS counting process, for a given subset I of the real line, is defined as

$$N(I) = \int N_2(I|t)\ dN_1(t) \tag{6.2}$$

or

$$N(I) = \sum_{\substack{T_n^* \text{ element of } N_1(.)}} N_2(I|T_n^*) \tag{6.3}$$

where $N_2(I|T_n^*)$ is the number of cluster members with center at T_n^* falling in the interval I.

The interval from (0,t) is a subset of the real line. Strictly speaking, the number of occurrences in this interval should be written as N((0,t)). However, the notation N(t) will be adopted here, as far as the lower limit in the interval is zero.

Orderliness is an important concept associated to the theory of point processes (Cox and Isham, 1980). A counting process N(t) is said to be orderly if for a small δ

$$P[N(t,t+\delta) > 1] = O(\delta) \tag{6.4}$$

where $O(\delta)$ is a magnitude such that

$$\lim_{\delta \to 0^+} \frac{O(\delta)}{\delta} = 0$$

The condition given in (6.4) practically excludes the possibility of multiple simultaneous occurrences, i.e., more than one point located in the same time instant. However, a process without multiple simultaneous occurrences is strictly defined by the condition

$$P[N(\{t\}) > 1] = 0 \tag{6.5}$$

where {t} is a set consisting of a singleton in the real line.

Before further clarification is given for (6.4) and (6.5), it is important to define when a process is simply stationary (Cox and Isham, 1980). Given an arbitrary interval I, the counting process N(.) is simply stationary if the distribution of N(I) is invariant under translation of the interval I. For a simply stationary process, the intensity is defined as

$$\rho = \lim_{\delta \to 0^+} \frac{E[N(\delta)]}{\delta} \tag{6.6}$$

so that the mean E[N(I)], if finite, is proportional to the Lebesgue measure $|I|$ (length, area, volume) of I

$$E[N(I)] = \rho|I| \tag{6.7}$$

For example, the Poisson process discussed in Chapter 4 is simply stationary and its intensity is precisely λ. In eqs. (6.6) and (6.7), any dependence of ρ on time is redundant, since these apply to simply stationary processes.

Another magnitude associated to orderliness is the occurrence parameter (Cox and Isham, 1980), defined as

$$\nu = \lim_{\delta \to 0^+} \frac{P[N(\delta) > 0]}{\delta} \tag{6.8}$$

The intensity of the processes, as defined in (6.6), applies to any kind of process, not necessarily orderly, since it includes multiple simultaneous occurrences. On the other hand, the occurrence parameter does not consider multiplicities, since it represents the rate or intensity of the process of instants at which points occur (Cox and Isham, 1980).

An important theorem, known as Korolyouk's theorem (Cox and Isham, 1980), is presented here as a result. It states that $\nu \leq \rho$ with $\nu = \rho$ only when the process has no multiple occurrences. Similarly, Dobrushin's lemma (Cox and Isham, 1980), states that if a process is simply stationary with finite intensity ρ, then orderliness is equivalent to no multiple occurrences and ρ can be computed, in addition to eq. (6.8), by one of the following ways:

$$\rho = \nu = E[N(1)]$$

$$\rho = \nu = \lim_{\delta \to 0^+} \frac{P[N(\delta) = 1]}{\delta} \tag{6.9}$$

For the NS counting process the property of orderliness is evident, given that it evolves on a continuous space (time) and the distance distribution for the cluster members, $f_T(t)$, corresponds to a continuous random variable. Also, as it will be clear later in the text, the NS process is simply stationary, has finite intensity ρ and therefore has no multiple occurrences. Under these assumptions (Cox and Isham, 1980), the process is viewed as one where the random variable $dN(t)$ or $N(t,t+\delta)$ takes values zero or one with probabilities:

$$\begin{aligned} P[N(t,t+\delta)=1] &= \rho\delta + O(\delta) \\ P[N(t,t+\delta)>1] &= O(\delta) \\ P[N(t,t+\delta)=0] &= 1 - \rho\delta + O(\delta) \end{aligned} \tag{6.10}$$

It can be shown that the intensity for the NS process is given by

$$\rho = \lambda\, E[N_2(.)] \tag{6.11}$$

where the conditioning of the secondary process on the arrival times of the primary process has been removed, due to the process being stationary and clusters sizes being independently and identically distributed. Note, based also on the stationarity of the process, that the sequence of times of cluster members, with respect to the cluster center, can be denoted as T_j, $j=1,..,N_2(.)$. The same consideration applies to the White Noise terms.

As stated before, the precipitation intensity process is never measured. Instead, cumulative amounts of precipitation over disjoint time intervals are recorded and is over this sample that estimation is performed. Given that the NS counting process is stationary, the cumulative amount of precipitation in any interval of length T is given by

$$Y_k = \sum_{i=0}^{N(T)} U_i \qquad (6.12)$$

where N(T) represents the NS counting process in the interval (0,T). The reasoning for eq. (6.12) is similar to the one presented for the PWN model. In order to derive properties for Y_k, properties for the random variable N(T) must be derived first. This topic will be covered in the next sections.

Before proceeding with the derivation of properties for N(T), it is important to bring into attention the fact that different distributions for the cluster size $N_2(.)$ are used in the literature. For example, Kavvas and Delleur (1975, 1981) and Ramirez and Bras (1982, 1985) assume a geometric distribution, while Rodriguez-Iturbe et al. (1984) use the Poisson distribution. Foufoula-Georgiou and Guttorp (1986) state that the selection of this distribution has been based on mathematical considerations, although it drastically affects estimated parameters for the model and derived properties for the precipitation process.

6.3 Probability generating functional and probability generating function

The probability generating functional (pgfl) for a random measure is seen by many authors (Cox and Isham, 1980; Waymire and Gupta, 1981b) as an extension of the concept of probability generating function (pgf) for a random vector.

Assume a space H where a point or counting process N(.) is evolving. The pgfl for the point process, denoted as $G[\xi]$, is defined as

$$G[\xi] = E\left[\exp\left(\int_H Ln[\xi(t)]\, dN(t) \right) \right] \tag{6.13}$$

for arbitrary real-valued functions $\xi(t)$ defined on H, satisfying

$$0 \le \xi(t) < 1, \text{ for all t element of H}$$
$$\xi(t) = 1 \text{ , for all t not an element of H} \tag{6.14}$$

Note that the argument for the pgfl is the function $\xi(.)$. A complete justification of the form of the pgfl is given in Cox and Isham (1980) and Waymire and Gupta (1981b). Cadavid et al. (1991) reproduce that justification.

The pgfl for any point process can be expressed also as (Cox and Isham, 1980; Waymire and Gupta, 1981b)

$$G[\xi] = E\left[\prod_{T_i \text{ element of } N(.)} \xi(T_i) \right] \tag{6.15}$$

where T_i, in this case, denotes arrival times for the counting process N(.). The point process in (6.15) admits multiple simultaneous occurrences at a point T_i.

For example, for an homogeneous Poisson process with rate λ, the pgfl is given by (Cox and Isham, 1980; Waymire and Gupta, 1981b)

$$G[\xi] = \exp\left\{ -\lambda \int_H [1 - \xi(t)]\, dt \right\} \tag{6.16}$$

A very simple proof of this result is found in the references given above.

The pgfl for the a cluster process in two levels, not necessarily the NS process, is given by (Waymire and Gupta, 1981b)

$$G[\xi] = G_1[G_2[\xi]] \tag{6.17}$$

where $G_2[\xi]$ is the pgfl for a cluster and $G_1[\ ]$ is the pgfl for the primary process.

Since in the NS process cluster centers follow a Poisson process with rate λ, eq. (6.16) in (6.17) yields

$$G[\xi] = \exp\left\{ - \lambda \int_{-\infty}^{+\infty} \left[1 - G_2[\xi(t)] \right] dt \right\} \qquad (6.18)$$

where the space H has been replaced by the real line. According to eq. (6.18), clusters arriving at any time can generate members located to the left of the cluster center.

The pgfl for a given cluster having center in t, t as specified in eq. (6.18), is

$$G_2[\xi(t)] = E\left[\exp\left(\int_{-\infty}^{+\infty} Ln[\xi(v)] \, dN_2(v|t) \right) \right]$$

It can be shown that this expression is equivalent to (Cadavid et al., 1991)

$$G_2[\xi(t)] = E\left[\prod_{i=1}^{N_2(.)} \xi(t+T_i) \right]$$

The conditionality of the cluster size on the arrival of cluster centers has been removed, since they are independent. Note the appearance of $t+T_i$, the occurrence time for a cluster member measured from $-\infty$.

Let $f_{N_2}(m) = g_m$ be the probability distribution function for the cluster size. Then

$$G_2[\xi(t)] = \sum_{m=1}^{\infty} g_m \, E\left[\prod_{i=1}^{m} \xi(t+T_i) \right]$$

where m has been constrained to take values greater or equal than 1. Since the times T_i are independently and identically distributed as $f_T(t)$, it follows from the previous equation

$$G_2[\xi(t)] = \sum_{m=1}^{\infty} g_m \prod_{i=1}^{m} E\left[\xi(t+T_i) \right]$$

$$G_2[\xi(t)] = \sum_{m=1}^{\infty} g_m \left(\int_{-\infty}^{+\infty} \xi(t+v)\ f_T(v)\ dv \right)^m$$

and finally

$$G_2[\xi(t)] = g_{N_2} \left(\int_{-\infty}^{+\infty} \xi(t+v)\ f_T(v)\ dv \right) \tag{6.19}$$

where $g_{N_2}(.)$ is the probability generating function (pgf) for the cluster size distribution.

Replacing (6.19) in eq. (6.18), the pgfl for the NS process becomes

$$G[\xi]=\exp \left\{ - \lambda \int_{-\infty}^{+\infty} \left[1 - g_{N_2} \left(\int_{-\infty}^{+\infty} \xi(t+v)\ f_T(v)\ dv \right) \right] dt \right\} \tag{6.20}$$

A less elaborated proof for this result is presented by Ramirez and Bras (1982).

Let $N(t_1)$ be the NS counting process for an interval $(0,t_1)$. The objective is to obtain the probability generating function (pgf) for $N(t_1)$, $g(z)$, from the pgfl for the NS process given in eq. (6.20), which is defined as

$$g(z) = \sum_{n=0}^{\infty} z^n\ f_{N(t_1)}(n)$$

with $|z| \leq 1$.

To obtain $g(z)$, take $\xi(.)$ to be of the form (Ramirez and Bras, 1982)

$$\xi(t+v) = 1 - (1 - z)\ I_{(0,t_1)}(t+v) \tag{6.21}$$

Under this selection, $G[\xi]$ becomes $g(z)$, with $|z| \leq 1$. It is easy to verify that $\xi(t+v)$, as defined in (6.21), takes the following values

$$\xi(t+v) = z \text{ for } 0 \leq t+v \leq t_1 \text{ or } -t \leq v \leq t_1 - t$$

$$\xi(t+v) = 1 \text{ for } -\infty < t+v < 0 \text{ or } -\infty < v < -t \tag{6.22}$$

$$\xi(t+v) = 1 \text{ for } t_1 < t+v < \infty \text{ or } t_1 - t < v < \infty$$

and therefore, the most inner integral in eq. (6.20) becomes

$$\int_{-\infty}^{+\infty} \xi(t+v)\ f_T(v)\ dv = 1 - (1-z)p_1(t)$$

with

$$p_1(t) = \int_0^{t_1} f_T(x-t)dx \tag{6.23}$$

where a new variable of integration, x=t+v, has been introduced. The term $p_1(t)$ is the probability that a cluster with center in t will have a member falling in the interval $(0,t_1)$.

Finally, the pgf of $N(t_1)$ is obtained as (Ramirez and Bras, 1982)

$$g(z)=\exp\left\{ - \lambda \int_{-\infty}^{t_1} \left[1 - g_{N_2} \left(1 - (1-z)p_1(t) \right) \right] dt \right\} \tag{6.24}$$

The upper limit in the integral was changed to t_1, since the evolution of the process is being considered up to that time.

Following a similar development and choosing the function $\xi(t+v)$ to be of the form

$$\xi(t+v) = 1 - \sum_{i=1}^{k} (1 - z_i)\ I_{(t_{i-1},t_i]}(t+v) \tag{6.25}$$

it can be shown that the multivariate pgf for the random vector $[N(0,t_1),N(t_1,t_2),..,N(t_{k-1},t_k)]$, specifying the counting process in disjoint intervals $(0,t_1),(t_1,t_2),..,(t_{k-1},t_k)$, is of the form

$$g(z_1,z_2,..,z_k)=\exp\left\{ - \lambda \int_{-\infty}^{t_1} \left[1 - g_{N_2}\left[1 - \sum_{i=1}^{k} (1-z_i)p_i(t)\right]\right]dt \right\} \tag{6.26}$$

with $t_0=0$ and $p_i(t)$ defined as

$$p_i(t) = \int_{t_{i-1}}^{t_i} f_T(x-t)dx \tag{6.27}$$

Results in eqs. (6.25) to (6.27) were taken from Ramirez and Bras (1982).

As mentioned previously, two different distributions have been used in the literature for the cluster size. Geometric and Poisson. The NS process with geometric distribution for the cluster size will be presented here, since it is the one used in this study. It was not possible to find an analytical closed solution for the pgf for the Poisson case.

At this point, the distribution for the cluster members location, with respect to cluster centers, is introduced as the exponential distribution with parameter α ($\alpha>0$)

$$f_T(t) = \alpha \exp(-\alpha t)\, I_{(0,\infty)}(t) \tag{6.28}$$

This distribution does not allow points falling to the left of the cluster center.

From (6.28), the term $p_1(t)$, defined in (6.23), is given by (Cadavid et al., 1991)

$$p_1(t) = \phi(t_1)\exp(\alpha t) I_{(-\infty,0)}(t) + \phi(t_1-t) I_{(0,t_1)}(t) \tag{6.29}$$

with $\phi(t)$ defined in general as

$$\phi(t) = 1-\exp(-\alpha t) \tag{6.30}$$

In words, $\phi(t)$ represents the cdf for the exponential distribution with parameter α.

Similarly, it is possible to show that $p_i(t)$, defined in (6.27) for $i\geq 2$, is of the form (Cadavid et al., 1991)

$$p_i(t) = [\exp(-\alpha t_{i-1})-\exp(-\alpha t_i)]\exp(\alpha t)\Big[I_{(-\infty,t_0)}(t) + \sum_{k=1}^{i-1} I_{(t_{k-1},t_k)}(t) \Big] + \phi(t_i-t)\, I_{(t_{i-1},t_i)}(t) \tag{6.31}$$

The geometric distribution with parameter p, for the cluster size, is written as (Parzen, 1964)

$$f_{N_2}(m) = p(1-p)^m \, I_{\{1,2,..\}}(m) \tag{6.32}$$

and its pgf is

$$g_{N_2}(y) = \frac{1}{1 - qy}, \quad |y| \leq 1, \quad q=1-p \tag{6.33}$$

Using (6.33) in (6.24) (Cadavid et al., 1991)

$$g(z) = \exp\left(\frac{-\lambda(1-z)t_1}{p+q(1-z)} \right) \left[\frac{p}{p+q(1-z)\phi(t_1)} \right]^{\frac{\lambda(1-z)}{\alpha[p+q(1-z)]}} \tag{6.34}$$

The derivation for the following result in Cadavid et al. (1991), for the pgf of the random vector $[N(t_0,t_1),N(t_1,t_2)]$, where z_1 is for the first interval and z_2 for the second, with $|z_1| \leq 1$ and $|z_2| \leq 1$:

$$g(z_1,z_2) = \exp\Bigg\{$$

$$\frac{\lambda}{\alpha q} \operatorname{Ln} \frac{p}{p+q\{(1-z_1)+(z_1-z_2)\exp[-\alpha(t_1-t_0)]-(1-z_2)\exp[-\alpha(t_2-t_0)]\}}$$

$$- \frac{\lambda(1-z_1)(t_1-t_0)}{p+q(1-z_1)} - \frac{\lambda(1-z_2)(t_2-t_1)}{p+q(1-z_2)} - \frac{\lambda p}{\alpha q[p+q(1-z_1)]}$$

$$\operatorname{Ln} \frac{p+q(1-z_1)+q\{(z_1-z_2)-(1-z_2)\exp[-\alpha(t_2-t_1)]\}}{p+q\{(1-z_1)+(z_1-z_2)\exp[-\alpha(t_1-t_0)]-(1-z_2)\exp[-\alpha\,(t_2-t_0)]\}}$$

$$- \frac{\lambda p}{\alpha q[p+q(1-z_2)]} \operatorname{Ln} \frac{p}{p+q(1-z_2)\{1-\exp[-\alpha(t_2-t_1)]\}} \Bigg\} \tag{6.35}$$

It is possible to find a general form for the pgf of the random vector $[N(0,t_1),N(t_1,t_2),..,N(t_{k-1},t_k)]$ (Cadavid et al., 1991), with the components in $\underline{z} = [z_1,z_2,..,z_k]$ $|z_i| \leq 1$, i=1,...,k, associated to each one of the intervals:

$$\theta_{i,j} = q\{\exp[-\alpha(t_{i-1}-t_j)]-\exp[-\alpha(t_i-t_j)]\} \tag{6.36}$$

$$\beta_i = q \exp[-\alpha(t_i-t_{i-1})]$$

$$g(\underline{z}) = \exp\Bigg\{ \lambda \Bigg[\frac{1}{\alpha\, q} \mathrm{Ln}(p) - \frac{1}{\alpha\, q} \mathrm{Ln}\Big(p + \sum_{i=1}^{k} \theta_{i,0}(1-z_i) \Big) +$$

$$\frac{p}{\alpha\, q} \sum_{i=1}^{k-1} \frac{1}{1-qz_i} \mathrm{Ln} \frac{1-qz_i-\beta_i(1-z_i)+\sum_{j=i+1}^{k} \theta_{j,i-1}(1-z_j)}{p+\sum_{j=i+1}^{k} \theta_{j,i}(1-z_j)} +$$

$$\frac{p}{\alpha\, q} \frac{1}{1-qz_k} \mathrm{Ln} \frac{1-qz_k-\beta_k(1-z_k)}{p} - \sum_{i=1}^{k} \frac{(1-z_i)(t_i-t_{i-1})}{1-qz_i} \Bigg] \Bigg\} \qquad (6.37)$$

with $t_0=0$.

6.4 Properties of the NS point process

The role of the probability generating function in regard to the NS point process and other point processes is evident from the expressions given below. First, the second order properties for the process are (Ramirez and Bras, 1982):

$$E[N(t_1)] = \frac{\partial g(z)}{\partial z} \Bigg|_{z=1} \qquad (6.38)$$

$$\mathrm{Var}(N(t_1)) = \frac{\partial g(z)}{\partial z} + \frac{\partial^2 g(z)}{\partial z^2} - \left[\frac{\partial g(z)}{\partial z} \right]^2 \Bigg|_{z=1} \qquad (6.39)$$

$$\mathrm{Cov}[N(0,t_1),N(t_{k-1},t_k)] = \frac{\partial^2 g(z_1,z_k)}{\partial z_1 \partial z_k} \Bigg|_{z_1=z_k=1}$$

$$- \left[\frac{\partial g(z_1,z_k)}{\partial z_1} \Bigg|_{z_1=z_k=1} \frac{\partial g(z_1,z_k)}{\partial z_k} \Bigg|_{z_1=z_k=1} \right] \qquad (6.40)$$

Equation (6.40) gives the covariance for the random variables counting arrivals in the intervals $(0,t_1)$ and (t_{k-1},t_k).

The factorial moment of order r for $N(t_1)$, $r \geq 0$ and r integer, is defined as (Mood et al., 1974)

$$M^*_{[r]} = E\left[\prod_{j=0}^{r} [N(t_1)-j] \right] \tag{6.41}$$

and they are related to the pgf g(z) by

$$M^*_{[r]} = \frac{d^{r+1} g(z)}{dz^{r+1}} \Bigg|_{z=1} \tag{6.42}$$

This result can be derived from the basic definition of the pgf g(z). Equation (6.38) is a particular case of (6.42) for r=0.

The finite dimensional pdf for the counting process is (Ramirez and Bras, 1982)

$$P[N(t_0,t_1)=j_1, N(t_1,t_2)=j_2, .., N(t_{k-1},t_k)=j_k] =$$

$$\prod_{i=1}^{k} \frac{1}{j_i!} \left(\frac{\partial^{j_1+j_2+..+j_k} g(\underline{z})}{\partial^{j_1} z_1 \partial^{j_2} z_2 .. \partial^{j_k} z_k} \Bigg|_{z_1=z_2=..=z_k=0} \right) \tag{6.43}$$

and for the univariate case

$$f_{N(t_1)}(n) = \frac{1}{n!} \frac{d^n g(z)}{dz^n} \Bigg|_{z=0} \tag{6.44}$$

The use of eqs. (6.34) and (6.35) in (6.38) to (6.40) would produce results for the second order properties of the NS process, when the cluster size distribution is geometric. However, obtaining the derivatives of eqs. (6.24) and (6.26), the last for k=2, will yield more general results, since in those expressions no particular distributions have been assumed. The reader can review the details of the derivations in Ramirez and Bras (1982):

$$E[N(t_1)] = \lambda \, E[N_2(.)] \, t_1 \tag{6.45}$$

$$\mathrm{Var}(N(t_1)) = \lambda \left\{ E[N_2^2(.)-N_2(.)] \int_{-\infty}^{t_1} p_1(t)^2 dt + E[N_2(.)] t_1 \right\} \quad (6.46)$$

$$\mathrm{Cov}[N(0,t_1), N(t_{k-1}, t_k)] = \lambda\, E[N_2^2(.)-N_2(.)] \int_{-\infty}^{t_k} p_1(t) p_k(t) dt \quad (6.47)$$

Introducing in eqs. (6.46) and (6.47) the exponential distribution with parameter α and using eqs. (6.29) and (6.31) for $p_1(t)$ and $p_k(t)$, the following set of expressions are obtained (Cadavid et al., 1991)

$$\mathrm{Var}(N(t_1)) = \lambda E[N_2^2(.)] t_1 - \frac{\lambda}{\alpha} E[N_2^2(.)-N_2(.)] \phi(t_1) \quad (6.48)$$

$$\mathrm{Cov}[N(0,t_1), N(t_{k-1}, t_k)] = \frac{\lambda}{2\alpha} E[N_2^2(.)-N_2(.)] \phi(t_1)$$

$$\{\exp[-\alpha(t_{k-1}-t_1)] - \exp[-\alpha(t_k-t_1)]\} \quad (6.49)$$

When the geometric distribution with parameter p is introduced (Parzen, 1964)

$$E[N_2(.)] = 1/p$$

$$E[N_2^2(.)] = (2 - p)/p^2 \quad (6.50)$$

eqs. (6.45), (6.48) and (6.49) transform into the following second order properties for the NS process with geometric cluster sizes (Cadavid et al. ,1991):

$$E[N(t_1)] = \lambda t_1 / p \quad (6.51)$$

$$\mathrm{Var}(N(t_1)) = \frac{\lambda(2-p)t_1}{p^2} - \frac{2\lambda q}{\alpha p^2} \phi(t_1) \quad (6.52)$$

$$\text{Cov}[N(0,t_1),N(t_{k-1},t_k)] = [(\lambda q)/(\alpha p^2)]\ \phi(t_1)$$

$$\{\exp[-\alpha(t_{k-1}-t_1)]-\exp[-\alpha(t_k-t_1)]\} \qquad (6.53)$$

6.5 Marginal probability distribution function for $N(t_1)$

The marginal probability distribution function for the number of occurrences in the interval $(0,t_1)$ is, at least theoretically, defined by the relationship in (6.44). Its application requires computation of the derivatives of g(z) as given in (6.34). For the determination of this distribution, full distributional assumptions are used, both for cluster size and for cluster members location with respect to the cluster center. As will be shown later, computing derivatives for eq. (6.34) is a formidable task, since no law of formation was obtained for those derivatives. For this reason, no general results, similar to eqs. (6.45) to (6.47), are given in this section.

Equation (6.34) is written in the following way

$$g(z) = \exp\left[\frac{\lambda(1-z)}{\alpha[p+q(1-z)]}\ \text{Ln}\ \frac{p}{p+q(1-z)\phi(t_1)} - \frac{\lambda(1-z)t_1}{p+q(1-z)}\right] \qquad (6.54)$$

Making z=0 in eq. (6.54), one obtains

$$f_{N(t_1)}(0) = \exp(-\lambda t_1)\left[\frac{p}{1-q\ \exp(-\alpha t_1)}\right]^{\lambda/\alpha} \qquad (6.55)$$

Ramirez and Bras (1982) arrive to eq. (6.55) starting from the more general form given in (6.24).

In order to obtain $f_{N(t_1)}(1)$, the first derivative of (6.54) with respect to z must be computed. The following result is obtained directly from eq. (6.54)

$$\frac{dg(z)}{dz} = g(z)\left[\frac{\lambda(1-z)}{\alpha[p+q(1-z)]}\ \frac{q\phi(t_1)}{p+q(1-z)\phi(t_1)} -\right.$$

$$\frac{\lambda p}{\alpha[p+q(1-z)]^2} \text{ Ln } \frac{p}{p+q(1-z)\phi(t_1)} + \frac{\lambda p t_1}{[p+q(1-z)]^2} \Bigg] \tag{6.56}$$

and making z=0 in (6.56)

$$f_{N(t_1)}(1) = \frac{\lambda}{\alpha} \left\{ \exp(-\lambda t_1) \left[\frac{p}{1-q \exp(-\alpha t_1)} \right]^{\lambda/\alpha} \right.$$

$$\left. \left[\frac{q\phi(t_1)}{1-q \exp(-\alpha t_1)} - p \text{ Ln } \frac{p}{1-q \exp(-\alpha t_1)} + \alpha p t_1 \right] \right\} \tag{6.57}$$

To obtain larger order derivatives for g(z), it is convenient to simplify notation in eq. (6.54) by rewriting it as

$$g(z) = \exp[\lambda f(z)] \tag{6.58}$$

with

$$f(z) = \frac{1-z}{\alpha(1-qz)} \text{ Ln } \frac{p}{p+q\phi(t_1)(1-z)} - \frac{(1-z)t_1}{1-qz} \tag{6.59}$$

Note that f(1)=0 and therefore g(1)=1 as expected.

Derivatives of order n for g(z) and f(z), evaluated at z=0, will be denoted as

$$\frac{d^n g(0)}{dz^n} \text{ and } \frac{d^n f(0)}{dz^n}$$

Based on eq. (6.58), derivatives of order up to five for g(z), evaluated at z=0, are

$$g(0) = \exp[\lambda f(0)] \tag{6.60}$$

$$\frac{dg(0)}{dz} = \lambda \, g(0) \, \frac{df(0)}{dz} \tag{6.61}$$

$$\frac{d^2 g(0)}{dz^2} = \lambda \, g(0) \left[\frac{d^2 f(0)}{dz^2} + \lambda \left(\frac{df(0)}{dz} \right)^2 \right] \tag{6.62}$$

$$\frac{d^3g(0)}{dz^3} = \lambda\, g(0) \left[\frac{d^3f(0)}{dz^3} + \lambda \frac{d^2f(0)}{dz^2} \frac{df(0)}{dz} + \lambda^3 \left(\frac{df(0)}{dz} \right)^3 \right] \tag{6.63}$$

$$\frac{d^4g(0)}{dz^4} = \lambda\, g(0) \left\{ \frac{d^4f(0)}{dz^4} + \lambda \left[4 \frac{d^3f(0)}{dz^3} \frac{df(0)}{dz} + 3 \left(\frac{d^2f(0)}{dz^2} \right)^2 \right] \right.$$

$$\left. + 6\,\lambda^2 \frac{d^2f(0)}{dz^2} \left(\frac{df(0)}{dz} \right)^2 + \lambda^3 \left(\frac{df(0)}{dz} \right)^4 \right\} \tag{6.64}$$

$$\frac{d^5g(0)}{dz^5} = \lambda g(0) \left\{ \frac{d^5f(0)}{dz^5} + \lambda \left[5 \frac{d^4f(0)}{dz^4} \frac{df(0)}{dz} + 10 \frac{d^3f(0)}{dz^3} \frac{d^2f(0)}{dz^2} \right] \right.$$

$$+ \lambda^2 \left[10 \frac{d^3f(0)}{dz^3} \left(\frac{df(0)}{dz} \right)^2 + 15 \left(\frac{d^2f(0)}{dz^2} \right)^2 \frac{df(0)}{dz} \right] +$$

$$\left. 10\,\lambda^3 \frac{d^2f(0)}{dz^2} \left(\frac{df(0)}{dz} \right)^3 + \lambda^4 \left(\frac{df(0)}{dz} \right)^5 \right\} \tag{6.65}$$

For this research, derivatives of this type were computed manually up to degree 11. They were checked and extended to 20 using a symbolic mathematics software package. The final result for the nth derivative of g(z), evaluated at zero, for $n \geq 3$, is

$$\frac{d^ng(0)}{dz^n} = \lambda\, g(0) \left\{ \frac{d^nf(0)}{dz^n} + \sum_{i=1}^{n-2} \lambda^i \left[\sum_{j=1}^{m_{n,i}} C_{n,i,j} \left(\prod_{k=1}^{i+1} \frac{d^{\ell_k}f(0)}{dz^{\ell_k}} \right) \right] \right.$$

$$\left. + \lambda^{n-1} \left(\frac{df(0)}{dz} \right)^n \right\} \tag{6.66}$$

The most exterior summation represents powers of λ. Each power of λ is multiplied by a second summation with $m_{n,i}$ terms. Each term is formed by a coefficient $C_{n,i,j}$ and by the product of i+1 derivatives of f(z), evaluated at zero, the order of these derivatives being

represented by ℓ_k. The collection of integer numbers $\{\ell_1,\ell_2,..,\ell_{i+1}\}$ is called here a sequence and has the property

$$\sum_{k=1}^{i+1} \ell_k = n \tag{6.67}$$

The reader can verify this formulation with eqs. (6.63) to (6.65). Values of ℓ_k, for given n and i, can be equal among themselves, generating powers of the derivatives of f(z).

The derivatives of f(z) evaluated at zero were investigated using the same symbolic mathematics package mentioned before. Taylor series expansion of f(z) around z=0 was done and the following law of formation was obtained for n≥2

$$\frac{d^n f(0)}{dz^n} = -\frac{n!q^n}{\alpha}\left[-\frac{p}{q}\left(\mathrm{Ln}\,\frac{p+q\phi(t_1)}{p}+\alpha t_1\right)-\frac{\phi(t_1)}{p+q\phi(t_1)}+\frac{1}{p+q\phi(t_1)}\sum_{i=1}^{n-1}\frac{1}{i}\left(\frac{\phi(t_1)}{p+q\phi(t_1)}\right)^i\left(\frac{p}{q}+\frac{\phi(t_1)}{i+1}\right)\right] \tag{6.68}$$

From eq. (6.59), for n=0,

$$f(0) = \frac{1}{\alpha}\,\mathrm{Ln}\,\frac{1}{p+q\phi(t_1)} - t_1 \tag{6.69}$$

and for n=1

$$\frac{df(0)}{dz} = -\frac{q}{\alpha}\left[-\frac{p}{q}\left(\mathrm{Ln}\,\frac{p+q\phi(t_1)}{p}+\alpha t_1\right)-\frac{\phi(t_1)}{p+q\phi(t_1)}\right] \tag{6.70}$$

In regard to eq. (6.66), it was not possible to formulate laws of formations for $m_{n,i}$, $C_{n,i,j}$ and ℓ_k. To solve the problem, with the mathematical package mentioned before, values were obtained and printed out. They were stored in a computer file, so that eq. (6.66) can be used to produce numerical results n between 3 and 20. For n values from 0 to 2, eqs. (6.60) to (6.62) are used.

Table 6.1 gives values for $m_{n,i}$, for n from 3 to 20, while Table 6.2 gives sequences and coefficients for n from 3 to 11. The amount of data required to compute 20 derivatives for g(z), evaluated at z=0, is evident from both tables. Sequences and coefficients for n from 13 to 20 are given in Cadavid et al. (1991).

Table 6.1 Number of terms $m_{n,i}$ for each power λ^i, to evaluate derivatives of order n for g(z) at z=0, from n=3 to n=20.

n / i:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	Total
3	1																		1
4	2	1																	3
5	2	2	1																5
6	3	3	2	1															9
7	3	4	3	2	1														13
8	4	5	5	3	2	1													20
9	4	7	6	5	3	2	1												28
10	5	8	9	7	5	3	2	1											39
11	5	10	11	10	7	5	3	2	1										54
12	6	12	15	13	11	7	5	3	2	1									75
13	6	14	18	18	14	11	7	5	3	2	1								99
14	7	16	23	23	20	15	11	7	5	3	2	1							133
15	7	19	27	30	26	21	15	11	7	5	3	2	1						174
16	8	21	34	37	35	28	22	15	11	7	5	3	2	1					229
17	8	24	39	47	44	38	29	22	15	11	7	5	3	2	1				295
18	9	27	47	57	58	49	40	30	22	15	11	7	5	3	2	1			383
19	9	30	54	70	71	65	52	41	30	22	15	11	7	5	3	2	1		488
20	10	33	64	84	90	82	70	54	42	30	22	15	11	7	5	3	2	1	625

The procedure to compute $f_{N(t_1)}(n)$, n=0,..,20, given parameters values for λ, p (q=1-p) and α is as follows. Using eqs. (6.68) to (6.70), compute derivatives of f(z) evaluated at z=0. With these values, eqs. (6.60), (6.61), (6.62) and (6.66) will give derivatives for g(z) at z=0. Finally, eq. (6.44) will provide the required probabilities.

