

Simplifying the analysis of Climatic data using GenStat

*A thesis submitted in partial fulfillment of the requirements of the award of the
degree of*

Master of Science in Applied Statistics

By

Jacob Otieno Ong'ala
Department of Mathematics and Applied Statistics
Maseno University
Private Bag
Maseno Kenya

©November 2010

Declaration

I certify that this thesis is my original work, and has not been presented in any other university for a degree. The work reported herein has been carried out by me and all sources of information have been accredited by means of references.

Sign _____

Date _____

Jacob Otieno Ong'ala

This thesis has been submitted for examination with our approval as the University Supervisors.

1. Dr. David Andrew Stern

Department of Mathematics and Applied Statistics

Maseno University

Sign _____

Date _____

2. Dr. Roger Daniel Stern

Statistical Services Center

University of Reading

Sign _____

Date _____

Acknowledgement

I would like to acknowledge a number of people without whose help this thesis would not have taken shape. First my thanks go to my supervisors; David Stern and Roger Stern who sacrificed their time to discuss this work with me at all the stages. Their criticism, comments and corrections helped me in the learning process. Secondly, thanks go to all my lecturers and colleagues in the department of Mathematics and Applied Statistics for their team spirit especially to the Chairman, Professor Omolo Ongati, and Mrs. Joyce Oduor for their continuous encouragement. Finally, I would like to thank all my friends and family for their continued support especially my dear wife Vivian for her understanding during my study.

God bless you all

Dedication

I dedicate this thesis to my Mom (Jane Auma), Dad (Caleb Ong'ala), Wife (Vivian Awuor) and Son (Emmanuel Jeffer) for their prayers and moral support during my education. "Ero uru kamano ahinya "

Abstract

Despite the rapid development of statistical packages, a lot of climatic data still remain unanalyzed due to lack of specialized routines in most of the packages. One package has a climatic menu though with limitation on complex analysis such as generalized linear models. Others can perform the Generalized Linear Model analysis but do not have a specialized menu for analyzing climatic data. There is no statistical package currently available which has a specialized capability to do climatic analysis easily and includes the use of generalized linear models. This study starts the work of creating a specialized menu in GenStat for analyzing climatic data by implementing Markov modeling of rainfall data. Four procedures have been written and corresponding dialogues were created to ease their use. Incorporating a climatic menu into GenStat package will support researchers in agricultural and many other fields that need an analysis of climatic data as part of their work.

Contents

Declaration	i
Acknowledgement	ii
Dedication	iii
Abstract	iv
Acronyms	viii
Definitions.	ix
List of Tables	x
List of figures	xi
List of Graphs	xiv
List of Appendixes	xv
Chapter 1: Introduction	1
1.1: Background information	1
1.2: Basic Concept	3
1.2.1: The Generalized Linear Model	3
1.2.2: Markov chain	4
1.2.3: The Markov Chain of first order	5
1.2.4: High Order Markov Chain	5
1.2.5: Modeling Rainfall data using Markov Chain	6
1.2.6: Fitting a first order Markov Model to rainfall data	6
1.3: Statement of the problem	8
1.4: Objective of the study	8
1.5: Significance of the study	8
Chapter 2: Literature review	9
Chapter 3: Fitting Markov Chain Models using Instat and MarkSim	11
3.1: Introduction	11
3.2: Instat.	11
3.2.1: Data preparation	11
3.2.2: Fitting the model	14
3.2.3: First Order Markov Model for rainfall occurrences	15
3.2.4: Second Order Markov Model for rainfall occurrences	18

3.2.5: Modeling rainfall amounts.-----	22
3.3: MarkSim-----	25
3.3.1: Introduction -----	25
3.3.2: The Parameter file -----	26
3.3.3: The Baseline Probits and the Lag Parameters -----	26
Chapter 4: GenStat Command Interface-----	30
4.1: Introduction -----	30
4.2: Command language -----	30
4.3: Compacting programs-----	32
4.3.1: Procedures-----	32
4.3.2: Macros -----	34
4.4: Creating Menus and Dialogs -----	34
Chapter 5: Markov modeling of Climatic data in GenStat -----	36
5.1: Introduction -----	36
5.2: COUNT Procedure. -----	37
5.2.1: Options for COUNT Procedure-----	37
5.2.2: Parameters for Count Procedure -----	38
5.2.3: Description of Count Procedure -----	38
5.3: PREPARE Procedure -----	40
5.3.1: Options for PREPARE procedure -----	40
5.3.2: Parameter for Prepare procedure-----	40
5.3.3: Description of Prepare Procedure-----	40
5.4: FITTING command procedures.-----	41
5.4.1: Options for FITTING procedure-----	41
5.4.2: Parameter for FITTING procedure -----	41
5.4.3: Description of Fitting Procedure-----	42
5.5: AMOUNTFITTING command procedures. -----	42
5.5.1: Options for FITTING procedure-----	42
5.5.2: Parameter for Fitting procedure-----	42
5.5.3: Description of Fitting Procedure-----	43
Chapter 6: Results and Discussion -----	44

6.1: Application of the model to Katumani Data-----	44
6.2: Description of the data-----	44
6.3: Organizing the data-----	44
6.4: The User Menu -----	49
6.5: Two- State Markov chain modeling-----	50
6.5.1: Order zero -----	50
6.5.2: Order one -----	57
6.6: Higher order two- state Markov Chain Modeling-----	61
6.6.1: Order Two Normal Markov Model-----	61
6.6.2: Order Two High Markov Model-----	62
6.6.3: Normal and High Markov models for order three and four-----	65
6.6.4: Order five-----	68
6.7: Markov Chain with Initial data -----	71
6.7.1: Order one -----	71
6.7.2: Order Two-----	72
6.8: Three-state Markov chain Modeling -----	73
6.9: Modeling Rainfall Amount-----	74
6.9.1: Modeling rainfall Amount of order zero -----	74
6.9.2: Modeling rainfall Amount of order one-----	75
6.9.3: Modeling High Markov model of order two for Rainfall Amount-----	78
Chapter 7: Conclusion and Further Work -----	80
References-----	81

Acronyms

w	Wet (used interchangeably with rain)
d	Dry
ww	Wet given wet (equivalent to rain given rain)
wd	Wet given dry the previous day
www	Wet given wet in the previous day given wet in the day before.
ddd	Dry given dry given dry
w2d	Wet given day in the two previous day
p_dd	Probability of dry given dry
p_www	Probability of wet given wet given wet.
f_dd	Fitted probability of rain given rain
GLM	Generalized Linear Model

Definitions.

Logit	The logarithm of the ratio of frequencies of two different categorical and mutually exclusive outcomes such as wet and dry
Macro	A set of GenStat commands.
Markov chain	A stochastic process where the probability of a state today given the previous state and the entire past depends only on the current state
Probit	An inverse cumulative distribution function (CDF), or quintile associated with the standard normal distribution.
Procedure	A series of complete GenStat statements
State	A condition of rainfall amount determined by the threshold value.
Threshold	A limit determining the boundary of state i.e. for a threshold value of one, the values of rainfall less than or equal one will be grouped as dry otherwise wet.

List of Tables

Table 1: <i>Probit values for the twelve months</i>	27
Table 2: <i>Probability values for the twelve months of Katumani Simulated data</i>	28
Table 3: <i>Summary of analysis of fitted probability of rain for 0-order with 2</i>	55
Table 4: <i>Estimate of parameters of fitted probability of rain for 0-order with 2</i>	56
Table 5: <i>Accumulated analysis of deviance of fitted probability of rain for 0-order with 2 harmonics</i>	56
Table 6: <i>Accumulated analysis of deviance of fitted probability of rain for 0-order with 3 harmonics</i>	56
Table 7: <i>Estimate of parameter for rainfall amount model (w given d) with 2 harmonics</i>	75
Table 8: <i>Estimate of parameter for rainfall amount model (w given w) with 2 harmonics</i>	76
Table 9: <i>Estimate of parameter for rainfall amount model (w given d) with 3 harmonics</i>	76
Table 10: <i>Estimate of parameter for rainfall amount model (w given w) with 3 harmonics</i>	77
Table 11: <i>Summary statistics for Katumani rainfall data</i>	85

List of figures

Figure 1: <i>Markov Modeling Menu</i>	11
Figure 2: <i>Counts/Totals Dialogue box</i>	12
Figure 3: <i>Counts of rainy days of order zero.</i>	12
Figure 4: <i>Prepare Dialogue Box</i>	13
Figure 5: <i>Probability of rainy day or order zero</i>	13
Figure 6: <i>Fitting Probability Dialogue box</i>	14
Figure 7: <i>Fitted Probabilities of order zero.</i>	15
Figure 8: <i>Count/Totals dialogue box for order one</i>	16
Figure 9: <i>Counts of rainy days of order one.</i>	16
Figure 10: <i>Probability of rainy days of order one</i>	17
Figure 11: <i>Fitting probabilities dialogue box for order one</i>	18
Figure 12: <i>Counts of rainy days of order two.</i>	19
Figure 13: <i>Probability of rainy days of order two</i>	19
Figure 14: <i>Fitting probabilities dialogue box for order two with four curves</i>	20
Figure 15: <i>Fitting probabilities dialogue box for order two with three curves</i>	21
Figure 16: <i>Count and totals dialogue box for rainfall amount of order zero</i>	23
Figure 17: <i>Total rain amounts of order zero.</i>	23
Figure 18: <i>Fitting rainfall amount for order zero</i>	24
Figure 19: <i>Parameter File for Katumani</i>	26
Figure 20: <i>Sample rainfall data</i>	30
Figure 21: <i>Dialogue for Summary by Groups</i>	31
Figure 22: <i>Command corresponding to Summary by Groups Dialogue.</i>	31
Figure 23: <i>Example that uses Day length procedure</i>	32
Figure 24: <i>Running the commands</i>	32
Figure 25: <i>Procedure Counts- sec 1</i>	33
Figure 26: <i>Procedure Counts- sec 2</i>	34
Figure 27: <i>Macro</i>	34
Figure 28: <i>A .grc file for creating Counts and Amount Daily data menu</i>	36
Figure 29: <i>A .gad file for attaching dialogue command</i>	36
Figure 30: <i>A sample rainfall data and results of count procedures.</i>	39

Figure 31: <i>Stalk column dialogue box</i>	45
Figure 32: <i>Stalked data</i>	45
Figure 33: <i>Fill column Year with Numerical Sequence dialogue box</i>	46
Figure 34: <i>Year in variate</i>	46
Figure 35: <i>Data with day, month and year</i>	47
Figure 36: <i>Calculate dialogue box for date</i>	47
Figure 37: <i>Date column calculated but with a fault message</i>	48
Figure 38: <i>Data set with 'DayOfYear' and 'Rain' column ready for analysis</i>	49
Figure 39: <i>User Menu</i>	49
Figure 40: <i>Count and Totals from daily Data dialogue box</i>	50
Figure 41: <i>Dialogue box for Counts and totals from Daily Data (Order Zero)</i>	51
Figure 42: <i>Counts of rainy days (order zero)</i>	51
Figure 43: <i>Rainfall amounts (order zero)</i>	52
Figure 44: <i>Dialogue box for Probability of rain (Order Zero)</i>	52
Figure 45: <i>Probabilities of rainy day form GenStat (order zero)</i>	52
Figure 46: <i>Dataset withy week number created (Order Zero)</i>	54
Figure 47: <i>Dialogue box for fitting Probability of rain (Order Zero)</i>	55
Figure 48: <i>First order counts and raindall amounts</i>	57
Figure 49: <i>Chance of rain given the state in the previous day</i>	58
Figure 50: <i>Dialogue box for Count (Normal Markov Chain of Order Two)</i>	61
Figure 51: <i>Dialogue box for Count (High orderMarkov Chain of Order Two)</i>	63
Figure 52: <i>Counts and amounts for 2nd -high order Markov chain(table structure)</i>	63
Figure 53: <i>Counts for 2nd -high order Markov chain (variate structure)</i>	64
Figure 54: <i>Probability of rain for 2nd -high order Markov chain</i>	64
Figure 55: <i>Counts of rainy days for High Markov Chain of order 5</i>	69
Figure 56: <i>Amount of rainfall for High Markov Chain of order 5</i>	69
Figure 57: <i>Probability of rain for High Markov Chain of order 5</i>	69
Figure 58: <i>Counts dialogue box with initial value set of a 1st order Markov chain</i>	71
Figure 59: <i>Counts for Two-state 1st order Markov chain with initial</i>	71
Figure 60: <i>Counts dialogue box with initial value set of a 2nd order Markov chain</i>	72
Figure 61: <i>Counts and amount table of 2nd order Markov chain with initial given</i>	72

Figure 62: <i>Counts dialogue zero order 3-state Markov chain</i>	73
Figure 63: <i>Dialogue box for fitting Amount</i>	74
Figure 64: <i>Rainfall data for Katumani imported directly from Instat</i>	84
Figure 65: <i>Dialogue command for Count Procedure</i>	101
Figure 66: <i>Dialogue command for Prepare Procedure</i>	102
Figure 67: <i>Dialogue command for Fitting probability Procedure</i>	102
Figure 68: <i>Dialogue command for Fitting Amounts Procedure</i>	102
Figure 69: <i>.gad file for count dialogue</i>	103
Figure 70: <i>.gad file for Prepare dialogue</i>	103
Figure 71: <i>.gad file for Fitting probability dialogue</i>	103
Figure 72: <i>.gad file for fitting amount dialogue</i>	103

List of Graphs

Graph 1: <i>Overall Chance of rain</i>	14
Graph 2: <i>Fitted Probability curve</i>	15
Graph 3: <i>Probability curves for order two</i>	17
Graph 4: <i>Fitted Probabilities for rainfall of order one</i>	18
Graph 5: <i>Probability curves for order two</i>	20
Graph 6: <i>Fitted Probabilities for rainfall of order two with four curves</i>	21
Graph 7: <i>Fitted Probabilities for rainfall of order two with three curves</i>	22
Graph 8: <i>Fitted mean rainfall amounts for order zero.</i>	24
Graph 9: <i>The Eight Different Probabilities of Rain across the Year</i>	29
Graph 10: <i>Probabilities of order zero (GenStat)</i>	53
Graph 9: <i>Overall chance of weekly rainfall of order zero (GenStat)</i>	54
Graph 10: <i>Fitted Probabilities of order zero (GenStat)</i>	57
Graph 11: <i>Probabilities of rain for order one (GenStat)</i>	59
Graph 12: <i>Probabilities of rain for order summarized on weekly basis</i>	60
Graph 13: <i>Fitted Probabilities of order one (GenStat)</i>	60
Graph 14: <i>Fitted Probabilities of rain for order two (Normal Markov)</i>	62
Graph 15: <i>Fitted Probabilities of order two-High Order (GenStat)</i>	65
Graph 16: <i>Fitted Probabilities of normal Markov chain of order three</i>	66
Graph 17: <i>Fitted Probabilities of high Markov chain of order three.</i>	67
Graph 18: <i>Fitted Probabilities of high Markov chain of order four</i>	68
Graph 19: <i>Fitted Probabilities of high Markov chain of order five.</i>	70
Graph 20: <i>Fitted Mean rainfall of order zero (GenStat)</i>	74
Graph 21: <i>Fitted Mean rainfall of order one with harmonic of order 2</i>	75
Graph 22: <i>Fitted Mean rainfall of order one with harmonic of order 3</i>	76
Graph 23: <i>Fitted Mean rainfall of High Markov of order two</i>	78
Graph 24: <i>Fitted Mean rainfall of order zero (GenStat)</i>	79

List of Appendixes

Appendix I: Rainfall data for Katumani imported directly from Instat	84
Appendix II: Summary Statistics for Katumani Rainfall data.	85
Appendix III: COUNT Procedure	86
Appendix IV: PREPARE Procedure.....	90
Appendix V: FITTING Probability Procedure.....	93
Appendix VI: AMOUNTFITTING Procedure.....	98
Appendix VII: Commands for Menus and dialogue (.grc files).....	101
Appendix VIII: Commands for Menus and dialogue (.gad files).....	103

Chapter 1: Introduction

1.1: Background information

Recent developments in statistical software have simplified the entire process of data manipulation and analysis. This permits the analysis of large and complicated data sets. Climatic data has been routinely collected in many locations for many years, but a lot of this data remains unanalyzed. In the past, climatic analysis was often left to specialists due to complexities of the data, such as the strong within and between year variability. The advances in the software bring routine analysis of climatic data within reach of non-specialists. This is important due to the current worldwide interest in climate particularly climate change.

Lack of climatic data analysis has made it difficult to understand the difference between climate change and natural climatic variability. Statistical analysis is therefore important as it aims at “identifying the systematic behavior in the data set and also searching for periodic variation, a quasi-periodic variation, a trend, persistence or extreme events in the climate element under analysis” (Buchdahl, 1999). This will enhance the understanding of whether there is a climate change or climatic variability within a given time frame.

Although a number of statistical packages have the capability to analyze complex statistical problems, most of them do not have specialized routines for analyzing climatic data. Instat was introduced in the early 1980s as a simple statistics package to help in the teaching of statistics. It was later improved by adding more components with particular interest for processing climatic data (Stern, Rijks, Dale, & Knock, 2006). It is the only available package specialized for analyzing climatic data.

The analysis of climatic data is easier and user friendly in Instat due to its additional Climatic Menu. Through this menu, most of the climatic analysis can be done e.g. summaries, analysis of events and Markov modeling of rainfall data, etc. Though Instat has the ability to analyze climatic data, it does not include the facilities, such as generalized linear models (GLMs), needed for a full analysis of climatic data. GenStat on the other hand is a powerful package which can handle complex GLM. It is a general statistics package that offers a wide range of high-quality statistical techniques and graphics.

GenStat has a flexible working environment where data can be analyzed using a user-friendly menu-based interface and a powerful command language from which you can create your own command procedures to perform the analyses. The command language permits the user to write his/her own programs that can cover situations where the standard analysis does not give them what they require. “So any GenStat analysis can be used in the construction of a new technique.” (Payne, et al., 2009). Despite the command language, most users are still reluctant in using it.

It is possible to analyze climatic data using GenStat even though it does not have the climate menu as in Instat (Gallagher & Stern, *Analyzing Climatic Data Using Genstat for Windows*, 2009). Currently, the climatic data analysis using GenStat is currently only possible if you know and understand the direction to take for the solution. This requires more understanding for users who might want to use the package to analyze their climatic data (Gallagher & Stern, 2009). The existing GenStat commands and procedures can be used to construct other procedures, to simplify their use and make them easily available to most users; in addition, the procedures can be used to form their corresponding menu and dialogues. A climatic menu can therefore be added to GenStat package to make it easier for many users to do their work without having to master the command language.

This study starts the implementation of a climatic menu into the GenStat package which will be able to handle high order Markov chains to be used in modeling rainfall. Chapter two of this report discusses the analysis of rainfall data using Instat and Marksim with particular interest on Markov modeling. The extensive use of the GenStat command language and its dialogue/menu interface is discussed in chapter three. Chapter four describes how rainfall data is prepared for fitting in Instat and how the same is implemented in GenStat using the existing and the newly developed procedures. Markov modeling of climatic data and how it is implemented in GenStat is then discussed in chapter five. Finally chapter six discusses how the GenStat procedures are used in a real and large climatic dataset.

1.2: Basic Concept

1.2.1: The Generalized Linear Model

The Generalized Linear Model was introduced by (Nelder & Wedderburn, 1972). It is an extension of the General Linear Model to include response variables that follow any probability distribution in the exponential family e.g. Binomial, Poisson, Multinomial, Gamma, Negative Binomial etc. It may be used where the response variable neither follows a Normal distribution nor have homogenous variances (Payne, et al., 2009). Comparing GLM and Multiple regression models (a form of general linear model) makes its features seen more clearly (Stern & Coe, 1984). The expression below can be used to define General linear models:

$$y_i = \beta_{0i} + \sum_{j=1}^p \beta_{ji}x_{ji} + \epsilon_i \quad (i = 1, \dots, n \text{ and } E(\epsilon_i) = 0) \quad (1-1)$$

These set of n equations can be written in the form of a compact model as shown in the below.

$$Y = X\beta + \epsilon \quad (1-2)$$

where

- a) Y is a vector of response variable and is a standard linear model meeting the Gauss-Markov condition and can be expressed as shown in equation 1-3.

$$E(Y) = \theta = X\beta \quad (1-3)$$

- b) X is the matrix explanatory variable (Covariate),
- c) β is a vector of unknown parameters (where β are estimated by solving the least-square equations(1-4) and ϵ is a vector of unobservable of errors corresponding to the observation.

$$X'X\hat{\beta} = X'Y \quad (1-4)$$

The approach used by (Nelder & Wedderburn, 1972) was to describe any given model in terms of its link function and its variance function. The variance function describes the

relationship between the mean and the variance of the dependent variable to allow for a proper calculation of the variance under non-normal conditions while the link function describes the non-linear relationship between the mean of the dependent variable and the linear right hand side.

GLMs are used for regression modeling for non-normal data. Suppose we generalize equation (1-3) with a linear predictor based on the mean of the outcome variable, then the function $g(\mu)$ will be called the link function.

$$g(\mu) = \theta = \mathbf{X}\boldsymbol{\beta} \tag{1-5}$$

The link function can be inverted as shown in equation 1-6

$$\mu = g^{-1}(\mathbf{X}\boldsymbol{\beta}) \tag{1-6}$$

If y_i binomially distributed with mean μ then the link function is logit as derived by (Nelder & Wedderburn, 1972) and expressed as:

$$g(\mu) = \log \left[\frac{\mu}{1 - \mu} \right] = \mathbf{X}\boldsymbol{\beta} \tag{1-7}$$

Then the μ can be expressed as:

$$\mu = \frac{\exp(\theta)}{1 + \exp(\theta)} \tag{1-8}$$

1.2.2: Markov chain

A Markov chain is a time ordered probabilistic process consisting of a finite number of **states** and some known probabilities p_{ij} , where p_{ij} is the probability of moving from state i to state j (Khamsi, 1999). That is, the probability at some point of time τ being in a certain state is conditioned on the states of the previous time, where the number of previous periods is termed as the order of the chain. A Markov chain is useful for analyzing events whose likelihood depends on what happened last.

1.2.3: The Markov Chain of first order

In the first-order Markov chain, the current state is dependent solely on the state of the immediate previous period and the chance that a process is in state j at time τ given that it was in state i at time $\tau - 1$ is represented by transitional probability P_{ij} which is expressed as follows

$$P_{ij,\tau} = Pr(X_{\tau} = j | X_{\tau-1} = i) \quad (1-9)$$

P_{ij} can be estimated as shown below (Siriwardena, Sriathan, & McMahan, 2002):

$$P_{ij} = \frac{q_{ij}}{\sum_{j=1}^c q_{ij}}; \quad i, j = 0, 1, 2, 3, \dots, C \quad (1-10)$$

Where q_{ij} = historical frequency of transition from state i to state j and

C = the maximum number of states

1.2.4: High Order Markov Chain

A Markov chain of order λ is referred to as high order Markov chain if λ greater than 1. The probability that on time τ will have a particular state depends on the states of the previous time $\tau - \lambda$. For example the Markov chains of order 2 and 3 satisfy the conditions in equation (1-11) and (1-12) respectively.

$$\begin{aligned} P_{i_2, i_1, j, \tau} &= Pr(X_{\tau} = j | X_{\tau-1} = i_1, X_{\tau-2} = i_2, \dots, X_0 = i_0) \\ &= Pr(X_{\tau} = j | X_{\tau-1} = i_1, X_{\tau-2} = i_2) \end{aligned} \quad (1-11)$$

Where $P_{i_2, i_1, j, \tau}$ is the transition probability of state j in time τ , given state i_1 in time $\tau - 1$ and state i_2 in time $\tau - 2$

$$\begin{aligned} P_{i_3, i_2, i_1, j, \tau} &= Pr(X_{\tau} = j | X_{\tau-1} = i_1, X_{\tau-2} = i_2, \dots, X_0 = i_0) \\ &= Pr(X_{\tau} = j | X_{\tau-1} = i_1, X_{\tau-2} = i_2, X_{\tau-3} = i_3) \end{aligned} \quad (1-12)$$

Modeling a high order Markov chain leads to a high-dimensional space of parameters (Vardi & Ju, 1999). A higher order Markov chain say of order 2 with two states will have four parameters, order 3 will have eight parameters order 4 will have sixteen parameters and order λ will have 2^λ parameters. Increasing the number of states increases the number of parameter in each order. Such models may not be accurate in situation where there may be no sufficient data to estimate them. However, (Longhai & Radford, 2008) and (Vardi & Ju, 1999) suggest that the high dimension of parameters can be reduced in such situations.

1.2.5: Modeling Rainfall data using Markov Chain

Since (Gabriel & Neumann, 1961) proposed Markov chain for modeling the sequence of wet and dry days, it gained popularity in representing daily precipitation occurrence processes (Mimikou, 1983). This has been contributed by the flexibility of the Markov chain model and the ease of estimating the parameters and obtaining the fitted model without resorting to simulation (Stern & Coe, 1984).

The occurrence of a wet state in day τ is dependent of what state that occurred in the previous day's $\tau-1$, $\tau-2$, $\tau-3$ The random variable X_τ represents the occurrence or non-occurrence of rain on day τ such that

$$X_\tau = \begin{cases} 0 & \text{if day } \tau \text{ is dry} \\ 1 & \text{if day } \tau \text{ is wet} \end{cases} \quad (1-13)$$

1.2.6: Fitting a first order Markov Model to rainfall data

The first order model assumes that the probability of rain occurring on any day depends only on whether it did or did not rain on the previous day. To fit this model, the parameter for transition probability $p_{i,\tau}$ is estimated over the year (Stern & Coe, 1982). The $P_{i,\tau}$ is the probability of rain in day τ given state i (for $i = 0,1$) in day $\tau - 1$. The estimate of $p_{i,\tau}$ is given by $r_{i,\tau}$ (Stern & Coe, 1984) which is the proportion of years with state i in their day $\tau - 1$ that had rain in their day τ . The $r_{i,\tau}$ is expressed as shown in equation (1-14).

$$r_{i,\tau} = \frac{n_{i1,\tau}}{n_{i1,\tau} + n_{i0,\tau}} \quad (1-14)$$

where

$n_{i1,\tau}$ is the number of years with rain on day τ

$n_{i0,\tau}$ is the number of years with no rain on day τ

The random variable $n_{i1,\tau}$ is assumed to be binomially distributed with the probability of success being $p_{i,\tau}$ and $(n_{i1,\tau} + n_{i0,\tau})$ is the number of trials. Therefore the model used is

$$p_{i,\tau} = g(\theta_{i\tau}) \quad (1-15)$$

where $g()$ is a logit link function connecting the probabilities $p_{i,\tau}$ to the function $\theta_{i\tau}$ which is a linear unknown parameters (Stern & Coe, 1984). The model is a generalized linear model since binomial is a member of the exponential family (Nelder & Wedderburn, 1972) $p_{i,\tau}$ is therefore expressed as

$$p_{i,\tau} = \frac{\exp(\theta_{i\tau})}{1 + \exp(\theta_{i\tau})} \quad (1-16)$$

(Stern & Coe, 1984) suggested that Fourier analysis may be used to express $\theta_{i\tau}$ as shown below:

$$\theta_{i\tau} = a_{i0} + \sum_{k=1}^m [a_{ik} \sin(k\tau') + b_{ik} \cos(k\tau')] \quad (1-17)$$

where $\tau' = 2\pi\tau/366$ and m is the number of harmonics

1.3: Statement of the problem

Markov chain models have been in existence in the past to describe rainfall occurrence data (Stern & Coe, 1982). Several advances are made in improving the Markov model fitting to rainfall data for example; the use of higher order Markov chains (Stern & Coe, 1984) and finding a suitable Markov order (Avik, Deepanwita, & Dashgupta, 2010). However little has been done to make the modeling of rainfall data using Markov chains accessible to users in statistical packages.