Table 6.2 Sequences $\{\ell_k,\ k=1,..,i+1\}$ and coefficients $C_{n,i,j}$ to evaluate derivatives of order n for g(z) at z=0, for n from 3 to 11.

n	i	j	<--Sequence ℓ_k, k=1,..,i+1-->	$C_{n,i,j}$
3	1	1	2 1	3
4	1	1	3 1	4
4	1	2	2 2	3
4	2	1	2 1 1	6
5	1	1	4 1	5
5	1	2	3 2	10
5	2	1	3 1 1	10
5	2	2	2 2 1	15
5	3	1	2 1 1 1	10
6	1	1	5 1	6
6	1	2	4 2	15
6	1	3	3 3	10
6	2	1	4 1 1	15
6	2	2	3 2 1	60
6	2	3	2 2 2	15
6	3	1	3 1 1 1	20
6	3	2	2 2 1 1	45
6	4	1	2 1 1 1 1	15
7	1	1	6 1	7
7	1	2	5 2	21
7	1	3	4 3	35
7	2	1	5 1 1	21
7	2	2	4 2 1	105
7	2	3	3 3 1	70
7	2	4	3 2 2	105
7	3	1	4 1 1 1	35
7	3	2	3 2 1 1	210
7	3	3	2 2 2 1	105
7	4	1	3 1 1 1 1	35
7	4	2	2 2 1 1 1	105
7	5	1	2 1 1 1 1 1	21
8	1	1	7 1	8
8	1	2	6 2	28
8	1	3	5 3	56
8	1	4	4 4	35
8	2	1	6 1 1	28
8	2	2	5 2 1	168
8	2	3	4 3 1	280
8	2	4	4 2 2	210
8	2	5	3 3 2	280
8	3	1	5 1 1 1	56
8	3	2	4 2 1 1	420
8	3	3	3 3 1 1	280
8	3	4	3 2 2 1	840
8	3	5	2 2 2 2	105
8	4	1	4 1 1 1 1	70
8	4	2	3 2 1 1 1	560
8	4	3	2 2 2 1 1	420
8	5	1	3 1 1 1 1 1	56
8	5	2	2 2 1 1 1 1	210
8	6	1	2 1 1 1 1 1 1	28
9	1	1	8 1	9
9	1	2	7 2	36
9	1	3	6 3	84
9	1	4	5 4	126
9	2	1	7 1 1	36
9	2	2	6 2 1	252
9	2	3	5 3 1	504
9	2	4	4 4 1	315
9	2	5	5 2 2	378
9	2	6	4 3 2	1260
9	2	7	3 3 3	280
9	3	1	6 1 1 1	84
9	3	2	5 2 1 1	756
9	3	3	4 3 1 1	1260
9	3	4	4 2 2 1	1890
9	3	5	3 3 2 1	2520
9	3	6	3 2 2 2	1260
9	4	1	5 1 1 1 1	126
9	4	2	4 2 1 1 1	1260
9	4	3	3 3 1 1 1	840
9	4	4	3 2 2 1 1	3780
9	4	5	2 2 2 2 1	945
9	5	1	4 1 1 1 1 1	126
9	5	2	3 2 1 1 1 1	1260
9	5	3	2 2 2 1 1 1	1260
9	6	1	3 1 1 1 1 1 1	84
9	6	2	2 2 1 1 1 1 1	378
9	7	1	2 1 1 1 1 1 1 1	36
10	1	1	9 1	10
10	1	2	8 2	45
10	1	3	7 3	120
10	1	4	6 4	210
10	1	5	5 5	126
10	2	1	8 1 1	45
10	2	2	7 2 1	360
10	2	3	6 3 1	840
10	2	4	5 4 1	1260
10	2	5	6 2 2	630
10	2	6	5 3 2	2520
10	2	7	4 4 2	1575
10	2	8	4 3 3	2100
10	3	1	7 1 1 1	120
10	3	2	6 2 1 1	1260
10	3	3	5 3 1 1	2520
10	3	4	4 4 1 1	1575
10	3	5	5 2 2 1	3780
10	3	6	4 3 2 1	12600
10	3	7	4 2 2 2	3150
10	3	8	3 3 3 1	2800

Table 6.2 Continued.

n	i	j	<--Sequence ℓ_k, k=1,..,i+1-->	$C_{n,i,j}$
10	3	9	3 3 2 2	6300
10	4	1	6 1 1 1 1	210
10	4	2	5 2 1 1 1	2520
10	4	3	4 3 1 1 1	4200
10	4	4	4 2 2 1 1	9450
10	4	5	3 3 2 1 1	12600
10	4	6	3 2 2 2 1	12600
10	4	7	2 2 2 2 2	945
10	5	1	5 1 1 1 1 1	252
10	5	2	4 2 1 1 1 1	3150
10	5	3	3 3 1 1 1 1	2100
10	5	4	3 2 2 1 1 1	12600
10	5	5	2 2 2 2 1 1	4725
10	6	1	4 1 1 1 1 1 1	210
10	6	2	3 2 1 1 1 1 1	2520
10	6	3	2 2 2 1 1 1 1	3150
10	7	1	3 1 1 1 1 1 1 1	120
10	7	2	2 2 1 1 1 1 1 1	630
10	8	1	2 1 1 1 1 1 1 1 1	45
11	1	1	10 1	11
11	1	2	9 2	55
11	1	3	8 3	165
11	1	4	7 4	330
11	1	5	6 5	462
11	2	1	9 1 1	55
11	2	2	8 2 1	495
11	2	3	7 3 1	1320
11	2	4	6 4 1	2310
11	2	5	5 5 1	1386
11	2	6	7 2 2	990
11	2	7	6 3 2	4620
11	2	8	5 4 2	6930
11	2	9	4 4 3	5775
11	2	10	5 3 3	4620
11	3	1	8 1 1 1	165
11	3	2	7 2 1 1	1980
11	3	3	6 3 1 1	4620
11	3	4	5 4 1 1	6930
11	3	5	6 2 2 1	6930
11	3	6	5 3 2 1	27720
11	3	7	4 3 3 1	23100
11	3	8	3 3 3 2	15400
11	3	9	4 4 2 1	17325
11	3	10	5 2 2 2	6930
11	3	11	4 3 2 2	34650
11	4	1	7 1 1 1 1	330
11	4	2	6 2 1 1 1	4620
11	4	3	5 3 1 1 1	9240
11	4	4	4 4 1 1 1	5775
11	4	5	5 2 2 1 1	20790
11	4	6	4 3 2 1 1	69300
11	4	7	3 3 3 1 1	15400
11	4	8	3 3 2 2 1	69300
11	4	9	3 2 2 2 2	17325
11	4	10	4 2 2 2 1	34650
11	5	1	6 1 1 1 1 1	462
11	5	2	5 2 1 1 1 1	6930
11	5	3	4 3 1 1 1 1	11550
11	5	4	4 2 2 1 1 1	34650
11	5	5	3 3 2 1 1 1	46200
11	5	6	3 2 2 2 1 1	69300
11	5	7	2 2 2 2 2 1	10395
11	6	1	5 1 1 1 1 1 1	462
11	6	2	4 2 1 1 1 1 1	6930
11	6	3	3 3 1 1 1 1 1	4620
11	6	4	3 2 2 1 1 1 1	34650
11	6	5	2 2 2 2 1 1 1	17325
11	7	1	4 1 1 1 1 1 1 1	330
11	7	2	3 2 1 1 1 1 1 1	4620
11	7	3	2 2 2 1 1 1 1 1	6930
11	8	1	3 1 1 1 1 1 1 1 1	165
11	8	2	2 2 1 1 1 1 1 1 1	990
11	9	1	2 1 1 1 1 1 1 1 1 1	55

The reader is invited to verify the following results for n=2 and n=3, using the procedure outlined before

$$f_{N(t_1)}(2) = \frac{\lambda}{2\alpha} \exp(-\lambda t_1) \left[\frac{p}{1-q \exp(-\alpha t_1)} \right]^{\lambda/\alpha} \left\{ -2pq \right.$$

$$\text{Ln} \frac{p}{1-q \exp(-\alpha t_1)} + 2\alpha pq t_1 + \frac{q\phi(t_1)}{[1-q \exp(-\alpha t_1)]^2} [3q-2-(2q-1)q \exp(-\alpha t_1)]$$

$$\left. + \frac{\lambda}{\alpha} \left[p \text{ Ln} \frac{1-q \exp(-\alpha t_1)}{p} + \frac{q\phi(t_1)}{1-q \exp(-\alpha t_1)} + \alpha p t_1 \right]^2 \right\} \qquad (6.71)$$

$$f_{N(t_1)}(3) = \frac{\lambda}{6\alpha} \exp(-\lambda t_1) \left[\frac{p}{1-q \exp(-\alpha t_1)} \right]^{\lambda/\alpha}$$

$$\left\{ -6pq^2 \text{ Ln} \frac{p}{1-q \exp(-\alpha t_1)} + 6\alpha pq^2 t_1 + \frac{q^2\phi(t_1)}{1-q \exp(-\alpha t_1)} \right.$$

$$\left[6q - \frac{3[2-q-q \exp(-\alpha t_1)]}{1-q \exp(-\alpha t_1)} - \frac{[3-2q-q \exp(-\alpha t_1)]\phi(t_1)}{[1-q \exp(-\alpha t_1)]^2} \right] +$$

$$\frac{3\lambda}{\alpha} \left\{ - 2pq \text{ Ln} \frac{p}{1-q \exp(-\alpha t_1)} + 2\alpha pq t_1 + \frac{q\phi(t_1)}{1-q \exp(-\alpha t_1)} \right.$$

$$\left. \left[2q - \frac{[2-q-q \exp(-\alpha t_1)]}{1-q \exp(-\alpha t_1)} \right] \right\} \left\{ - p \text{ Ln} \frac{p}{1-q \exp(-\alpha t_1)} + \alpha p t_1 + \right.$$

$$\left. \frac{q\phi(t_1)}{1-q \exp(-\alpha t_1)} \right\} + \frac{\lambda^2}{\alpha^2} \left\{ - p \text{ Ln} \frac{p}{1-q \exp(-\alpha t_1)} + \alpha p t_1 + \right.$$

$$\left. \frac{q\phi(t_1)}{1-q \exp(-\alpha t_1)} \right\}^3 \Bigg\} \qquad (6.72)$$

6.6 Factorial moments for the NS process

Factorial moments up to order 3 for the NS process will be needed later to derive properties of the NSWN model. The relationship between factorial moments and the pgf was defined in eq. (6.42).

Factorial moment of order zero is the mean of the process. In order to obtain factorial moments of order 1, 2 and 3, derivatives of the pgf must be computed and evaluated at 1. Equations (6.60) to (6.63) can be used for this purpose, changing the evaluation at z=0 for evaluation at z=1.

Following the same approach as that described for the derivatives of f(z) evaluated at zero, the law of formation found for the derivatives of this function at z=1, for n≥2, is:

$$\frac{d^n f(1)}{dz^n} = \frac{n! q^{n-1}}{\alpha p^n} \left[\alpha t_1 - \sum_{i=1}^{n-1} \frac{\phi(t_1)^i}{i}\right] \tag{6.73}$$

For n=1, the expression is

$$\frac{df(1)}{dz} = \frac{t_1}{p} \tag{6.74}$$

Using eqs. (6.73) and (6.74) in the corresponding derivatives for g(z), evaluated at z=1, and introducing full distributional assumptions, the required factorial moments are (Cadavid et al., 1991)

$$M^*_{[1]} = \frac{\lambda}{\alpha p^2} \left[2q(\alpha t_1 - \phi(t_1)) + \lambda \alpha t_1^2\right] \tag{6.75}$$

$$M^*_{[2]} = \frac{\lambda}{\alpha p^3} \left\{3q^2[2(\alpha t_1 - \phi(t_1)) - \phi(t_1)^2] + 6\lambda t_1 q(\alpha t_1 - \phi(t_1)) + \lambda^2 \alpha t_1^3\right\} \tag{6.76}$$

$$M^*_{[3]} = \frac{\lambda}{\alpha p^4} \left\{24q^3\left(\alpha t_1 - \phi(t_1) - \frac{\phi(t_1)^2}{2} - \frac{\phi(t_1)^3}{3}\right) + \frac{12\lambda q}{\alpha}[(\alpha t_1 - \phi(t_1))\right.$$

$$\left.(3\alpha t_1 - \phi(t_1)) - \alpha t_1 \phi(t_1)^2] + 12\lambda^2 t_1^2 q(\alpha t_1 - \phi(t_1)) + \lambda^3 \alpha t_1^4\right\} \tag{6.77}$$

6.7 Properties of the NSWN precipitation process

Equation (6.12) gives the cumulative amount of precipitation for the NSWN model, for any interval of length T. Some of the moments for Y_k will be derived in this section, as well as its marginal distribution. Some of the properties are presented in the most general way possible, so that they can be used with other variations of the model, for example, with the cluster size following a Poisson distribution (Rodriguez-Iturbe et al., 1984). Distributional assumptions are introduced one at a time.

The White Noise terms are assumed independently and identically distributed as $f_U(u)$, with mean E[U] and variance σ_U^2. Their common distribution will be assumed to be the exponential distribution with parameter μ (Mood et al., 1974), so that

$$E[U] = 1/\mu \tag{6.78}$$

$$\sigma_U^2 = 1/\mu^2 \tag{6.79}$$

As before for the PWN model, μ has units 1/length, 1/in. in this case.

Although the gamma distribution was considered with the PWN model, it will not be used here. First, a high level of complication has been already obtained for the NSWN model for precipitation. Second, this distribution did not improve the PWN model performance in any way and similar results should be expected for the NSWN model.

Following similar procedures as those presented for the PWN precipitation process in Section 4.3, the mean for Y_k, μ_Y, is given by

$$\mu_Y = E[N(T)]\ E[U] = \lambda\ E[N_2(.)]\ T\ E[U] \tag{6.80}$$

based on eq. (6.45). When the geometric distribution for the cluster size is introduced, eq. (6.80) becomes

$$\mu_Y = \lambda T \, E[U] \, / \, p \tag{6.81}$$

according to eqs. (6.50). Finally, using the exponential distribution for the White Noise terms

$$\mu_Y = \lambda T \, / \, (\mu p) \tag{6.82}$$

The main step in computing the variance of Y_k is conditioning its second raw moment on the counting process. After doing this and performing algebraic manipulations (Cadavid et al., 1991)

$$E[Y_k^2] = E[N(T)] \, E[U^2] + M^*_{[1]} \, E^2[U] \tag{6.83}$$

and

$$\sigma_Y^2 = E[N(T)] \, \sigma_U^2 + Var(N(T)) \, E^2[U]$$

Replacing in this expression eqs. (6.45) and (6.48)

$$\sigma_Y^2 = \lambda E[N_2(.)]T\sigma_U^2 + \lambda E^2[U]\{E[N_2^2(.)]T - E[N_2(.)(N_2(.)-1)]\phi(T)/\alpha\} \tag{6.84}$$

Using eqs. (6.50) in (6.84), the variance of Y_k for the cluster size following a geometric distribution is

$$\sigma_Y^2 = \frac{\lambda T}{p} \, \sigma_U^2 + \lambda \, E^2[U]\{ \frac{2 - p}{p^2} \, T - \frac{2q}{\alpha p^2} \, \phi(T)\} \tag{6.85}$$

Finally, using the exponential distribution for the White Noise element in the model, eq. (6.85) becomes

$$\sigma_Y^2 = 2\lambda \, [\alpha T - q\phi(T)] \, / \, (\alpha p^2 \mu^2) \tag{6.86}$$

As shown for the PWN model, the covariance of lag δ for the amount of precipitation in disjoint time intervals is ($\delta \geq 1$)

$$\gamma_Y(\delta) = E\left[\sum_{i=0}^{N(T)} U_i \sum_{j=0}^{M(T)} U_j \right] - \mu_Y^2$$

with M(T) representing $N(\delta T,(\delta+1)T)$. This case is different to the PWN model in the sense that N(T) and M(T) are dependent. The expected value in the above equation is computed as

$$E\left[\sum_{i=0}^{N(T)} U_i \sum_{j=0}^{M(T)} U_j \right] = E_{N(T),M(T)} \left\{ E\left[\sum_{i=0}^{N(T)} \sum_{j=0}^{M(T)} U_i U_j \mid N(T),M(T) \right] \right\}$$

where the most external expectation is taken over the joint pdf for N(T) and M(T). Since the White Noise terms are independent, the previous equation becomes

$$E\left[\sum_{i=0}^{N(T)} U_i \sum_{j=0}^{M(T)} U_j\right] = E^2[U]\, E_{N(T),M(T)}\left[N(T)\, M(T)\right]$$

$$E\left[\sum_{i=0}^{N(T)} U_i \sum_{j=0}^{M(T)} U_j\right] = E^2[U]\, E\left[N(T)\, M(T)\right]$$

and therefore

$$\gamma_Y(\delta) = E^2[U]\mathrm{Cov}[N(0,T),N(\delta T,(\delta+1)T)] \tag{6.87}$$

Using the result in (6.49) in eq. (6.87)

$$\gamma_Y(\delta) = E^2[U]\ \lambda\ E[N_2{}^2(.)-N_2(.)]\phi(T)^2\exp[-\alpha T(\delta-1)]/(2\alpha) \tag{6.88}$$

Introducing the geometric distribution for $N_2(.)$

$$\gamma_Y(\delta) = E^2[U]\ \lambda q\ \phi(T)^2\exp[-\alpha T(\delta-1)]/(\alpha p^2) \tag{6.89}$$

Finally, when the White Noise terms are exponential with parameter $1/\mu$

$$\gamma_Y(\delta) = \lambda q\ \phi(T)^2\exp[-\alpha T(\delta-1)]/(\alpha p^2\mu^2) \tag{6.90}$$

Based on eqs. (6.90) and (6.86), when full distributional assumptions are in use, the autocorrelation function of lag δ for the amount of precipitation in disjoint intervals of length T, for the NSWN model, is given by

$$\rho_Y(\delta) = \frac{q\phi(T)^2\exp[-\alpha T(\delta-1)]}{2[\alpha T-q\phi(T)]}\ ,\ \delta\geq 1 \tag{6.91}$$

The autocorrelation function for the NSWN model is controlled by two parameters, p and α, which are parameters for the NS arrival process, with no influence from the instantaneous amounts of precipitation. It represents a Markovian process (Foufoula-Georgiou and Guttorp, 1986). Finally, the autocorrelation function in (6.91)

is based on the possibility of a given cluster having members in two disjoint time intervals.

Although eq. (6.91) represents a Markovian process, it is possible to show that it generates lag-1 and lag-2 autocorrelation coefficients satisfying the following constraints

$$\rho_Y(2) \leq \rho_Y(1)$$

$$\frac{[1-\rho_Y(2)/\rho_Y(1)]^2}{2[\mathrm{Ln}(\rho_Y(1)/\rho_Y(2)) - 1+\rho_Y(2)/\rho_Y(1)]} - \rho_Y(1) > 0$$

These constraints define the same feasible region as that presented by Obeysekera et al., (1987) for the NSWN model with Poissonian cluster size, which is a subset of the feasible region for an ARMA (1,1) model.

The skewness coefficient of Y_k, denoted g_Y, is computed using the general expression

$$g_Y = \left(E\left[Y_k^3 \right] - 3\mu_Y E\left[Y_k^2 \right] + 2\mu_Y^3 \right) / \sigma_Y^3 \qquad (6.92)$$

It can be shown that the third raw moment for Y_k is given by (Cadavid et al., 1991)

$$E[Y_k^3] = E[N(T)]\ E[U^3] + 3\ M^*_{[1]}\ E[U^2]\ E[U] + M^*_{[2]} E^3[U] \qquad (6.93)$$

Note the link between the raw moments of Y_k (eqs. (6.83) and (6.93)) and the factorial moments for the counting process $M^*_{[r]}$.

Replacing eqs. (6.93) and (6.83) in (6.92), using eqs. (6.75) and (6.76) for the factorial moments, dividing by the standard deviation to the power of three and taking into account that for the exponential distribution with parameter μ

$$E[U^i] = i! / \mu^i \qquad (6.94)$$

the skewness coefficient for Y_k becomes

$$g_Y = \frac{3}{2^{3/2}} \frac{\alpha^{1/2}}{\lambda^{1/2}} \frac{q[2(\alpha T-\phi(T))(1+p)-q\phi(T)^2]+2\alpha p^2 T}{[\alpha T-q\phi(T)]^{3/2}} \tag{6.95}$$

The next property is the kurtosis coefficient for the amount of precipitation in the time interval of length T, κ, defined as

$$\kappa = \left(E\left[Y_k^4 \right] - 4\mu_Y E\left[Y_k^3 \right] + 6\mu_Y^2 E\left[Y_k^2 \right] - 3\mu_Y^4 \right) / \sigma_Y^4 \tag{6.96}$$

The derivation of the fourth raw moment for Y_k is given in Cadavid et al., (1991). The result is

$$E[Y_k^4] = E[N(T)]\ E[U^4] + M^*_{[1]}\ (4E[U^3] + 3E^2[U^2]) +$$

$$6M^*_{[2]}\ E[U^2]\ E^2[U] + M^*_{[3]}\ E^4[U] \tag{6.97}$$

Performing all the operations indicated in eqs. (6.96) and (6.97) and using eqs. (6.75) to (6.77), (6.86) and (6.94), the resulting expression for κ is

$$\kappa = \alpha\{3q(\alpha T-\phi(T))[q(2+4p+\lambda T-\lambda\phi(T)/\alpha)+2p(\lambda T+3p)]-q^2\phi(T)^2[3(1+2p)$$

$$-2q\phi(T)]+3\alpha p^2 T(\lambda T+2p)\}/(\lambda[\alpha T-q\phi(T)]^2) \tag{6.98}$$

The last property to be considered for the amount of precipitation in an interval of length T is its marginal pdf. Given that Y_k for the NSWN model is a compound process, $f_Y(y)$ is written as

$$f_Y(y) = f_{N(T)}(0) I_{\{0\}}(y) + \left(\sum_{n=1}^{\infty} f_U^{(n)}(y)\ f_{N(T)}(n) \right) I_{(0,\infty)}(y)$$

Analytical expressions for $f_{N(T)}(n)$ were computed for n from 0 to 3. For n between 4 and 20, exact numerical values can be obtained. For n beyond 20 no results are available. Later in this chapter, an empirical method for extrapolating $f_{N(T)}(n)$ will be presented. This is needed to approximate $f_Y(y)$ and other required distributions. Meanwhile, the only possible step is to replace the distribution of

the sum of n independent White Noise terms by a gamma distribution with scale parameter μ and shape parameter n, giving the result

$$f_Y(y)=f_{N(T)}(0)I_{\{0\}}(y)+\left[\sum_{n=1}^{\infty}\frac{\mu^n y^{n-1}}{(n-1)!}f_{N(T)}(n)\right]I_{(0,\infty)}(y) \qquad (6.99)$$

6.8 Parameter estimation by method of moments

In order to estimate parameters, λ, p, α and μ, for the NSWN model, four moment properties for the amount precipitation, among those presented before, must be chosen. They are equated to sample estimates and the set is solved for the parameters. In principle, any four properties could be used. The traditional method includes mean, variance and correlation coefficients of lag 1 and 2. Other lags could be selected as well. However, this would produce a correlation function passing exactly through those two lags, ignoring for example lag-1, and this may not be a convenient situation.

If maximum likelihood estimation was out of reach for the PWN model, the argument is much stronger for the NSWN model, in which the degree of complication is larger. In the NSWN model, amounts of precipitation in disjoint intervals are not independent. Besides, up to this point, the pdf for the amount of precipitation in the NSWN model has been approximated and no judgement has been presented about the goodness of that approximation.

The reader may recall, from Chapter 2, the maximum likelihood estimator used by Foufoula-Georgiou and Guttorp (1986), based on binary series of occurrences derived from precipitation samples. This estimation procedure is not used here, based mainly on two considerations. First, the primary objective of this research is not

examination of parameter estimation methods. Second, since the beginning, all computations have been oriented to moment estimation.

As a final observation, Obeysekera et al. (1987) suggest the use of a moment estimator based on sample lag-1 correlation coefficients estimated at two different temporal scales, but they also conclude that this estimator is inferior to the traditional method. Besides, being the main objective of this research disaggregation of precipitation records, one should try to keep estimation procedures at only one aggregation level.

In conclusion, as a first attempt, the traditional method of moments will be used here. From eqs. (6.82), (6.86) and (6.91) for δ=1 and δ=2, and recalling that

$$\phi(T) = 1-\exp(-\alpha T); \quad q=1-p$$

moment estimators for the monthly parameters for the NSWN model are:

$$\hat{\alpha}_\tau = \mathrm{Ln}(r_\tau(1)/r_\tau(2))/T \tag{6.100}$$

$$\hat{p}_\tau = \frac{1-2r_\tau(1)\hat{\alpha}_\tau T}{[1-\exp(-\hat{\alpha}_\tau T)][1-\exp(-\hat{\alpha}_\tau T)+2r_\tau(1)]} \tag{6.101}$$

$$\hat{\mu}_\tau = 2\bar{Y}_\tau\{\hat{\alpha}_\tau T-(1-\hat{p}_\tau)[1-\exp(-\hat{\alpha}_\tau)]\}/(\hat{\alpha}_\tau\hat{p}_\tau T\, S_\tau^2) \tag{6.102}$$

$$\hat{\lambda}_\tau = \bar{Y}_\tau\hat{p}_\tau\hat{\mu}_\tau /T \tag{6.103}$$

where the index τ stands for the monthly partition adopted for the precipitation recording stations used in this study.

Equations (6.100) to (6.103) are constrained. First, if $r_\tau(1)$ or $r_\tau(2)$ are negative, $\hat{\alpha}_\tau$ is not defined. Also, if $r_\tau(2)$ is larger than $r_\tau(1)$, $\hat{\alpha}_\tau$ is negative. For $\hat{p}_\tau$, eq. (6.101) may yield negative values,

the case being the same for μ_τ and λ_τ, depending on sample statistics and previously estimated parameters for the model. A failure is produced whenever one of the above constraints is violated.

6.9 Results for parameters estimated by method of moments.

The first data set to which the method of moments was applied is a simulated sample of a NSWN process, representing 49 years worth of data. The simulation, referred to as simulation 1, was obtained with the set of parameters listed in Table 6.3, which correspond to values estimated using optimization techniques (Sections 6.10 and 6.11) for the Denver Wsfo Ap station, for T=60 min., month 05. Although the process of simulation is continuous, the highest resolution sample obtained was hourly. Larger aggregation scale samples were obtained from the hourly simulated data.

Monthly statistics were estimated for simulation 1 at two aggregation levels, T=60 min. and T_a=1440 min.. From these, parameters were estimated using eqs. (6.100) to (6.103). Table 6.3 gives a summary of results for estimated parameters and Figures 6.2 and 6.3 present theoretical and estimated correlograms. Plotted correlograms verify the goodness of the simulation algorithm, as well as estimation procedures discussed in Chapter 3.

In regard to Figures 6.2 and 6.3 it is important to mention that correlograms estimated from simulated samples were observed to be similar to the counterparts for recorded precipitation samples, in terms of shapes, values and sample variability. Sample variability in the correlogram for T_a=1440 min. is evident. In this sense, one can state that the NSWN model provides an adequate description of the

Table 6.3 Summary of results for parameters estimated from NSWN simulated precipitation series, using method of moments, for simulation 1, for different aggregation levels.

Parameter: λ, Units: 1/min. x 10^{-3}, Population Value: 0.10232

	T=60 min.	T_a=1440 min.
Mean	0.10261	0.09077
Standard Deviation	0.02714	0.00763
Maximum	0.14203	0.09663
Minimum	0.05941	0.07765
Number of failures	0	5

Parameter: p, Population Value: 0.07208

	T=60 min.	T_a=1440 min.
Mean	0.07223	0.24722
Standard Deviation	0.00180	0.08931
Maximum	0.09858	0.37038
Minimum	0.04703	0.13761
Number of failures	0	5

Parameter: α, Units: 1/min. x 10^{-2}, Population Value: 0.22100

	T=60 min.	T_a=1440 min.
Mean	0.22997	0.16274
Standard Deviation	0.07134	0.06001
Maximum	0.38353	0.26678
Minimum	0.13783	0.09766
Number of failures	0	3

Parameter: μ, Units: 1/in., Population Value: 24.36232

	T=60 min.	T_a=1440 min.
Mean	23.9924	7.00805
Standard Deviation	1.05575	2.39853
Maximum	26.2089	9.90631
Minimum	22.7070	3.97527
Number of failures	0	5

precipitation process. However, in many other cases the NSWN model was not able to fit a given correlogram.

Two observations are in order in regard to Table 6.3. Method of moments did not present any failure for T=60 min., but the number of failures increased to 5 out of 12 for T_a=1440 min.. Again,

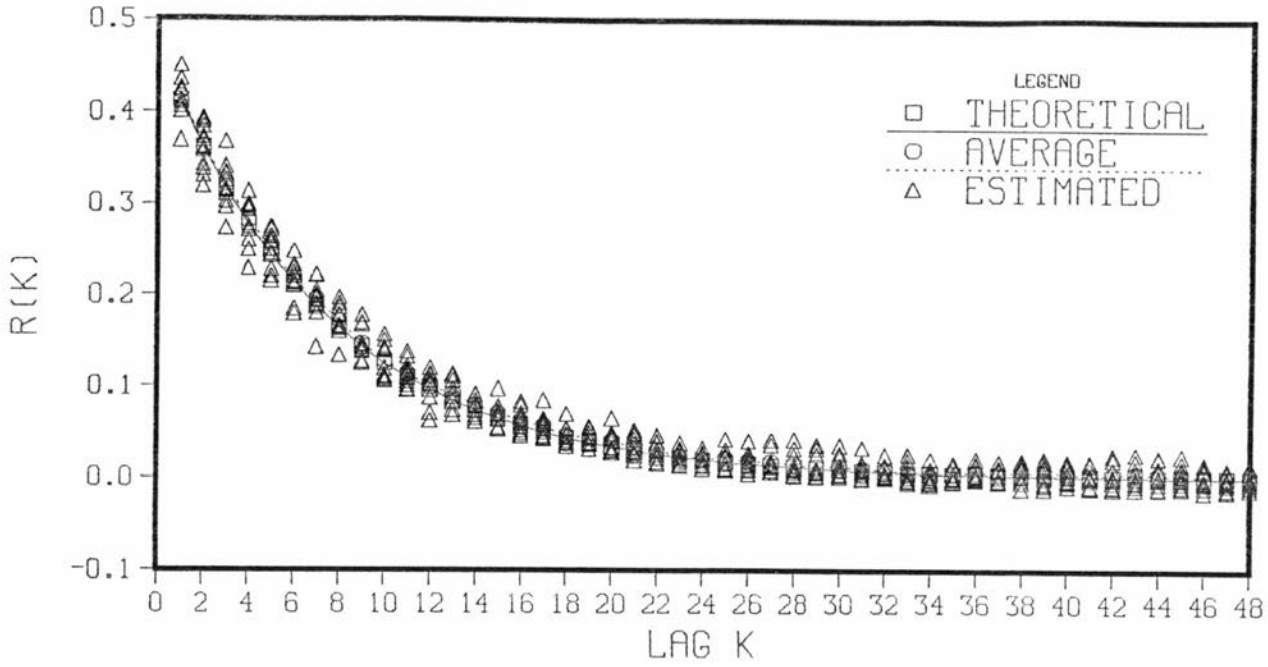


Figure 6.2 Theoretical and estimated correlograms for NSWN simulated process, for simulation 1, for T=60 min..

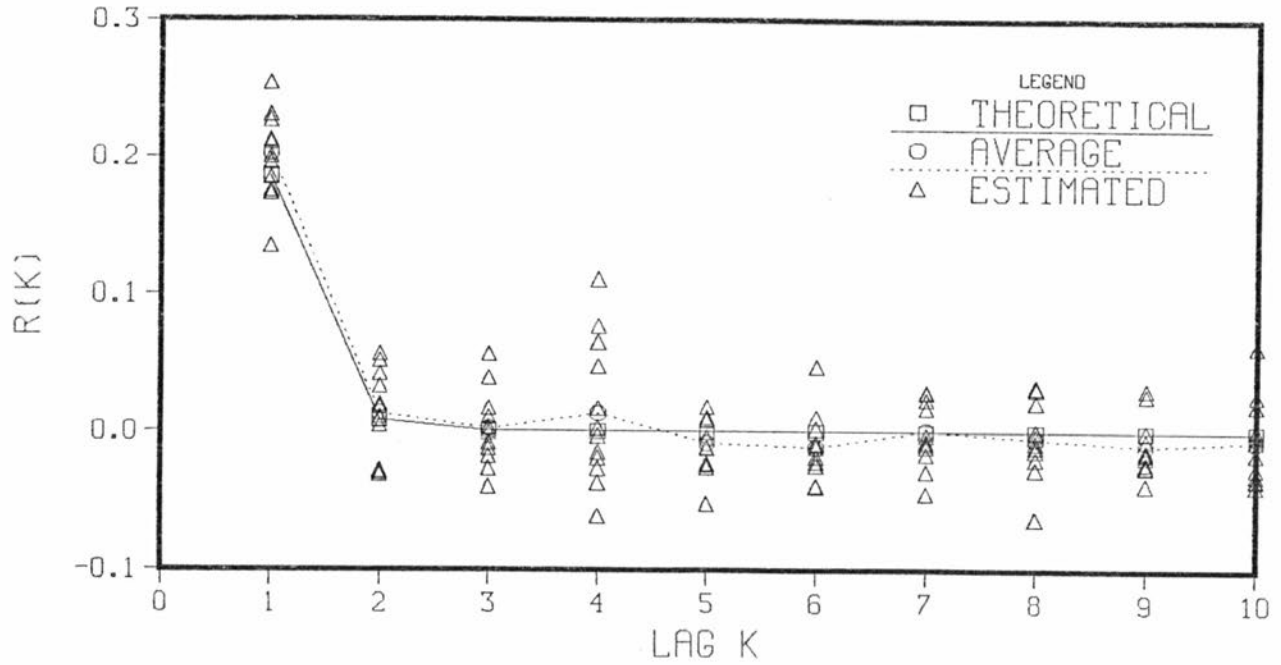


Figure 6.3 Theoretical and estimated correlograms for NSWN simulated process, for simulation 1, for T=1440 min..

incompatibility of the model at different temporal scales is evident in Table 6.3, since parameter estimates are not the same in terms of mean values.

Method of moments was also applied to the precipitation recording stations used in this study, for two different aggregation levels for each station. Results are given in Tables 6.4 to 6.7, in the order in which parameters are estimated using eqs. (6.100) to (6.103). In each table, whenever is required, observations are presented on the reason producing estimation failures.

Tables 6.4 to 6.7 show how method of moments tends to perform better when the level of aggregation is the hour. The number of failures increases appreciably for T=5 min. and T_a=1440 min. For the hourly case, annual periodicity is observed, in such a way that parameter estimates increase for summer or summer transition months. This means that during summer months, storms are more frequent, with smaller number of cluster members, located closer to the cluster center, with each member producing a larger precipitation yield, as compared to winter months, when storms are predominantly of larger duration and lower intensity. These characteristics agrees with the observed precipitation behavior in Colorado.

One could attempt to apply moment estimators to monthly precipitation data. Two reasons inhibit this application. First, the sample correlation available for monthly data is from month to month and not within the season. Second, inclusion skewness and kurtosis in the estimation procedure makes solution of equations heavy and difficult to obtain.

Table 6.4 Estimated parameters for the NSWN model for Denver Wsfo Ap station, for different temporal aggregation levels, using method of moments.

T=60 min.

τ	$\hat{\alpha}_\tau$ (1/min.) $\times 10^{-1}$	$\hat{p}_\tau$	$\hat{\mu}_\tau$ (1/in.)	$\hat{\lambda}_\tau$ (1/min.) $\times 10^{-3}$	Observations
1	0.03793	0.020899	326.727	0.07736	
2	0.05288	0.025276	313.127	0.11981	
3	0.04486	0.013351	361.300	0.14194	
4	0.04914	0.025294	126.317	0.13150	
5	0.05597	0.094360	32.7854	0.18026	
6	0.08928	0.118514	26.9073	0.12764	
7	0.16897	0.371362	9.30086	0.14005	
8	0.15585	0.341402	9.87705	0.10779	
9	0.04288	0.182538	17.3309	0.08372	
10	0.05675	0.049054	94.9364	0.09740	
11	0.04256	0.018782	296.940	0.10641	
12	0.03189	0.014515	310.525	0.05938	

T=1440 min.

τ	$\hat{\alpha}_\tau$ (1/min.) $\times 10^{-2}$	$\hat{p}_\tau$	$\hat{\mu}_\tau$ (1/in.)	$\hat{\lambda}_\tau$ (1/min.) $\times 10^{-4}$	Observations
1	0.14591	0.44608	13.3305	0.67404	
2					Correlation negative
3					Correlation negative
4					Correlation negative
5	0.17045	0.23476	7.25019	0.99123	
6	0.12698	0.36565	5.00631	0.73234	
7	0.04736	0.57017	4.12029	0.95255	
8	0.16223	0.14315	15.3899	0.70394	
9	0.13082	0.30558	8.50132	0.68687	
10					Correlation negative
11					Correlation negative
12					Correlation negative

Table 6.5 Estimated parameters for the NSWN model for Greenland 9 SE station, for different temporal aggregation levels, using method of moments.

T=60 min.

τ	$\hat{\alpha}_\tau$ (1/min.) x 10^{-1}	$\hat{p}_\tau$	$\hat{\mu}_\tau$ (1/in.)	$\hat{\lambda}_\tau$ (1/min.) x 10^{-3}	Observations
1	0.03438	0.13495	80.33567	0.089042	
2	0.03015	0.09141	90.82454	0.100455	
3	0.01261	0.03889	59.03361	0.058902	
4	0.04401	0.08481	55.36718	0.147885	
5	0.05747	0.16331	27.22535	0.225754	
6	0.09979	0.24240	15.85714	0.174484	
7	0.19860	0.30461	11.15179	0.210256	
8	0.21101	0.18670	16.48421	0.173846	
9	0.11274	0.26360	22.28655	0.129106	
10	0.03952	0.08594	60.47824	0.074920	
11	0.03422	0.08104	65.22412	0.084373	
12	0.06884	0.02799	60.73068	0.018651	

T=1440 min.

τ	$\hat{\alpha}_\tau$ (1/min.) x 10^{-2}	$\hat{p}_\tau$	$\hat{\mu}_\tau$ (1/in.)	$\hat{\lambda}_\tau$ (1/min.) x 10^{-3}	Observations
1					Correlation negative
2					Correlation negative
3					Correlation negative
4					Correlation negative
5					Correlation negative
6	0.07448	0.42892	5.358821	0.104050	
7	0.08220	0.65302	3.780899	0.152116	
8					α negative
9	0.09270	0.36766	9.254666	0.074172	
10	0.17320	0.10240	35.35516	0.051751	
11	0.24398				p negative
12					Correlation negative

Table 6.6 Estimated parameters for the NSWN model for Idaho Springs station, for different temporal aggregation levels, using method of moments.

T=5 min.

τ	$\hat{\alpha}_\tau$ (1/min.)	$\hat{p}_\tau$	$\hat{\mu}_\tau$ (1/in.)	$\hat{\lambda}_\tau$ (1/min.) x 10^{-3}	Observations
1					α negative
2					α negative
3					α negative
4	0.00721	0.02588	286.664	0.23754	
5	0.04249	0.13575	249.462	0.55672	
6	0.03595	0.06285	188.790	0.38754	
7	0.06637	0.03312	235.957	0.39043	
8	0.05125	0.02149	257.648	0.36860	
9	0.12980	0.16036	10.1437	0.07828	
10	0.06460	0.22786	118.933	0.25094	
11					α negative
12					α negative

T=60 min.

τ	$\hat{\alpha}_\tau$ (1/min.) x 10^{-1}	$\hat{p}_\tau$	$\hat{\mu}_\tau$ (1/in.)	$\hat{\lambda}_\tau$ (1/min.) x 10^{-3}	Observations
1	0.06532	0.02332	354.9704	0.41076	
2	0.01284	0.00157	478.7013	0.07178	
3	0.08573	0.05746	307.3935	0.41569	
4	0.04302	0.02114	230.2212	0.15268	
5	0.04157	0.08171	114.4487	0.14809	
6	0.14402	0.15120	58.60070	0.27768	
7	0.17494	0.38133	16.11613	0.29976	
8	0.12317	0.16535	19.69441	0.21166	
9	0.30740	0.87375	1.855267	0.07616	
10	0.13728	0.49910	49.58866	0.22316	
11	0.01809	0.00666	237.0252	0.12123	
12	0.01602	0.00373	973.7752	0.19676	

Table 6.7 Estimated parameters for the NSWN model for Ward station, for different temporal aggregation levels, using method of moments.

T=5 min.

τ	$\hat{\alpha}_\tau$ (1/min.) x 10^{-1}	$\hat{p}_\tau$	$\hat{\mu}_\tau$ (1/in.)	$\hat{\lambda}_\tau$ (1/min.) x 10^{-2}	Observations
1					α negative
2					α negative
3					α negative
4					α negative
5	0.04196	0.00956	125.977	0.01435	
6					α negative
7	0.70520	0.01680	400.508	0.05439	
8	0.64710	0.04676	165.258	0.05405	
9					α negative
10	0.99659	0.19198	133.125	0.13096	
11					α negative
12					α negative

T=60 min.

τ	$\hat{\alpha}_\tau$ (1/min.) x 10^{-1}	$\hat{p}_\tau$	$\hat{\mu}_\tau$ (in.) x 10^{-1}	$\hat{\lambda}_\tau$ (1/min.) x 10^{-3}	Observations
1	0.09332	0.03473	664.156	0.17401	
2	0.02120	0.00647	413.858	0.21145	
3	0.04539	0.03405	262.477	0.19805	
4	0.03783	0.01697	114.891	0.16967	
5	0.08596	0.17412	24.8478	0.49772	
6	0.32832				p negative
7	0.07558	0.32353	13.6774	0.33502	
8	0.03983	0.20115	15.5153	0.21019	
9	0.06818	0.26999	29.9649	0.27454	
10	0.06075	0.12367	45.5581	0.28088	
11	0.03529	0.00442	859.603	0.18827	
12	0.03111	0.02135	368.989	0.09044	

6.10 Estimation of parameters using Weighted Least Squares method

The method of moments, as presented in the previous section, exhibits serious shortcomings, due mainly to the small amount of information used by the method. For example, in cases when correlations of lag 1 or 2 are negative, coefficients for other lags could be used to obtain model parameters. Also, when negative estimated values for α_τ are produced, due to $r_\tau(2)$ being larger than $r_\tau(1)$, more information on the trend of the correlogram could produce feasible estimated parameter values. A partially increasing correlogram with the lag could be the result of sample variability, and this variability is not included in the traditional moment estimation procedure.

The objective of this section is to design a parameter estimation procedure for the NSWN model, based on moment properties, which includes more information on those properties and which has the ability to include, at least in a subjective fashion, some type of indicator about sample variability. Perhaps the method will be consistent for different aggregation scales.

The technique selected is based on weighted least squares (WLS). Denote model properties, for a given month, by ψ_i^τ, where i denotes a given property and τ is the month. Properties are functions of model parameters, according to eqs. (6.82), (6.86), (6.91), (6.95) and (6.98). The problem statement is

$$\underset{\lambda_\tau, p_\tau, \alpha_\tau, \mu_\tau}{\text{Min}} \quad z = \sum_{i=1}^{n} w_i (\psi_i^\tau / \hat{\psi}_i^\tau - 1)^2 \, I_{(0,\infty)}(\hat{\psi}_i^\tau) \tag{6.104}$$

$$\text{subject to:} \quad \lambda_\tau > 0 \; ; \quad 1 > p_\tau > 0; \quad \alpha_\tau > 0; \quad \mu_\tau > 0$$

where z is the objective function, $\hat{\psi}_i^{\tau}$ is the monthly estimated property and w_i is a weighting factor. The solution to this minimization problem will yield estimates for model parameters. A very similar approach is used by Islam et al. (1988) in order to estimate parameters for a space-time model for cumulative precipitation. Burlando (1989) uses a very similar methodology to estimate parameters for NS type precipitation models.

The index i indicates model properties as follows: i=1, mean; i=2, variance; i=3, coefficient of skewness; i=4, coefficient of kurtosis; i=5,..,n, correlation coefficients of lag i-4. The configuration of the objective function, eq. (6.104), is defined by the user, starting by the number of terms n, which must be greater or equal than 5. In the next step, the user selects which of the first four properties are included in the objective function.

The weights w_i are defined according to two different methods. In the first, weights for all properties are equal, so that the solution to the minimization problem will be the same as that obtained without weights. In the second method, the user defines weights ($w_i < 1.0$) for the first four properties to be included in the model. If one of the first four properties is not included, the corresponding weight is made equal to zero. The remaining weights, for the correlation coefficients, are computed in such a way that all weights add up to 1.0 and in such way that they decrease with the lag, according to an inverse law. The following expressions are obtained for computation of weights associated to correlation coefficients:

$$w_i = a/i, \quad i \geq 5 \tag{6.105}$$

$$a = \left(\sum_{i=1}^{4} w_i\right) / \left[1 - \sum_{i=1}^{n} (1/i)\right] \qquad (6.106)$$

The last term in eq. (6.104) is an indicator function for the ith estimated property. If this is negative, model property and associated weight are eliminated from the estimation procedure.

The fundamental assumption is that definition of the number of terms in the objective function and weights provides the estimation procedure with subjective information about sample variability on estimated model properties. Also, by having the possibility to define and redefine objective function configurations, the user is able to avoid unfeasible parameter estimates violating the constraints given in eq. (6.104).

The estimation procedure described above was translated into a FORTRAN 77 computer code. The numerical algorithm used to solve the optimization problem is known as downhill simplex method in multidimensions (Press et al., 1986). No major details on the minimization algorithm are given here, except to highlight some of its characteristics. First, the algorithm is unconstrained, despite the constraints appearing in eq. (6.104). The algorithm used by Islam et al. (1988) is also unconstrained, although Burlando (1989) uses a constrained method. A great deal of difficulty was experienced in the implementation of constrained algorithms and the method mentioned above performed very well. A very limited number of runs, 1% or less, gave failure due to unfeasible parameter values. For a given optimization run, the algorithm converges when the difference between two consecutive objective function evaluations is smaller than a given tolerance. After that, the same optimization run is reinitialized in the neighborhood of the previous solution. The final solution is

obtained when optimal parameter values for two different initializations are as close as a given tolerance. This last feature gives a larger guarantee of the solution being at a global maximum.

Although it was stated in Chapter 1 that no computer codes are listed or described in this report, a brief account is given in the following. The computer program is interactive in most of the information required, specially for objective function configuration. The program provides the user with on-screen plots for historical and fitted correlograms. By having access to a plot of the historical correlogram, the user is able to decide on the number of terms in the objective function and to observe the degree of scatter of the historical correlogram. For a given set of estimated properties, the program allows up to nine optimization runs and one of them could be reserved for moment estimation, as described in Section 6.8. From one run to another, number of terms and weighting method in the objective function can be changed. The user decides on parameters to enter the optimization process and to which of them fixed values are given. Finally, with a summary of results for all optimization runs, provided on screen by the program, the user decides on final estimated parameter values for a given set of estimated properties.

It is recommended to include at least mean and variance in the objective function. As shown before, correlation coefficients depend on p_τ and α_τ. Mean and variance include the four parameters. Although the problem, as formulated, includes skewness and kurtosis, no success was obtained in performing optimization with these two model properties. They systematically yielded negative estimated parameter values. For this reason, along with those presented for the

method of moments, optimal estimation technique was not applied to precipitation data aggregated to the monthly level, for the recording stations used in this study.

6.11 Results for parameters estimated using Weighted Least Squares method

As explained before for the method of moments, the first data set to which the WLS technique was applied corresponds to simulation 1, as described in Section 6.9. A summary of results is given in Table 6.8. Similar comments, as those presented for Table 6.3, are in order here. There are failures in the estimation procedure for T_a=1440 min., although these are reduced from 5 to 2. Second, the incompatibility of the model at different time scales is still present and therefore WLS does not solve this problem. A more detailed comparison of the two estimation procedures is given in Section 6.12.

In the next step, WLS was applied to the precipitation recording stations used in this study, for two different aggregation levels for each station. Results are given in Tables 6.9 to 6.12. A similar analysis, as that presented for moment estimates, is valid for the results in Tables 6.9 to 6.12, in regard to annual periodicity. Since the set of results here is more complete, in terms of aggregation scale, than it was for the method of moments, one concludes that the same periodic behavior is kept at both aggregation scales. The complementary role of for the two estimation methods is present in Tables 6.9 to 6.12.

Results presented in Tables 6.9 to 6.12 confirm once again the incompatibility of the model at different temporal scales. In general, the trend of estimated parameters with the aggregation scale

Table 6.8 Summary of results for parameters estimated from NSWN simulated precipitation series using WLS method, for simulation 1, for different aggregation levels.

Parameter: λ, Units: 1/min. x 10^{-3}, Population Value: 0.10232

	T=60 min.	T_a=1440 min.
Mean	0.09882	0.09230
Standard Deviation	0.00824	0.01081
Maximum	0.11065	0.10181
Minimum	0.08685	0.06353
Number of failures	0	2

Parameter: p, Population Value: 0.07208

	T=60 min.	T_a=1440 min.
Mean	0.07071	0.33588
Standard Deviation	0.00719	0.15635
Maximum	0.08636	0.60586
Minimum	0.06004	0.13763
Number of failures	0	2

Parameter: α, Units: 1/min. x 10^{-2}, Population Value: 0.22100

	T=60 min.	T_a=1440 min.
Mean	0.21657	0.11346
Standard Deviation	0.01899	0.06201
Maximum	0.24136	0.19613
Minimum	0.18186	0.01442
Number of failures	0	2

Parameter: μ, Units: 1/in. , Population Value: 24.36232

	T=60 min.	T_a=1440 min.
Mean	23.8279	5.94845
Standard Deviation	1.51525	3.19751
Maximum	25.7049	11.0966
Minimum	20.9650	2.45725
Number of failures	0	2

Table 6.9 Estimated parameters for the NSWN model for Denver Wsfo Ap station, for different temporal aggregation levels, using WLS method.

T=60 min.

τ	$\hat{\lambda}_\tau$ (1/min.) x 10^{-3}	$\hat{p}_\tau$	$\hat{\alpha}_\tau$ (1/min.) x 10^{-1}	$\hat{\mu}_\tau$ (1/in.)	Moment estimates?
1	0.07173	0.01776	0.03548	356.401	No
2	0.08983	0.02275	0.03697	260.750	No
3	0.09706	0.02080	0.02560	158.563	No
4	0.09475	0.02273	0.03257	101.301	No
5	0.10232	0.07208	0.02210	24.3623	No
6	0.09591	0.15116	0.03692	15.8523	No
7	0.13272	0.37747	0.13490	8.67127	No
8	0.10273	0.42671	0.08048	7.53120	No
9	0.08694	0.20246	0.03952	16.2271	No
10	0.06450	0.05546	0.02719	55.5995	No
11	0.08812	0.02919	0.02979	158.231	No
12	0.05224	0.01358	0.02751	292.150	No

T_a=1440 min.

τ	$\hat{\lambda}_\tau$ (1/min.) x 10^{-4}	$\hat{p}_\tau$	$\hat{\alpha}_\tau$ (1/min.) x 10^{-2}	$\hat{\mu}_\tau$ (1/in.)	Moment estimates?
1	0.67404	0.44608	0.14591	13.3305	Yes
2	0.52107	0.42530	0.01249	8.09060	No
3	0.94172	0.45162	0.10660	7.08708	No
4	0.85741	0.72958	0.03564	2.85678	No
5	0.99123	0.23476	0.17045	7.25019	Yes
6	0.73234	0.36565	0.12698	5.00631	Yes
7	0.95255	0.57017	0.04736	4.12029	Yes
8	0.70394	0.14315	0.16223	15.3899	Yes
9	0.68687	0.30558	0.13082	8.50132	Yes
10	0.56785	0.32587	0.10203	8.33183	No
11	0.54823	0.35254	0.02345	8.15397	No
12	0.43754	0.61646	0.03624	5.39038	No

Table 6.10 Estimated parameters for the NSWN model for Greenland 9 SE station, for different temporal aggregation levels, using WLS method.

T=60 min.

τ	$\hat{\lambda}_\tau$ (1/min.) x 10^{-3}	$\hat{p}_\tau$	$\hat{\alpha}_\tau$ (1/min.) x 10^{-1}	$\hat{\mu}_\tau$ (1/in.)	Moment estimates?
1	0.08904	0.13495	0.03438	80.3357	Yes
2	0.10045	0.09141	0.03015	90.8245	Yes
3	0.09129	0.05409	0.02253	65.7821	No
4	0.10516	0.07000	0.02610	47.7075	No
5	0.16210	0.11666	0.03790	27.3657	No
6	0.15437	0.26902	0.05672	12.6412	No
7	0.20116	0.40974	0.11011	7.93233	No
8	0.17174	0.27600	0.15800	11.0162	No
9	0.11855	0.30920	0.06431	17.4471	No
10	0.05444	0.08006	0.02087	47.1704	No
11	0.07375	0.06634	0.03096	69.6402	No
12	0.04810	0.06081	0.02320	72.0987	No

T=60 min.

τ	$\hat{\lambda}_\tau$ (1/min.) x 10^{-3}	$\hat{p}_\tau$	$\hat{\alpha}_\tau$ (1/min.) x 10^{-2}	$\hat{\mu}_\tau$ (1/in.)	Moment estimates?
1	0.03682	0.26535	0.01153	16.9616	No
3	0.06501	0.36284	0.03832	7.01383	No
4	0.07622	0.46914	0.03239	5.23848	No
5	0.08066	0.31808	0.02573	5.03605	No
6	0.10405	0.42892	0.07448	5.35882	Yes
7	0.15211	0.65302	0.08220	3.78090	Yes
8	0.12553	0.70344	0.04666	3.16697	No
9	0.07417	0.36766	0.09270	9.25467	Yes
10	0.05175	0.10240	0.17320	35.3552	Yes
12	0.04522	0.56988	0.08132	7.26181	No

Table 6.11 Estimated parameters for the NSWN model for Idaho Springs station, for different temporal aggregation levels, using WLS method.

T=5 min.

τ	$\hat{\lambda}_\tau$ (1/min.) x 10^{-3}	$\hat{p}_\tau$	$\hat{\alpha}_\tau$ (1/min.)	$\hat{\mu}_\tau$ (1/in.)	Moment estimates?
1	0.16418	0.01474	0.00210	218.637	No
2	0.06003	0.00142	0.00106	432.556	No
3	0.36977	0.06877	0.00643	222.743	No
4	0.12339	0.01426	0.00334	270.194	No
5	0.20010	0.06335	0.00701	192.121	No
6	0.35633	0.06600	0.02911	165.306	No
7	0.38819	0.06049	0.05376	128.452	No
8	0.30084	0.03030	0.03557	149.125	No
9	0.07828	0.16036	0.12980	10.1437	Yes
10	0.25094	0.22786	0.06460	118.933	Yes
11	0.11688	0.00478	0.00175	313.195	No

T=60 min.

τ	$\hat{\lambda}_\tau$ (1/min.) x 10^{-3}	$\hat{p}_\tau$	$\hat{\alpha}_\tau$ (1/min.) x 10^{-1}	$\hat{\mu}_\tau$ (1/in.)	Moment estimates?
1	0.11376	0.02270	0.01061	100.991	No
2	0.06155	0.00193	0.01077	335.112	No
3	0.31796	0.07977	0.04838	169.370	No
4	0.06565	0.01589	0.01505	131.646	No
5	0.12311	0.07449	0.03109	104.375	No
6	0.25678	0.22345	0.08953	36.6672	No
7	0.29976	0.38133	0.17494	16.1161	Yes
8	0.21166	0.16535	0.12317	19.6944	Yes
9	0.07616	0.87375	0.30740	1.85527	Yes
10	0.22316	0.49910	0.13728	49.5887	Yes
11	0.12123	0.00666	0.01809	237.025	Yes
12	0.03828	0.00305	0.00225	231.220	No

Table 6.12 Estimated parameters for the NSWN model for Ward station, for different temporal aggregation levels, using WLS method.

T=5 min.

τ	$\hat{\lambda}_\tau$ (1/min.) x 10^{-3}	$\hat{p}_\tau$	$\hat{\alpha}_\tau$ (1/min.) x 10^{-1}	$\hat{\mu}_\tau$ (1/in.)	Moment estimates?
1	0.11184	0.06897	0.02971	210.611	No
2	0.21883	0.00923	0.02154	294.386	No
3	0.18197	0.03250	0.04009	246.652	No
4	0.11714	0.00436	0.02448	303.154	No
5	0.66511	0.05063	0.18009	110.208	No
6	0.27984	0.13884	0.60244	27.0031	No
7	0.54393	0.01680	0.70520	400.508	Yes
8	0.31758	0.04452	0.28310	101.997	No
9	0.44569	0.09902	0.35504	129.333	No
10	0.27012	0.06096	0.06043	86.4765	No
11	0.11087	0.00748	0.01756	294.705	No
12	0.05489	0.02092	0.01248	225.078	No

T=60 min.

τ	$\hat{\lambda}_\tau$ (1/min.) x 10^{-3}	$\hat{p}_\tau$	$\hat{\alpha}_\tau$ (1/min.) x 10^{-1}	$\hat{\mu}_\tau$ (1/in.)	Moment estimates?
1	0.11979	0.07508	0.04080	211.516	No
2	0.19124	0.01087	0.01737	222.789	No
3	0.18002	0.02488	0.04294	326.609	No
4	0.09455	0.01674	0.01771	64.9193	No
5	0.45728	0.16347	0.07477	24.3165	No
6	0.20873	0.17979	0.24207	16.3376	No
7	0.33502	0.32353	0.07558	13.6774	Yes
8	0.41447	0.26557	0.19848	23.1732	No
9	0.21553	0.23946	0.03535	26.5230	No
10	0.21506	0.08576	0.04701	50.2992	No
11	0.11516	0.00725	0.01941	320.881	No
12	0.08603	0.02726	0.02703	274.901	No

is as follows: $\hat{\lambda}_\tau$, $\hat{\alpha}_\tau$ and $\hat{\mu}_\tau$ decrease with an increase in the temporal scale, although the variation in $\hat{\lambda}_\tau$ is mild. On the other hand, $\hat{p}_\tau$ is subject to a steep increase. The same type of behavior can be observed in Table 6.4 for the means of the estimated parameters.

6.12 Comparison of estimation methods

In the previous sections, two different estimation methods have been used to obtain parameters for the NSWN precipitation model. Although they both are based on moment properties of the process, the first one was named method of moments and the second one Weighted Least Squares (WLS).

The first point of comparison is given in Table 6.13, where biases, root mean square errors (RMSE) and number of failures are given for simulation 1 of the NSWN precipitation process. In deriving these indicators, a total of 12 sample points, minus number of failures, were used for each parameter. Table 6.13 is illustrated further by Figures 6.4 and 6.5, where estimation results for each parameter have been plotted, along with average and population values. Note that for the parameter associated to the distribution of the White Noise terms, the inverse value is plotted.

For T=60 min., method of moments performs better in terms of bias, while WLS does in terms of RMSE. For T_a=1440 min., method of moments appears superior. However, WLS reduces considerably the number of failures in estimation. WLS incorporates information not used by the method of moments and since this information carries larger sample variability, one should expect, as it happened, a larger

Table 6.13 Bias, Root Mean Square Error (RMSE) and number of failures for parameters estimated for NSWN simulation 1, for two different temporal scales

T=60 min.

Parameter	Method of Moments Bias	RMSE	Fail.	WLS Bias	RMSE	Fail.
λ (1/min.)	2.87×10^{-7}	2.60×10^{-5}	0	-3.50×10^{-6}	8.64×10^{-6}	0
p	2.46×10^{-4}	1.71×10^{-2}	0	-1.38×10^{-3}	7.02×10^{-3}	0
α (1/min.)	8.97×10^{-5}	6.89×10^{-4}	0	-4.43×10^{-5}	1.87×10^{-4}	0
μ (1/in.)	-3.69×10^{-1}	1.08	0	5.34×10^{-1}	1.55	0

T=1440 min.

Parameter	Method of Moments Bias	RMSE	Fail.	WLS Bias	RMSE	Fail.
λ (1/min.)	1.15×10^{-5}	1.35×10^{-5}	5	1.00×10^{-5}	1.43×10^{-5}	2
p	-1.75×10^{-1}	1.94×10^{-1}	5	-2.64×10^{-1}	3.30×10^{-1}	2
α (1/min.)	5.83×10^{-4}	8.12×10^{-4}	3	1.08×10^{-3}	1.23×10^{-3}	2
μ (1/in.)	-1.73×10^{1}	1.34×10^{1}	5	-1.84×10^{1}	1.70×10^{1}	2

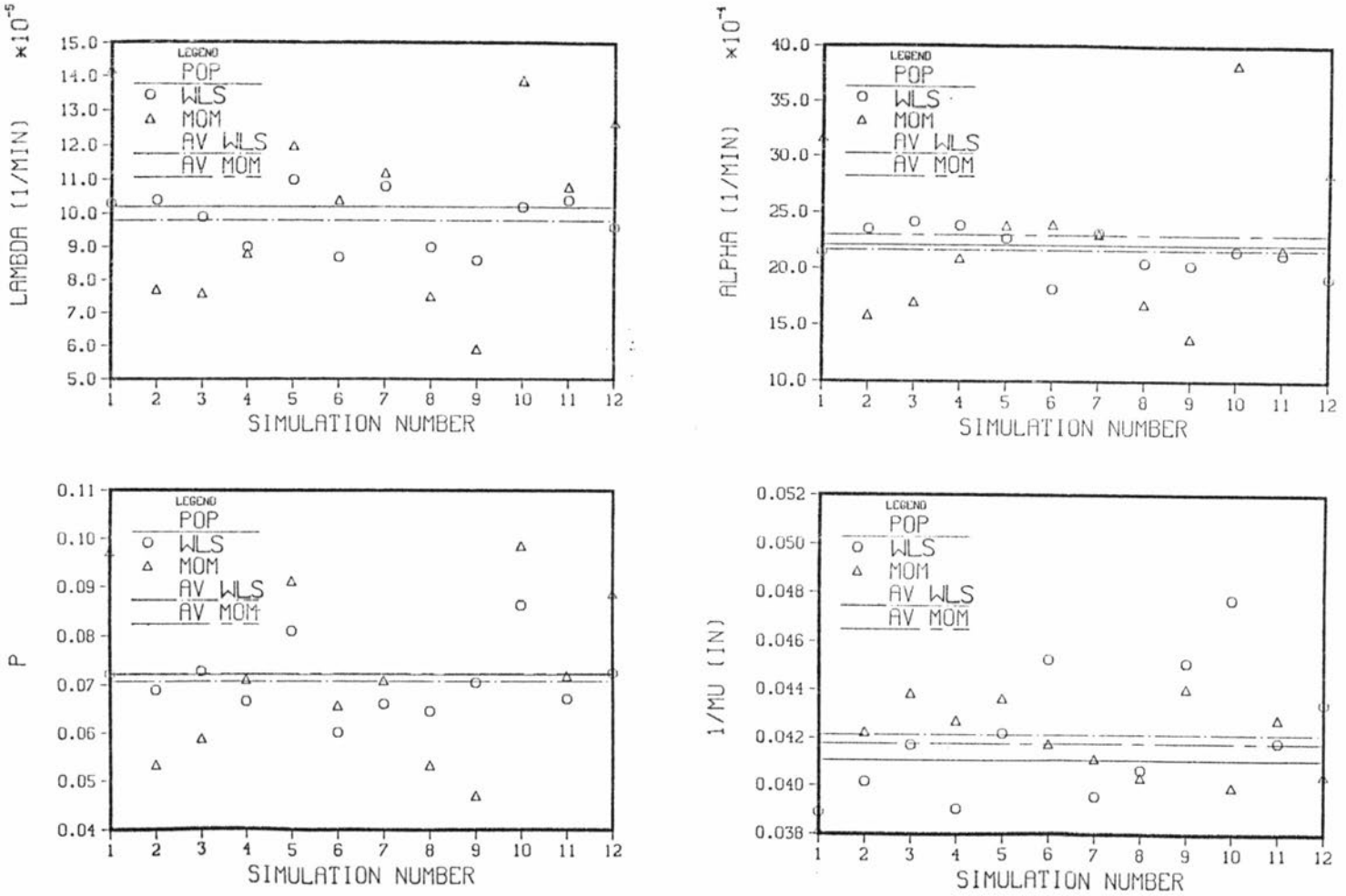


Figure 6.4 Estimation results for NSWN process simulation 1, for T= 60 min.

(MOM: method of moments, OPT: weighted least squares method).

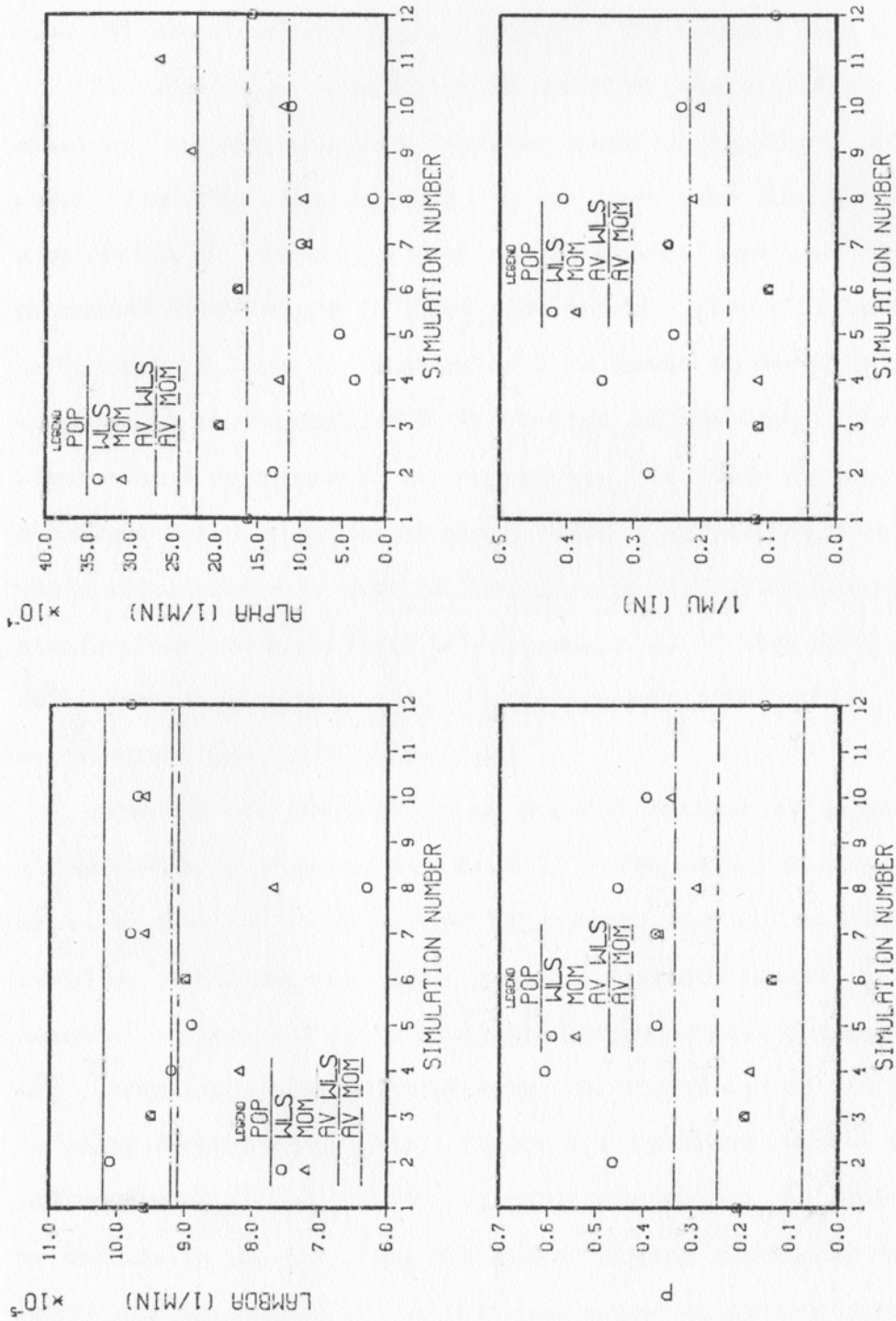


Figure 6.5 Estimation results for NSWN process simulation 1, for T_a = 1440 min.

(MOM: method of moments, OPT: weighted least squares method).

variability on parameter estimates. The ability of WLS to increase the number of feasible solutions is also seen when simulation and historical results are lumped together. When method of moments is used, 33 out of 120 cases gave failure. WLS reduced this number to 5.

Two additional simulations of the NSWN precipitation process were obtained, in this case with periodic monthly parameters. They are not used often in this report, since they gave similar results to simulation 1. They are mentioned because some of the figures presented later belong to these simulations. They will be referred to as simulation 2 and 3. Simulation 2 is based on monthly parameters estimated for Greenland 9 SE station at the hourly level, while simulation 3 corresponds to parameters for Idaho Springs station, obtained for 5 min. measurement scale. Simulation 2 has 49 years, while simulation 3 is made of 5 years. An important result for the simulations, both periodic and not periodic, is that they are similar to historical samples, in terms of monthly statistics, especially correlograms and their variability.

Some of the properties of the WLS estimation procedure are illustrated in Figures 6.6 to 6.11. The upper plot in Figure 6.6 shows a case in which method of moments and WLS produce similar results, while in the lower plot WLS performs better than method of moments. Figure 6.7 shows a typical estimated correlogram for T_a=60 min., with large sample variability, for simulation 2, and the benefit in using decreasing weights. Figure 6.8 illustrates the use of the different weighting methods, along with change in the number of terms to include in the objective function. Figure 6.9 shows the type of correlograms fitted using WLS when method of moments does not work. Figure 6.10 gives examples of fitted correlograms using WLS technique

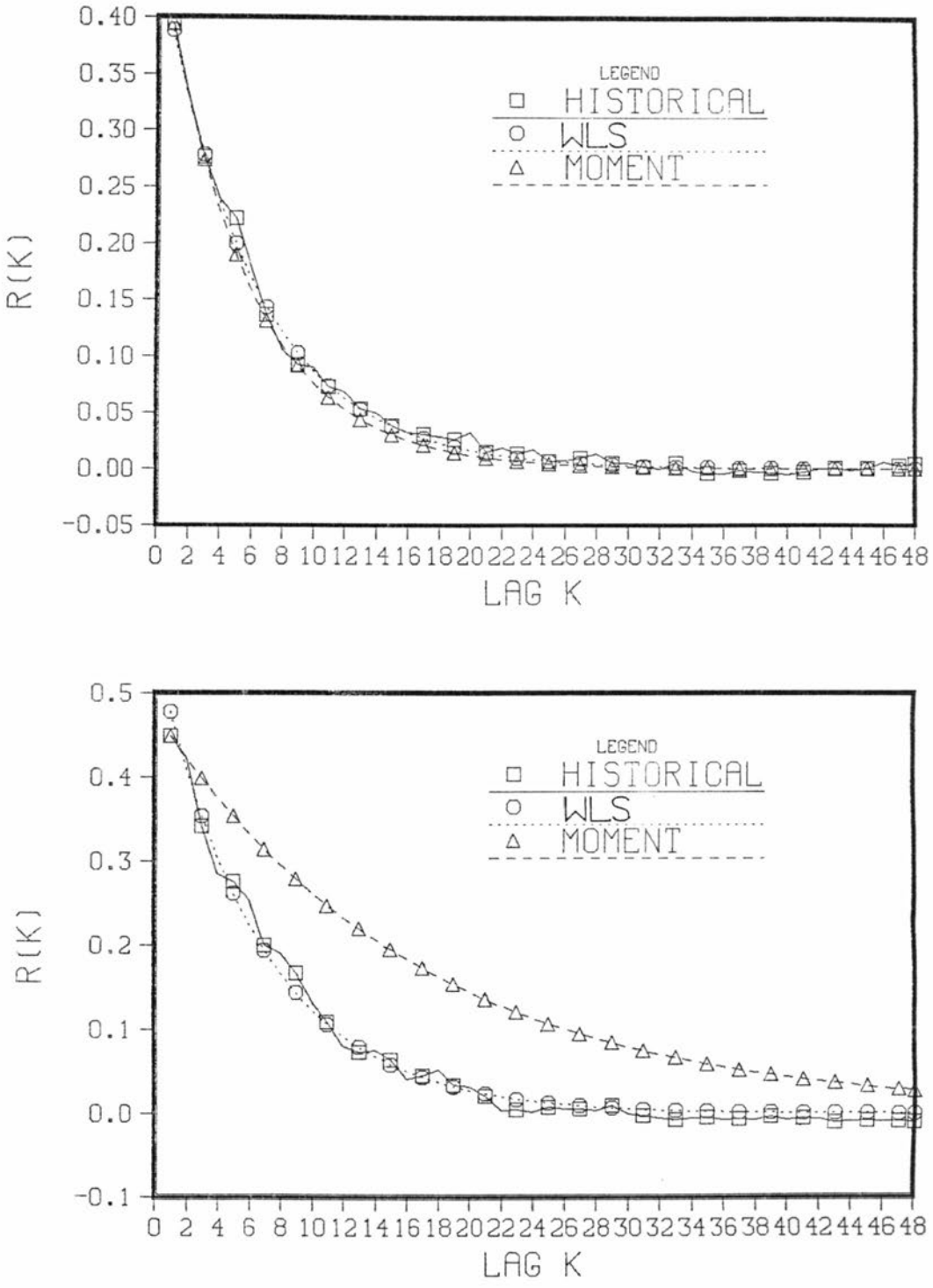


Figure 6.6 Historical and fitted correlograms for NSWN simulation 2, for T=60 min., for months 11 and 12.