A powerful statistical package with specialized routine for analyzing climatic data is needed to handle such models. The introduction of Instat package was meant to simplify the modeling of rainfall data and make it accessible to user though it cannot handle complex analyses like the Generalized Linear Models.

GenStat is another powerful package with the capability of climatic data analysis (Gallagher & Stern, 2009) but it doesn't have a climatic menu. Despite its potential and availability of a guide for climatic analyses, it is still not easy to use it for climatic analysis due to lack of a specialized climatic menu.

1.4: Objective of the study

Analysis in GenStat can be done through the command language if it cannot be achieved through dialogues and menus. It is possible to build the commands into a procedure that can perform a particular task. These procedures can then be used to develop special menus and dialogues. This study starts the implementation of an extension of the Instat climatic menu into the GenStat package through creating procedures, then developing them into menus. These are used to implement Markov chain modeling of rainfall data in Genstat.

1.5: Significance of the study

Simplifying climatic data analysis will enable more researchers to be able to do the analysis and hence include climatic information in their work. Non specialists will have an opportunity to perform simple climatic analysis on their data without necessarily consulting the data specialists. This is intended to encourage more researchers to utilize climatic data in their research. The Markov modeling analysis will incorporate the Generalized Linear Model that is readily available in GenStat.

Chapter 2: Literature review

The study on the sequence of daily rainfall occurrence was started by (Gabriel & Neumann, 1961). They assumed that the probability of rainfall on any day depended only on whether the previous day was wet or dry, that is whether rainfall did or did not occur. Given the event on the previous day, then, the probability of rainfall is assumed independent of events of further preceding days. They classified days as wet or dry according to whether there had or had not been recorded at least 0.1mm of precipitation in the 24-hour from 8.00 a.m. to 8.00 a.m. the following day. They found that the daily rainfall occurrence for the Tel Aviv data was successfully fitted with the first-order stationary Markov chain model. The model parameters were the two conditional probabilities:

$$p_1 = p(\text{Wet day/previousday wet}) \quad (2-1)$$

$$p_0 = p(\text{Wet day/previousday dry}) \quad (2-2)$$

Given the Markov chain and estimates of its two basic probabilities, they derived various properties of rainfall occurrence patterns. These included probabilities of rainfall i days after a wet or dry day. A wet spell of length k was defined as a sequence of k wet days preceded and followed by dry days. Dry spells were defined correspondingly. Weather cycles were defined as combinations of a wet spell and an adjacent dry spell. The distribution of spell lengths was found to be geometric. They then derived probabilities of a wet spell of length k and dry spell of length m , probability of a weather cycle of n days and the probability of exactly s wet days among n following a wet day. For large n , the distribution of the number of wet days tends to normality. The fit of the model was tested on data of daily rainfall in Tel Aviv (Gabriel & Neumann, 1961) for the 27 rainy seasons 1923/24 – 1949/50. Chi-square tests of goodness of fit showed no significant deviations from the actual observations on rainfall so the model was said to hold.

Rainfall usually occurs in seasons of various lengths. They did not have any further coefficients of periodicity, persistence or any other parameters. The model was not suggested in terms of actual rainfall occurrence but as a statistical description of the observations. In

their paper (Kottegoda, Natale, & Raiteri, 2004), they reported that the first order of the Markov chain model was found to fit the observed data in Italy successfully. This model was based on the assumption that there was a dependency of the daily rainfall occurrence to that of the previous day.

In the analysis of Samaru, Nigeria rainfall (Stern R. D., 1979), he modeled the occurrence of rainfall as a first order Markov chain. The probability of rainfall was not constant within each period of interest instead the probability of rainfall is a function of the time of the year. At the beginning and end of the season, the probability of rain is greater, if there was rain on the previous day, than if the previous day was dry. In the middle of the season, the probability of rain appears to be approximately the same irrespective of the state of the previous day. A linear logistic model was fitted to the data. However, to estimate the transition probabilities in the Markov chain model, lots of data is required hence statistical software to ease the analysis.

A climatic Guide (Stern, Rijks, Dale, & Knock, 2006) described the use of Instat Software to analyze climatic data. The guide explained step by step how to analyze most of the element of climate. This simplifies the analysis of climatic data due its existing dialogue and menus specialized for analyzing climatic data. However there are instances when it is not sufficient. For example, situation where higher order Markov modeling (of more than order 2) is required, Instat cannot achieve. Alternative software to Instat is Marksim which is based on a stochastic weather generator and uses a third-order Markov process to model daily weather data (Jones & Thornton, 1999)

The widely used stochastic model in rainfall modeling is the GLM (Chandler & Wheeler, 2002) due to the stochastic nature of rainfall (Mimikou, 1983) this has not been achieved in Instat. GenStat software is a more powerful tool because it can be used to perform the more complex analysis of data especially the GLM. GenStat climatic Guide (Gallagher & Stern, 2004) describes how GenStat software can be used to analyze climatic data.

Chapter 3: Fitting Markov Chain Models using Instat and MarkSim

3.1: Introduction

Markov models are fitted for rainfall data collected from Katumani Experimental Research Station in Kenya (longitude-1°35'S, Longitude-37°14'E, Altitude-1600mm) for the years 1961-2001 (41 years). The data is already available in the Instat library, and is used to fit a zero, first and second order Markov models using Instat package. The same data will be simulated using MarkSim software which also fits a third order Markov model to the data. The two approaches (Instat and MarkSim) are used to establish the appropriate order of the Markov model of rainfall data for Katumani Experimental Research Station.

3.2: Instat.

Instat package has a specialized menu for Markov modeling of rainfall data. Using the Climatic Menu in Instat, Markov modeling sub menu involves three stages. These are; data preparation, fitting the model and using the model. Figure 1 below shows the “Climatic => Markov Modeling” Menu in Instat.

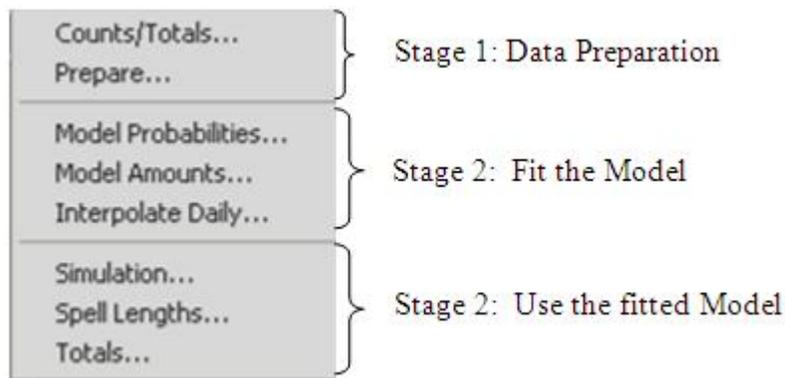


Figure 1: *Markov Modeling Menu*

3.2.1: Data preparation

The data preparation stage summarizes the rainfall data into counts and totals for the model fitting. The Menu “Climatic => Markov Modeling => Counts/Totals” (Figure 1) is used. It is possible to model a zero, first and second order Markov chains for rainfall occurrences in Instat. When the zero option is selected in the ‘Order for Count’ section, it

calculates the total number of wet and dry days for each day of the year (two variables only; *dry* and *wet*), first order calculates the number of days with a specific state given the state in the previous day (four variables i.e. 'ww', 'wd', 'dw' and 'dd') while second order calculates number of days with a specific state given the state in the two previous days (eight variables i.e. 'www', 'wwd', 'wdw', 'dww', 'ddw', 'dwd', 'wdd', 'ddd'). Figure 2 is a dialog box in Instat that is used to generate the counts/totals. (We illustrate these stages with a zero order Markov model)

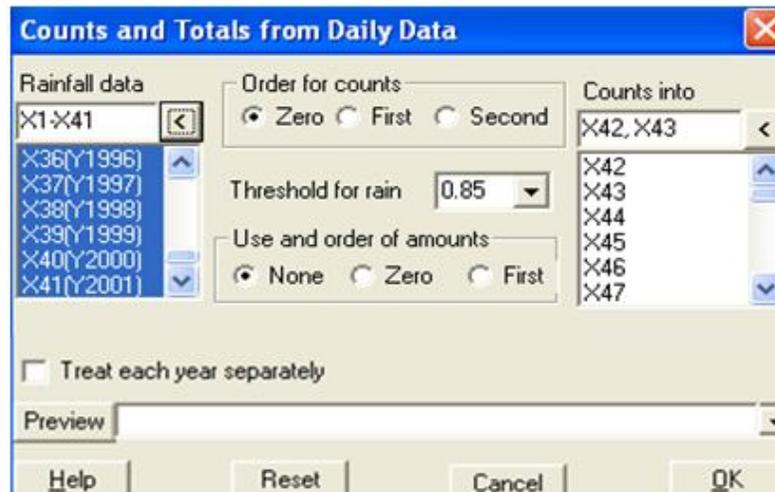


Figure 2: Counts/Totals Dialogue box

The result for count is displayed in the spreadsheet (Figure 3) for order zero. Remember the result did not display the amount of rainfall since the radio button *none* for amount was selected as shown in Figure 2.

Current Worksheet - Katumani.wor				
Columns	Constants	Strings	Labels	Title: K
	X42	X43	X44	X
	dry	rain		
1	35	6		
2	33	8		
3	31	10		
4	34	7		
5	33	8		
6	29	12		

Figure 3: Counts of rainy days of order zero.

The column X42 is renamed as '**dry**' and X43 as '**rain**'. They contain the number of occasions, in the 41 years, that each day is dry or rainy. For example, 8 years out of 41years

had rainfall on day 5 (5th January), while, only 6 out of 41 years received rainfall on the 1st January. From the counts, the chance of rainfall on a particular day can be calculated, for example using the result in figure 3, the chance of rain on 1st January. is $\frac{6}{41}$ on 2nd January. $\frac{8}{40}$, on 3rd January $\frac{10}{41}$ etc. These probabilities are calculated using the Menu “*Climate => Markov Modeling => Prepare*”. The dialog (Figure 4) has the option of calculating the summary of rainfall counts and amounts on a daily basis or otherwise.



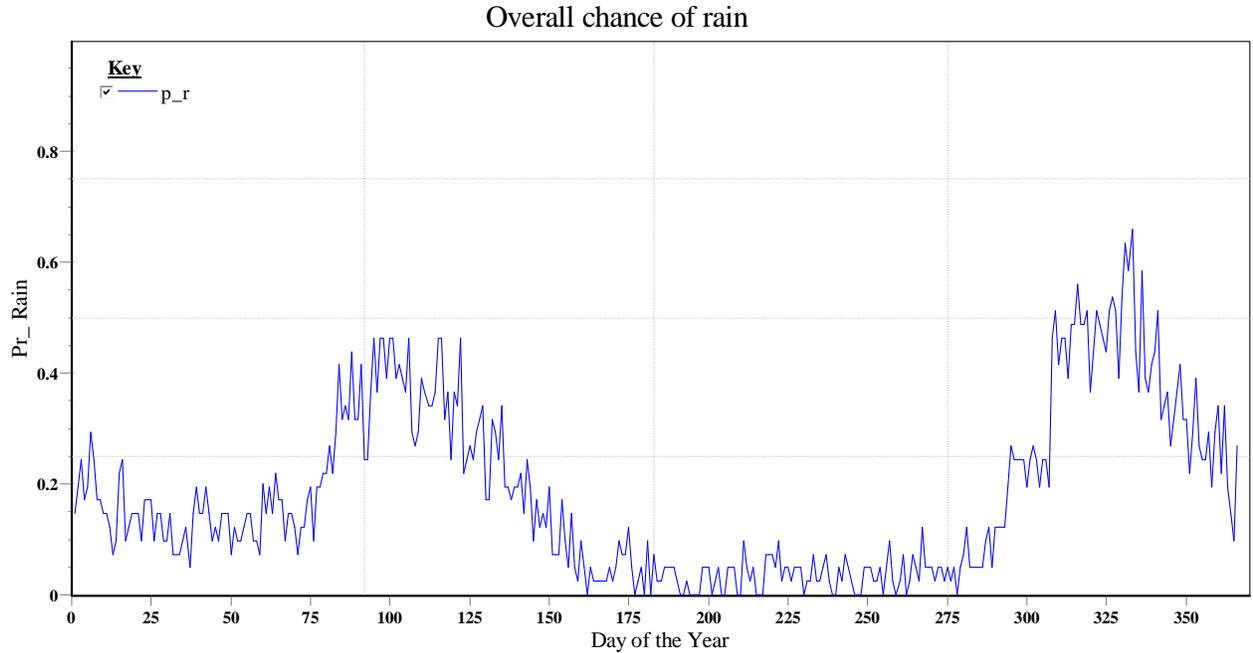
Figure 4: Prepare Dialogue Box

The Prepare dialogue box sets up the columns needed to fit the models (Figure 5) and also produces a graph of the chance of rain (Graph 1).