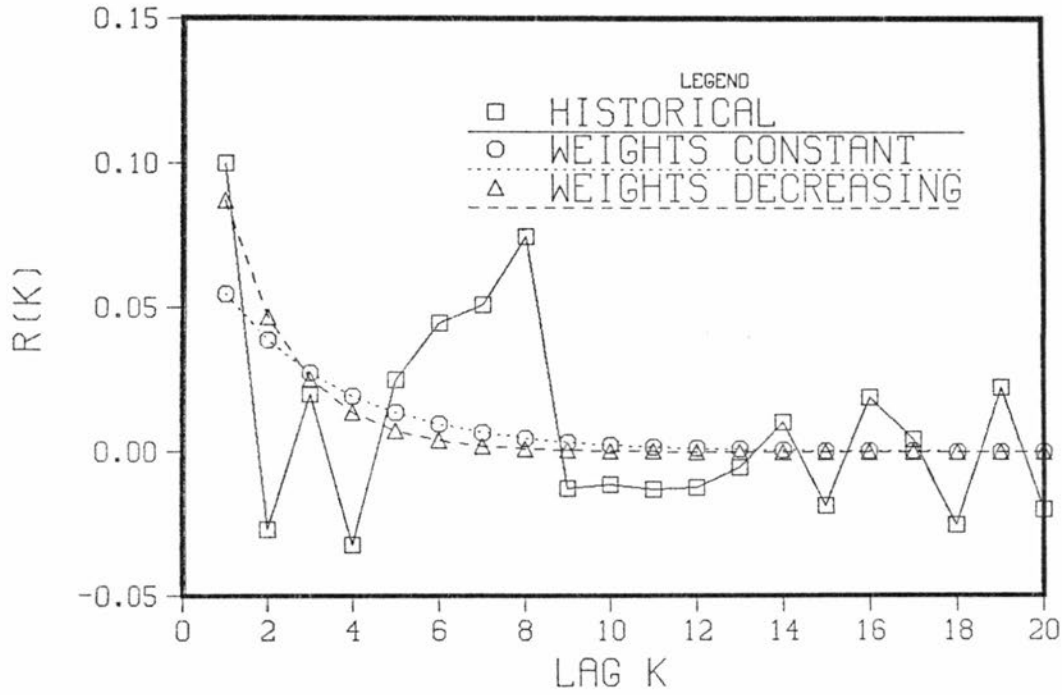


Figure 6.7 Role of weighting method on fitted correlograms (T_a=1440 min., simulation 2).

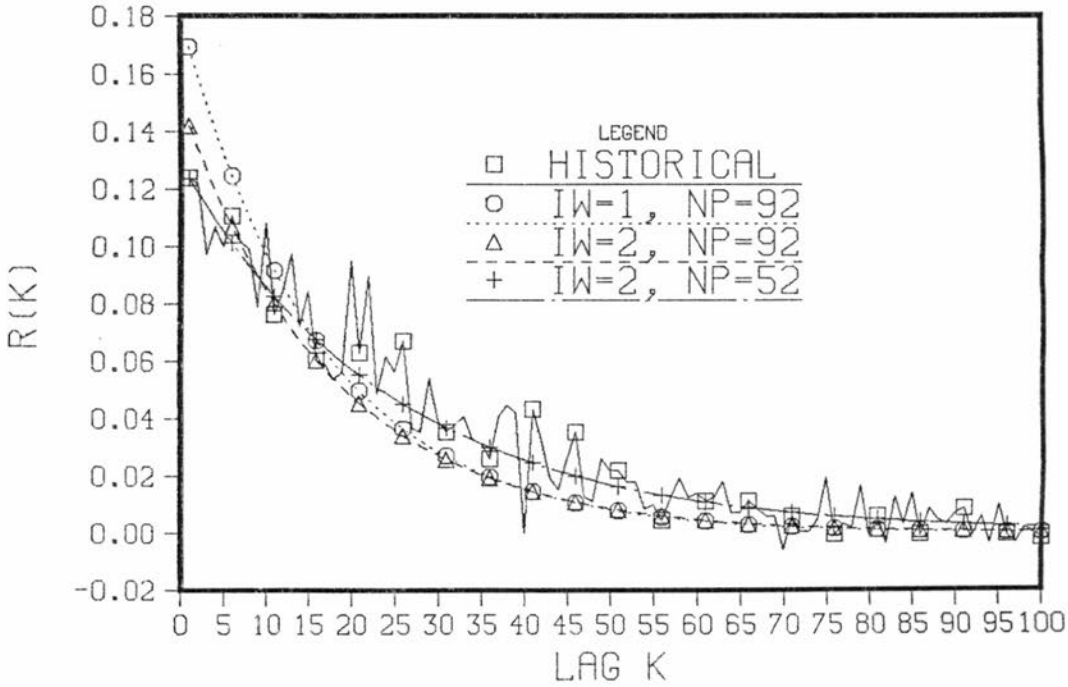


Figure 6.8 Role of weighting methods and number of terms in the objective function on fitted correlograms (T=5 min., simulation 3): IW=1, weights are constant; IW=2, weights decrease with lag; NP: number of terms in objective function.

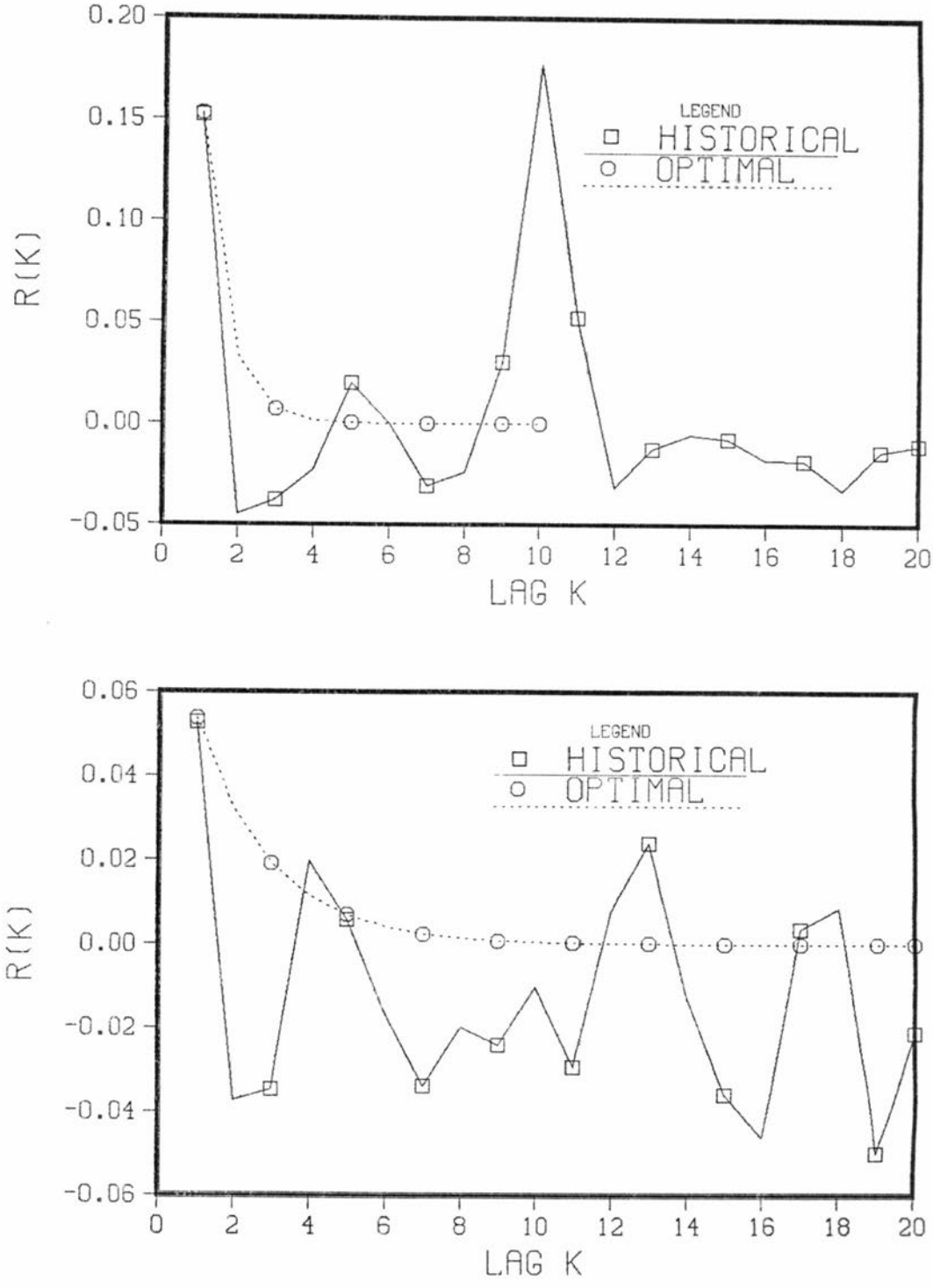


Figure 6.9 Historical and NSWN fitted correlograms (WLS), for Denver Wsfo Ap station, for months 03 and 04, for T_a=1440 min..

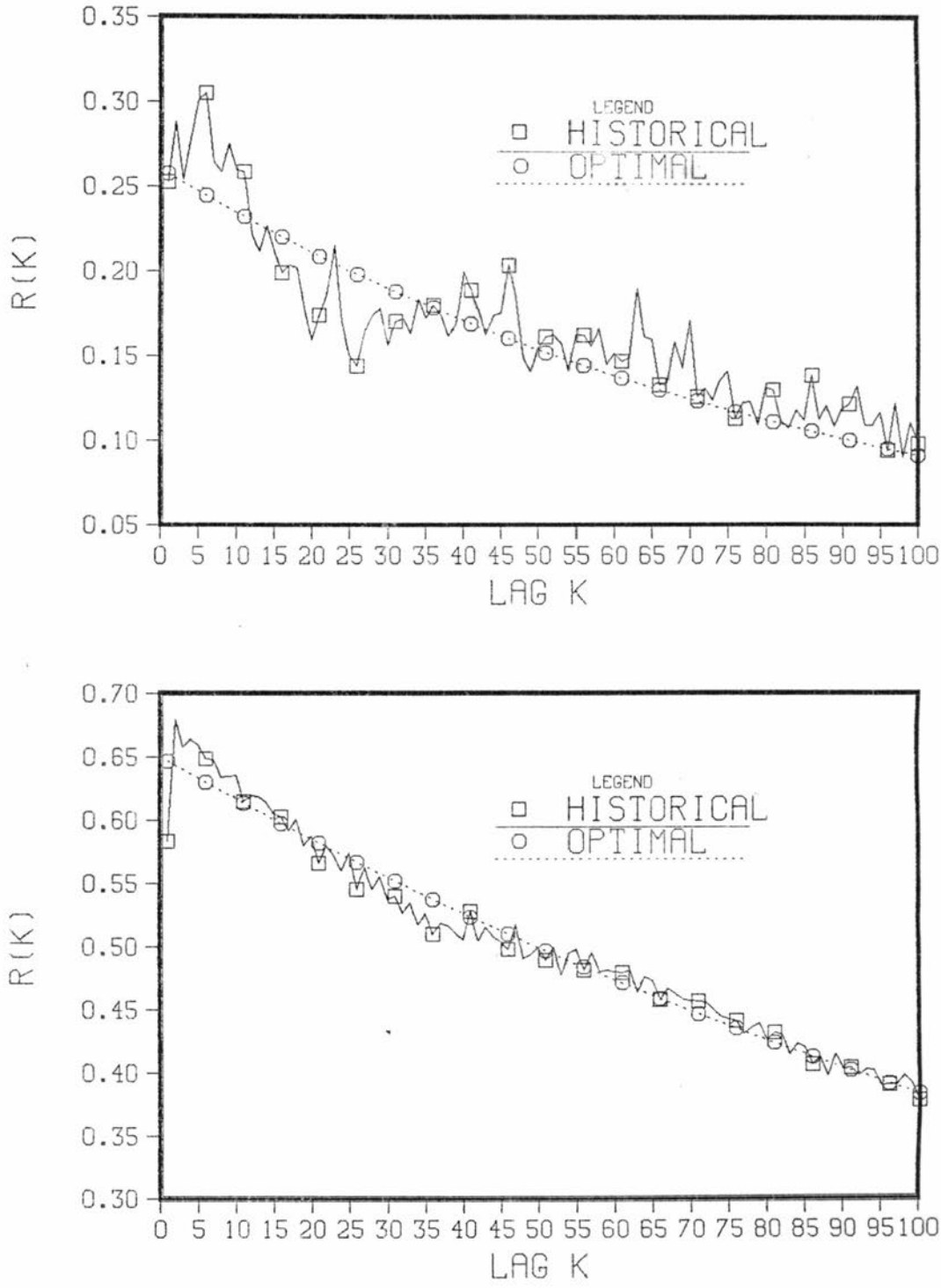


Figure 6.10 Historical and NSWN fitted correlograms (WLS), for Idaho Springs station, for months 01 and 02, for T=5 min..

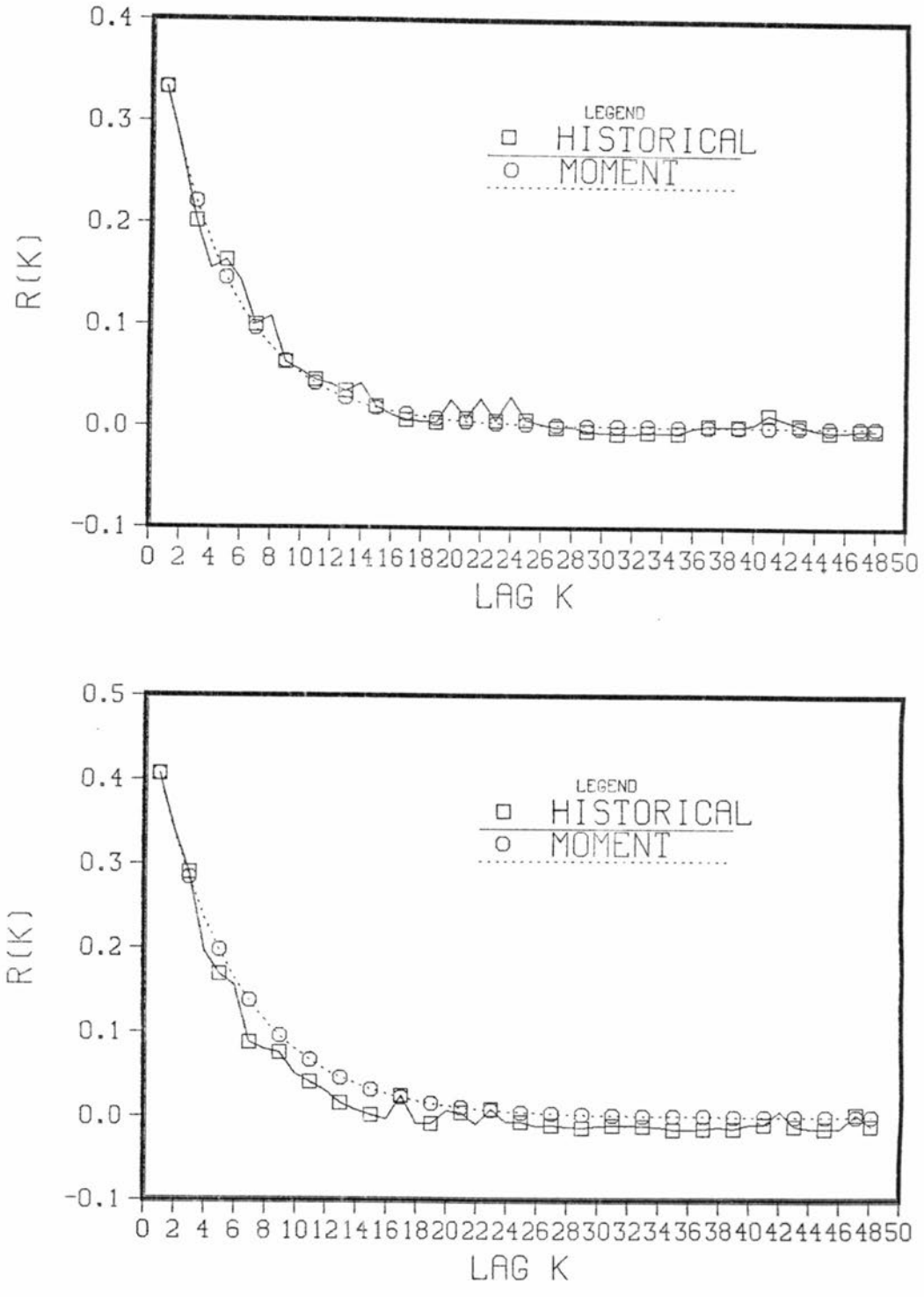


Figure 6.11 Historical and NSWN fitted correlograms (moment), for Greenland 9 SE station, for months 01 and 02, for T=60 min..

when method of moments fails, due to a partially increasing correlogram with the lag. Finally, Figure 6.11 is an advocate of the method of moments, since in this case method of moments works better than WLS.

WLS is adopted as the standard method for this study. Despite the fact that method of moments seems to exhibit smaller bias and, in some cases, smaller variability on estimated parameters, optimal estimation technique is viewed as a smoothing tool, which tends to fit closer correlograms to the sample counterpart, uses more information, is able to take into account sample variability in the historical statistics and exhibits, consequently, a larger rate of success. However, results presented above show that method of moments is useful in many cases, and for this reason it has been built as an option within the WLS procedure. Both methods are complementary and they must be used in conjunction.

6.13 Extrapolation of the marginal distribution for N(T)

In Section 6.5 a procedure to compute the marginal pdf of the number of occurrences in a time interval of length T for the NS process was described. However, this procedure is limited to values of the random variable between 0 and 20.

When parameter estimates obtained in Section 6.11 are used, it is possible to approximate, using $f_{N(T)}(n)$, n from 0 to 20, mean and variance of the number of occurrences in a time interval of length T. These approximations can be compared to the exact values obtained by means of eqs. (6.51) and (6.52).

The procedure described above was applied to all samples available for this research, precipitation recording stations and NSWN

simulations. It was observed that biases in E[N(T)] and Var(N(T)) were large enough, approximately 40%, especially for large aggregation scales (60 min. and 1440 min.), as to require some improvement in the number of terms for which $f_{N(T)}(n)$ could be computed. It was decided to formulate and test some models to extrapolate the distribution of N(T) beyond 20.

Two models were considered. The first one is empirical and it was chosen after examination of many plots of $Ln[f_{N(T)}(n)]$ versus n. The model is of the form

$$f_{N(T)}(n) = \exp[-(An^{\theta} + Bn + C)] \qquad (6.107)$$

where θ is an exponent and A, B and C are model coefficients. The purpose of θ is to mimic some slight curvature observed in the plots of $Ln[f_{N(T)}(n)]$ versus n. Coefficients A, B and C, for a given value of θ, are computed using standard linear regression techniques, after eq. (6.107) is transformed to the logarithmic space.

In order to estimate coefficients in the empirical model, besides of defining the value of θ, one must choose the number of points, n_p, to enter the regression procedure. It is important to clarify that the purpose of (6.107) is to approximate the tail of the distribution, and therefore the points are counted backwards from the last point, n=20. For example, if 15 points were to be used to estimate coefficients, these would be from n=6 to n=20.

Several regression runs were made with values of n_p ranging from 3 to 20 and values of θ ranging from 0.01 to 2.00. After examination of all runs, for recorded and simulated data, it was decided that θ=0.05 and n_p=14 provided an adequate fit for all cases. The goodness of fit was judged by the multiple correlation coefficient and by the standard error or standard deviation of residuals. These two

indicators are computed for the transformed data. A third indicator, called standard error in the original space, was used to measure the standard deviation of the residuals for untransformed data. Multiple correlations coefficients were in all cases very close to 1.0, within the 5th or 6th decimal place.

Although attempts were made in order to find an optimal value for θ, using least square theory, results showed that such estimate tends to either 0.0 or 1.0, but at these points expressions used for coefficient computations become indeterminate. Furthermore, the empirical model in (6.107) was always superior to its linear version, obtained by making θ=0.0.

Once the model in eq. (6.107) is defined, $f_{N(T)}(n)$ is extrapolated up to a value n_m, in such a way that the bias in the variance of N(T) is smaller than a given tolerance an that $f_{N(T)}(n_m)$ is smaller than another tolerance value. Bias in Var(N(T)) dominates bias in E[N(T)]. Next, values of $f_{N(T)}(n)$ computed according to (6.107), for n between 21 and n_m, are rescaled in such way that the following condition is met:

$$\sum_{n=0}^{n_m} f_{N(T)}(n) = 1.0$$

The second model tested is probabilistic and it represents the binomial negative distribution with parameters p an r (Mood et al., 1974). Assuming that N(T) follows this model

$$f_{N(T)}(n) = \left(\begin{matrix} n+r-1 \\ n \end{matrix} \right) p^r(1-p)^n \, I_{\{0,1,2,..\}}(n) \qquad (6.108)$$

where $0<p<1$ and $r>0$. Note that r is a real number. The coefficient in front of eq. (6.108) is an extension of the binomial coefficient for integer numbers and is computed as

$$\binom{n+r-1}{n} = \frac{1}{n!}\prod_{i=0}^{n-1}(n+r-1-i) \qquad (6.109)$$

Mean and variance for the negative binomial distribution are

$$E[N(T)] = r(1-p)/p \qquad (6.110)$$

$$Var(N(T)) = r(1-p)/p^2 \qquad (6.111)$$

Using computed values for E[N(T)] and Var(N(T)) with eqs. (6.51) and (6.52), parameters r and p are obtained from eqs. (6.110) and (6.111). The same extrapolation procedure as that described for the empirical model is used with the negative binomial model.

The basis for selecting the negative binomial model as an option rests on two points. First, it is a true discrete marginal pdf. Second, it resembles some of the features present in the empirical model, like an exponential tail.

In order to compare both models, same indicators as those used for the empirical model were determined for the binomial model: correlation coefficient and standard errors in the transformed and untransformed spaces. Biases for mean and variance of N(T) were also computed for both models after extrapolation,. Finally, the ratio between $f_{N(T)}(21)$ and $f_{N(T)}(20)$ was obtained. This ratio is expected to be less than 1.0.

The negative binomial model dominated in terms of biases. In terms of regression indicators, the empirical model performed better than the negative binomial model. These results were expected, since they represent the way in which each model is fitted. However, for the ratio of probabilities evaluated at 21 and 20, the negative binomial model failed several times to give values smaller than 1.0.

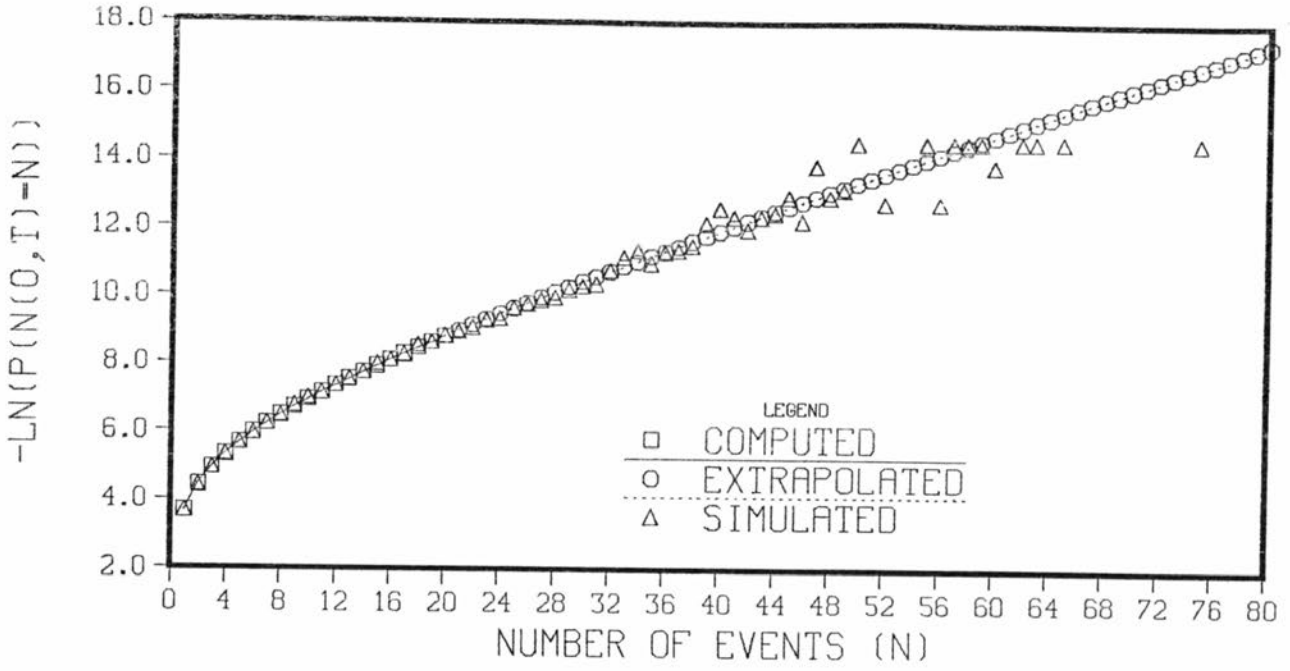
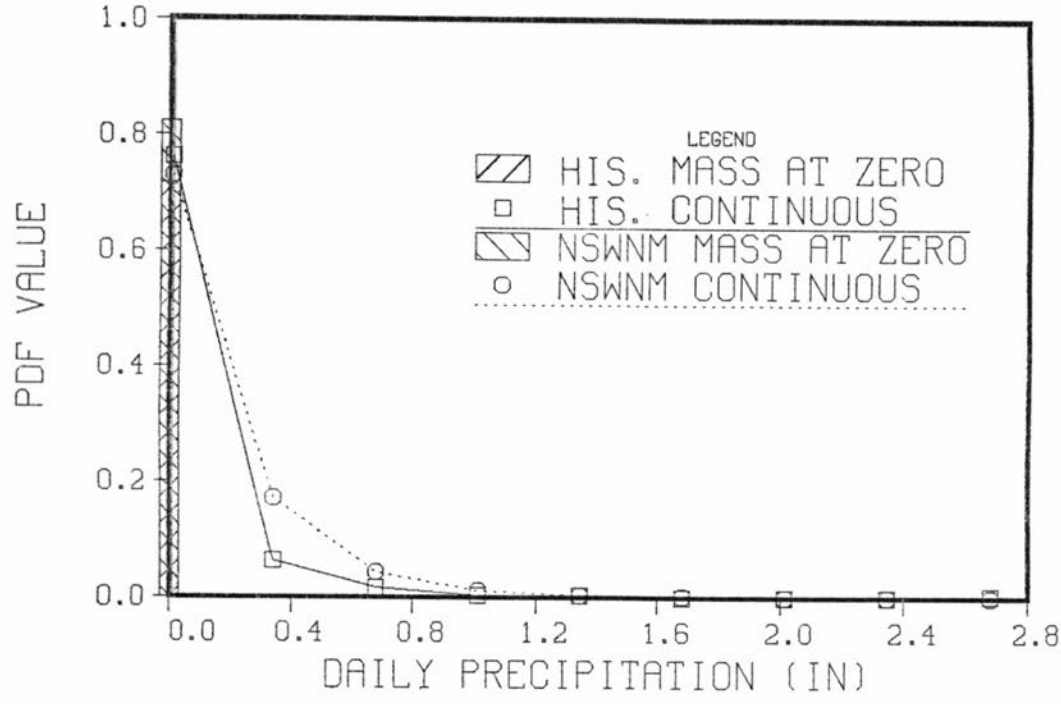


Figure 6.12 Example of computed, empirically extrapolated and simulated distributions for the N(T) in the NS process, for λ=0.00009475 1/min., p=0.02273, α=0.03257 1/min. and T=60 min..

On the basis of this analysis, the empirical model was selected as the one to be used to extrapolate the pdf of the number of epics in the interval of length T for the NS process. Figure 6.12 reinforces this conclusion, where examples of computed, empirically extrapolated and simulated distributions are plotted in the transformed space.

The empirical model was tested by computing the distribution of the number of occurrences, the distribution of the total amount of precipitation in the interval and the distribution of the number of epics conditional on the total amount of precipitation, for all the data sets, recorded and simulated, available for this investigation. All distributions behaved well. Substantial changes are introduced in the last two distributions when the empirical extrapolation model is used.

As given before for the PWN model, a last point is presented here to judge the goodness of fit of the NSWN precipitation model to the precipitation data used in this study. Figures 6.13 and 6.14 give examples of sample and computed pdf for the amount of precipitation for some months in the precipitation recording stations, for the aggregation scale. The theoretical pdf for the amount of precipitation is computed by means of eq. (6.99). In general, observed fits are good. Similar results were obtained for simulated samples.



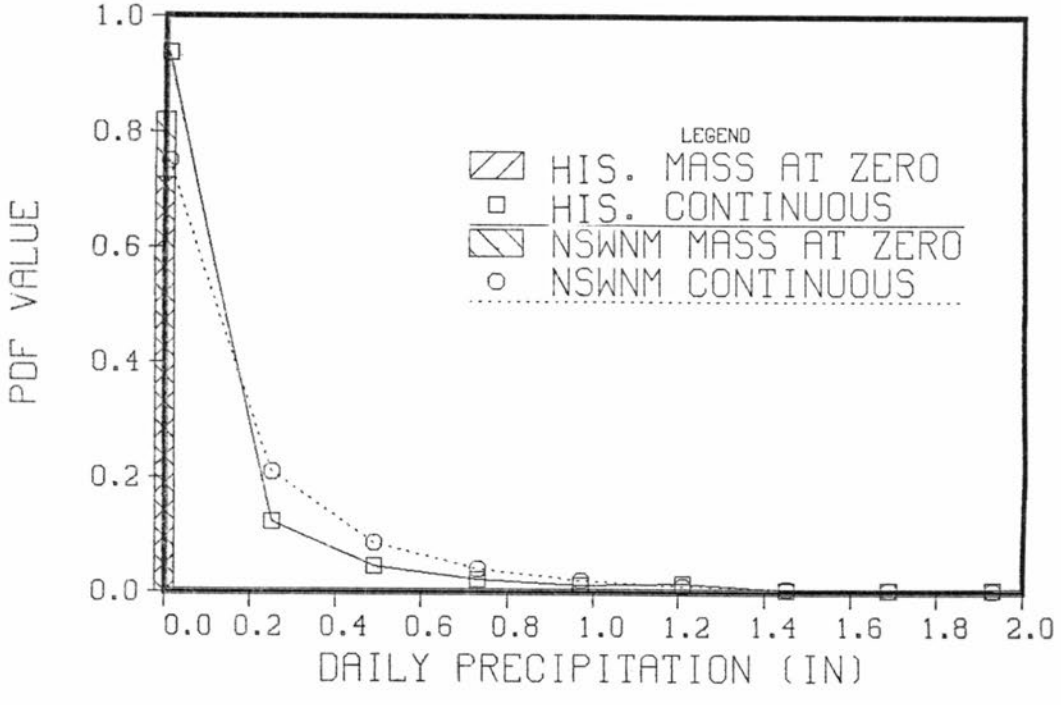


Figure 6.13 Historical and NSWN computed probability distribution functions for the amount of precipitation, for Denver Wsfo Ap station, for months 03 and 08 and for T_a = 1440 min..

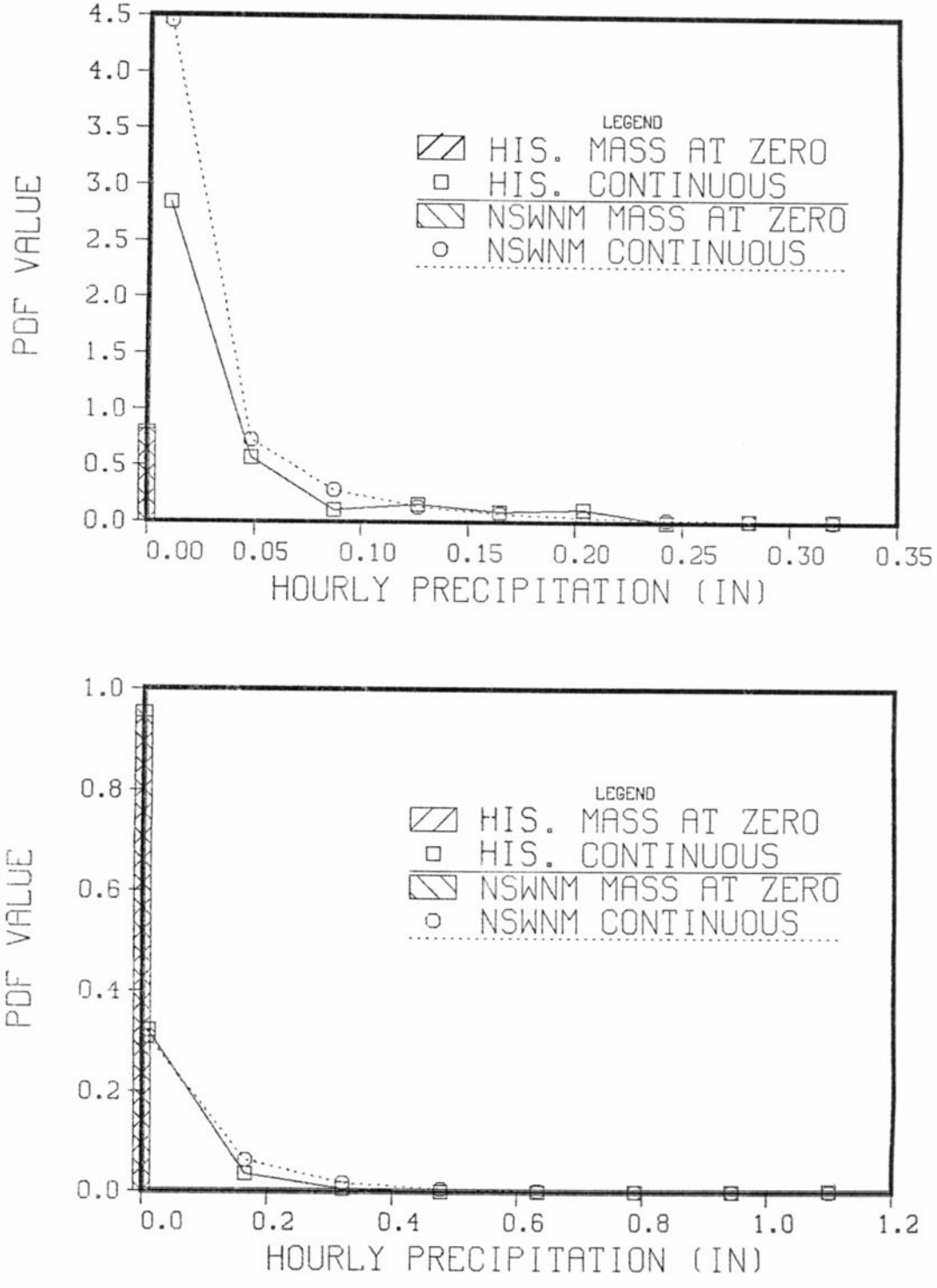


Figure 6.14 Historical and NSWN computed probability distribution functions for the amount of precipitation, for Idaho Springs station, for months 02 and 08 and for T_a = 60 min..

Chapter 7

NEYMAN-SCOTT WHITE NOISE DISAGGREGATION MODEL FOR PRECIPITATION

7.1 Introduction

Based on the results given in Chapter 6, a Neyman-Scott White Noise Disaggregation (NSWND) model for precipitation is developed in this chapter. The outline is similar to Chapter 5, based on the consideration that both models, Poisson White Noise and Neyman-Scott White Noise, can be expressed as compound processes. Disaggregation procedure and disaggregation model structure are the same, the same type of distributions are required and therefore notation is the same, as compared to the PWN disaggregation model. The difference stems from the fact that both models are controlled by different counting processes. In this regard, the reader is invited to review Sections 5.1 and 5.2, where the disaggregation procedure is described and model structure is presented.

In this chapter, the required distributions for the NSWND model are presented. The disaggregation algorithm is explained. The model is tested using samples drawn from a NSWN process and the data for the precipitation recording stations used in this study. Finally, some alternatives are explored in order to improve model performance.

The NSWND model developed in this chapter is not unique. It is entirely based on the univariate distribution for the number of occurrences in a time interval of length T. Different models could be obtained if the history of the process up to a given time is

considered, for example, the bivariate distribution for the number of occurrences in two consecutive disjoint time intervals. So far, the reader must be aware of the difficulties encountered with the distribution of the number of occurrences in the interval of length T. These difficulties would be increased enormously if the mentioned characteristics were to be included in a different NSWND model. In this sense, new light must be added to the analytical treatment of the NS process.

7.2 Distribution of N(T) conditional on Y

The distribution of the number of occurrences in a time interval of length T, conditional on the precipitation recorded in the same interval, is obtained from eq. (5.5)

$$f_{N(T)|Y}(n) = \frac{f_U^{(n)}(y)\ f_{N(T)}(n)}{\sum_{k=1}^{\infty} f_U^{(k)}(y)\ f_{N(T)}(k)}\ I_{\{1,2,..\}}(n) \qquad (7.1)$$

where, for this case, N(T) is the Neyman-Scott (NS) counting process and $f_U^{(n)}(.)$ is the n fold convolution of the common distribution of the White Noise terms. Since no general analytical expressions are available for $f_{N(T)}(n)$, the next feasible step in eq. (7.1) is to replace the convolution by a gamma distribution with scale parameter μ and shape parameter n, under the assumption that White Noise terms are independently and identically distributed following an exponential distribution with parameter μ. The final result is

$$f_{N(T)|Y}(n) = \frac{\frac{\mu^n}{(n-1)!}\ y^{n-1}\ f_{N(T)}(n)}{\sum_{k=1}^{\infty} \frac{\mu^k}{(k-1)!}\ y^{k-1}\ f_{N(T)}(k)}\ I_{\{1,2,..\}}(n) \qquad (7.2)$$

In order to compute the distribution in (7.2), values for $f_{N(T)}(n)$ are obtained using the procedure described in Section 6.13, which includes the empirical extrapolation model.

7.3 Distribution of U_i, conditional on Y, N(T) and $\underline{U}_{i-1}$

The objective in this section is to present the distribution of the ith White Noise term, U_i, given the recorded precipitation in the interval of length T, Y, given N(T), and given that i-1 previous White Noise terms, U_j, j=1,..,i-1, have taken values u_j, j=1,..,i-1. This distribution was already derived in Section 5.4. The general result, when the White Noise terms are independently and identically distributed as $f_U(u)$, is given in eq. (5.25)

$$f_{U_i|(Y,\underline{U}_{i-1})}(u_i) = f_U(u_i) \frac{f_U^{(n-i)}(R_{i+1})}{f_U^{(n-i+1)}(R_i)} I_{(0,R_i)}(u_i) \tag{7.3}$$

with

$$R_i = y - \sum_{j=1}^{i-1} u_j \tag{7.4}$$

$$R_1 = y$$

For i=1, eq. (7.3) simplifies to

$$f_{U_1|Y}(u_1) = f_U(u_1)\, f_U^{(n-1)}(y-u_1) \,/\, f_U^{(n)}(y)\, I_{(0,y)}(u_1) \tag{7.5}$$

When the White Noise terms follow the exponential distribution with parameter μ, the following set of expressions is obtained

$$f_{U_i|(Y,\underline{U}_{i-1})}(u_i) = (n-i)\,(R_i-u_i)^{n-i-1} \,/\, R_i^{\,n-i}\, I_{(0,R_i)}(u_i) \tag{7.6}$$

$$f_{U_1|Y}(u_1) = (n-1)\,(y-u_1)^{n-2} \,/\, y^{n-1}\, I_{(0,y)}(u_1) \tag{7.7}$$

$$F_{U_i|(Y,\underline{U}_{i-1})}(u_i) = \frac{R_i^{n-i} - \left(R_i - u_i \right)^{n-i}}{R_i^{n-i}} I_{(0,R_i]}(u_i) \qquad (7.8)$$

$$u_i = R_i - \left[R_i^{n-i} - (1 - F_i) \right]^{1/(n-i)} \qquad (7.9)$$

where F_i is a value assigned to the cdf, $0 \leq F_i \leq 1$ and eq. (7.9) gives the corresponding quantile. As before for the PWND model, eq. (7.9) facilitates simulation procedures.

7.4 Distribution of $\underline{T}_n$ conditional on N(T)

Finding the distribution of the vector of arrival times $\underline{T}_n$ in the interval (0,T), conditional on n arrivals in the same interval, was an easy task for the PWND model. However, for the NSWND model this distribution is far from easy to find. No solution is provided for this problem in the literature reviewed for this research.

The approach followed here to find the distribution of $\underline{T}_n$ is based on the one-to-one relationship between the counting specification and the arrival times specification for a point process. Analytical derivations were carried out successfully for n=1 and n=2. For n=3, although a result was obtained for the joint pdf of $\underline{T}_3$, it was considered too long and too difficult to verify and complete it to the required distributions. For $n \leq 2$, complete results are presented and used in the NSWND model. For n=3, the result is presented for sake of completeness. In order to provide arrivals times for $n \geq 3$, simulation and sampling procedures were designed and tested.

Derivations for the joint distributions of $\underline{T}_n$ for different n values are lengthy. Details are given in Cadavid et al. (1991). The general procedure is described in the following paragraph.

Let $\underline{T}_n = [T_1, T_2, .., T_n]$ be the vector of random variables representing arrival times in the interval (0,T). Let $\underline{t}_n = [t_1, t_2, .., t_n]$ be another vector such that $0 \leq t_1 \leq t_2 \leq .. \leq t_n \leq T$. In general, the joint pdf for $\underline{T}_n$ is obtained via the operation (Taylor and Karlin, 1984)

$$f_{\underline{T}_n | N(T)}(\underline{t}_n) = \lim_{\Delta t_1, \Delta t_2, .., \Delta t_n \to 0} \left\{ \frac{1}{\Delta t_1 \Delta t_2 .. \Delta t_n \, f_{N(T)}(n)} \right.$$

$$P\Big[N(0,t_1)=0,\ N(t_1,t_1+\Delta t_1)=1,\ N(t_1+\Delta t_1,t_2)=0, ..,$$

$$\left. N(t_{n-1},t_n+\Delta t_n)=1,\ N(t_n+\Delta t_n,T)=0 \Big] \right\} \qquad (7.10)$$

In order to clarify eq. (7.10), let the event within brackets be denoted by A. This event involves a total of 2n+1 intervals. The key step in the derivation of (7.10) is to realize that the joint probability of A and N(T)=n is precisely the probability of A.

The term P[A] is computed from the joint probability generating function for 2n+1 intervals, using the results given in eqs. (6.36) and (6.37), with k=2n+1 and t_k=T, and the relationship defined by eq. (6.43). In this last expression, care must be exercised in taking the derivatives, a total of n, for those z_j associated with intervals having one arrival.

The distribution of T_1, conditional on one arrival in the interval (0,T), is

$$f_{T_1 | N(T)=1}(t_1) = \frac{1}{H_1} \left[\frac{\alpha q \exp(-\alpha t_1)}{1-q \exp(-\alpha T)} + \frac{\alpha p q \exp[-\alpha(T-t_1)]}{1-q \exp[-\alpha(T-t_1)]} \right.$$

$$\left. + \alpha p \right] I_{(0,T)}(t_1) \qquad (7.11)$$

with

$$H_1 = \frac{q\phi(T)}{1-q\ \exp(-\alpha T)} - p\ \mathrm{Ln}\ \frac{p}{1-q\ \exp(-\alpha T)} + \alpha p T \qquad (7.12)$$

Recall that $\phi(T) = 1-\exp(-\alpha T)$ and $q=1-p$. It is easy to verify how for $p=1$ (Poisson process) eq. (7.11) collapses into the uniform distribution in $(0,T)$, as expected. Integration of eq. (7.11) yields the cdf for T_1

$$F_{T_1|N(T)=1}(t_1) = \frac{1}{H_1}\left[\frac{q\phi(t_1)}{1-q\ \exp(-\alpha T)} - p\ \mathrm{Ln}\ \frac{p}{1-q\ \exp(-\alpha t_1)} + \alpha p t_1\right] I_{(0,T]}(t_1) + I_{(T,\infty)}(t_1) \qquad (7.13)$$

For simulation purposes eq. (7.13) has to be inverted numerically.

The probability distribution function of $\underline{T}_2$, conditional on two arrivals in the interval $(0,T)$, is

$$f_{\underline{T}_2|\ N(T)=2}(\underline{t}_2) = \frac{2}{H_2}\left\{\frac{\alpha^2 q^2 \exp[-\alpha(t_1+t_2)]}{[1-q\ \exp(-\alpha T)]^2} + \frac{\alpha^2 pq\ \exp[-\alpha(t_2-t_1)]}{\{1-q\ \exp[-\alpha(T-t_1)]\}^2} + \lambda\alpha \prod_{i=1}^{2}\left[\frac{q\ \exp(-\alpha t_i)}{1-q\ \exp(-\alpha T)} + \frac{pq\ \exp[-\alpha(T-t_i)]}{1-q\ \exp[-\alpha(T-t_i)]} + p\right]\right\} I_{(0,t_2)}(t_1)\ I_{(t_1,T)}(t_2) \qquad (7.14)$$

with

$$H_2 = \left\{-2pq\ \mathrm{Ln}\ \frac{p}{1-q\ \exp(-\alpha T)} + 2\alpha pqT + \frac{q\phi(T)}{[1-q\ \exp(-\alpha T)]^2}[3q-2-(2q-1)q\ \exp(-\alpha T)] + \frac{\lambda}{\alpha}\left[p\ \mathrm{Ln}\ \frac{1-q\ \exp(-\alpha T)}{p} + \frac{q\phi(T)}{1-q\ \exp(-\alpha T)} + \alpha p T\right]^2\right\} \qquad (7.15)$$

The cumulative distribution function for $\underline{T}_2$, conditional on N(T) being 2, is given in Cadavid et al., (1991).

The marginal pdf for T_1 is computed as

$$f_{T_1|\ N(T)=2}(t_1) = \int_{t_1}^{T} f_{\underline{T}_2|\ N(T)=2}(\underline{t}_2)\ dt_2$$

and the final result is

$$f_{T_1|\ N(T)=2}(t_1) = \frac{2}{H_2}\left\{ \frac{\alpha q^2 \exp(-\alpha t_1)[\exp(-\alpha t_1)-\exp(-\alpha T)]}{[1-q\ \exp(-\alpha T)]^2} \right.$$

$$+\alpha pq\ \frac{\phi(T-t_1)}{\{1-q\ \exp[-\alpha(T-t_1)]\}^2} + \lambda \left[\frac{q[\exp(-\alpha t_1)-\exp(-\alpha T)]}{1-q\ \exp(-\alpha T)} \right.$$

$$\left. + p\ \mathrm{Ln}\ \frac{1-q\ \exp[-\alpha(T-t_1)]}{p} + \alpha p(T-t_1) \right] \left[\frac{q\ \exp(-\alpha t_1)}{1-q\ \exp(-\alpha T)} + \right.$$

$$\left. \left. \frac{pq\ \exp[-\alpha(T-t_1)]}{1-q\ \exp[-\alpha(T-t_1)]} + p \right] \right\} I_{(0,T)}(t_1) \qquad (7.16)$$

From eq. (7.16), the marginal cdf for T_1, conditional on two arrivals in the interval, is

$$F_{T_1|\ N(T)=2}(t_1) = \frac{2}{H_2}\left\{ \frac{q^2\phi(t_1)[1-2\exp(-\alpha T)+\exp(-\alpha t_1)]}{2[1-q\ \exp(-\alpha T)]^2} + \right.$$

$$pq\ \mathrm{Ln}\ \frac{1-q\ \exp(-\alpha T)}{1-q\ \exp[-\alpha(T-t_1)]} - \frac{p^2 q\ \exp(-\alpha T)[\exp(\alpha t_1)-1]}{[1-q\ \exp(-\alpha T)]\{1-q\ \exp[-\alpha(T-t_1)]\}}$$

$$+\alpha pq t_1 + \frac{\lambda}{2\alpha}\left[\frac{q\phi(t_1)}{1-q\ \exp(-\alpha T)} + p\ \mathrm{Ln}\ \frac{1-q\ \exp(-\alpha T)}{1-q\ \exp[-\alpha(T-t_1)]} \right.$$

$$\left. \left. + \alpha p t_1 \right]^2 \right\} I_{(0,T]}(t_1) + I_{(T,\infty)}(t_1) \qquad (7.17)$$

The distribution of T_2 conditional on T_1, both random variables conditional on N(T)=2, is given by the expression

$$f_{T_2|\ [T_1,N(T)=2]}(t_2) = f_{\underline{T}_2|N(T)=2}(\underline{t}_2) \ / \ f_{T_1|N(T)=2}(t_1) \qquad (7.18)$$

Replacing eqs. (7.14) and (7.15) in (7.18) and integrating gives the following result for the cdf of T_2 conditional on T_1 and conditional on N(T)=2

$$F_{T_2|\ [T_1,N(T)=2]}(t_2) = \frac{2}{H_2\ f_{T_1|\ N(T)=2}(t_1)} \Bigg\{$$

$$\frac{\alpha q^2 \exp(-\alpha t_1)[\exp(-\alpha t_1)-\exp(-\alpha t_2)]}{[1-q\ \exp(-\alpha T)]^2} + \frac{\alpha pq\phi(t_2-t_1)}{\{1-q\ \exp[-\alpha(T-t_1)]\}^2} +$$

$$\lambda \left[\frac{q[\exp(-\alpha t_1)-\exp(-\alpha t_2)]}{1-q\ \exp(-\alpha T)} + p\ \mathrm{Ln}\ \frac{1-q\ \exp[-\alpha(t_2-t_1)]}{p} + \right.$$

$$\left. \alpha p(t_2-t_1) \right] \left[\frac{q\ \exp(-\alpha t_1)}{1-q\ \exp(-\alpha T)} + \frac{pq\ \exp[-\alpha(T-t_1)]}{1-q\ \exp[-\alpha(T-t_1)]} \right.$$

$$\left. + p \right] \Bigg\} I_{(t_1,T]}(t_2) + I_{(T,\infty)}(t_2) \qquad (7.19)$$

The procedure to simulate T_1 and T_2, when N(T)=2, is as follows. First, with a given value of the cdf, invert eq. (7.17) to obtain t_1. Second, with t_1 and a different value for the cdf, inversion of eq. (7.19) gives t_2. Numerical methods must be used to invert eqs. (7.17) and (7.19).

Finally, the pdf for $\underline{T}_3$, conditional on N(T)=3, is given here as an incomplete result, since it was impossible to verify it and derive the conditional and marginal distributions required for simulation. The result is

$$f_{\underline{T}_3 \mid N(T)=3}(\underline{t}_3) = \frac{\lambda}{\alpha} \frac{f_{N(T)}(0)}{f_{N(T)}(3)} \left\{ \frac{2\alpha^3 q^3 \exp[-\alpha(t_1+t_2+t_3)]}{[1-q\exp(-\alpha T)]^3} + \right.$$

$$2pq^2\alpha^3 \frac{\exp[-\alpha(t_2-t_1)]\exp[-\alpha(t_3-t_1)]}{\{1-q\exp[-\alpha(T-t_1)]\}^3} + \lambda\alpha \left[\frac{q\exp(-\alpha t_1)}{1-q\exp(-\alpha T)} + \right.$$

$$\left. \frac{pq\exp[-\alpha(T-t_1)]}{1-q\exp[-\alpha(T-t_1)]} + p \right] \left[\frac{\alpha q^2 \exp[-\alpha(t_2+t_3)]}{[1-q\exp(-\alpha T)]^2} + \right.$$

$$\left. \frac{\alpha pq\exp[-\alpha(t_3-t_2)]}{\{1-q\exp[-\alpha(T-t_2)]\}^2} \right] + \lambda\alpha \left[\frac{q\exp(-\alpha t_2)}{1-q\exp(-\alpha T)} + \right.$$

$$\left. \frac{pq\exp[-\alpha(T-t_2)]}{1-q\exp[-\alpha(T-t_2)]} + p \right] \left[\frac{\alpha q^2 \exp[-\alpha(t_1+t_3)]}{[1-q\exp(-\alpha T)]^2} + \right.$$

$$\left. \frac{\alpha pq\exp[-\alpha(t_3-t_1)]}{\{1-q\exp[-\alpha(T-t_1)]\}^2} \right] + \lambda\alpha \left[\frac{q\exp(-\alpha t_3)}{1-q\exp(-\alpha T)} + \right.$$

$$\left. \frac{pq\exp[-\alpha(T-t_3)]}{1-q\exp[-\alpha(T-t_3)]} + p \right] \left[\frac{\alpha q^2 \exp[-\alpha(t_1+t_2)]}{[1-q\exp(-\alpha T)]^2} + \right.$$

$$\left. \frac{\alpha pq\exp[-\alpha(t_2-t_1)]}{\{1-q\exp[-\alpha(T-t_1)]\}^2} \right] + \lambda^2\alpha \prod_{i=1}^{3} \left[\frac{q\exp(-\alpha t_i)}{1-q\exp(-\alpha T)} + \right.$$

$$\left. \left. \frac{pq\exp[-\alpha(T-t_i)]}{1-q\exp[-\alpha(T-t_i)]} + p \right] \right\} I_{(0,T)}(t_1) I_{(t_1,T)}(t_2) I_{(t_2,T)}(t_3) \quad (7.20)$$

Figures 7.1 to 7.3 give examples of the cumulative distribution functions represented by eqs. (7.13), (7.17) and (7.19).

The terms H_1 and H_2 given before in eqs. (7.12) and (7.15) are directly related to $f_{N(T)}(n)$, for n=1 and n=2. Note the presence of

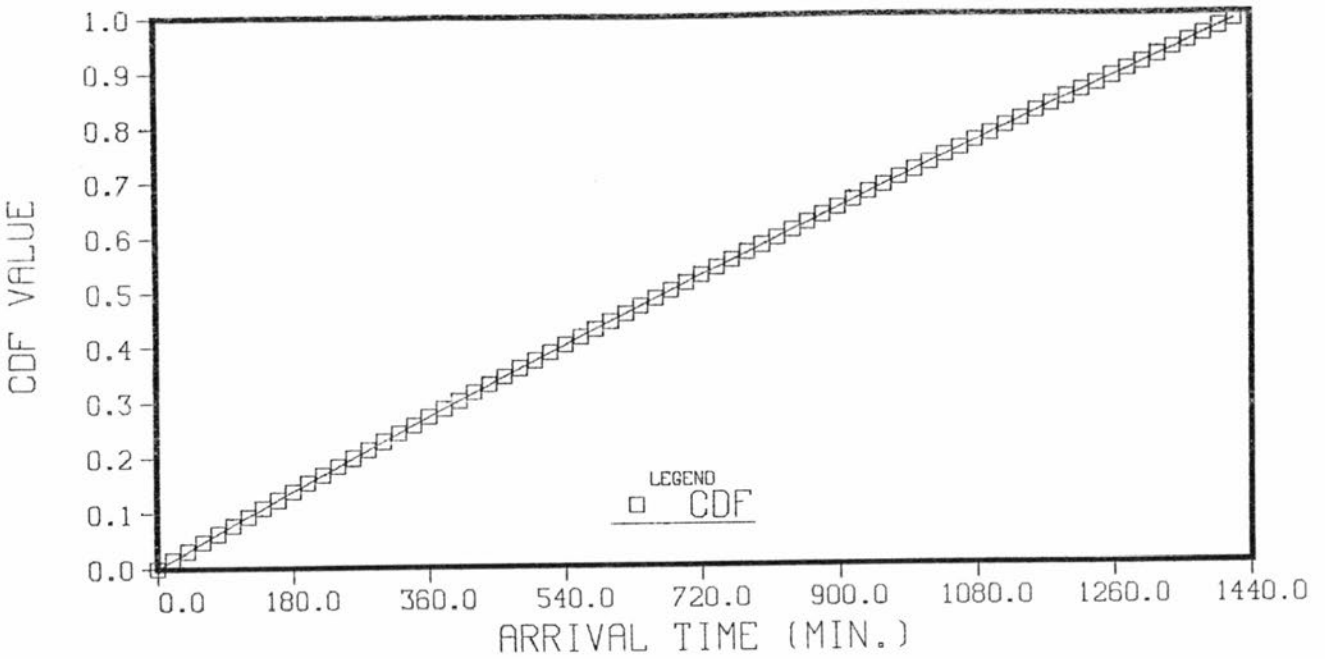


Figure 7.1 Example of the cdf for the first arrival time, conditional on one arrival in the interval, for p=0.730, α=0.000356 1/min. and T=1440 min..

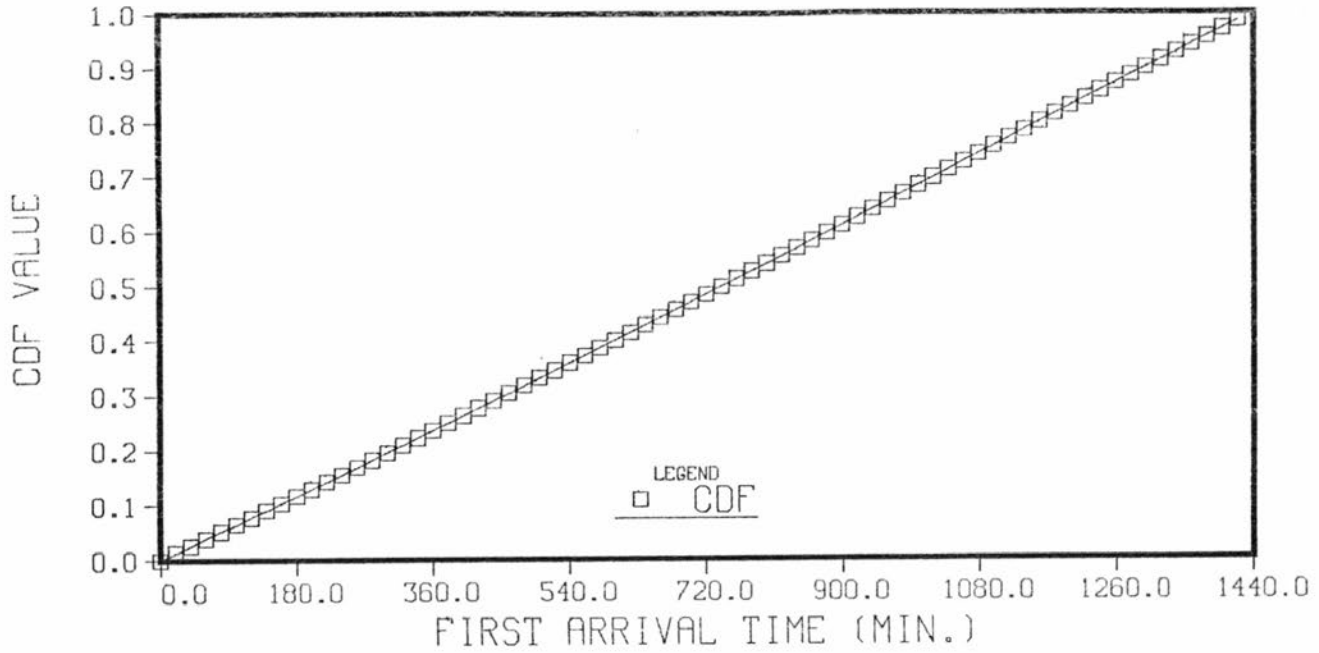


Figure 7.2 Example of the cdf for the first arrival time, conditional on two arrivals in the interval, for λ=0.000086 1/min, p=0.730, α=0.000356 1/min. and T=1440 min..

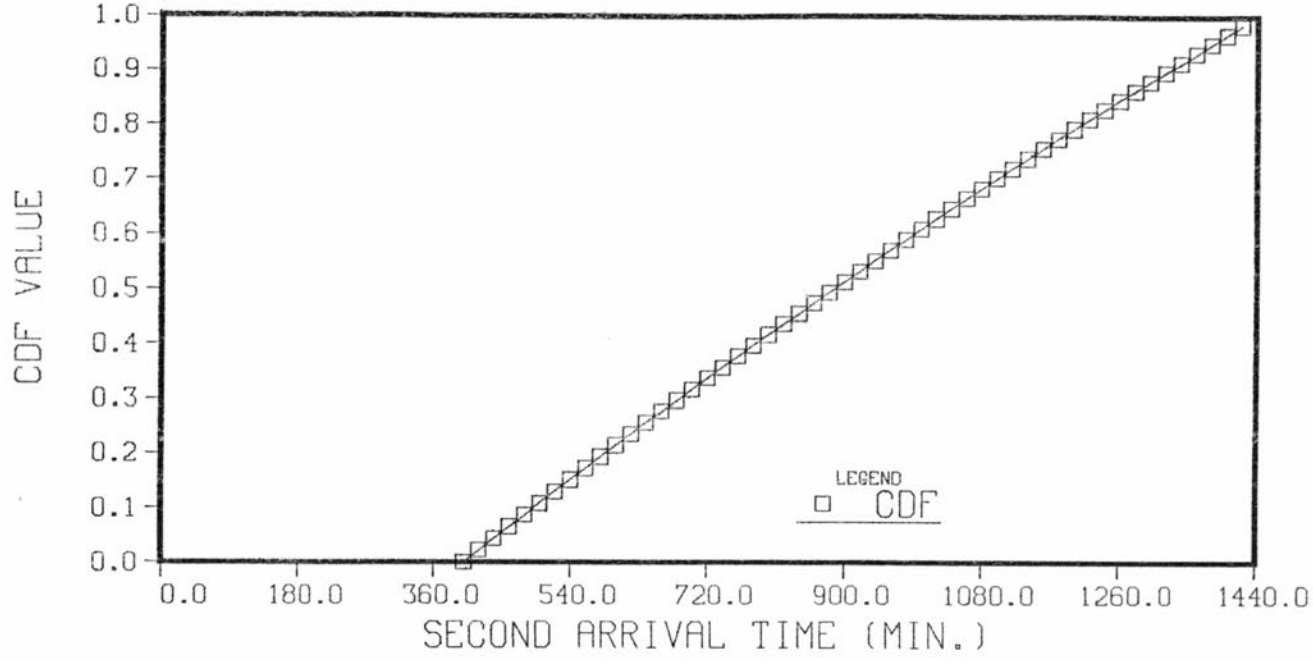


Figure 7.3 Example of the cdf for the second arrival time, conditional on two arrivals in the interval, for λ=0.000086 1/min, p=0.730, α=0.000356 1/min., T=1440 min. and t_1=400 min..

$f_{N(T)}(3)$ and $f_{N(T)}(0)$ in eq. (7.20). This is the reason for which analytical expressions for these probabilities, for n from 0 to 3, were given in Section 6.5.

7.5 Simulation and sampling of arrival times

During the operation of the NSWND model it may happen that more than two arrival times are required in any interval of length T. However, from the results presented in the previous section, any attempt to derive more general distributions for the arrival times does not seem feasible. In order to solve this problem, simulation and sampling algorithms for arrival times were designed. They are described in the following paragraphs.

To simulate arrival times, values for the NSWN model parameters, λ, p and α, are required. Also, the aggregation scale T_a is required. The simulation algorithm proceeds as follows:

1. Let the arrival time for the last cluster center be denoted as t^*_{k-1}. Initially, t^*_0 is made equal to zero.
2. Sample an interarrival time, τ_k, for the cluster arrival process, from an exponential distribution with parameter λ. The arrival time for the kth cluster center is $t^*_k = t^*_{k-1} + \tau_k$.
3. Sample the kth cluster size value, $n_2(k)$, from the geometric distribution with parameter p.
4. Draw a total of $n_2(k)$ relative cluster member locations from the exponential distribution with parameter α. Denote them by $t_{k,j}$, $j=1,..,n_2(k)$.
5. A subsample of size $n_2(k)$ of arrival times for the NS process is computed as

$$t_j = t^*_k + t_{k,j}, \quad j=1,..,n_2(k)$$

Steps 1 to 5 are repeated for as many values of k as required. In the final step of the simulation, all subsamples are assembled in a common vector. This vector is ranked from smallest to largest. Then, the vector becomes a sample of arrival times for the NS process. A generic element of this vector is denoted by t_i, i=1,..., n_s, where n_s is the total number of arrival times.

The number of intervals of length T_a in the vector of arrival times is

$$n_a = \text{int}\left(t_{n_s} / T_a \right) + 1$$

where int (.) is the integer part operator. An additional vector of length n_a is created. A generic element of this vector, i_j, gives the number of arrival times in the jth interval.

With the information provided by the simulation, the algorithm to draw n arrival times in the interval (0,T), $n \geq 3$, works in the following way:

1. If the jth interval has n arrivals (i_j=n), the arrival times to use in the disaggregation process are:

   $$t'_k = t_k - (j-1)T_a, \text{ for all } t_k \text{ such that } (j-1)T_a \leq t_k \leq jT_a$$

   After this, make i_j=0, so that the same interval is not used twice.

2. If no interval with n arrival times is found, two empirical options are available. In the first one, n arrival times are drawn from a uniform population in the interval $(0,T_a)$. In the second option, the sample of arrival times is split and rescaled. A uniform number τ_o in the interval (t_1, t_{n_s-2n-1}) is drawn and n consecutive arrival times, greater than τ_o, are

selected. Draw a second uniform number, τ_f, in the interval (t_{n_s-n-1}, t_{n_s-n}). The required arrival times are computed as

$$t'_k = \frac{t_k - \tau_o}{\tau_f - \tau_o} T_a, \text{ for all } t_k \text{ such that } \tau_o \leq t_k \leq \tau_f$$

The uniform sampling defined in step 2 is, in a certain extent, justified by the linear shape of the distributions presented in Figures 7.1 to 7.3.

7.6 Disaggregation model algorithm

In order to operate the NSWND model, the following information must be provided: aggregation and disaggregation scales, model parameters, definition of the empirical model to extrapolate the distribution of the number of occurrences, recorded precipitation sample, seeds for generation of random numbers, tolerances and number of iterations for numerical solution of equations, and sampling method for arrival times. As before for the PWND model, the NSWND model is operated on the sample for a given month.

The ratio between aggregation and disaggregation scales is

$$R = T_a/T_d$$

and the result must be an integer number. The aggregated or recorded sample is denoted by y_i, $i=1,..,N_a$, and the disaggregated series is denoted by y_j^d, $j=1,..,RN_a$, where N_a is the total sample size, collection of all years, available for the specific month at the aggregation scale.

Before proceeding to disaggregate of values, the following operations are performed in the order given: computation of the coefficients in the empirical extrapolation model, computation and

extrapolation of the pdf for the number of occurrences in the interval, and simulation of arrival times for the NS process.

To disaggregate the ith recorded precipitation value y_i, the model proceeds as follows:

- If $y_i = 0$, the value is disaggregated into zero values:
  $y_j^d = 0$, $j=(R-1)i+1,..,Ri$
- If y_i is missing ($y_i<0$), make the disaggregated series, y_j^d, equal to a negative value representing missing data, for j from $(R-1)i+1$ up to Ri.
- If y_i is positive, the algorithm operates in the following way:
  1. Sample a value for $N(T_a)$, n, conditional on y_i, from the distributions given in eq. (7.2).
  2. If n=1, make $u_1 = y_i$. Otherwise, sample n-1 White Noise values from the distribution defined by eqs. (7.4) and (7.9). The last White Noise value is
     $$u_n = y_i - \sum_{k=1}^{n-1} u_k$$
  3. If n=1, use eq. (7.13) to generate one arrival time. If n=2, eqs. (7.17) and (7.19) are used to sample two arrival times. For $n\geq 3$, use the sampling algorithm described in Section 7.5 Arrival times in $(0, T_a)$ are denoted by t'_k.
  4. The sequence of ordered pairs $((i-1)T_a+t'_k, u_k)$, $k=1,..,n$, is a sample realization of the process in the interval $((i-1)T_a, iT_a)$, which preserves the NSWN process and the recorded amounts of precipitation. From this sequence, disaggregated values are computed using the expression
     $$y_j^d = \sum_{k=1}^{n} u_k I_{[(j-1)T_d, jT_d)}((i-1)T_a+t'_k),\ j=(i-1)R+1,..,Ri$$

7.7 Testing the NSWND model

As explained before for the PWND model, the procedures and results presented in this section are not intended as exact statistical tests of hypothesis for the NSWND model. They are oriented to asses the goodness in model operation and its performance, when the underlying process controlling precipitation formation is a NSWN process.

In the first trial, the NSWND model was applied systematically to all months comprising simulation 1, as described in Section 6.9, to disaggregate from daily values (T_a=1440 min.) to hourly values (T_d=60 min.). Disaggregation runs were made with parameters estimated at the aggregation scale using the weighted least squares (WLS) method.

The first part of the analysis is focused on correlogram reproduction. Figure 7.4 presents plots of the theoretical correlogram for simulation 1, computed using eq. (6.91), with parameter values given in Table 6.3, and the sample correlograms estimated from disaggregated series. For these last series, the average correlogram is also plotted. According to Figure 7.4, results are far from good.

Since one of the problems observed in this study and reported in the literature is the incompatibility of the model at different temporal scales, it was decided to obtain new disaggregated samples for NSWN simulation 1, with parameters estimated at the disaggregation scale (T_d=60 min.), using the WLS technique. Figure 7.5 compares theoretical and sample correlograms for the disaggregated series. Improvement of results is outstanding, as shown in Figures 7.4 and 7.5.