Columns	Constants	Strings	Labels	Title: Katumani Experimental Research Station - Daily Rainfall (mm) 1961-2001 Lat Lon									
	X45	X46	X47	X48	X49	X50	X51	X52	X53	X54	X55	X56	
	p_r	circ	ds1	dc1	ds2	dc2	ds3	dc3	ds4	dc4	ds5	dc5	
1	0.1463415	-3.124425	716633E-02	-0.9998527	0.0343276	0.9994106	147876E-02	-0.9986741	861474E-02	0.9976432	-0.0857305	-0.9963183	
2	0.1951219	-3.107258	-0.0343276	-0.9994106	861474E-02	0.9976432	-0.102821	-0.9946999	0.1369061	0.990584	-0.1708297	-0.9853005	
3	0.2439024	-3.090091	147876E-02	-0.9986741	0.102821	0.9946999	-0.1538906	-0.9880879	0.2045521	0.9788557	-0.2546712	-0.9670277	
4	0.1707317	-3.072924	861474E-02	-0.9976432	0.1369061	0.990584	-0.2045521	-0.9788557	0.2712339	0.9625134	-0.3366373	-0.9416344	
5	0.1951219	-3.055757	-0.0857305	-0.9963183	0.1708297	0.9853005	-0.2546711	-0.9670277	0.3366373	0.9416344	-0.4161247	-0.9093075	
6	0.2926829	-3.03859	-0.102821	-0.9946999	0.2045521	0.9788557	-0.3041148	-0.9526354	0.4004539	0.9163169	-0.492548	-0.8702853	
7	0.2439024	-3.021422	-0.1198812	-0.9927883	0.2380333	0.971257	-0.3527521	-0.9357168	0.462383	0.8866803	-0.5653447	-0.8248548	
8	0.1707317	-3.004255	-0.1369061	-0.990584	0.2712339	0.9625134	-0.4004539	-0.9163169	0.5221326	0.8528643	-0.6339785	-0.7733507	
9	0.1707317	-2.987088	-0.1538906	-0.9880879	0.3041148	0.9526354	-0.4470938	-0.8944871	0.5794211	0.8150284	-0.6979442	-0.7161522	

Figure 5: Probability of rainy day or order zero

The columns set in Figure 5 above X45 shows the probability of rain while X46-X56 shows the finals that are used for fitting the curve.



Graph 1: Overall Chance of rain

3.2.2: Fitting the model

The second stage of Markov modeling is fitting the Markov model to the rainfall data. It uses the model probabilities dialogue which is obtained through, “*Markov Modeling => Model Probabilities*” menu which opens fitting the probabilities dialogue box (Figure 6) This produces the fitted probabilities, which are given in a column called 'f_r' (Figure 7) and a curve of the fitted probabilities, Graph 2

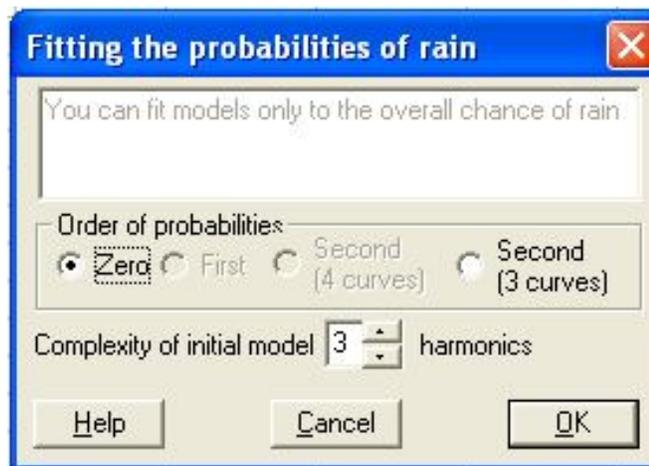
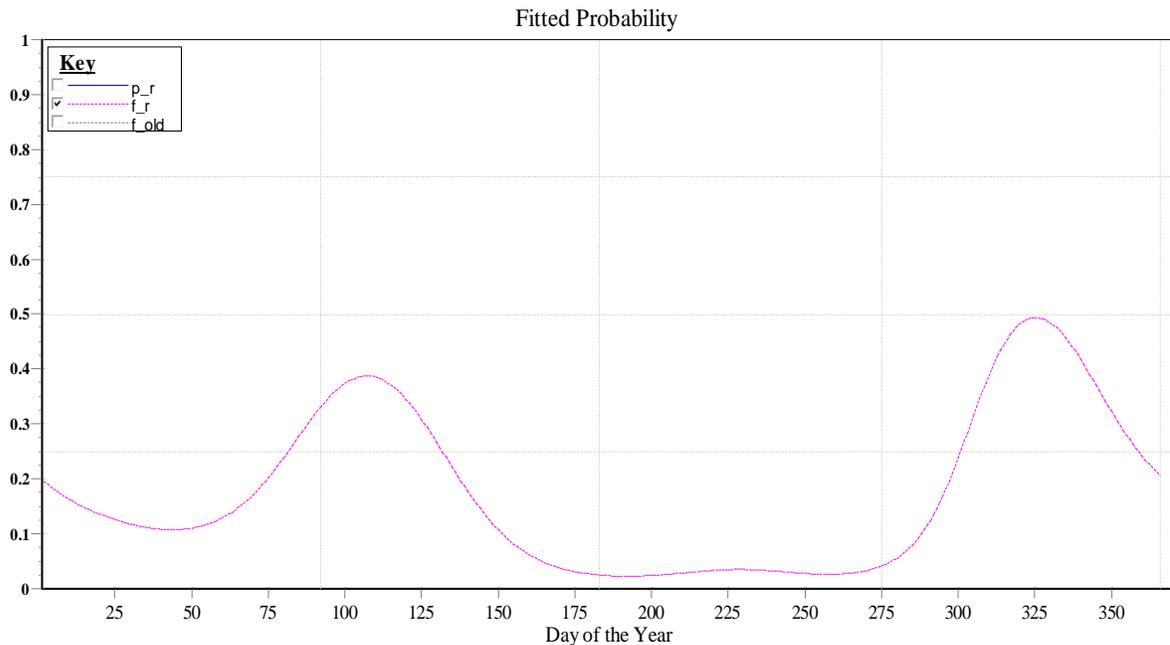


Figure 6: Fitting Probability Dialogue box

sheet - KATUMANI workin		
Strings	Labels	Title:
X58	X59	
f_r	estr	
0.2406601	-1.973053	
0.2314891	-0.5935322	
0.2225461	-1.020109	
0.2138521	-0.9957734	
0.2054253	-0.2795608	
0.1972812	0.5144213	
0.1894324	-0.1340795	
0.1818888		
0.174658		

Figure 7: Fitted Probabilities of order zero.



Graph 2: Fitted Probability curve

3.2.3: First Order Markov Model for rainfall occurrences

Instat have the ability to fit Markov chains for counts of up to order two to rainfall data. A first order Markov chain is modeled by using the dialogue box for Count/Totals is shown below (Figure 8).

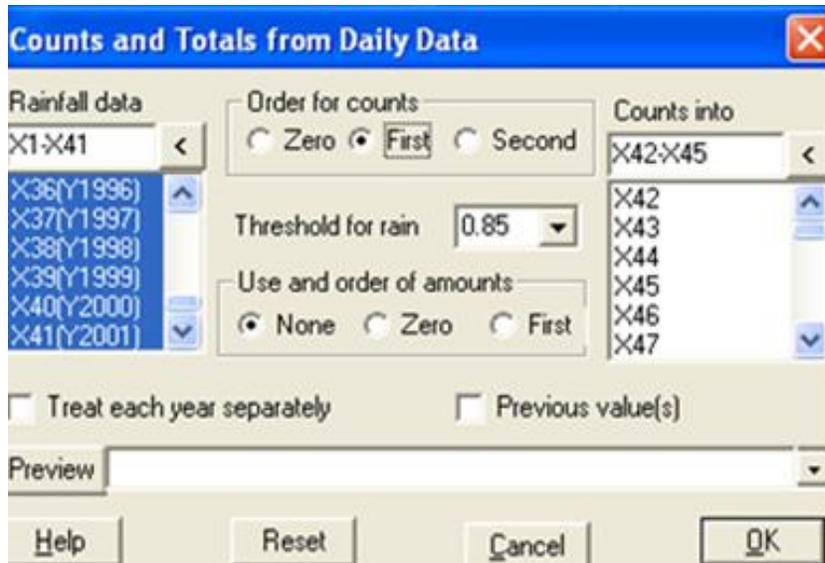


Figure 8: Count/Totals dialogue box for order one.

The results for the first order is displayed in figure 9. Four variable are calculated these include 'dd', 'dr', 'rd', 'rr'.

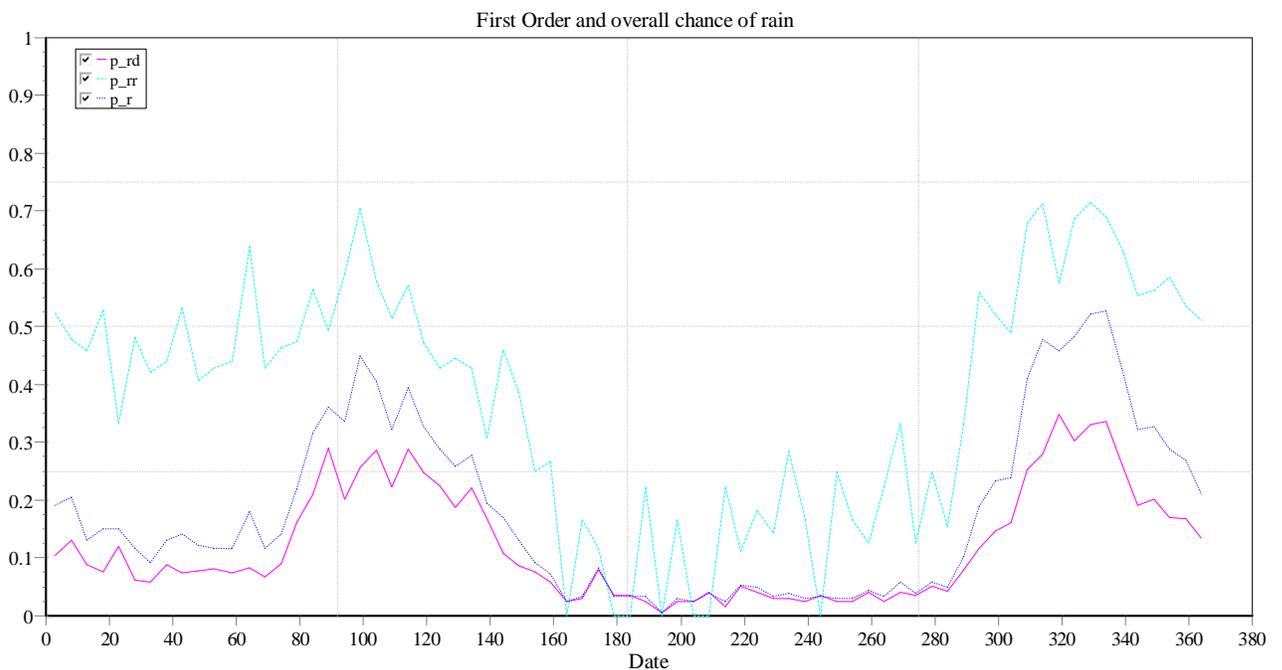
Current Worksheet - Katumani.wor					
Columns	Constants	Strings	Labels	Title: Katumani Experimental F	
	X42	X43	X44	X45	X46
	dd	dr	rd	rr	
1	27	7	2	4	
2	29	4	6	2	
3	29	2	4	6	
4	29	5	2	5	
5	31	2	3	5	
6	26	3	7	5	
7	25	6	4	6	
8	29	5	2	5	
9	29	5	5	2	

Figure 9: Counts of rainy days of order one.

The probability of rainfall are computed using the prepare dialogue and the results obtained are give in columns (Figure 10) and a plot is produced (Graph 5)

	X46	X47	X48	X49	X50	X51	X52	X53	X54
	Date	_d	p_rd	_r	p_rr	dry	rain	tot	p_r
1	3	162	0.1049383	42	0.5238096	165	39	204	0.1911765
2	8	161	0.1304348	44	0.4772727	163	42	205	0.204878
3	13	181	839779E-02	24	0.4583333	178	27	205	0.1317073
4	18	171	602339E-02	34	0.5294118	174	31	205	0.1512195
5	23	175	0.12	30	0.3333333	174	31	205	0.1512195
6	28	178	179775E-02	27	0.4814815	181	24	205	0.1170732
7	33	186	913978E-02	19	0.4210526	186	19	205	268293E-02
8	38	180	888889E-02	25	0.44	178	27	205	0.1317073
9	43	175	428572E-02	30	0.5333334	176	29	205	0.1414634
10	48	178	865169E-02	27	0.4074074	180	25	205	0.1219512
11	53	184	152174E-02	21	0.4285714	181	24	205	0.1170732

Figure 10: Probability of rainy days of order one



Graph 3: Probability curves for order two

The 'Fitting the probability' dialog box is used as shown in Figure 11 is used to fit the model.