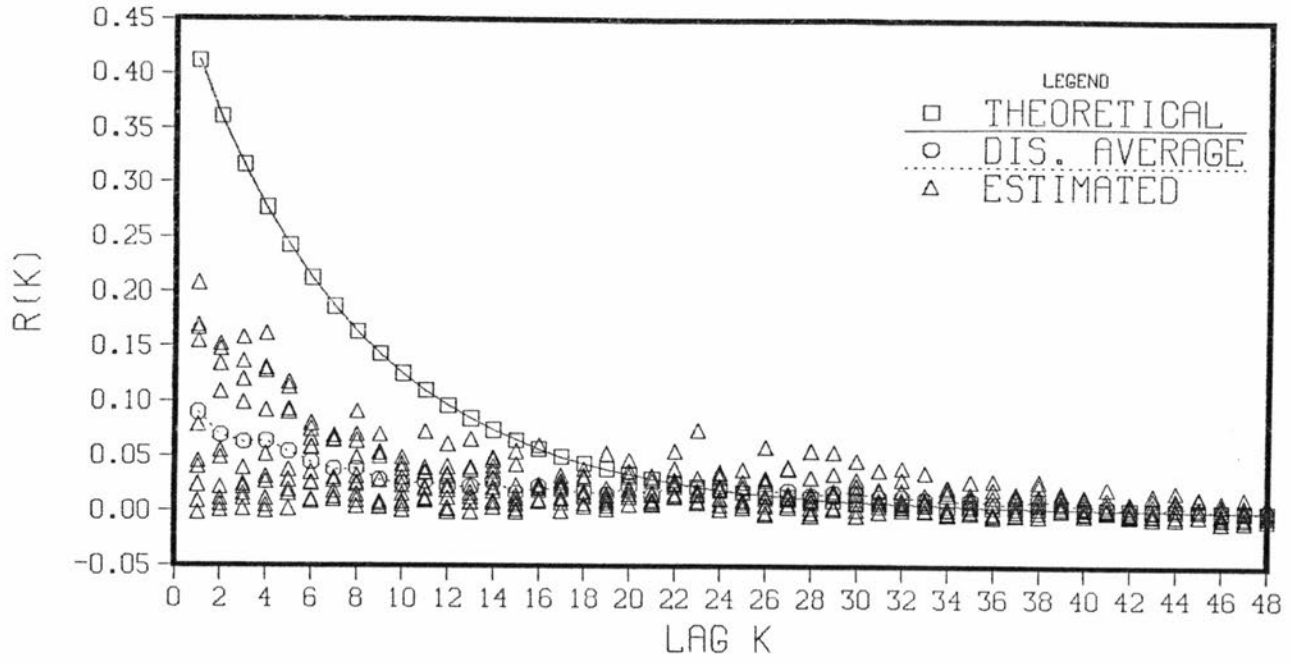


Figure 7.4 Theoretical and estimated correlograms for NSWN disaggregated series, for simulation 1, using parameters estimated at T_a=1440 min..

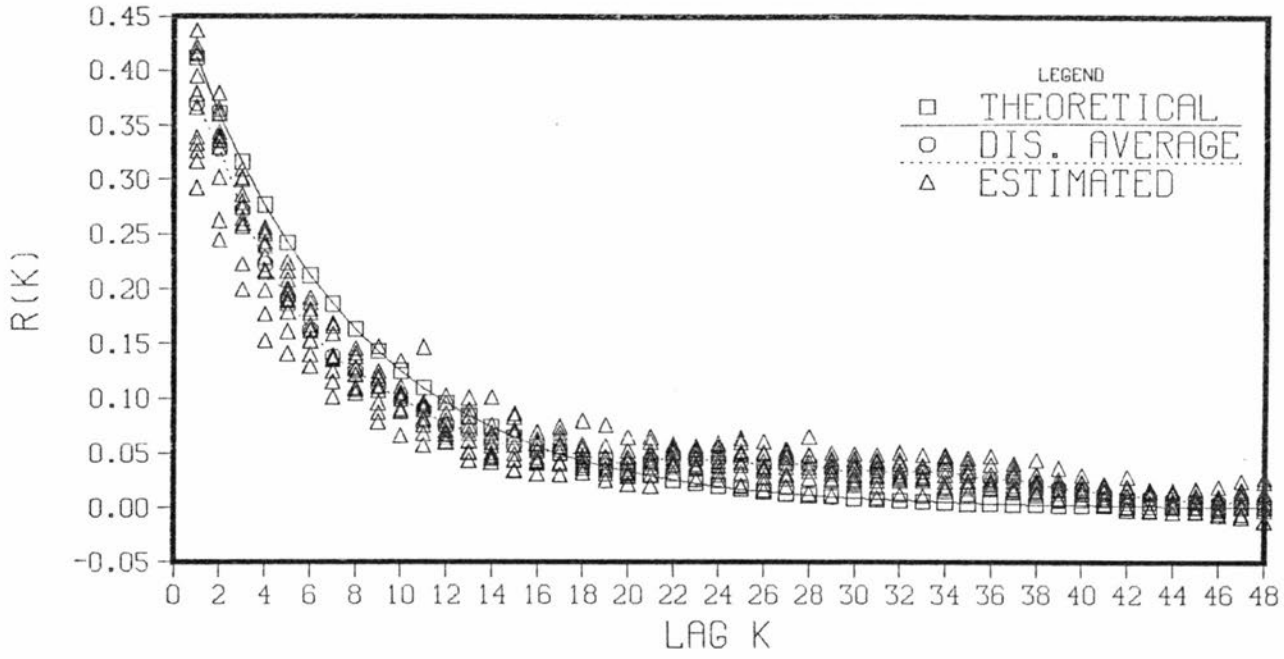


Figure 7.5 Theoretical and estimated correlograms for NSWN disaggregated precipitation series, for simulation 1, using parameters estimated at T=60 min..

In a third attempt, the same disaggregation procedure was followed using population parameters, values for which are given in Table 6.3. Figure 7.6 presents the corresponding results. Although a slight improvement is obtained, bias in the estimated correlograms, similar to that appearing in Figure 7.5, is still present.

Several of the assumptions made for NSWND model development were investigated as possible causes for the bias present in Figures 7.5 and 7.6, by making additional disaggregation runs, all of them with population parameters for simulation 1. In the first step, the exponent θ and the number of points n_p defining the empirical extrapolation model for the pdf of the number of occurrences were changed. No improvement was produced.

The next step consisted in testing the methodologies used to draw arrival times, when the required number is not available in the simulated sample. In fact, Figure 7.6 was obtained with samples disaggregated by splitting and rescaling. Results for uniform sampling are given in Figure 7.7. Comparison of both figures does not show any improvement. Therefore, no differentiation will be made on these two methodologies in the sequel.

A possible explanation for the bias in the correlogram is in the concept of cluster. Although the disaggregation model has been built around a clustering process, the model itself does not recognize the presence of a cluster. In other words, when working on a given interval, the model ignores the state of the process in previous and future intervals. As stated before, developing a model including this feature, given the current state of the art, would represent enormous difficulties. A second possible explanation is in the simulation of arrival times. However, since no theoretical results are available

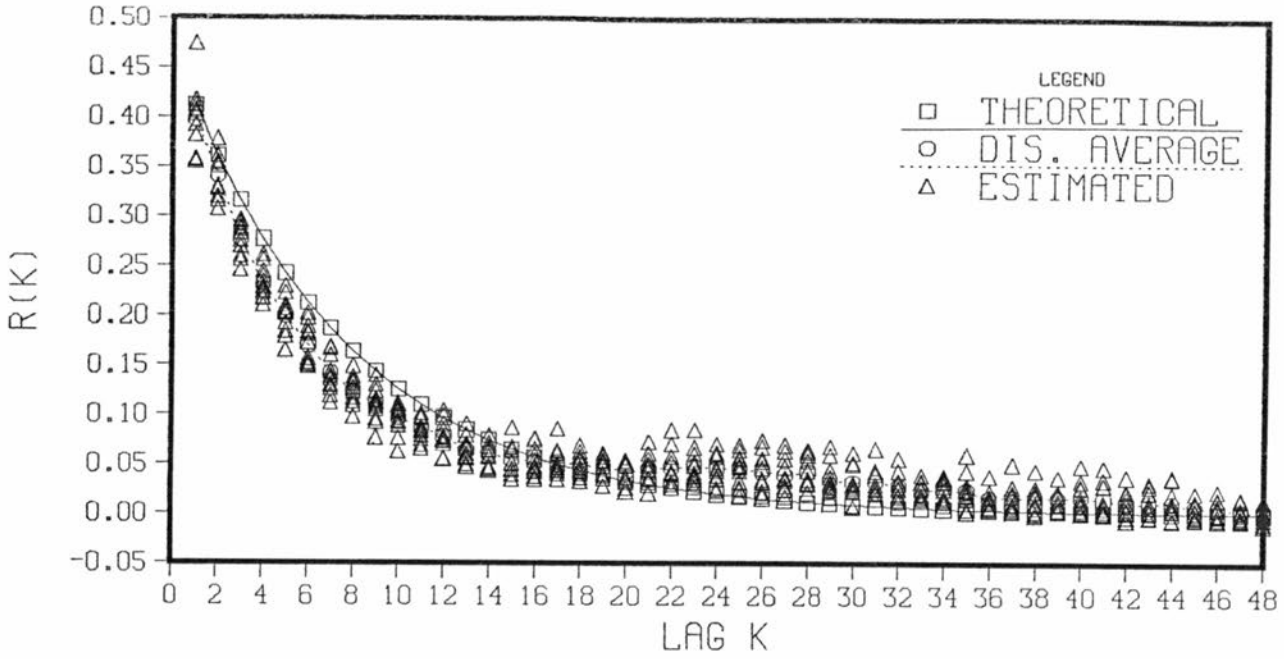


Figure 7.6 Theoretical and estimated correlograms for NSWN disaggregated precipitation series, for simulation 1, using population parameters (Sampling by splitting and rescaling).

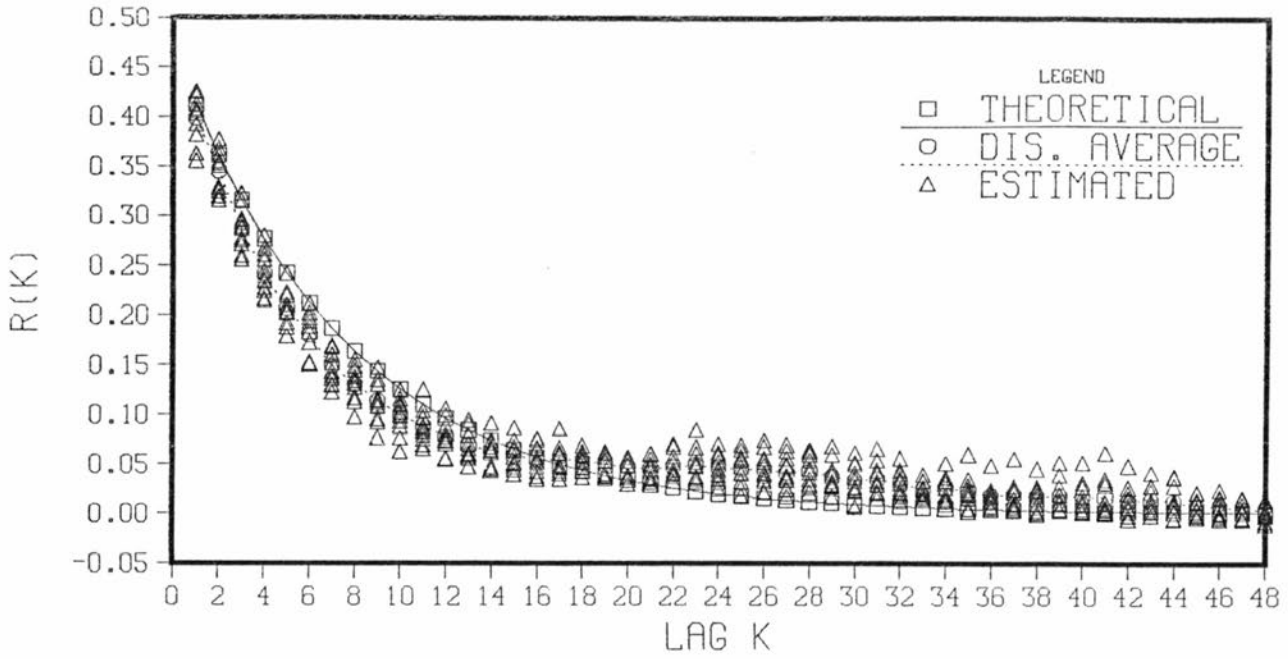


Figure 7.7 Theoretical and estimated correlograms for NSWN disaggregated precipitation series, for simulation 1, using population parameters (Uniform sampling).

for the joint distribution of 3 or more of these times, this statement must be left as hypothetical. Therefore, it is accepted that the model, as built, is yielding its maximum, as in Figures 7.6 and 7.7, and no further improvement appears feasible at this point.

In order to complete the performance analysis for the NSWND model, Table 7.1 gives a summary for statistics estimated for the different disaggregated samples for simulation 1. In this table, population values and statistics estimated from the original series are also presented. These two sets confirm the goodness of the simulation procedure, as mentioned in Section 6.9. Note the perfect agreement in the mean, which is the only statistic preserved exactly by the disaggregation algorithm. No preservation is obtained with parameters estimated at T_a=1440 min.. However, when parameters estimated at T=60 min. or population parameters are use, the degree of preservation improves appreciably, with reduction in the standard deviation for sample statistics. Adequate preservation of statistics is detected up to order 3. Kurtosis is not preserved at all by the disaggregation model.

The forced conclusion in this section is that the NSWND model performs well, as long as disaggregation runs are made with parameter values similar to those controlling the process at the disaggregation scale. As stated by Foufoula-Georgiou and Guttorp (1986), the NSWN model is scale dependent. However, this result is different from the one obtained for the PWND model. The PWND model performed well when the underlying process was PWN. For the NSWND model, even though the underlying process is NSWN, no good performance is obtained with estimation at any temporal aggregation level.

Table 7.1 Summary of statistics estimated from NSWN simulation 1 disaggregated hourly precipitation series, for different choices of parameters to use in the disaggregation.

Sample	$\bar{Y}$ $\times 10^{-2}$ (in.)	$\hat{S}$ $\times 10^{-1}$ (in.)	$\hat{g}_Y$	k	r(1)	r(2)
1	0.34960	0.22836	10.43	594.00	0.41200	0.36084
2						
Mean	0.35392	0.23282	10.50	154.93	0.41533	0.36222
St. Dev.	0.02605	0.01149	0.734	28.72	0.02024	0.02527
Max.	0.39782	0.25636	12.17	228.40	0.44968	0.39318
Min.	0.31718	0.22013	9.86	134.24	0.36854	0.31939
3						
Mean	0.35392	0.39443	17.90	485.47	0.08912	0.06839
St. Dev.	0.02514	0.07700	4.57	307.62	0.07693	0.06050
Max.	0.39782	0.50882	26.96	1236.9	0.20789	0.15147
Min.	0.31718	0.28156	11.57	167.45	-0.00193	0.00069
4						
Mean	0.35392	0.23640	11.93	262.69	0.37010	0.32834
St. Dev.	0.02605	0.01473	2.97	208.88	0.04783	0.03994
Max.	0.39782	0.26481	18.4	773.84	0.43727	0.37933
Min.	0.31718	0.21710	9.22	111.81	0.29259	0.24475
5						
Mean	0.35392	0.23295	11.00	186.62	0.39432	0.33806
St. Dev.	0.02605	0.01640	1.74	92.312	0.01892	0.01873
Max.	0.39782	0.27416	14.71	397.23	0.41922	0.38336
Min.	0.31718	0.21475	9.71	127.53	0.35846	0.31293

Samples
1: Population
2: Sample simulated series
3: Disaggregated series with parameters estimated at T_a= 1440 min.
4: Disaggregated series with parameters estimated at T= 60 min.
5: Disaggregated series with population parameters

7.8 Application of the NSWND model to historical precipitation samples

In order to further investigate the applicability of the NSWND model, disaggregation was performed on a total of 22 months for the precipitation recording stations used in this study. Table 7.2 gives a summary of months and temporal scales used. Disaggregation for the precipitation stations was performed with parameters obtained at the aggregation scale using the WLS method. Tables 7.3 to 7.6 compare historical monthly statistics with those estimated from disaggregated series, for the four precipitation recording stations.

Table 7.2 Description of historical samples for application of the NSWND model.

Station	Months	T_a (min.)	T_d (min.)	R
Denver Wsfo Ap	03,06,07,08,11	1440	60	24
Greenland 9 SE	01,06,07,08,09,10	1440	60	24
Idaho Springs	02,06,07,08,11	60	5	12
Ward	04,06,07,08,11,12	60	5	12

Results presented in Tables 7.3 to 7.6 are discouraging. Only in a few months an acceptable preservation of statistics is observed: Denver Wsfo Ap station, month 08; Greenland 9 SE station, month 10; Idaho Springs station, months 02 and 11; Ward station, months 11 and 12. For the remaining months, qualitatively speaking, results are not very different from those obtained with the PWND model. Occasionally, preservation of isolated statistics, within certain months, is detected.

Table 7.3 Comparison of statistics for historical and NSWN disaggregated hourly series, for Denver Wsfo Ap station. T_a= 1440.0 min., T_d= 60.0 min., R=24

	Historic	Model	Historic	Model
Month τ:	03		06	
$\bar{Y}_\tau \times 10^{-3}$ (in.)	1.7654	1.7654	2.4015	2.4004
$S_\tau \times 10^{-3}$ (in.)	9.9400	20.8690	21.8696	29.2218
$\hat{g}_\tau$	10.1243	18.9587	19.6414	21.7464
$r_\tau(1)$	0.75611	0.02398	0.44566	0.06938
$\hat{p}_o^\tau$	0.93552	0.98413	0.96046	0.98222
No of positive values	1679	413	996	448
Maximum value (in.)	0.27	0.89	1.11	1.51
No of missing values	744	744	732	720
Month τ:	07		08	
$\bar{Y}_\tau \times 10^{-3}$ (in.)	2.4328	2.4328	1.9180	1.9172
$S_\tau \times 10^{-3}$ (in.)	29.1912	32.8376	25.5126	18.1306
$\hat{g}_\tau$	23.6784	20.4192	29.2155	18.2313
$r_\tau(1)$	0.20808	0.00435	0.22711	0.22099
$\hat{p}_o^\tau$	0.96886	0.98556	0.97337	0.97372
No of positive values	811	376	713	704
Maximum value (in.)	1.59	1.42	1.55	0.89
No of missing values	744	744	12	0
Month τ:	11			
$\bar{Y}_\tau \times 10^{-3}$ (in.)	1.1448	1.1443		
$S_\tau \times 10^{-3}$ (in.)	7.4128	16.1538		
$\hat{g}_\tau$	10.0448	21.0141		
$r_\tau(1)$	0.72793	0.00254		
$\hat{p}_o^\tau$	0.95601	0.98978		
No of positive values	1088	265		
Maximum value (in.)	0.18	0.70		
No of missing values	12	0		

Table 7.4 Comparison of statistics for historical and NSWN disaggregated hourly series, for Greenland 9 SE station. T_a= 1440.0 min., T_d= 60.0 min., R=24

	Historic	Model	Historic	Model
Month τ:	01		06	
$\bar{Y}_\tau \times 10^{-3}$ (in.)	0.4928	0.4909	2.7236	2.7162
$S_\tau \times 10^{-3}$ (in.)	4.4552	7.7273	24.6846	32.7559
$\hat{g}_\tau$	14.6383	25.1824	20.5936	18.8968
$r_\tau(1)$	0.33363	0.00356	0.29861	0.01147
$\hat{p}_o^\tau$	0.97756	0.99300	0.96379	0.98526
No of positive values	581	182	877	358
Maximum value (in.)	0.10	0.47	1.08	1.20
No of missing values	892	792	1698	1632
Month τ:	07		08	
$\bar{Y}_\tau \times 10^{-3}$ (in.)	3.7136	3.6964	3.3891	3.3810
$S_\tau \times 10^{-3}$ (in.)	36.0284	41.8252	34.4460	44.7681
$\hat{g}_\tau$	26.7805	20.4987	21.0165	23.9666
$r_\tau(1)$	0.23827	0.06284	0.30741	0.02149
$\hat{p}_o^\tau$	0.95490	0.98122	0.96367	0.98551
No of positive values	1114	466	943	377
Maximum value (in.)	2.24	2.23	1.58	2.67
No of missing values	2083	1968	830	768
Month τ:	09		10	
$\bar{Y}_\tau \times 10^{-3}$ (in.)	1.3185	1.3079	0.8649	0.8576
$S_\tau \times 10^{-3}$ (in.)	14.4451	17.5244	7.8730	8.4143
$\hat{g}_\tau$	26.5882	25.1378	13.3750	14.9211
$r_\tau(1)$	0.28296	0.05227	0.46125	0.26717
$\hat{p}_o^\tau$	0.97560	0.98837	0.97767	0.98175
No of positive values	606	291	592	488
Maximum value (in.)	0.96	1.18	0.29	0.30
No of missing values	1089	888	272	48

Table 7.5 Comparison of statistics for historical and NSWN disaggregated 5 min. series, for Idaho Springs station. T_a= 60.0 min., T_d= 5.0 min., R=12

	Historic	Model	Historic	Model
Month τ:	02		06	
$\bar{Y}_\tau \times 10^{-3}$ (in.)	0.4874	0.4769	0.1633	0.1567
$S_\tau \times 10^{-3}$ (in.)	2.5347	2.7972	1.9787	3.0526
$\hat{g}_\tau$	5.9543	7.6663	19.9078	30.5880
$r_\tau(1)$	0.58321	0.51282	0.49601	0.03034
$\hat{p}_o^\tau$	0.95951	0.96447	0.98948	0.99494
No of positive values	1069	959	391	196
Maximum value (in.)	0.03	0.05	0.09	0.19
No of missing values	13917	13332	6035	4464
Month τ:	07		08	
$\bar{Y}_\tau \times 10^{-3}$ (in.)	0.2589	0.2439	0.3329	0.3250
$S_\tau \times 10^{-3}$ (in.)	3.3356	5.4438	4.0560	6.0065
$\hat{g}_\tau$	27.2398	44.0213	28.5674	28.7941
$r_\tau(1)$	0.65573	0.12074	0.71317	0.11390
$\hat{p}_o^\tau$	0.98765	0.99462	0.98389	0.99322
No of positive values	524	234	494	213
Maximum value (in.)	0.19	0.51	0.28	0.29
No of missing values	2198	1176	13972	13224
Month τ:	11			
$\bar{Y}_\tau \times 10^{-3}$ (in.)	0.3904	0.3837		
$S_\tau \times 10^{-3}$ (in.)	2.1816	2.5417		
$\hat{g}_\tau$	6.7794	8.6837		
$r_\tau(1)$	0.45688	0.31757		
$\hat{p}_o^\tau$	0.96514	0.97195		
No of positive values	1086	889		
Maximum value (in.)	0.05	0.06		
No of missing values	12050	11508		

Table 7.6 Comparison of statistics for historical and NSWN disaggregated 5 min. series, for Ward station. T_a= 60.0 min., T_d= 5.0 min., R=12

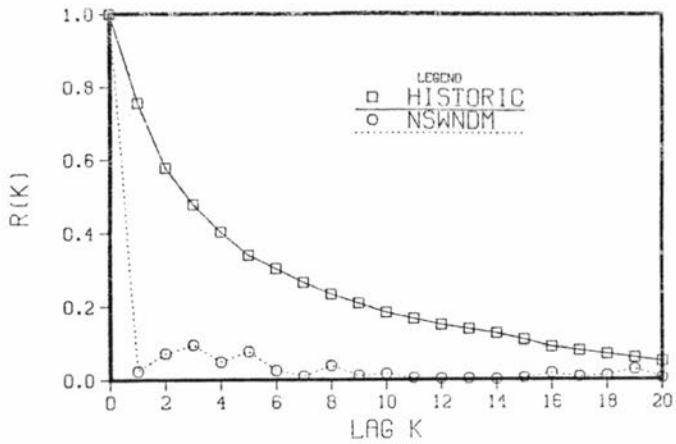
	Historic	Model	Historic	Model
Month τ:	04		06	
$\bar{Y}_\tau \times 10^{-3}$ (in.)	0.4434	0.4350	0.3732	0.3553
$S_\tau \times 10^{-3}$ (in.)	2.6457	4.2293	7.1452	6.6769
$\hat{g}_\tau$	8.3023	14.2457	89.7420	41.7037
$r_\tau(1)$	0.66053	0.17750	0.41144	0.33532
$\hat{p}_o^\tau$	0.96540	0.98288	0.98023	0.99207
No of positive values	942	475	586	247
Maximum value (in.)	0.07	0.14	0.94	0.61
No of missing values	15977	15456	13562	12072
Month τ:	07		08	
$\bar{Y}_\tau \times 10^{-3}$ (in.)	0.4041	0.3786	0.3497	0.3368
$S_\tau \times 10^{-3}$ (in.)	4.5347	7.1870	4.0986	5.7660
$\hat{g}_\tau$	26.3664	35.4589	29.1518	27.7817
$r_\tau(1)$	0.71801	0.06131	0.60646	0.08684
$\hat{p}_o^\tau$	0.98014	0.99219	0.98251	0.99180
No of positive values	791	332	532	259
Maximum value (in.)	0.23	0.54	0.31	0.32
No of missing values	4819	2136	14215	13044
Month τ:	11		12	
$\bar{Y}_\tau \times 10^{-3}$ (in.)	0.2516	0.2477	0.0583	0.0574
$S_\tau \times 10^{-3}$ (in.)	1.6431	1.8343	0.7704	0.8550
$\hat{g}_\tau$	7.0898	9.0486	13.4217	17.3165
$r_\tau(1)$	0.32006	0.29825	0.05710	0.06854
$\hat{p}_o^\tau$	0.97595	0.97922	0.99423	0.99499
No of positive values	778	683	170	150
Maximum value (in.)	0.04	0.04	0.02	0.03
No of missing values	10845	10332	15153	14676

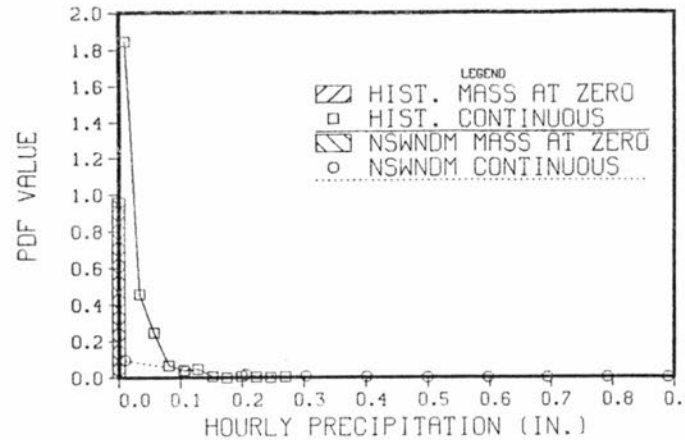
In general, the NSWND model tends to produce standard deviation, skewness coefficient and probability of zero precipitation larger than those estimated from historical samples. Lag-1 correlation coefficient is usually smaller. Operationally speaking, parameters in the model force the distribution of the number of occurrences, conditional on precipitation in the interval, to be mostly concentrated on the first three points (n=1, 2, and 3). The model draws small sample values for N(T), causing the behavior of statistics described before. For example, not a sufficient number of arrival times is generated to preserve correlation. Further evidence of this is found in the small number of positive occurrences and in the maximum precipitation value being larger than the historical one.

Figures 7.8 to 7.11 complement results given in Tables 7.3 to 7.6. They show estimated correlograms and probability distribution functions for the amount of precipitation, for some of the months. In a few of them, as stated before, adequate preservation is observed for shapes and values. However, in most of them, results are quite bad. In the case of probability distribution functions, the model does not reproduce large observed frequencies for small precipitation values.

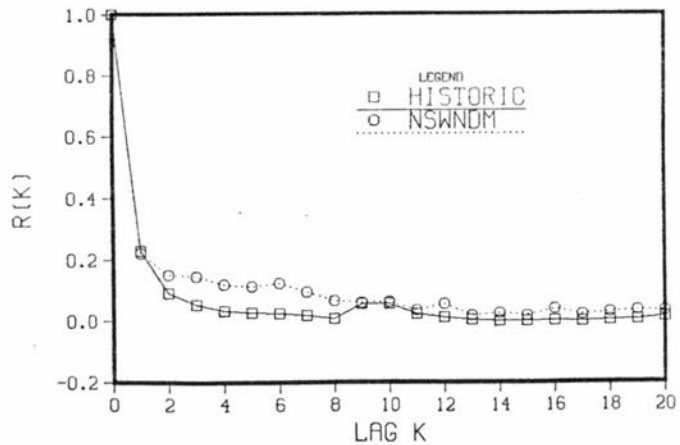
As stated before, the NSWN model provides an adequate representation of the precipitation process for some of the months in the recorded samples. However, for other cases the NSWN model does not fit the correlogram or estimated values for the parameters do not exist. Under the assumption that the NSWN model fits the monthly samples (Figures 6.13 and 6.14), incompatibility of model parameter estimates at different temporal scales is found to be the reason for the poor performance of the model. As shown in Section 7.6, the model could generate adequate disaggregated series, resembling the

Month 03





Month 08



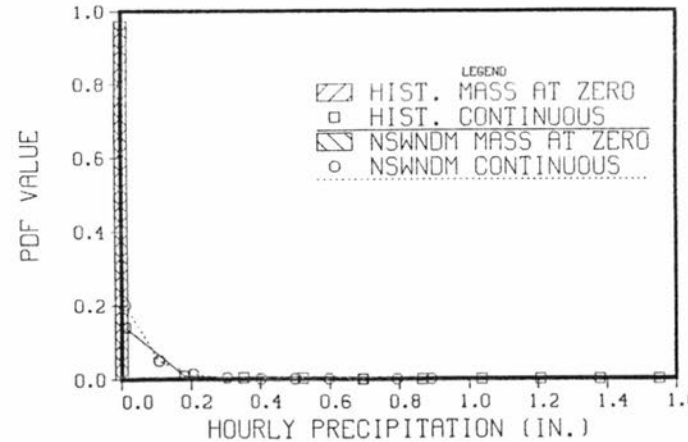
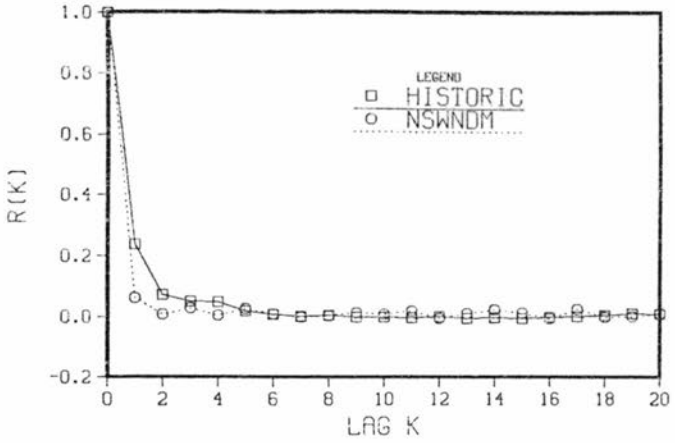
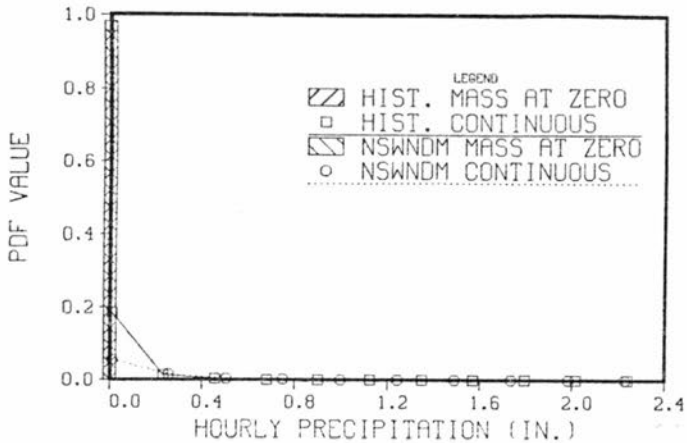


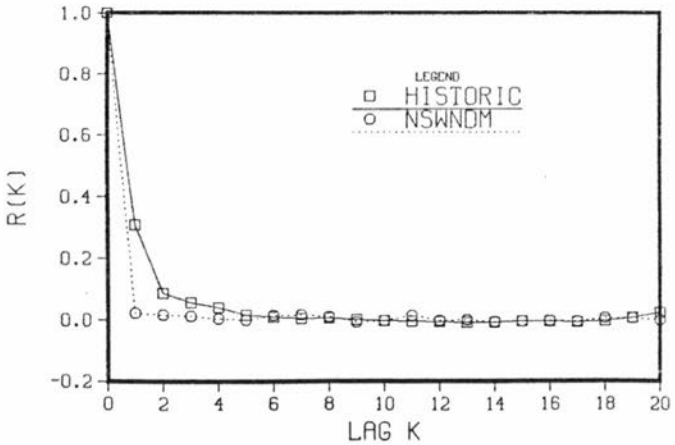
Figure 7.8 Correlograms and sample probability distribution functions for hourly historic and NSWN disaggregated precipitation series, for Denver Wsfo Ap station, for months 03 and 08 (T_d=60 min.).

Month 07





Month 08



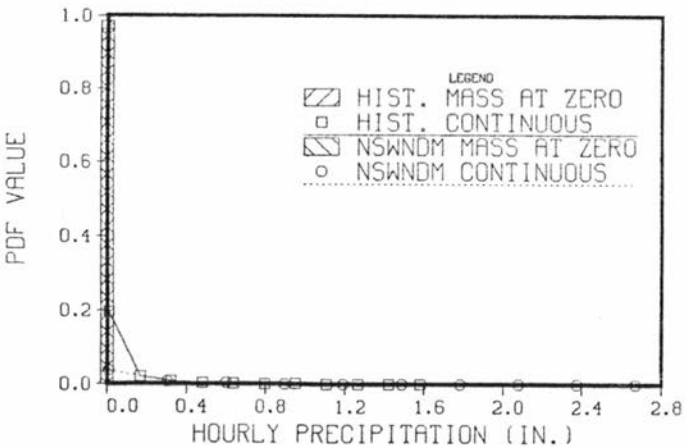
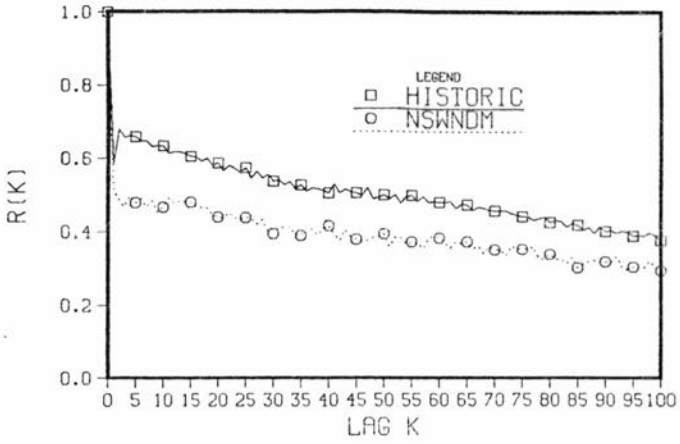
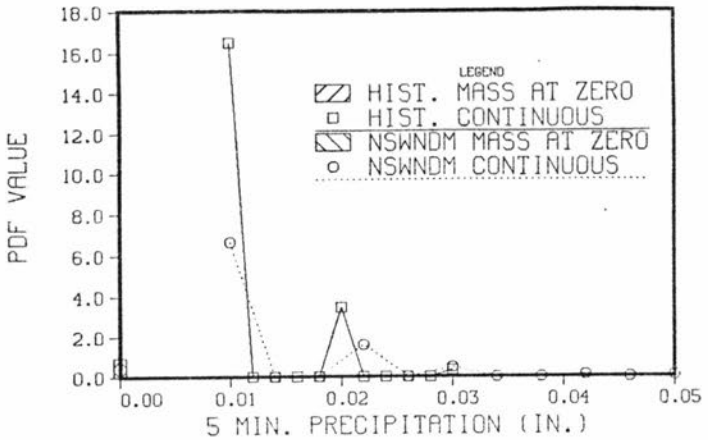


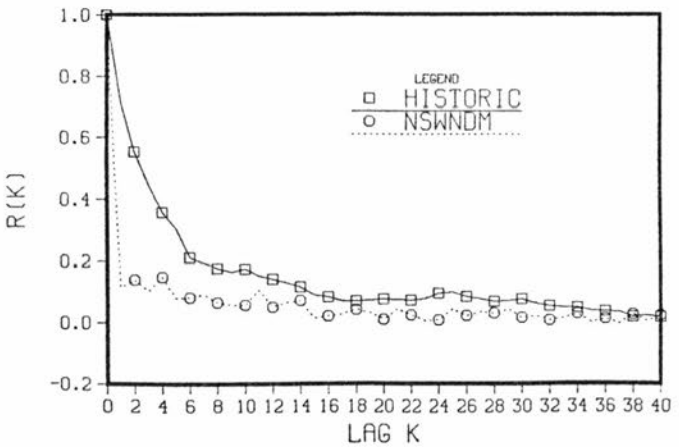
Figure 7.9 Correlograms and sample probability distribution functions for hourly historic and NSWN disaggregated precipitation series, for Greenland 9 SE station, for months 07 and 08 (T_d=60 min.).