It corresponding results is shown in the

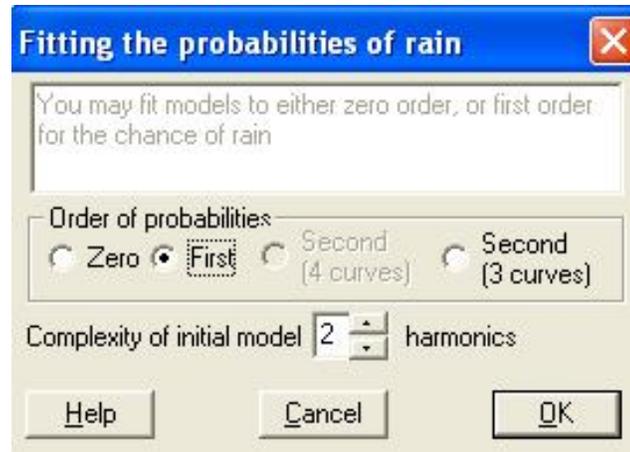
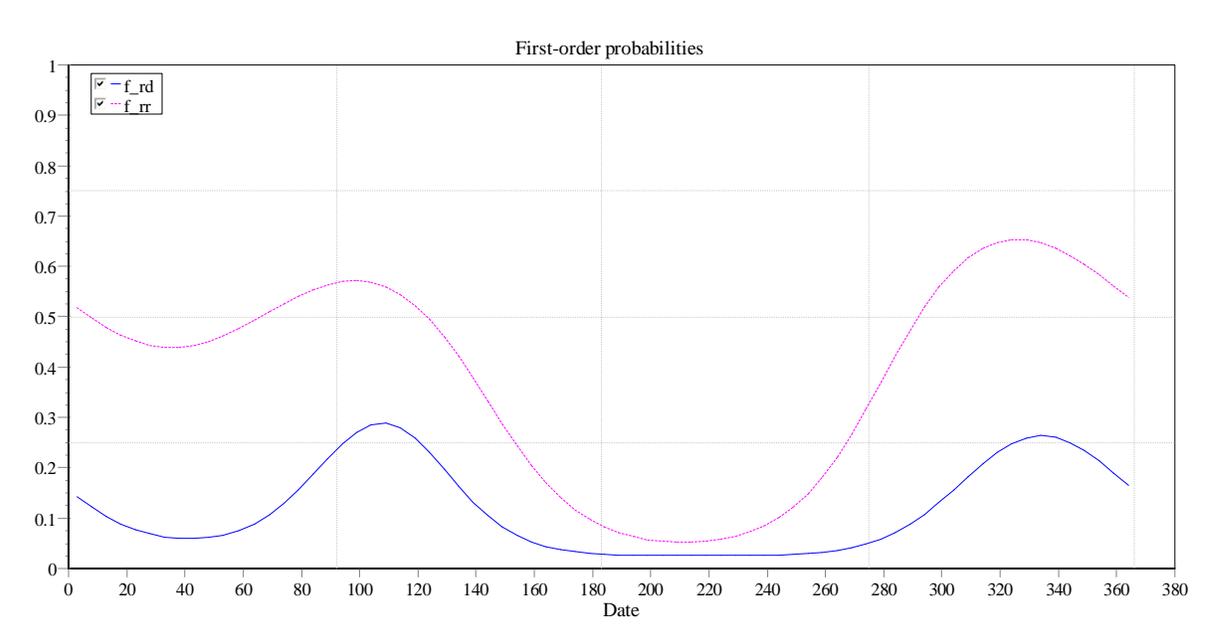


Figure 11: *Fitting probabilities dialog box for order one*



Graph 4: *Fitted Probabilities for rainfall of order one*

3.2.4: Second Order Markov Model for rainfall occurrences

To prepare data for second order Markov chain model, then ‘second-Order of Count’ radio button is selected from the ‘Count and Total from Daily Data’ dialog box. The results for second order gives 8 columns for counts namely ‘ddd’, ‘ddr’, ‘drd’, ‘drr’, ‘rdd’, ‘rdr’, ‘rrd’, ‘rrr’ (Figure 12). It is pointed out that out of 41 years, 26 years had a dry day on 3rd January given dry on 2nd January given dry on 1st January being (red circle on Figure 12). On

the other hand, 2 out of 41 years had a rainy day on 7th January given dry on 6th January given rain on 5th January (green circle on Figure 12).

	X42	X43	X44	X45	X46	X47	X48	X49
	ddd	ddr	drd	drr	rdd	rdr	rrd	rrr
1	28	0	6	1	2	0	1	3
2	23	6	1	3	5	1	1	1
3	26	3	1	1	3	1	5	1
4	27	2	2	3	2	0	2	3
5	27	4	1	1	2	1	1	4
6	24	2	1	2	7	0	2	3
7	24	1	4	2	2	2	3	3
8	24	5	2	3	1	1	2	3
9	27	2	2	3	2	3	0	2
10	27	4	4	0	2	1	1	2
11	29	4	1	1	2	0	2	2

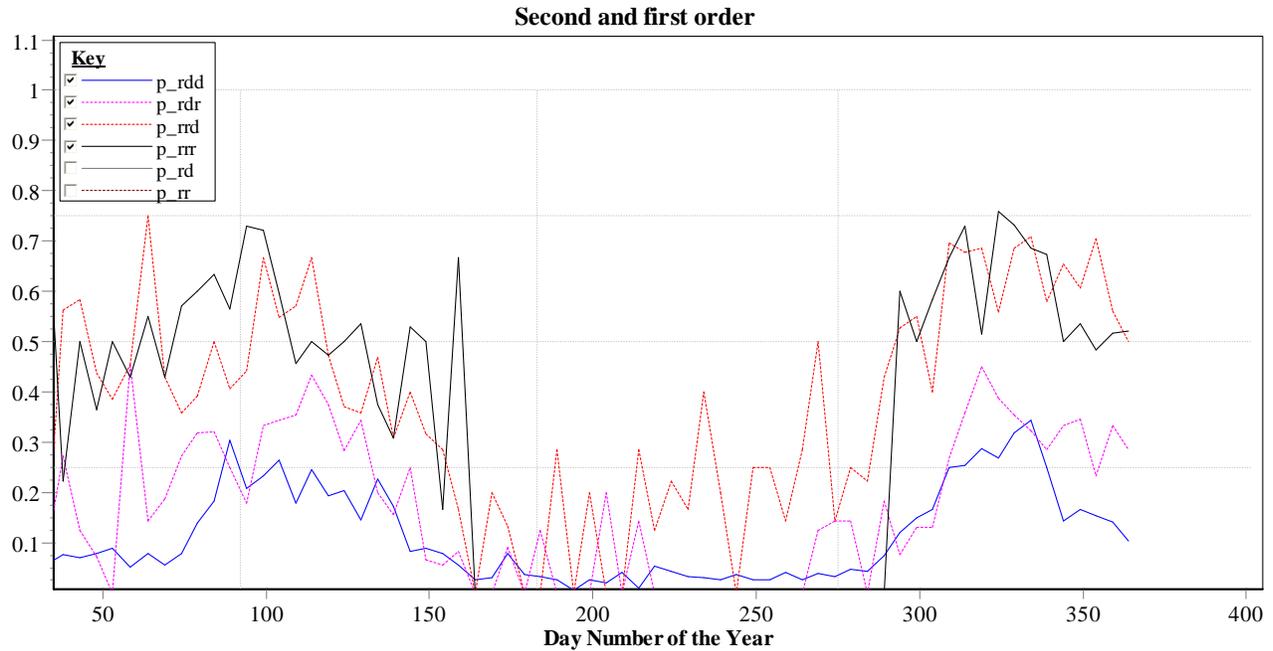
Figure 12: Counts of rainy days of order two.

The probability of rainfall are computed using the prepare dialogue and the results obtained are give in columns ‘p_rdd’, ‘p_rdr’, ‘p_rrd’, ‘P_rrr’ (Figure 13) and a plot produced (Graph 5)

	X51	X52	X53	X54	X55	X56	X57	X58	X59
	_dd	p_rdd	_dr	p_rdr	_rd	p_rrd	_rr	p_rrr	dd
1	29	896552E-02	0 *		7	0.1428571	4	0.75	27
2	27	0.1851852	7	0.1428571	2	0.5	4	0.25	28
3	29	0.1034483	4	0.25	6	0.8333333	2	0.5	29
4	29	896552E-02	2	0	4	0.5	6	0.5	29
5	29	896552E-02	5	0.2	2	0.5	5	0.8	31
6	31	0.2258064	2	0	3	0.6666667	5	0.6	26
7	26	692308E-02	3	0.6666667	7	0.4285714	5	0.6	25
8	25	0.04	6	0.1666667	4	0.5	6	0.5	29
9	29	896552E-02	5	0.6	2	0	5	0.4	29
10	29	896552E-02	5	0.2	5	0.2	2	1	31

Figure 13: Probability of rainy days of order two

The results for 2nd order Markov chain plots two sets of curves; the 1st order and 2nd order probability curves. The curves in Graph 5 do not have clear distinctions and therefore do not present the probabilities well.



Graph 5: Probability curves for order two

The fitted model of order two is now obtained by using ‘Fitting Probability dialog box shown in Figure 14. The order of probabilities option to be used at this stage is ‘Second (4 curves)’ if you want all the curves for rain given the conditions of the two previous days (‘rrr’, ‘rrd’, ‘rdr’ and ‘rdd’).

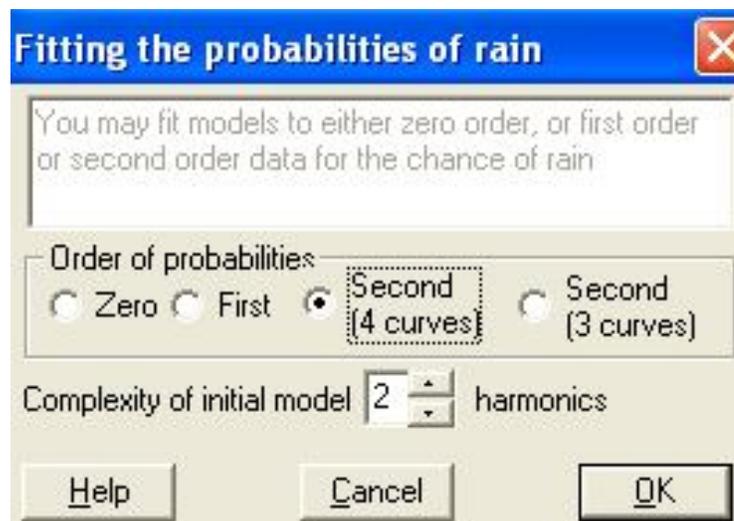
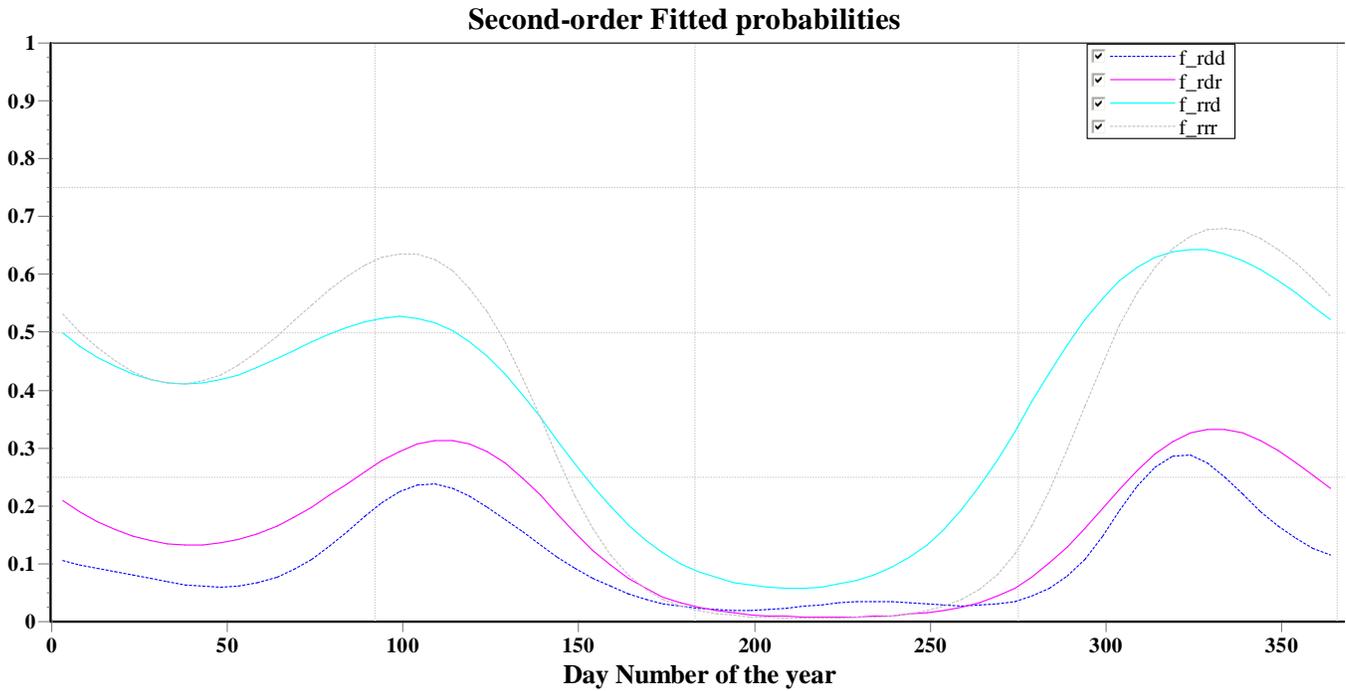


Figure 14: Fitting probabilities dialogue box for order two with four curves

Graph 6 shows the fitted probabilities of order two. There is no clear distinction between f_{rrr} and f_{rrd} curves and therefore there is no need to have both of them.