Month 02





Month 08



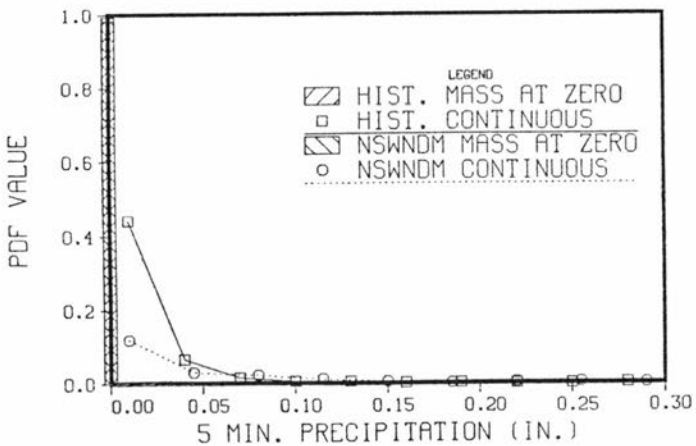
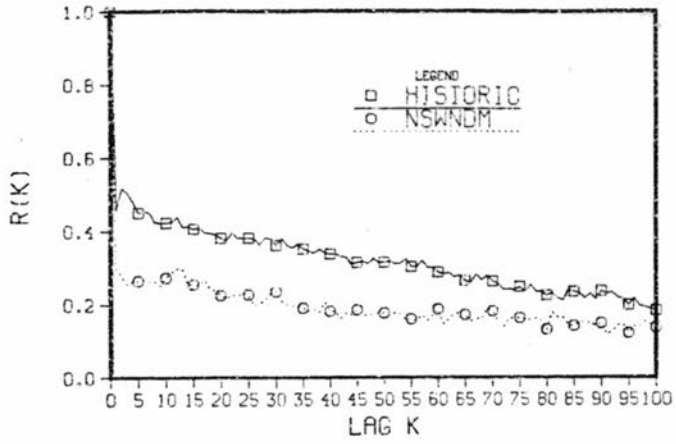
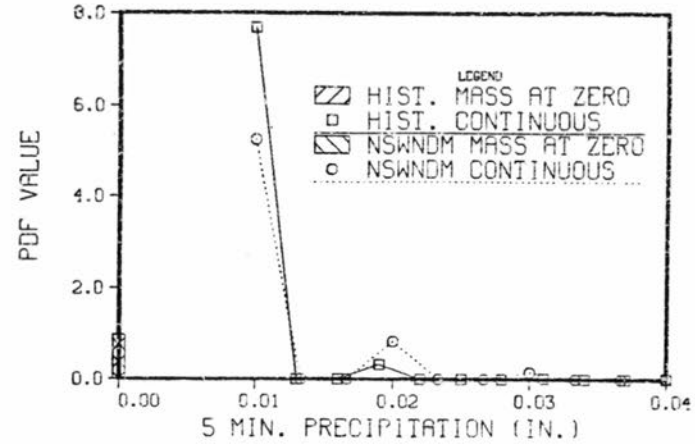


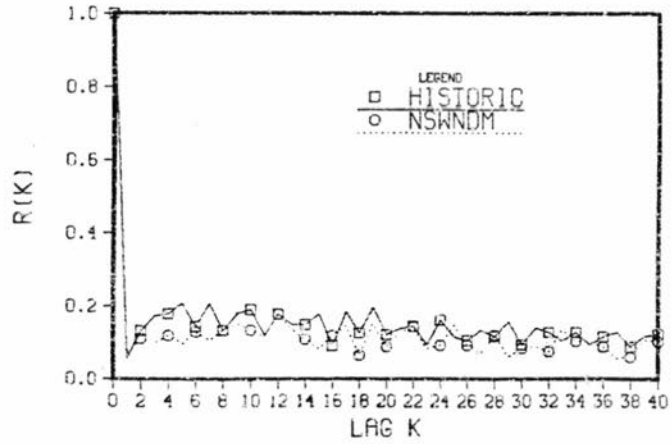
Figure 7.10 Correlograms and sample probability distribution functions for 5 min. historic and NSWN disaggregated precipitation series, for Idaho Springs station, for months 02 and 08 (T_d=5 min.).

Month 11





Month 12



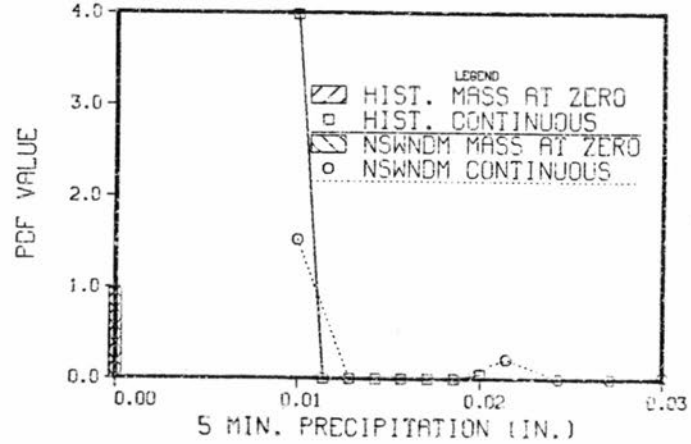


Figure 7.11 Correlograms and sample probability distribution functions for 5 min. historic and NSWN disaggregated precipitation series, for Ward station, for months 11 and 12 (T_d=5 min.).

historical ones, as long as parameters are similar to those obtained at the disaggregation scale. Examination of Tables 6.9 to 6.11 show that this same conclusion is valid for historical samples. For those monthly samples where the disaggregation model performed fairly good, parameter estimates at the aggregation and disaggregation scales tend to be similar. Therefore, an alternative to improve NSWND model applicability to historical precipitation samples is to improve parameter estimation. In this regard, further analysis is given in the closing sections of this chapter.

An important point is the impossibility to link bad performance of the disaggregation model to the months to which disaggregation was applied. This issue arises from the literature on temporal precipitation models, where these models are mainly applied to summer or summer transition months. In general, different seasons were covered in the estimation and disaggregation processes. Results suggest that the model tends to perform better for winter or winter transition months. This is considered a sample result explained by the similarity in parameter estimates. To the knowledge of the author, there is no statistical evidence in this investigation to differentiate models for different months or seasons, beyond the seasonal partition adopted.

In Chapter 3, a discussion on sample statistics variability and dependence on aggregation scale was presented. Correlograms were shown to be quite different at the aggregation and disaggregation levels. For this reason, not only the adequacy of the NSWN model and the model structure but the statistics used to estimate parameters must be held responsible for the kind of results obtained here.

Perhaps, the use of more consistent statistics and larger information in the samples could yield better results.

7.9 Variation of parameter estimates with the aggregation scale

Given that the observed precipitation process is well described by the NSWN model, the reason for the bad performance of the model is the use of parameter values which are not similar to those estimated at the disaggregation scale. Figures 7.12 and 7.13 show the variation of estimated parameter values with the aggregation scale for NSWN simulation 1, for the method of moments and the WLS method. Figures 7.14 and 7.15 present similar results for Denver Wsfo Ap station and Greenland 9 SE station, respectively, for the WLS method. Note that in Figures 7.12 to 7.15 the inverse of the parameter in the exponential distribution for the White Noise terms, μ, is plotted instead of the parameter itself. Also, plots in Figures 7.12 and 7.13 represent samples from the same population, since simulation 1 was drawn from a stationary process, while Figures 7.14 and 7.15 are for monthly estimated values. In all cases the scales considered were 60, 120, 240, 360, 720 and 1440 min.. No plots of this type were obtained for samples at 5 min. measurement scale.

Although this point was not further investigated, Figures 7.12 to 7.15 show how the largest variability (around mean value) for estimated parameter values appears for the smallest considered aggregation scale. This behavior is more marked for λ and α, parameters which depend directly on the temporal scale T. Working with eqs. (6.100) to (6.103), it is possible to show that p and μ depend on T only through variation of sample statistics with T, i.e., T does not appear in the expression used to compute those parameters.

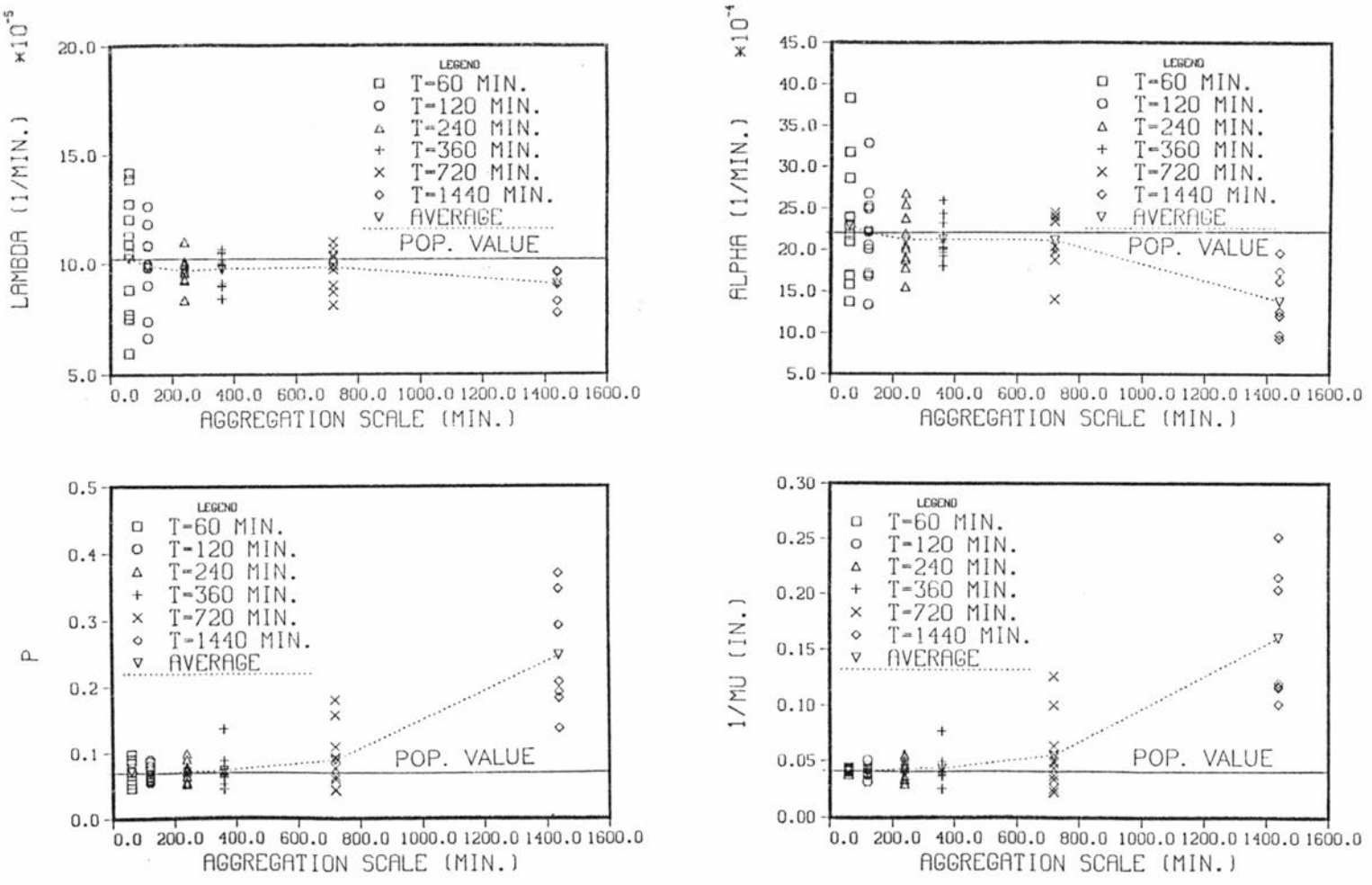


Figure 7.12 Variation of estimated parameter values with the aggregation scale for NSWN simulation 1 (Method of moments).

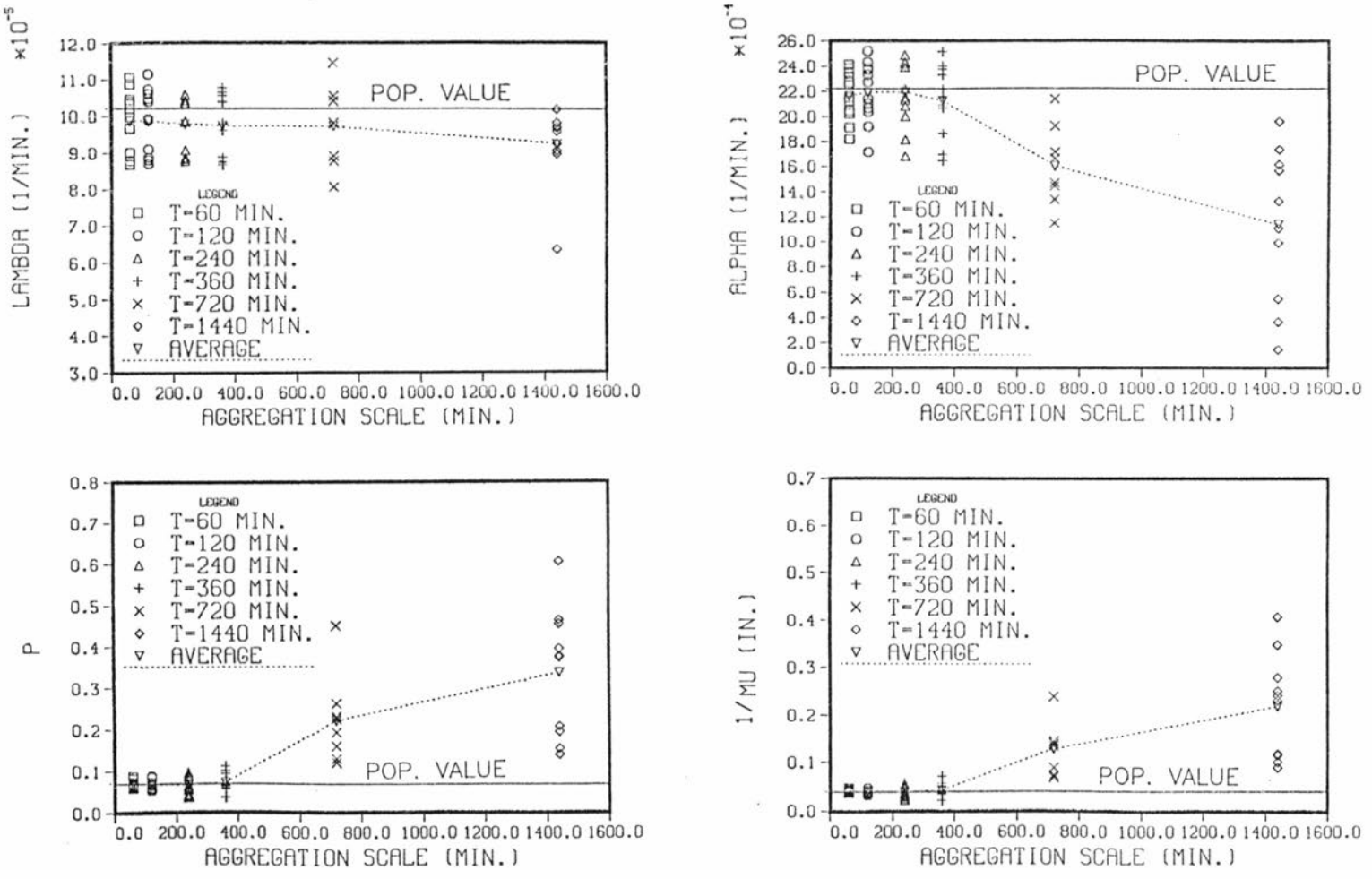


Figure 7.13 Variation of estimated parameter values with the aggregation scale for NSWN simulation 1 (WLS method).

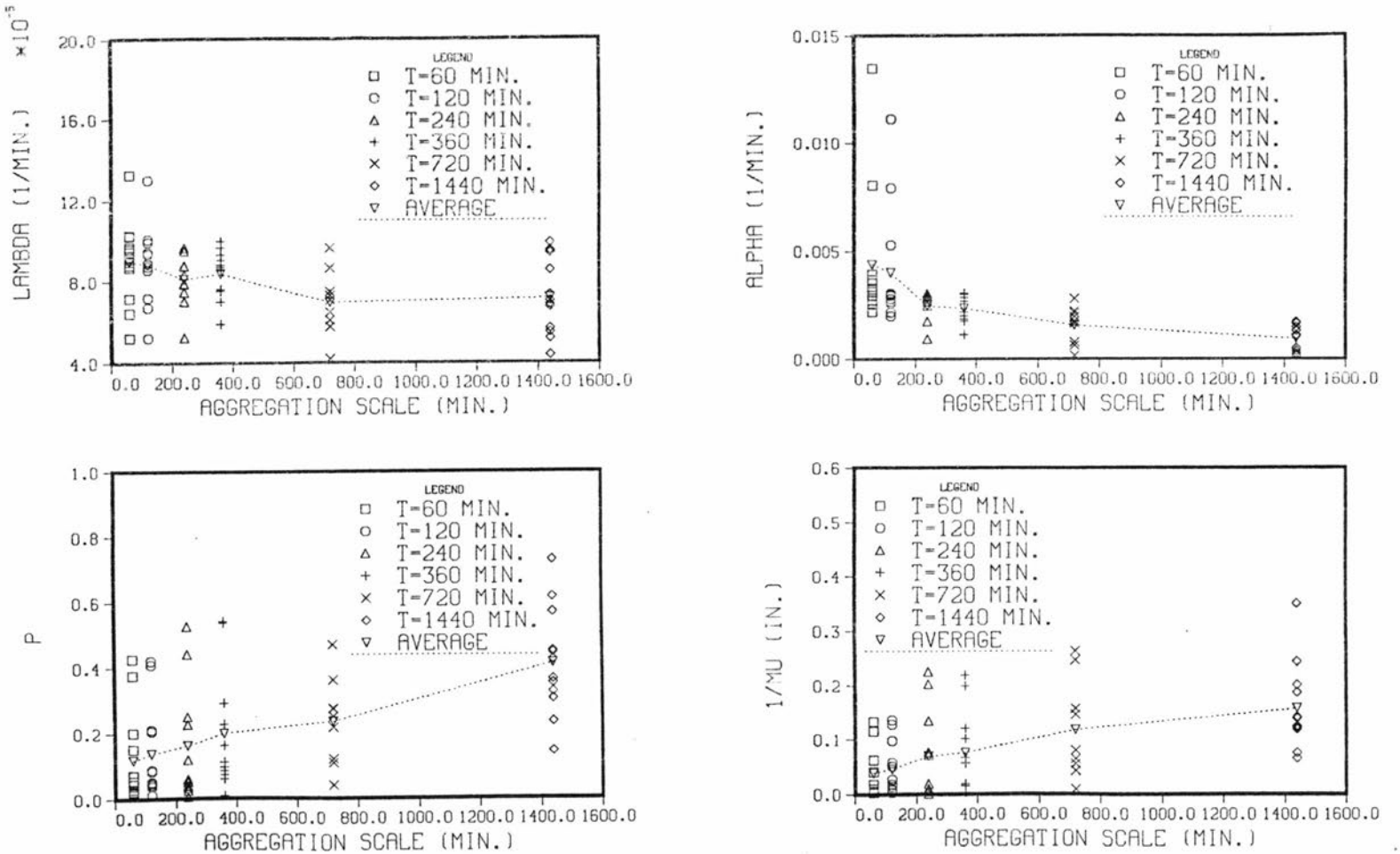


Figure 7.14 Variation of monthly estimated parameter values for the NSWN model with the aggregation scale, for Denver Wsfo Ap station (WLS method).

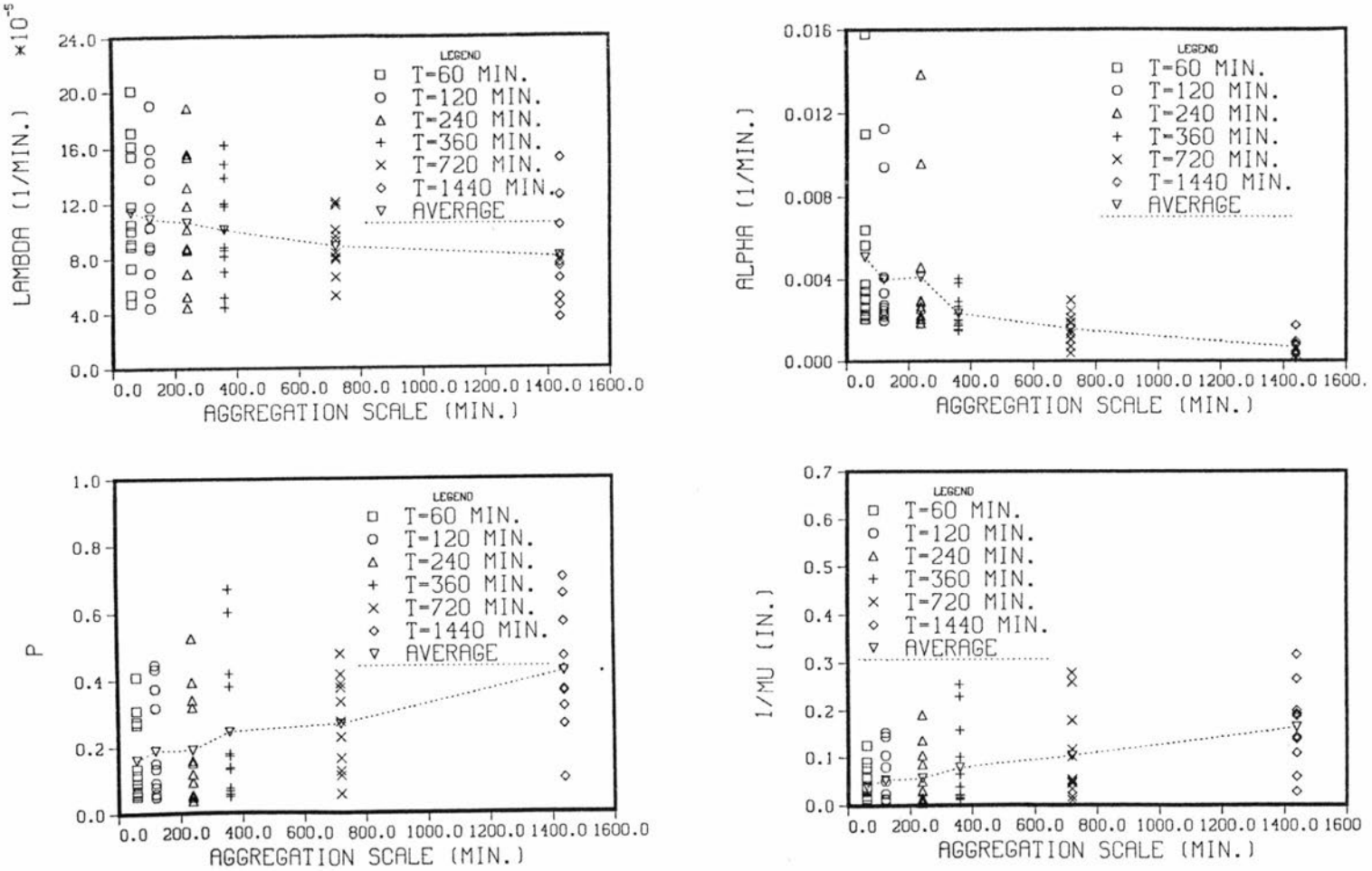


Figure 7.15 Variation of monthly estimated parameter values for the NSWN model with the aggregation scale, for Greenland 9 SE station (WLS method).

Plots presented in Figures 7.12 to 7.15 agree in the trend of variation of estimated parameters with the aggregation scale. For λ and α, estimated values tend to decrease with the aggregation scale. For p and $1/\mu$, estimated values increase with the aggregation scale. In terms of model performance, these trends translate in the model recognizing a smaller number of storms, with a smaller number of cluster members, located farther from the cluster center, each member yielding a larger instantaneous amount of precipitation, as the aggregation scale increases.

Plots derived for simulation 1 show that the variation of the cluster arrival rate is uniform for the range of aggregation scales investigated. The remaining parameters, p, α and μ, are characterized by a moderate variation followed by a steep increase or decrease. Although similar observations can be drawn from the plots for the two precipitation recording stations, care must be exercised in the sense that Figures 7.14 and 7.15 present seasonal estimated parameter values.

Although parameter estimation has not been one of the main objectives in this research, the course of the investigation has been forced in this direction. No further analysis is given here in regard to variation of parameter estimates with the aggregation scale, due mainly to space and time limitations. However, an important feature is observed in Figures 7.12 and 7.13. The existence of a region in the temporal scale domain where estimated parameter values appear highly consistent. For example, for simulation 1, this region could be between 60 and 360 min.. The existence of the same region appears feasible for the precipitation recording stations, as seen in Figures 7.14 and 7.15. Further investigation of this issue, using a more

complete set of simulated and recorded precipitation data and a wider range of aggregation scales, could generate a stronger basis for the applicability of the NSWN model and its corresponding disaggregation version and a better definition of the trends described above. For instance, the parameter in the geometric distribution, p, is expected to increase asymptotically to 1.0, as the aggregation scale increases.

7.10 Improvement of parameter estimates using information at another station

The last point examined in this research, in order to improve parameter estimation, is based on the plots presented in the previous section, where the trend of variation of estimated values with the aggregation scale is shown to maintain a similar shape for simulations and precipitation recording stations. Assume that disaggregation of precipitation records is required at a station, called the problem station, with records available at the aggregation scale. At the same time, at a nearby station or at a station with a similar precipitation regime, called the satellite station, records are available at both the aggregation and disaggregation scales, T_a and T_d, respectively. Monthly statistics are computed and parameters for the NSWN model are estimated using for example the WLS method, for the three samples defined above. Let $\hat{\theta}_P(T)$ and $\hat{\theta}_S(T)$ denote any of the parameters in the NSWN model estimated at a given aggregation scale for the problem and satellite stations, respectively. It is assumed that parameters to use in the disaggregation process at the problem station can be obtained using the following simple linear scheme

$$\hat{\theta}_P(T_d) = \hat{\theta}_P(T_a) \left[\hat{\theta}_S(T_d)/\hat{\theta}_S(T_a) \right] \qquad (7.21)$$

In words, eq. (7.21) states that ratios of estimated values at the aggregation and disaggregation scales are the same at both stations for a given parameter.

The procedure described above was applied to a given month for each one of the precipitation recording stations used in the study. These months were selected from the ones used previously. Table 7.7 gives problem and satellite stations, months and aggregation and disaggregation scales used in the disaggregation process.

Table 7.7 Description of historical samples for application of the NSWND model using information at another station.

Problem	Satellite	Month	T_a (min.)	T_d (min.)	R
Denver Wsfo Ap	Greenland 9 SE	06	1440	60	24
Greenland 9 SE	Denver Wsfo Ap	06	1440	60	24
Idaho Springs	Ward	08	60	5	12
Ward	Idaho Springs	04	60	5	12

Table 7.8 lists parameter values included in the disaggregation process, computed according to eq. (7.21), with estimated values given for every month and every station in Tables 6.9 to 6.12.

Table 7.8 Parameter values for application of the NSWND model using information at another station.

Station	Month	λ (1/min.) $\times 10^{-3}$	p	α (1/min.) $\times 10^{-2}$	μ (1/in.)
Denver Wsfo Ap	06	0.10865	0.22934	0.96697	11.8096
Greenland 9 SE	06	0.13626	0.17731	0.21657	16.9685
Idaho Springs	08	0.16218	0.02772	1.75685	86.6855
Ward	04	0.17772	0.01521	0.39336	423.999

The disaggregation process was applied as described before. Table 7.9 compares statistics for historical and NSWND disaggregated series and Figures 7.16 to 7.19 present correlograms and sample probability distribution functions for the same samples.

Table 7.9 Comparison of statistics for historical and NSWN disaggregated hourly series, for some months of the precipitation recording stations, when disaggregation is performed using information at a nearby station.

	Denver Wsfo Ap		Greenland 9 SE	
	Historic	Model	Historic	Model
Month τ:	06		06	
$\bar{Y}_\tau \times 10^{-3}$ (in.)	2.4015	2.4004	2.7236	2.7162
$S_\tau \times 10^{-3}$ (in.)	21.8696	25.9850	24.6846	20.6980
$\hat{g}_\tau$	19.6414	18.3861	20.5936	12.5958
$r_\tau(1)$	0.44566	0.40695	0.29861	0.21842
$\hat{p}_o^\tau$	0.96046	0.97880	0.96379	0.96665
No of positive values	996	534	877	810
Maximum value (in.)	1.11	1.05	1.08	0.81
No of missing values	732	720	1698	1632

	Idaho Springs		Ward	
	Historic	Model	Historic	Model
Month τ:	08		04	
$\bar{Y}_\tau \times 10^{-3}$ (in.)	0.3329	0.3250	0.4434	0.4350
$S_\tau \times 10^{-3}$ (in.)	4.0560	3.8925	2.6457	3.3311
$\hat{g}_\tau$	28.5674	22.2482	8.3023	10.5313
$r_\tau(1)$	0.71317	0.40425	0.66053	0.36100
$\hat{p}_o^\tau$	0.98389	0.98689	0.96540	0.97614
No of positive values	494	412	942	662
Maximum value (in.)	0.28	0.20	0.07	0.08
No of missing values	13972	13224	15977	15456

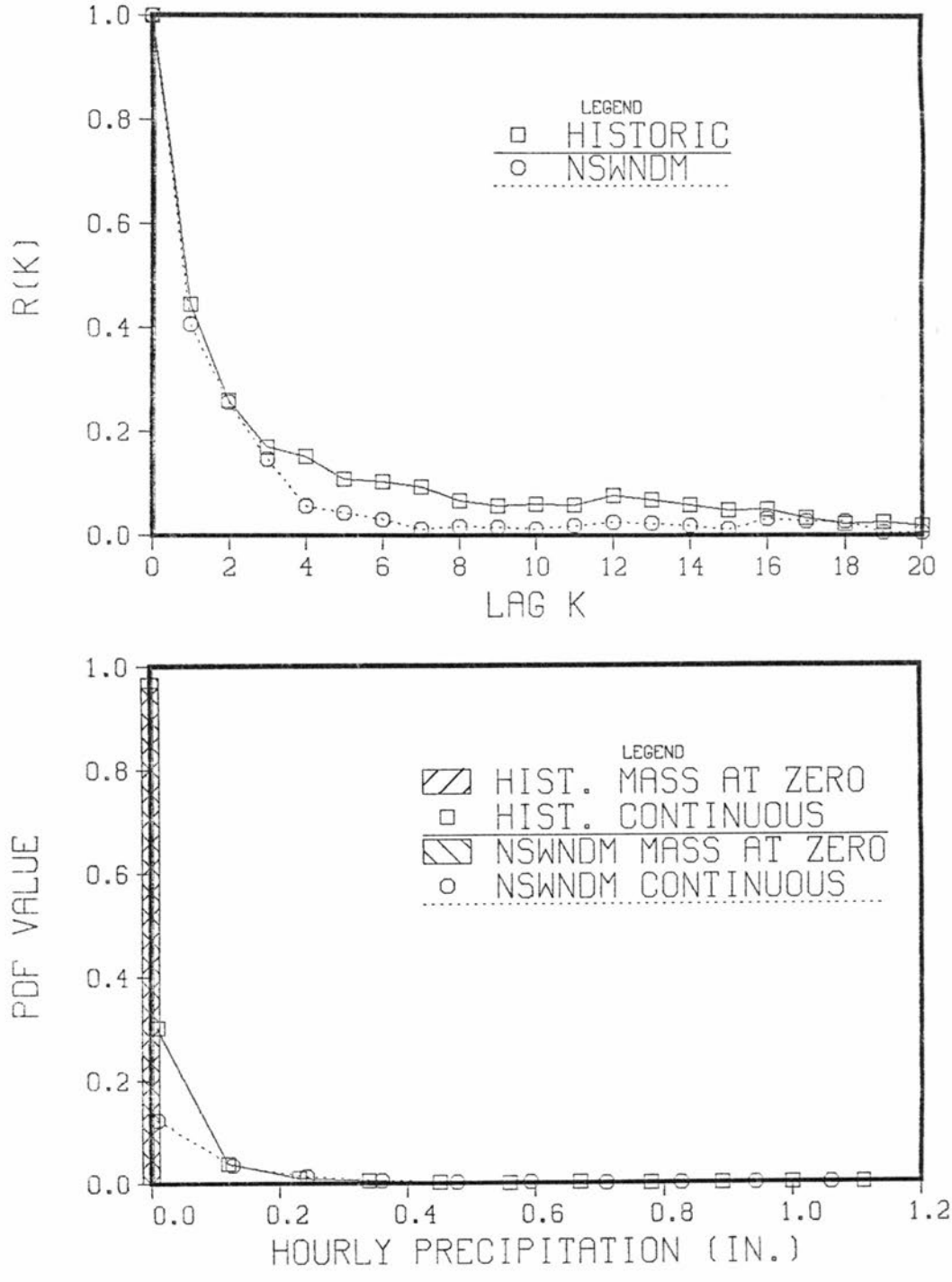


Figure 7.16 Correlogram and sample probability distribution function for hourly historic and NSWN disaggregated precipitation series, for Denver Wsfo Ap station, for month 06, obtained using information at Greenland 9 SE station (T_d=60 min.).