Graph 6: Fitted Probabilities for rainfall of order two with four curves

It is important to note that Instat in reducing the number of the fitted parameters for higher order Markov chain when the option of ‘second order with three curves’ is chosen in the Fitting Probability Dialogue box as shown in Figure 15. The fitted probabilities are shown in Graph 7.

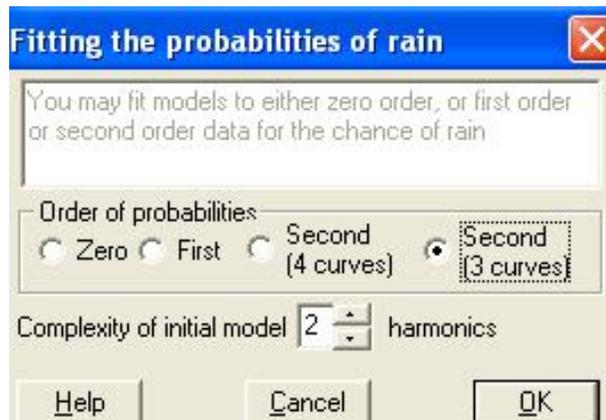
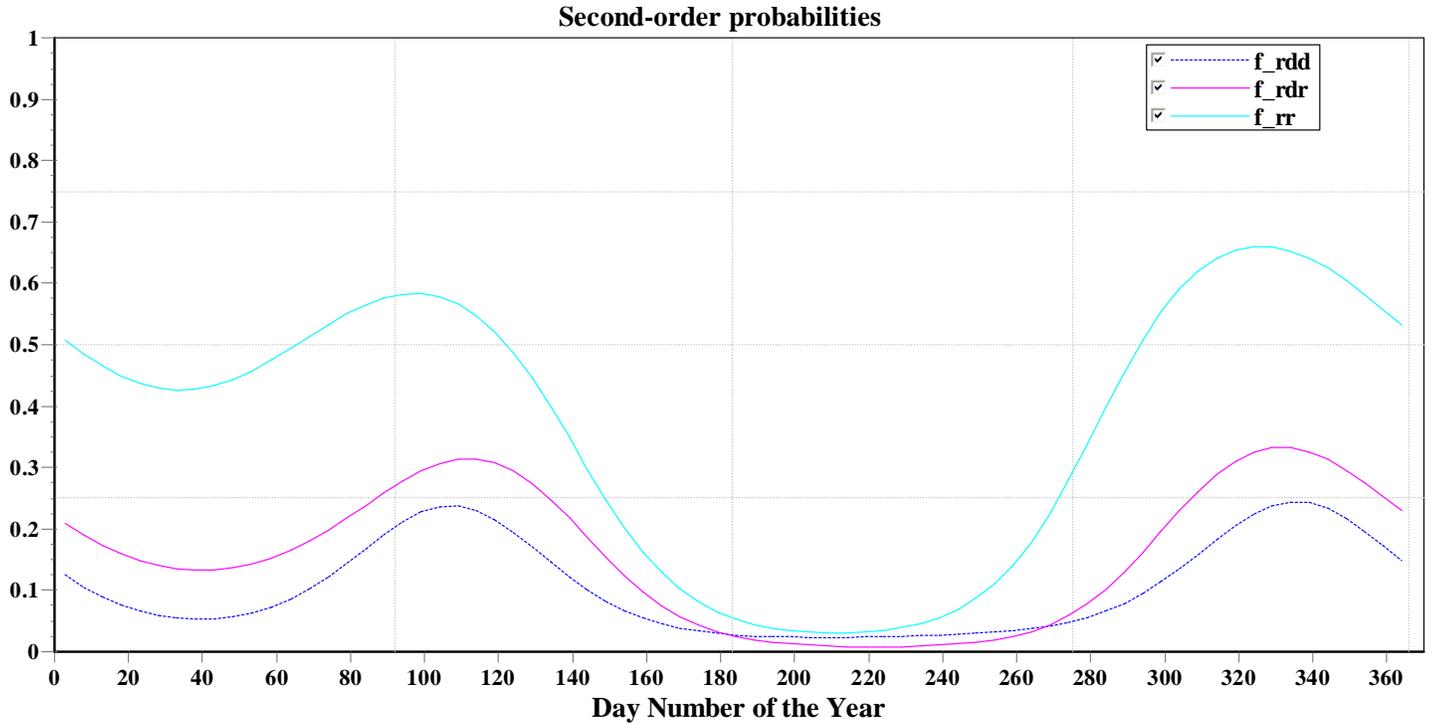


Figure 15: Fitting probabilities dialogue box for order two with three curves



Graph 7: *Fitted Probabilities for rainfall of order two with three curves*

It is important to note that when the number of parameters is reduced to three from four, then f_{rrd} and f_{rrr} are renamed to f_{rr} which is a first order. The current version of Instat the Markov command is limited to second-order Markov chains, however rainfall data for some sites are better modeled using higher order Markov chains.

3.2.5: Modeling rainfall amounts.

Instat can also model rainfall amounts on rainy days. Modeling the amounts involves the three stages as discussed in Section 3.2: Climatic => Markov Modeling => Counts/Totals menu is used. In the Count and Total dialogue box we choose zero order for amount as shown in

Figure 16 (circle in red).

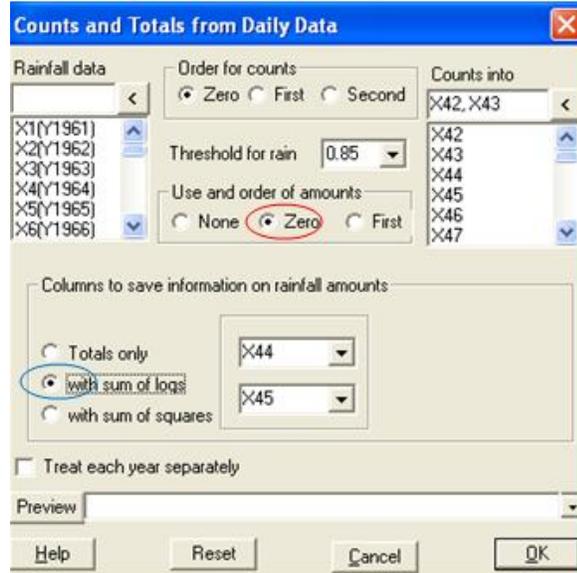


Figure 16: Count and totals dialogue box for rainfall amount of order zero

The result after clicking OK button will generate results with three columns; dry, wet, tr and lr (Figure 17). The 'tr' column indicate the total rainfall amount for the rainy days minus the total threshold values while the 'lr' column is the natural logarithm of the tr. The two columns are computed using the equations (3-1 and (3-2

$$tr_{\tau} = \sum_{y=1}^n (A_{\tau} - T)_y \tag{3-1}$$

$$lr_{\tau} = \ln(tr_{\tau}) \tag{3-2}$$

Where A_{τ} is the rainfall amount on day τ and T is the threshold value and y is the year number.

	dry	rain	tr	lr
1	35	6	22.2	3.5530
2	33	8	116.7	19.3970
3	31	10	101	16.1049
4	34	7	78.85	13.0649
5	33	8	87.2	13.6700
6	29	12	101	20.4840
7	31	10	73.7	11.7355

Figure 17: Total rain amounts of order zero.

The dry and wet columns in the figure above are given by default as the number of rainy days. The procedure that was followed when modeling the chance if rain is maintained to when modeling amounts. When fitting the model for amount, Climatic => Markov Modeling => Model Amounts menu is used and it fits the mean rain per rainy day.

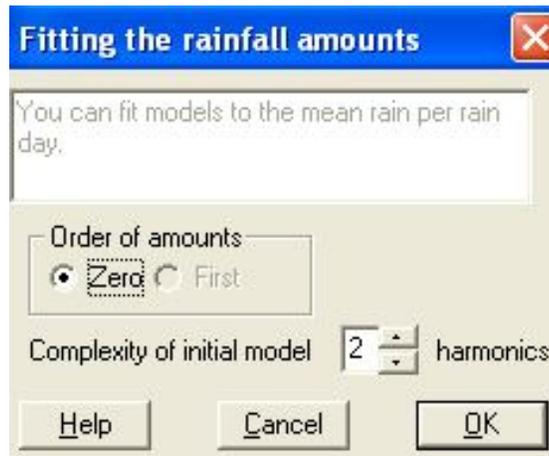
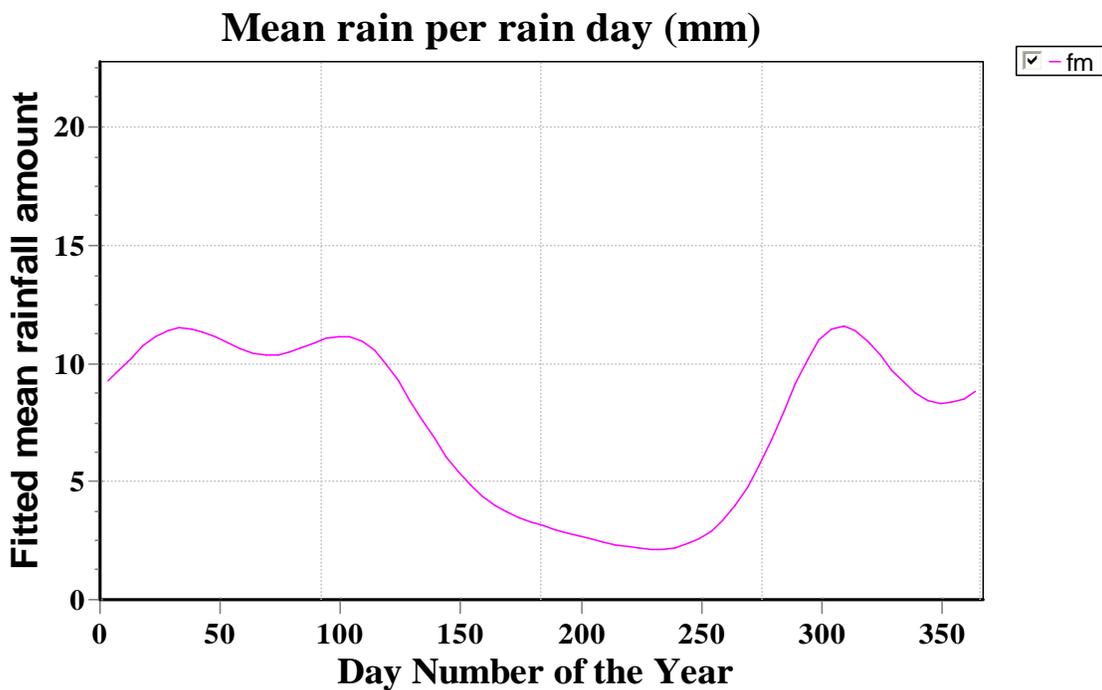


Figure 18: Fitting rainfall amount for order zero



Graph 8: Fitted mean rainfall amounts for order zero.

The fitted curve is as shown the Graph 8. The current version of Instat can only fit Markov model up to order 1.

3.3: MarkSim

3.3.1: Introduction

MarkSim is a computer tool that generates simulated data for crop modeling and risk assessment (Jones and Thornton 1999). Rainfall in the MarkSim model is modeled based on the third orders Markov chain in two stages. The first stage is transitional probability to determine whether any particular day is wet. Since it is a third order Markov chain, this wet day will depend on whether there was any rainfall in the previous three days. The second stage of the MarkSim model determines the amount of rainfall. When using MarkSim, a parameter file is created and then used to simulate daily climatic data.

MarkSim defines the probability of day i being wet as

$$P(W_i/D_1D_2D_3) = \phi^{-1}(b_1 + a_{i-1}d_1 + a_{i-2}d_2 + a_{i-3}d_3) \quad (3-3)$$

where:

ϕ^{-1} is the inverse normal probability function (probit)

b_i is the monthly baseline probit of a wet day following three consecutive dry days

a_m are binary coefficients of rain (1) or no rain (0) on a day

d_m are lag constants

When estimating the mean and the shape parameters of the distribution for each month, the method of maximum likelihood is used. In the process of generating rainfall records, the monthly baseline probabilities are interpolated to daily probabilities by using 12 point Fourier transform.

3.3.2: The Parameter file

The information that is required to create a parameter file is the longitude, latitude and elevation of the site of interest. The longitude (1.58333 degrees), latitude (37.2333 degrees) and elevation (1600 meters) for Katumani when entered, MarkSim will now create a parameter file for the simulated data.