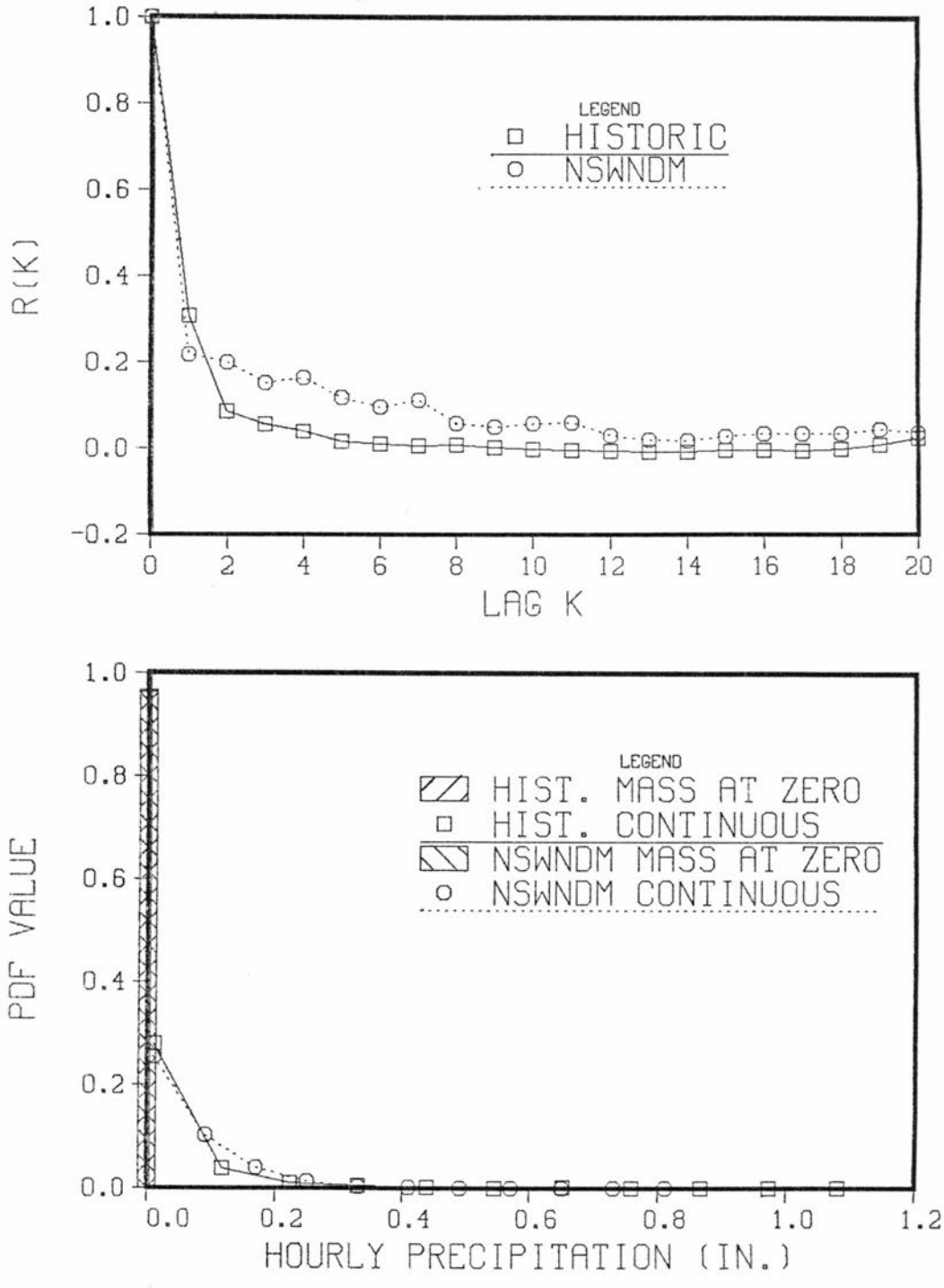


Figure 7.17 Correlogram and sample probability distribution function for hourly historic and NSWN disaggregated precipitation series, for Greenland 9 SE station, for month 06, obtained using information at Denver Wsfo Ap station (T_d=60 min.).

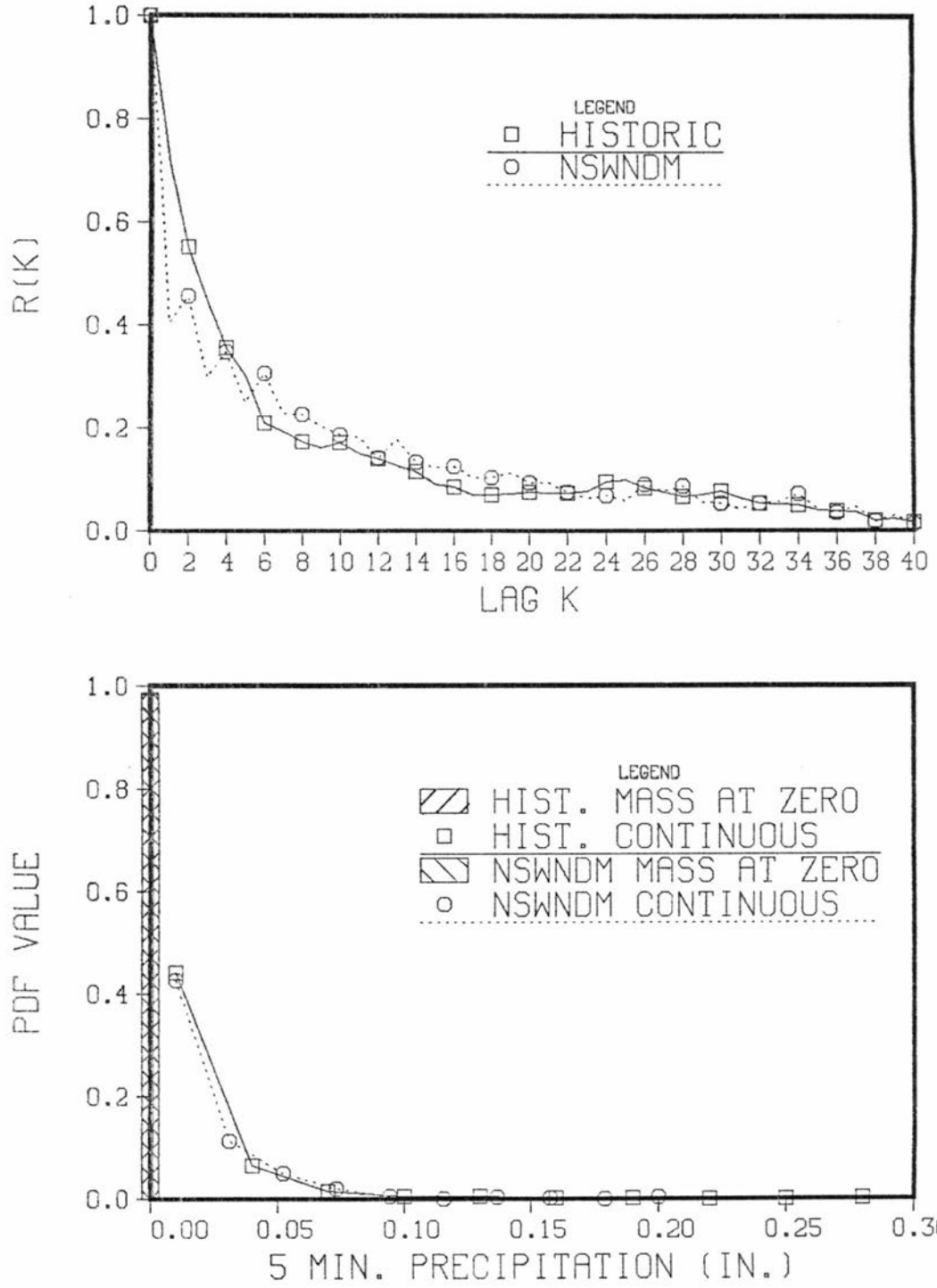


Figure 7.18 Correlogram and sample probability distribution function for 5 min. historic and NSWN disaggregated precipitation series, for Idaho Springs station, for month 08, obtained using information at Ward station T_d=5 min.).

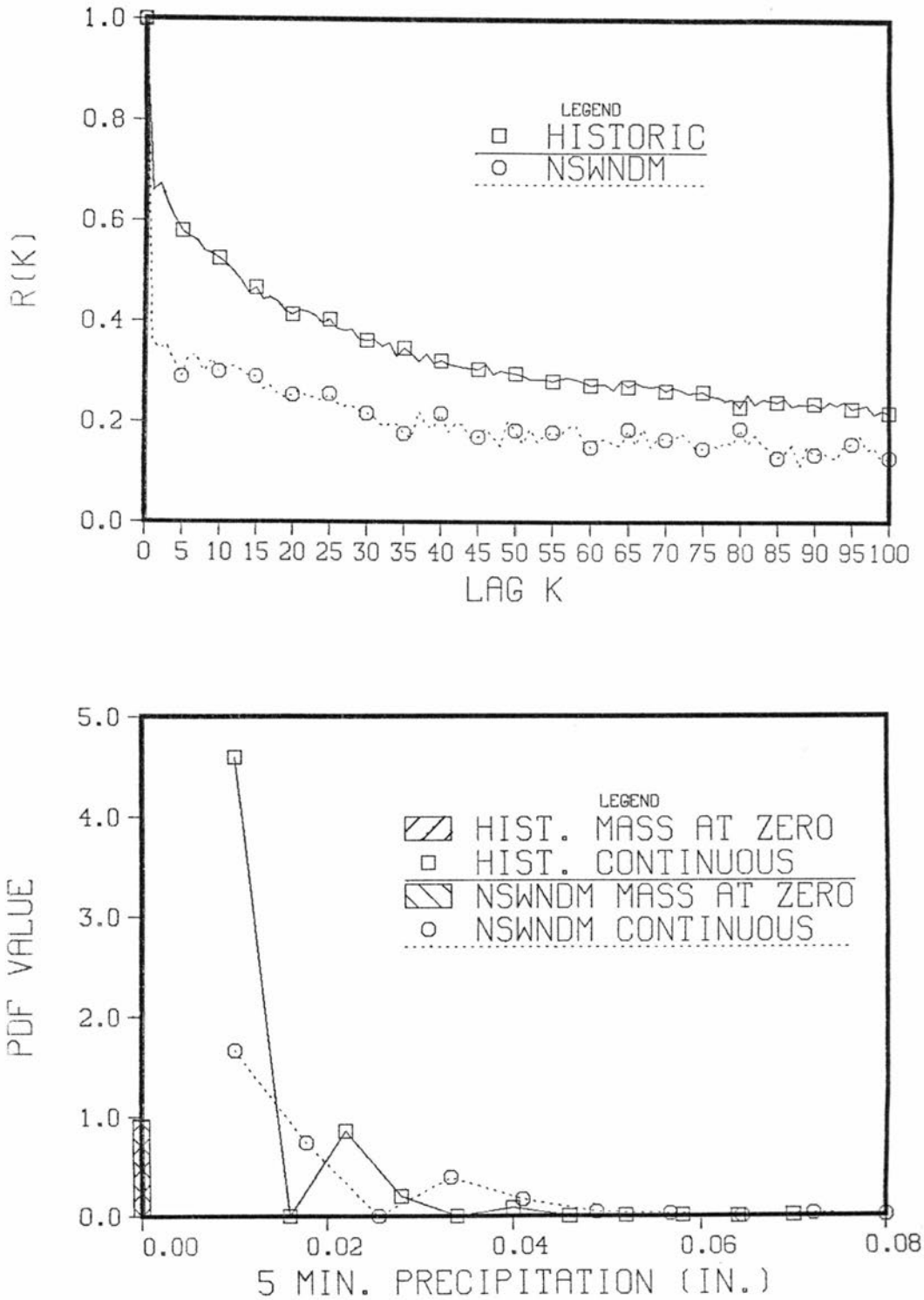


Figure 7.19 Correlogram and sample probability distribution function for 5 min. historic and NSWN disaggregated precipitation series, for Ward station, for month 04, obtained using information at Idaho Springs station (T_d=5 min.).

Values given in Table 7.9, when compared to values listed in Tables 7.3 to 7.6 for the corresponding months, show that a good improvement in disaggregation results is obtained when parameters to use in the disaggregation process are corrected according to eq. (7.21). Similar conclusions are drawn from Figures 7.16 to 7.19. Note that improvement of results is obtained for the four months in the four precipitation stations.

Although the technique illustrated above appears highly promising, the following points must be brought into attention. First, it is completely empirical. More complex versions of the scheme in eq. (7.21) could be investigated. Second, the definition of nearby stations is not clear so far and it does not follow a sound criteria. As stated in Chapter 3, spatial location was not considered in the selection of the precipitation recording stations. Spatial correlation between them is unknown and the way in which they were paired before obeys more to the stations having the same measurement scales than having high correlation or similar precipitation regimes. These two points could be attacked by formulation and use of space-time models and more in depth analysis of similarities and spatial correlation between stations.

Chapter 8

CONCLUSIONS AND RECOMMENDATIONS

8.1 Summary

This research has dealt with disaggregation of precipitation records taken at short sampling intervals. These records are defined as precipitation samples where intermittence is present. The need for precipitation disaggregation algorithms is real in the sense that investigation of certain hydrology processes, the study of some water resources related problems and the solution of design problems may require, in many instances, precipitation samples at a finer resolution than those available or recorded. High resolution precipitation records are scarce, costly and time consuming to process, store and maintain.

Survey of the current state of the art in precipitation modeling shows that only a few models have been built to perform disaggregation. These models are empirical and they have not been used in practice. Within the reviewed literature, two classes of models appear suitable for formulation of disaggregation algorithms: point process models and time series models. Within the first class, continuous and discrete formulations are distinguished. The class of point process evolving in continuous time is selected for this investigation. Initially, four models are adopted: Poisson White Noise (PWN), Poisson Rectangular Pulse (PRP), Neyman-Scott White Noise (NSWN) and Neyman-Scott Rectangular Pulse (NSRP).

The working definition for disaggregation considers a stochastic process governing precipitation intensity formation and a given recorded sample. The idea is to produce precipitation samples at a finer temporal scale which preserve the stochastic process, recorded amounts of precipitation and statistical properties at two or more temporal scales.

The general approach adopted in this research started with the simplest model (PWN model) and work toward more complicated formulations. However, tremendous difficulties were faced in the analytical derivation of some of the required distributions to formulate disaggregation methods based on Rectangular Pulse type models. Therefore, the set of models analyzed in this study was reduced to PWN and NSWN models.

Four precipitation recording stations located in Colorado were selected to test the disaggregation models. A monthly partition was adopted to account for within the year periodicity. For the four stations, monthly statistics up to the fourth order, including correlograms, were estimated at different temporal aggregation levels. Sample probability distribution functions for the amount of precipitation were considered also. The role of the aggregation scale on estimated statistics, specially the correlogram, was investigated. Results show that clustering and monthly periodicity are important characteristics found in precipitation samples.

In order to perform disaggregation as defined above, and given the selected models, distributions are required for the following random variables, conditional on the recorded amount of precipitation in a given time interval: number of occurrences, White Noise terms and arrival times.

The PWN model is the simplest. Derivation of properties and distributions is relatively easy, given the independence between occurrences in disjoint time intervals for the Poisson process. An important limitation of the model is the lack of serial correlation for the amount of precipitation. Assuming that the PWN model describes the recorded precipitation process, a second shortcoming is given by the incompatibility of the model at different aggregation scales, in the sense that parameter values estimated at different aggregation scales are significantly different.

The NSWN model is based on a clustering stochastic process to specify the arrival of precipitation in time. The degree of complication in derivations is increased considerably as compared to the PWN model. The distribution for the number of events in a given interval is derived up to 20 occurrences. For larger number of arrivals, an empirical model is proposed and tested. Traditional method of moments is used to estimate model parameters. This method usually gives a large number of estimation failures, specially for large aggregation scales, and shows incompatibility in parameter estimates. A second method is implemented based on moment properties of the process and numerical minimization. The Weighted Least Squares (WLS) method uses more information than method of moments and accounts for sample variability or scattering of properties included in the objective function. The main contribution of this method is a decrease in the number of failures. However, the price paid is larger variability in parameter estimates. This last method does not solve the problem of model incompatibility in parameter estimates at different aggregation scales.

From the analytical point of view, the formulation of a disaggregation algorithm is successful for the Neyman-Scott White Noise model. However, a major difficulty is found in the derivation of the distribution of arrival times, conditional on a given number of occurrences in the interval. The problem is solved for one and two arrivals. For larger number of occurrences, simulation and sampling algorithms are implemented and tested. The disaggregation model is tested on NSWN simulated series and on monthly samples for the four precipitation recording stations. Model incompatibility at different temporal scales inhibits good performance of the model. However, analysis of the variation of parameter estimates with the aggregation scale suggests the existence of a feasible region for application of the NSWN model, where parameter estimates appear highly compatible. Furthermore, use of precipitation information at a nearby station is shown to be a promising alternative to improve parameter estimation. The information at the other station is in the form of variation of parameter estimated values with the temporal scale.

8.2 Conclusions

The conclusions given bellow pertain to the Poisson White Noise (PWN) and the Neyman-Scott White Noise (NSWN) models. Conclusions are presented for the model itself and for the disaggregation algorithm derived from each model. Finally, some general conclusions are given.

1. From theoretical and practical points of view, the PWN model is not suitable for description of the precipitation process.

This conclusion is based on its lack of serial correlation and in the incompatibility of the model at different aggregation scales.

2. The PWN model played an important analytical role in this investigation. It was helpful in devising the disaggregation algorithm. Some of the analytical results obtained for the PWN model can be directly applied to any model with White Noise elements, especially the Neyman-Scott White Noise model.

3. The PWN model with exponential White Noise terms performed very well when applied to PWN simulated series, both in estimation of parameters and disaggregation of samples. In this sense, the model is recommended for its use in other fields of hydrology or other areas of science, for which the Poissonian assumption can be sustained. No improvement was found in treating the White Noise terms as gamma distributed.

4. The NSWN model is conceptually sound, at least in regard to modeling precipitation arrivals. Derivation of properties and distributions for the NSWN model was difficult and in some cases no analytical solutions were found. The model is complex, specially in regard to its application to precipitation disaggregation.

5. The incompatibility of parameter estimates for the NSWN model at different temporal aggregation scales makes it of limited application in precipitation modeling. However, for a given

scale, with parameters estimated at that scale, the model is adequate for description of the precipitation process and its use is recommended, for instance, for simulation purposes. For this particular scale, the NSWN model is able to reproduce the type of correlation found in precipitation records. This conclusion is based on the observation that sample statistics, specially correlograms, behaved similarly for NSWN simulated series and recorded precipitation samples.

6. The disaggregation version of the NSWN model performed well as long as parameter values used were similar to the ones estimated at the disaggregation scale. This behavior was observed for both simulated and recorded precipitation samples. Given the above requirement, the model has the ability to reproduce sample precipitation statistics up to the third degree, including correlograms, for the aggregation and disaggregation scales. Preservation of sample probability distribution functions was also observed. In this sense, the NSWND model is considered a good model and a contribution to precipitation disaggregation.

7. Analysis of the variation of estimated parameters with the aggregation scale for the NSWN model suggested the existence of a region in the temporal scale domain where the NSWN model appears compatible. This region could be a feasible region for application of the model. However, this point requires further investigation.

8. In regard to parameter values to be used in NSWN precipitation disaggregation process, the inclusion of information at a nearby station or at a station with a similar precipitation regime has been shown to improve results. As before, this issue needs additional investigation.

9. The Weighted Least Squares method implemented in this research decreases the number of estimation failures as compared to the traditional method of moments, for the NSWN model. Method of moments complements the optimal estimation technique and they should be used in this way. The Weighted Least Squares method performed well for two of the reasons that motivated its development: use of more information and subjective incorporation of sample variability or scattering of statistics in the estimation procedure. In this sense, the use of the Weighted Least Squares method makes the NSWN model applicable to more precipitation samples in a wider range of temporal scales.

10. Incompatibility of model parameters was observed for the NSWN model for simulated and precipitation samples. Therefore, the incompatibility is generated by the model itself, by the statistics used in the estimation process and by the amount of aggregation present in the samples and in the statistics.

The general conclusion, found in almost every study of this type, is that the problem of precipitation disaggregation has not been exhausted or solved completely. This research provides additional

exploration. Results are not completely satisfactory, although under certain circumstances, the developed disaggregation models have been shown to perform well. Important contributions have been given in the analytical description and practical application of the Neyman-Scott process.

Introduction of the NSWN model generates correlations by means of clustering. The use of White Noise terms remains unrealistic. In this sense, one should not expect the NSWN model to be universal, fitting any precipitation regime. Many cases shown in this report exemplify good and bad fits. Goodness of fit is usually measured through reproduction of correlation at one or several temporal scales. When the NSWN fits well a recorded precipitation sample, the disaggregation version performs properly conditional on the use of parameter values similar to those estimated at the disaggregation scale.

The PWN and NSWN models were applied to limited ranges of temporal scales. The PWN model was applied from 5 min. to monthly scales. Attempts of disaggregation from monthly to daily values were based mainly on the observed lack of correlation at these two temporal scales. The NSWN model was not applied to the monthly case since correlation is required for parameter estimation. Applications for the NSWN model were constrained to the range 5 min. to 1 day.

The PWN and NSWN models belong to the class of temporal precipitation models, which include the temporal scale as a variable to represent the process. However, as shown above, practical considerations impose a range of application on the temporal scale. Further constraints are found when model parameters are estimated. These may not exist or if they exist incompatibility may arise from

loss of information as samples are aggregated or from poor fit of the model to precipitation samples.

Investigation of the variation of parameter estimates with the temporal scale for the NSWN model has suggested the existence of the feasible region in the temporal domain where parameter estimates appear compatible. It seems reasonable to expect the feasible region to be smaller than the limited range of application of the model.

In addition to the above concepts, there is a component of impossibility in precipitation modeling (Smith, 1990). As aggregation proceeds, from no model and with no method of estimation can one accurately extract the underlying precipitation process, as defined and used in this investigation. As mentioned earlier, for example, it is not possible to obtain reasonable NSWN parameter estimates from precipitation samples aggregated to a monthly level.

8.3 Recommendations

The recommendations given here arise from the whole examination of this research. Investigation of precipitation modeling and disaggregation of precipitation in particular must be continued. Consideration should be given to disaggregation procedures based on Discrete Autoregressive Moving Average (DARMA) models, discrete point processes theory and more causal or Black Box type models.

Analytical improvement of the Neyman-Scott White Noise disaggregation model does not seem feasible given the current state of the art. If this model and in general temporal precipitation models continue to be used, priority must be given to the topics of parameter estimation, model compatibility and feasible region in the temporal scale domain for model application. Modified estimation approaches

could include the probability of zero precipitation as an additional property. New estimation approaches could try to fit the theoretical pdf for the amount of precipitation to its sample counterpart. In regard to estimation techniques, additional research is required in relation to sample variability.

As shown before, the investigation of variation of parameter estimates with the aggregation scale has the potential to improve results. As a first point, results presented in this research should be expanded to include more simulations, more precipitation recorded data and a wider range of aggregation scales. This will verify the existence of the feasible region in the temporal scale domain for application of the NSWN model. Success in the first point could direct the research toward the investigation, using NSWN simulated samples, of dimensionless curves showing the variation of parameter estimated values with aggregation scale and their possible use in parameter estimation and precipitation disaggregation. Parameters with dimensions 1/time could be made dimensionless by multiplying by the aggregation scale. A major difficulty will be posed by the parameter indexing the distribution of the White Noise terms, since this will be dependent on dimensions. These curves must be indexed by population parameters and sample size.

In order to apply disaggregation models, estimation of parameters in this research has been kept at the aggregation scale, without using information at other precipitation recording stations. However, it has been shown that use of precipitation information at a nearby station or at a station with similar precipitation regime could generate better results. As stated before, tests performed in this research did not go beyond pairing stations according to the

measurement scale. Necessarily, in order to expand this point, a space-time formulation of the process would be required, which would pose large difficulties. Topics to investigate would include the spatial correlation values required to improve parameter estimation and formulation of joint estimation of parameters for two or more stations.

The variation of parameter estimates with the aggregation scale presented in this research justified the use of a simple linear scheme to correct parameter values to be used in the NSWND model. In this sense, the investigation of more complex schemes could be coupled to some of the recommendations given before.

Disaggregation algorithms developed in this research are applicable to long samples of recorded precipitation, although nothing inhibits their application to isolated precipitation storms. An event based Neyman-Scott White Noise disaggregation model could be implemented assuming that it is possible to isolate clusters or storms in the precipitation record. The isolation is subjective. Distributions for the following random variables would be required: total precipitation depth or total yield per storm, storm duration, cluster size (number of White Noise terms) conditional on the total cluster yield, kth White Noise term conditional on the cluster size, on k-1 previous White Noise terms and on the total amount of precipitation in the cluster. For example, location times for the White Noise terms with respect to the cluster center arrival time are the order statistics of an exponential distribution.

Traditional method of moments when applied to temporal precipitation models is characterized by a large number of estimation failures. Extension of the Weighted Least Squares method to other

models for which estimation is based on fitting sample correlograms could expand the applicability of these models. Candidate models for this are: Poisson Rectangular Pulse, Neyman-Scott Rectangular Pulse and Barttlet-Lewis Rectangular pulse. Within this same line of research, a systematic investigation of the variation of parameter estimates with the aggregation scale would allow the verification and extension of the concept of feasible region for model application to these other models.

REFERENCES

Amorocho, J. and B. Wu, Mathematical Models for the Simulation of Cyclonic Storm Sequences and Precipitation Fields, J. Hydro., 32, 329-345, 1977.

Bell, T. L., A Space-Time Stochastic Model of Rainfall for Satellite Remote-Sensing Studies, J. Geoph. Res., 92(D8), 1987.

Berndtsson, R. and J. Niemczynowicz, Spatial and Temporal Scales in Rainfall Analysis - Some Aspects and Future Perspectives, J. Hydro., 100, 293-313, 1988.

Bras R. L. and I. Rodriguez-Iturbe, Rainfall Generation: A Nonstationary Time-Varying Multidimensional Model, Water Resour. Res., 12(3), 450-456, 1976.

Burlando P., Modelli Stocastici per la Previsione e la Simulazione della Precipitazione nel Tempo, Dissertazione, Politecnico di Milano, Istituto di Idraulica, Milano, 1989.

Cadavid L. G., D. C. Boes and J. D. Salas, Disaggregation of Short Term Precipitation Records, Report, Civ. Eng. Dept., Col. St. Univ., Ft. Collins, Colorado, 1991.

Chang T. J., M. L. Kavvas and J. W. Delleur, Daily Precipitation Modeling by Discrete Autoregressive Moving Average Processes, Water Resour. Res., 20(5), 565-580, 1984.

Corotis R. B., Stochastic Considerations in Rainfall Modeling, J. ASCE, HY7, 865-879, 1976.

Cox D. R. and V. Isham, Point Processes, Chapman and Hall, London, 1980.

Daley D. J. and D. Vere-Jones, A Summary of the Theory of Point Processes, in Stochastic Point Processes, ed. by P. A. W. Lewis, 299-383, 1972.

Eagleson P. S., N. M. Fennesey, W. Qinliang and I. Rodriguez-Iturbe, Application of Spatial Poisson Models to Air Mass Thunderstorm Rainfall, J. Geoph. Res., 92(D8), 9661-9678, 1987.

Entekhabi D., I. Rodriguez-Iturbe and P. S. Eagleson, Probabilistic Representation of the Temporal Rainfall Process by a Modified Neyman-Scott Rectangular Pulse Model: Parameter Estimation and Validation, Water Resour. Res., 25(4), 1989.

Foufoula-Georgiou E. and P. Guttorp, Compatibility of Continuous Rainfall Occurrence Models with Discrete Rainfall Observations, Water Resour. Res., 22(8), 1316-1322, 1986.

Foufoula-Georgiou E. and D. P. Lettenmaier, A Markov Renewal Model of Rainfall Occurrences, Water Resour. Res., 23(5), 1987.

Garcia-Bartual, R. and J. Marco, A Design Pluviograph Based on Stochastic Structure of Precipitation, XII IAHR Cong. and 4th International Conf. on Urban Storm Drainage, 41-46, 1987.

Georgakakos K. P. and R. L. Bras, A Hydrologically Useful Precipitation Model: 1. Formulation, Water Resour. Res., 20(11), 1585-1596, 1984a.

Georgakakos K. P. and R. L. Bras, A Hydrologically Useful Precipitation Model: 2. Case Studies, Water Resour. Res., 20(11), 1597-1610, 1984b.

Giambelluca T. W. and D. S. Oki, Temporal Disaggregation of Monthly Rainfall Data for Water Balance Modeling, Proc. of the Vancouver Symp., IAHS Pub. No. 168, 255-267, 1987.

Guttorp, P., On Binary Time Series Obtained from Continuous Time Processes Describing Rainfall, Water Resour. Res., 22(6), 1986.

Guttorp, P., Analysis of Event Based Precipitation Data with a View Toward Modeling, Water Resour. Res., 24(1), 35-43, 1988.

Hershenhorn J. and D. A. Woolhiser, Disaggregation of Daily Rainfall, J. Hydro., 95, 299-322, 1987.

Islam S., R. L. Bras and I. Rodriguez-Iturbe, Multidimensional Modeling Of Cumulative Rainfall: Parameter Estimation and Model Adequacy trough a Continuum of Scales, Water Resour. Res., 24(7), 985-992, 1988.

Jacobs B. L., I. Rodriguez-Iturbe and P. S. Eagleson, Evaluation of a Homogeneous Point Process Description of Arizona Thunderstorm Rainfall, Water Resour. Res., 24(7), 1988.

Kavvas M. L. and J. W. Delleur, The Stochastic and Chronologic Structure of Rainfall Sequences - Application to Indiana, Tech. Rep. 57, Purdue Univ. Water Resour. Ctr., 1975.

Kavvas M. L. and J. W. Delleur, A Stochastic Cluster Model of Daily Rainfall Sequences, Water Resour. Res., 17(4), 1151-1160, 1981.

Kedem B. and L. S. Chiu, Are Rain Rate Processes Self-Similar ?, Water Resour. Res., 23(10), 1816-1818, 1987.

Lawrance A. J. Some Models for Stationary Series of Univariate Events, in Stochastic Point Processes, ed. by P. A. W. Lewis, 199-256, 1972.

Le Cam L. A Stochastic Description of Precipitation, Proc. 4th Berkeley Symp. on Math., Stat. and Prob., Univ. of Calif. Press, 3, 165-186, 1961.

Liu N. and J. D. Salas, The Impact of Aggregation Levels on Parameter Estimation for Three Stochastic Rainfall Models, Unpublished Manuscript, Col. St. Univ., Ft. Collins, Colorado, 1988.

Mood A. H., F. A. Graybill and D. C. Boes, Introduction to the Theory of Statistics, McGraw Hill Book Co., 1974.

Moss M. E. and H. F. Lins, A Study of the Implications of Climatic Uncertainty, U. S. Geological Survey, Circular 1030,1988.

Moyal J. E., The General Theory of Stochastic Population Processes, Acta Mathematica, 108, 1-31, 1962.

Neyman J. and E. L. Scott, A Theory of the Spatial Distribution of Galaxies, Astrophysical J., 116, 144-163, 1952.

Neyman J. and E. L. Scott, Statistical Approach to Problems of Cosmology, J. Roy. Stat. Soc., Ser. B, 20(1), 1-43,1958.

Nguyen V. T. V. and J. Rousselle, A Stochastic Model for the Time Distribution of Hourly Precipitation, Water Resour. Res., 17(2), 399-409, 1981.

Obeysekera J. T. B., G. Q. Tabios III and J. D. Salas, On Parameter Estimation of Temporal Rainfall Models, Water Resour. Res., 23(10), 1837-1850, 1987.

Ormsbee L. E., Rainfall Disaggregation Model for Continuous Hydrologic Modeling, J. Hydra. Eng., ASCE, 115(4), 507-525, 1989.

Parzen E., Stochastic Processes, Holden-Day, Inc., 1964.

Press W. H., B. P. Flannery, S. A. Teukolsky and W. T. Vetterling, Numerical Recipes: The Art of Scientific Computing, Cambridge University Press, 1986.

Ramirez J. A. and R. L. Bras, Optimal Irrigation Control using Stochastic Cluster Point Process for Rainfall Modelling and Forecasting, Ralph M. Parsons Laboratory, Rep. 275, MIT, 1982.

Ramirez J. A. and R. L. Bras, Conditional Distributions of Neyman-Scott Models for Storm Arrivals and their use in Irrigation Control, Water Resour. Res., 21(3), 317-330, 1985.

Raudkivi A. J. and N. Lawgon, Simulation of Rainfall Sequences, J. Hydro., 22, 271-294, 1974.

Rodriguez-Iturbe I., V. K. Gupta and E. Waymire, Scale Considerations in the Modeling of Temporal Rainfall, Water Resour. Res., 20(11), 1611-1619, 1984.

Rodriguez-Iturbe I., Scale of Fluctuation of Rainfall Models, Water Resour. Res., 22(9), 15S-37S, 1986.

Rodriguez-Iturbe I. and P. S. Eagleson, Mathematical Models of Rainstorm Event in Space and Time, Water Resour. Res., 23(1), 181-190, 1987.

Rodriguez-Iturbe I., D. R. Cox and V. Isham, Some Models for Rainfall based on Stochastic Point Processes, Proc. Roy. Soc. of London, Ser. A, 410, 269-288, 1987a.

Rodriguez-Iturbe I., B. Febres de Power and J. B. Valdes, Rectangular Pulses Point Process Models for Rainfall: Analysis of Empirical Data, J. Geoph. Res., 92(D8), 9645-9656, 1987b.

Roldan J. and D. A. Woolhiser, Stochastic Daily Precipitation Models: 1. A Comparison of Occurrence Processes, Water Resour. Res., 18(5), 1451-1459, 1982.

Salas J. D., J. W. Delleur, V Yevjevich and W. L. Lane, Applied Modeling of Hydrologic Time Series, Water Resour. Pub., 1985.

Salas J. D., R. A. Smith and G. Tabios, Notes for the Course CE-622, Statistics in Water Resources, Col. St. Univ., Ft. Collins, Colorado, Fall 1987.

Santos E. G., Disaggregation Modeling of Hydrologic Time Series, Dissertation, Col. St. Univ., Ft. Collins, Colorado, 1983.

Sivapalan M. and E. F. Wood, A Multidimensional Model of Nonstationary Space-Time Rainfall at the Catchment Scale, Water Resour. Res., 23(7), 1987.

Smith J. A. and A. F. Karr, Statistical Inference for Point Process Models of Rainfall, Water Resour. Res., 21(1), 73-79, 1985.

Smith J. A., Statistical Modeling of Daily Rainfall Sequences, Water Resour. Res., 23(5), 885-893, 1987.

Taylor, M. H. and S. Karlin, An Introduction to Stochastic Processes, Academic Press Inc., 1984.

Todorovic P. and Yevjevich V., Stochastic Process of Precipitation, Hydro. Pap. 35, Col. St. Univ., Ft. Collins, Colorado, 1969.

Vere-Jones D., Some Applications of Probability Generating Functionals to the Study of Input-Output Streams, J. Roy. Stat. Soc., Ser. B, 30(2), 321-333, 1968.

Vere-Jones D., Stochastic Models for Earthquake Occurrences, J. Roy. Stat. Soc., Ser. A, 32(1), 1-62, 1970.

Waymire E. and V. K. Gupta, The Mathematical Structure of Rainfall Representations: 1. A Review of Stochastic Rainfall Models, Water Resour. Res., 17(5), 1261-1272, 1981a.

Waymire E. and V. K. Gupta, The Mathematical Structure of Rainfall Representations: 2. A Review of the Theory of Point Processes, Water Resour. Res., 17(5), 1273-1285, 1981b.

Waymire E. and V. K. Gupta, The Mathematical Structure of Rainfall Representations: 3. Some Applications of the Point Process Theory to Rainfall Processes, Water Resour. Res., 17(5), 1287-1294, 1981c.

Waymire E., V. K. Gupta and I. Rodriguez-Iturbe, A Spectral Theory of Rainfall Intensity at the Meso-β Scale, Water Resour. Res., 20(10), 1453-1465, 1984.

Westcott M., The Probability Generating Functional, The J. Australian Math. Soc., 14, Part 4, 448-466, 1972.

Woolhiser D. A. and J. Roldan, Stochastic Daily Precipitation Models: 1. A Comparison of Distributions of Amounts, Water Resour. Res., 18(5), 1461-1468, 1982.

Woolhiser D. A. and H. B. Osborn, A Stochastic Model of Dimensionless Thunderstorm Rainfall, Water Resour. Res., 21(4), 511-522, 1985a.

Woolhiser D. A. and H. B. Osborn, Point Storm Disaggregation - Seasonal and Regional Effects, 4th Inter. Hydro., 105-120, Ft. Collins, Colorado, 1985b.

Woolhiser D. A. and W. Econopouly, Stochastic Characteristics of Rainfall Events, Proc. the Sixth Annual AGU Front Range Branch HYDROLOGY DAYS, 25-36, Ft. Collins, Colorado, 1986.