Figure 19 shows the parameter file created from entering the longitude, latitude and elevation of Katumani into MarkSim.

```

Editing [C:\Program Files\CIAT\MarkSim\clxgen.log]
Files Edit

1
1
Row= 214 Col= 332 Record= 89792 Cluster= 139
KATUMANI Interpolated 1.583 37.233 1600

MONTH AV P BETA RAINDAYS S.E.
1 4.6 0.669 -1.279 0.141 0.32593
2 5.4 0.532 -1.111 0.201 0.30257
3 8.1 0.571 -1.018 0.253 0.29538
4 12.3 0.419 -0.672 0.431 0.27587
5 6.3 0.496 -0.955 0.291 0.29761
6 4.4 0.654 -1.201 0.168 0.32214
7 5.0 0.561 -1.031 0.238 0.31086
8 4.5 0.615 -1.154 0.188 0.32291
9 4.5 0.637 -1.138 0.191 0.31696
10 6.5 0.577 -1.014 0.255 0.30324
11 9.6 0.434 -0.695 0.418 0.28395
12 5.5 0.547 -0.987 0.274 0.30664
DI-3 0.7610 0.1980 0.2150 N= 2 Cluster 139 Phase 0.877
rain 21. 31. 67. 142. 64. 22. 36. 26. 25. 55. 114. 50.
temp 18.3 19.0 19.7 19.5 19.0 18.1 17.5 17.4 17.8 18.6 18.3 17.9
rang 15.0 15.9 13.7 11.7 11.9 12.8 12.2 12.1 13.8 13.5 12.1 12.4
radn 20.8 20.5 20.1 19.8 20.5 19.9 18.1 17.7 19.0 20.2 19.7 19.9
Output CLX file written: KATUMANI

```

Figure 19: Parameter File for Katumani

3.3.3: The Baseline Probits and the Lag Parameters

The probability of rain given the state in the three previous days is calculated from the parameter file using the equation 3-1. The monthly baseline probit values show the probability

of rain, given that it was dry in the three previous days as $\varphi^{-1}(b_1)$ i.e. when $\forall d_i = 0$. The baseline probit value in January was -1.279 (Figure 19: *Parameter File for Katumani* Figure 19) implying that in January the $P_{r/ddd} = \varphi^{-1}(-1.279)$ or 0.1004.

In third-order model, eight different rainfall probabilities are considered. Let r stand for a rainy day and d for a dry day, then the eight different probabilities of rain are ‘ddd’, ‘ddr’, ‘drr’, ‘drd’, ‘rrr’, ‘rrd’, ‘rdd’ and ‘rdr’. The eight different probabilities for each month can be calculated using the 12 baseline probit (BETA) values and the three lag coefficients (D1-3). MarkSim therefore uses just 15 parameters to calculate the probability of rain. The probability of rain in January given that the three previous days were dry, rainy, rainy, for example is

$$P_{r/drr} = \varphi^{-1}(-1.279 + (0)0.7610 + (1)0.1980 + (1)0.2150) = \varphi^{-1}(-0.866)$$

$$= 0.1922.$$

The rest of the probabilities are calculated in a similar way and are listed (

Table 2)

<i>Months</i>	<i>‘ddd’</i>	<i>‘ddr’</i>	<i>‘drr’</i>	<i>‘drd’</i>	<i>‘rrr’</i>	<i>‘rrd’</i>	<i>‘rdd’</i>	<i>‘rdr’</i>
<i>Jan</i>	-1.279	-1.064	-0.866	-1.081	-0.105	-0.32	-0.518	-0.303
<i>Feb</i>	-1.111	-0.896	-0.698	-0.913	0.063	-0.152	-0.35	-0.135
<i>Mar</i>	-1.018	-0.803	-0.605	-0.82	0.156	-0.059	-0.257	-0.042
<i>Apr</i>	-0.672	-0.457	-0.259	-0.474	0.502	0.287	0.089	0.304
<i>May</i>	-0.955	-0.74	-0.542	-0.757	0.219	0.004	-0.194	0.021
<i>Jun</i>	-1.201	-0.986	-0.788	-1.003	-0.027	-0.242	-0.44	-0.225
<i>Jul</i>	-1.031	-0.816	-0.618	-0.833	0.143	-0.072	-0.27	-0.055
<i>Aug</i>	-1.154	-0.939	-0.741	-0.956	0.02	-0.195	-0.393	-0.178
<i>Sep</i>	-1.138	-0.923	-0.725	-0.94	0.036	-0.179	-0.377	-0.162
<i>Oct</i>	-1.014	-0.799	-0.601	-0.816	0.16	-0.055	-0.253	-0.038
<i>Nov</i>	-0.695	-0.48	-0.282	-0.497	0.479	0.264	0.066	0.281
<i>Dec</i>	-0.987	-0.772	-0.574	-0.789	0.187	-0.028	-0.226	-0.011

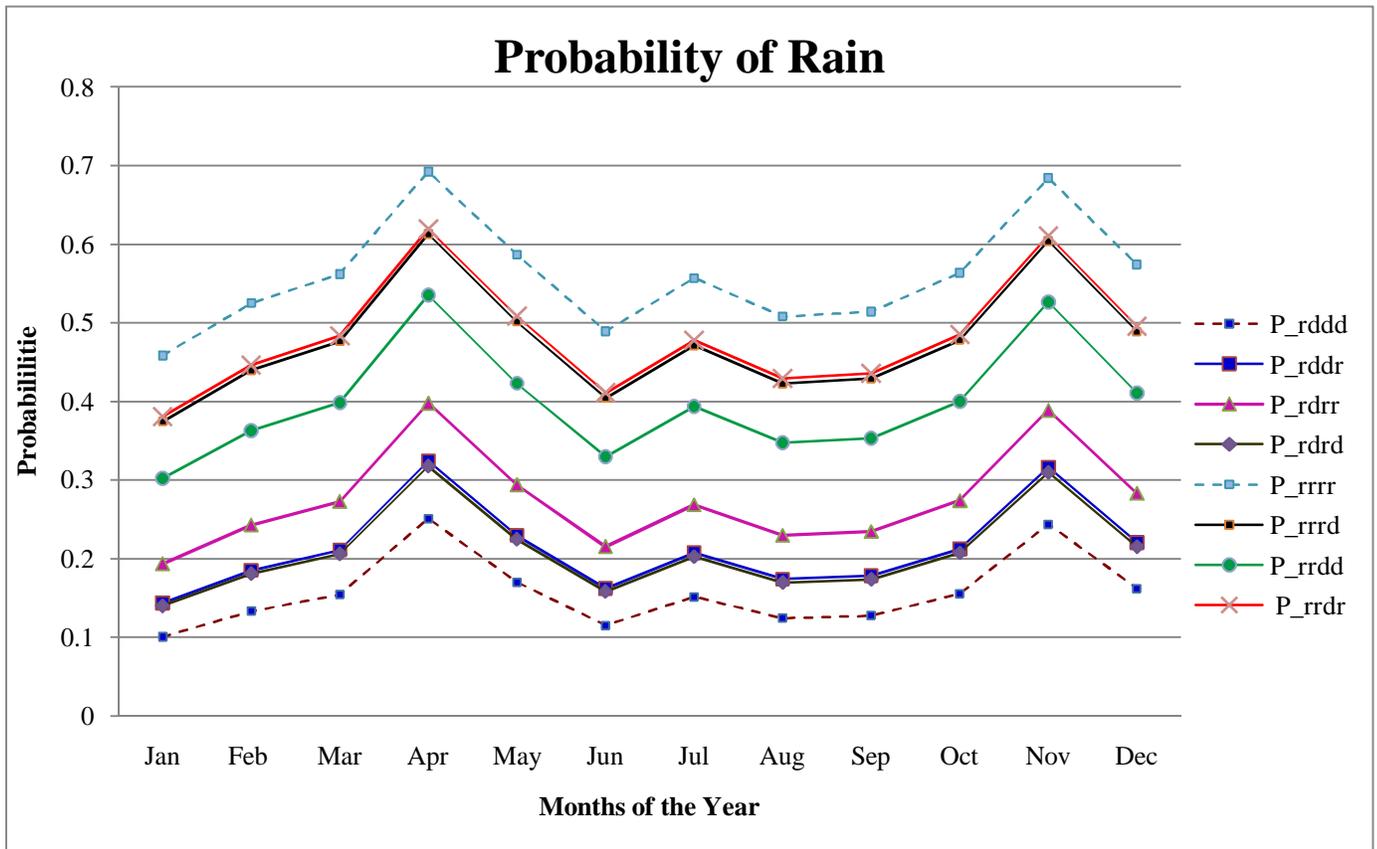
Table 1: *Probit values for the twelve months*

After the probit values have been computed, their corresponding values of probability are then computed. This can be archived well using MS-Excel.

	<i>P_r/ddd</i>	<i>P_r/ddr</i>	<i>P_r/drr</i>	<i>P_r/drd</i>	<i>P_r/rrr</i>	<i>P_r/rrd</i>	<i>P_r/rdd</i>	<i>P_r/rdr</i>
<i>Jan</i>	0.100449	0.143664	0.193245	0.139849	0.458188	0.374484	0.302229	0.380945
<i>Feb</i>	0.133284	0.185126	0.242589	0.180621	0.525117	0.439593	0.363169	0.446306
<i>Mar</i>	0.154339	0.210987	0.272589	0.206108	0.561983	0.476476	0.398589	0.483249
<i>Apr</i>	0.250792	0.323836	0.397818	0.31775	0.692166	0.612944	0.535459	0.619436
<i>May</i>	0.169789	0.22965	0.293909	0.224525	0.586675	0.501596	0.423088	0.508377
<i>Jun</i>	0.114876	0.162067	0.215348	0.15793	0.48923	0.40439	0.329969	0.41099
<i>Jul</i>	0.15127	0.20725	0.268288	0.202422	0.556855	0.471301	0.39358	0.478069
<i>Aug</i>	0.12425	0.173865	0.229347	0.169536	0.507978	0.422696	0.34716	0.429361
<i>Sep</i>	0.12756	0.178004	0.234226	0.173609	0.514359	0.428969	0.353087	0.435653
<i>Oct</i>	0.155291	0.212145	0.27392	0.20725	0.563559	0.478069	0.400134	0.484844
<i>Nov</i>	0.243528	0.315614	0.388972	0.309595	0.684031	0.60411	0.526311	0.610645
<i>Dec</i>	0.161821	0.220057	0.282984	0.215056	0.57417	0.488831	0.410601	0.495612

Table 2: *Probability values for the twelve months of Katumani Simulated data*

The value of the lag coefficient represents the extent of separation of the probit values which translates to the difference in the rainfall probabilities. For example, a large first lag coefficient value would represent a large difference between the probability of rain given the previous day was rainy and the probability of rain given that the previous day was dry. A small lag coefficient would represent a small separation of the rainfall probabilities at that coefficient. A very small third lag coefficient would suggest that a second-order model would be appropriate.



Graph 9: The Eight Different Probabilities of Rain across the Year

Figure 1 plots the eight rainfall probabilities over the year. With the large D1 value of 0.5295 there are two distinct groups of probabilities. In the first group the previous day was rainy and in the second group the previous day was dry. The probability of rain is higher when the previous day was rainy. The eight curves on the probit scale would be exactly parallel with the differences in probabilities varying slightly through the year.

Basically, Marksim was designed as computer tool that generates simulated rainfall data amongst other for crop modeling and risk assessment and to model a third order Markov chain for the simulated rainfall data. The assumption of third order models for different sites is not practical to some sites and therefore Marksim models can only be used well for rainfall data that require third order Markov chain model. GenStat capability to model any order of Markov chains is utilized in this work to develop a special routine for doing the same.

Chapter 4: GenStat Command Interface

4.1: Introduction

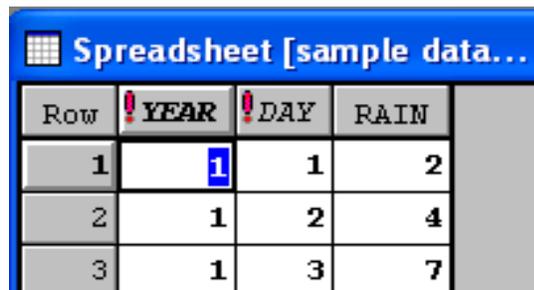
GenStat is a computer program for statistical analysis, with all the facilities of a general-purpose statistics package. It has a user friendly interface which makes it easy to use by selecting option from menus and dialogue box. Apart from being a collection of pre-programmed commands for selecting from fixed options of available analyses, it has a flexible command language which enables users to write their own commands and use them in situations where the available commands cannot solve their problems.

4.2: Command language

GenStat's own commands forms a language and is called GenStat Command Language. It is through this command language that is used to create procedures and macros within GenStat. The commands give you complete control over what is printed from analysis.

Whenever a dialogue box is completed, and [OK] button pressed, GenStat itself produces commands corresponding to the analysis process which are written in the Input Log and sent to the GenStat Server for processing. The results are sent to the output window once the Server has processed the commands. If there were any error in the commands, then the description is send to the Error Log, as well as to the Output window.

We use a sample data shown in Figure 20 to illustrate the command language.



Row	YEAR	DAY	RAIN
1	1	1	2
2	1	2	4
3	1	3	7

Figure 20: *Sample rainfall data*

To calculate the maximum and minimum rainfall per year 'Start=> Summary Statistics => Summaries of Groups (Variate)' menu is used to obtain the dialogue box selected as shown in Figure 21. The corresponding command for the analysis is generated in the input log window shown in **Figure 22**.

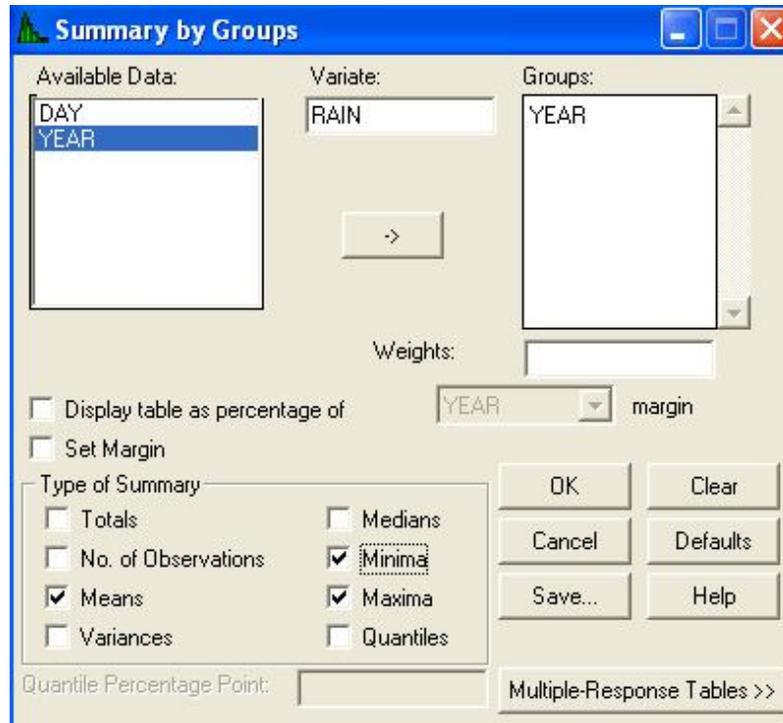


Figure 21: *Dialogue for Summary by Groups*

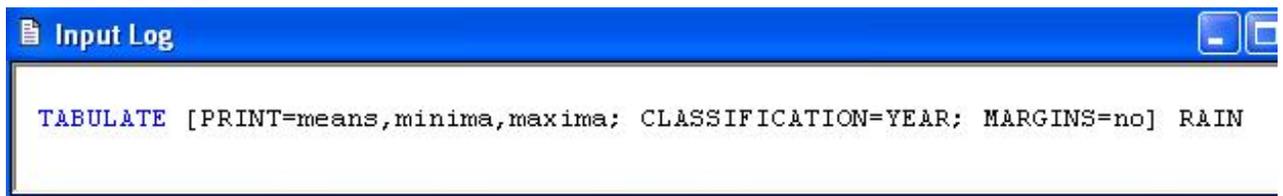


Figure 22: *Command corresponding to Summary by Groups Dialogue.*

The GenStat dialogues and menu make use of command without necessarily have to type them. It is important to learn beyond GenStat menu and make use of its commands since there are some facilities that are not available in the menu yet are in command. For example a simple procedure that calculate day lengths at any latitude and any day of the year (Gallagher & Stern, 2009). As a command it is called ‘daylength’, but there is no equivalent menu option. This command can be used to calculate the day lengths through the command interface.

An example on how ‘daylength’ is used is shown in Figure 23

```

Input Window;1*
PRINT      'Example of how to use procedure DAYLENGTH'
VARIATE    Dayno; !(1...31); DECIMALS=0
DAYLENGTH  [LATITUDE=52.205] DAYNUMBER=Dayno; DAYLENGTH=Daylength
CALCULATE  Sunset = 12 + Daylength / 2
CALCULATE  Sunrise = 12 - Daylength / 2
PRINT      'Sunrise and Sunset times for January at Wellesbourne,UK.'
PRINT      Dayno,Sunrise,Sunset; DECIMALS=2

```

Figure 23: Example that uses Day length procedure

The user has an option of submitting these commands line by line or a group of lines or the whole window by using the ‘Run=>Submit line’, ‘Run=>Submit Selection’ and ‘Run=>Submit window’ menus respectively as shown in Figure 24. Once the window has been submitted, the result is displayed in the output window.

The screenshot shows the 'Run' menu with the following options:

- Submit Line (Ctrl+L)
- Submit to Current Line (Ctrl+Shift+K)
- Submit from Current Line (Ctrl+Shift+M)
- Submit Selection (Ctrl+M)
- Submit Window (Ctrl+W)**
- Recycle Window (Ctrl+Shift+W)
- Submit Clipboard (Ctrl+K)
- Submit File... (Ctrl+B)
- Submit R Script...
- Submit BUGS Script...
- Restart Session
- Restart Server
- Flush Server

The 'Output' window displays the following text:

```

Sunrise and Sunset times for January at Wellesbourne,UK.
34 PRINT      Dayno,Sunrise,Sunset; DECIMALS=2
Dayno    Sunrise    Sunset
1.00     8.09     15.91
2.00     8.07     15.93
3.00     8.06     15.94
4.00     8.05     15.95
5.00     8.04     15.96
6.00     8.02     15.98
7.00     8.01     15.99
8.00     7.99     16.01
9.00     7.97     16.02

```

Figure 24: Running the commands

4.3: Compacting programs

Compacting programs in GenStat involves storing statements (commands) made by users so that they can refer to them later or use them repeatedly. There are two ways of compacting programs in GenStat namely: Macros and Procedures (Payne, et al., 2009).

4.3.1: Procedures

When statements are compacted using Procedure, they must be self-contained, apart from those explicitly defined as options or parameters of the procedure. That is, all the data

structures that they use are accessible only within the procedure. Once you write programs for complicated tasks, GenStat allows you to keep them for future use. You may also wish to use procedures written by other people by simply giving the name of the procedure, and specifying options and parameters as required.

When GenStat meets a statement with a name that it does not recognize as one of the standard GenStat directives, it first looks to see whether you have a procedure of that name already stored in your program. Then it looks it up in any procedure library that you may have attached explicitly to your program, taking these in order of their channel numbers. After locating the required procedure, GenStat reads it in, if necessary, and then executes it. Information is transferred to and from a procedure only by means of its options and parameters.

Below is a new procedure (*Counts*) which counts the number of dry and wet days on a given Markov chain order. It also calculates the mean amount of rainfall of rainy days.

```
PROCEDURE [restore=case] 'Counts'
OPTION NAME= \
'STATES' , "(I: Scalar giving the number of states of the chain)"\
'LIMITS' , "(I: Scalar or vector giving the values to determine the states)"\
'ORDER' , "(I: Order of the Markov Chain)"\
'CLASS' , "(I: Factor over which the data are to be totalled)"\
'HIGH' , "(I: YES/NO indicating if high order, default is NO)"\
'LABELS' , "(I: YES/NO on whether to label the tables of results, default is YES)"\
'SPREAD' , "(I: YES/NO on whether to display the results in spreadsheets)"\
'INITIAL' ; "(I: Scalar or variate - length =ORDER with initial values)"\
MODE=4(P),3(T),P; SET=3(yes),no,3(yes),no; DECLARED=8(yes);\
PRESENT=8(yes) ;\
DEFAULT=2,!(0.85),1,*,'NO','YES','NO',*;\
TYPE=!T('scalar'),!T('scalar','variate'),!T('scalar'),!T('factor'),\
3(!T('text')),!T('scalar','variate');\
VALUE=*,*,*,*,2(!T('YES','NO')),*,*
PARAMETER NAME= \
'DATA' , "(I: vector from which spell lengths are to be formed" \
'COUNTS' , "(O: tables containing the frequencies)" \
'AMOUNTS' , "(O: table containing the sum of the amounts)"\
'CATEGORIES' , "(O: variate containing the category of each observation)"\
'SCOUNTS' ; "(O: pointer to tables for frequencies)"\
MODE=P; SET=yes,4(no); DECLARED=yes,4(no); \
TYPE=!T('variate'),2(!T('table')),!T('variate'),!T('pointer'); \
PRESENT=yes,4(no)
```

Figure 25: Procedure Counts- sec 1

The procedure 'counts' in Figure 25 can now be executed through the command interface by referring to the procedure name. The command that make use of the procedure name is as shown in the table below.

```

COUNTS [order=2;states=2;class=DayOfYear; spread='YES';\
high='yes'; labels='yes';limits=!(0.2,0.5);\
INITIAL=!(0,0)]Rain;cat=c;counts=counts;a=amounts

```

Figure 26: *Procedure Counts- sec 2*

In the command above, the options are define in the brackets [] these are; *order, states, class, spread, High, labels, limits*. While the parameters are defined outside the brackets, these are *Rain, categories counts, amounts*. The rules for using the procedure are identical. Using a procedure does not require the knowledge of how the program inside operates, what data structures it contains, nor what directives it uses so long as you specify the correct options and parameters (Payne, et al., 2009).

4.3.2: Macros

A macro is created when GenStat program is placed into a text. Once a macro is formed, you can use the macro to execute the program that is built within it by typing a pair of hash characters ## followed by the macro name. An example of a macro is shown in the Figure 27

```

text[VALUES='COUNTS [order=2;states=3;class=DayOfYear; spread=YES; high=NO; \
labels=yes;limits=!(0.2,0.5);INITIAL=!(0,0)]Rain;cat=c;counts=counts;a=amounts']trial
## trial

```

Figure 27: *Macro*

The program shown in Figure 26 can be executed within a macro shown in the Figure 27. First the command is defined within the single quotes (‘’) of the text, then the macro is executed by running command line ##trial. The two method of compacting programs are made use of to built more complex commands.

4.4: Creating Menus and Dialogs

GenStat allow its users to create their own menus for existing procedures. Any menu created is added in the main menu called *user* .Once the procedure is completed, they are saved with the extension *.gpi*, in a GenStat folder called *AddIn* found in *C:\Program files\Gen12ed\AddIn*. The codes are then run using *Run=>Submit* window menu. The procedure is the added in the library using *Tools=>Procedure Library=>Build*. (Gallagher & Stern, 2009)

Commands for creating the menu are then created in a new input window. The layout and the content structure are created using the GenStat resource language basic structure is described below.

```
DIALOG id, width, height  
CAPTION text  
PROCEDURE text  
BEGIN  
...  
... dialogue control (layout and content instruction  
---  
END
```

DIALOG statement defines the *id* of the dialogue and its size of the dialogue box. CAPTION statement specifies the title that will appear on the dialogue title. The PROCEDURE statement specifies the procedure that will be created in the dialogue. The BEGIN-END is a block of statement that specifies the body structure of the Dialogue box. The file is then saved with extension *.grc*. (Gallagher & Stern, 2009).

The *.grc* file is then attached into GenStat. The layout of the commands used for attaching the *.grc* file is described below.

```
[AddInX]  
Type = DIALOG  
ProgID = my dialogue id  
MenuTitle = name in menu  
ResourceFile = dialogue code.grc
```

The underlined text needs to be change by the user. The command is also created in a new input window then saved with extension *.gad* .

A sample of the dialogue box created in this study is shown in Figure 28. All the commands and procedures are included in the appendix.

```

DIALOG dialog1, 240, 240
CAPTION "Counts and Totals from Daily Data"
PROCEDURE "COUNTS"
BEGIN
    PUSHBUTTON "&Run", IDRUN, 20, 220, 45, 15, BS_RUN
    PUSHBUTTON "&Reset", IDCLEAR, 95, 220, 45, 15, BS_CLEAR
    PUSHBUTTON "&Cancel", IDCANCEL, 160, 220, 45, 15, BS_CANCEL
    "PUSHBUTTON "&Default", IDDEFAULT, 210, 180, 30, 15, BS_DEFAULTS"
    LTEXT "&Available Data:", ITEXT, 5, 5, 55, 10
    LISTBOX IDLIST, 5, 20, 90, 120, GW_AVAILABLE
    LTEXT "&Data:", ISTATIC, 115, 5, 60, 10
    EDITTEXT DATA, 115, 15, 80, 14, GW_AVAILABLE | DT_TYPPVAR | GW_PARAM | EC_SET
    LTEXT "&Class for totals:", ISTATIC, 115, 35, 55, 10
    EDITTEXT CLASS, 115, 45, 80, 14, GW_AVAILABLE | DT_TYPPFAC | GW_OPT | EC_SET
    LTEXT "&Type of Markov:", ISTATIC, 115, 65, 60, 10
    COMBOBOX HIGH, 175, 65, 60, 40, CB_DROPLIST | GW_OPT
    SLIST HIGH, "Normal", "High"
    RLIST HIGH, "NO", "YES"
    LTEXT "&Number of States:", ISTATIC, 115, 90, 60, 10
    EDITTEXT STATES, 215, 90, 20, 11, GW_OPT
    LTEXT "&Threshold Values:", ISTATIC, 115, 105, 60, 10
    EDITTEXT LIMITS, 200, 105, 35, 11, GW_OPT
    LTEXT "&Initial Value:", ISTATIC, 115, 120, 60, 10
    EDITTEXT INITIAL, 200, 120, 35, 11, GW_OPT
    LTEXT "&Markov Chain Order:", ISTATIC, 115, 135, 80, 10
    EDITTEXT ORDER, 215, 135, 20, 11, GW_OPT
    GROUPBOX "Save result in", IDSTATIC, 40, 150, 170, 40
    TOGGLEBOX "Amount", a, 55, 165, 40, 14, GW_PARAM, "amount", "NO"
    TOGGLEBOX "Counts", counts, 110, 165, 40, 14, GW_PARAM, "counts", "NO"
    TOGGLEBOX "Class", cat, 155, 165, 40, 14, GW_PARAM, "c", "NO"
    TOGGLEBOX "Display in Spreadsheet", SPREAD, 100, 200, 90, 14, GW_OPT, "YES", "NO"

END

```

Figure 28: A *.grc* file for creating Counts and Amount Daily data menu

```

AddIn1]
Type=DIALOG
ProgID=dialog1
MenuTitle=Counts and Totals from Daily Data
ResourceFile=counts.grc

```

Figure 29: A *.gad* file for attaching dialogue command

Chapter 5: Markov modeling of Climatic data in GenStat

5.1: Introduction

The potentials of GenStat are exploited to provide routines that can be analyze rainfall data using Markov model and make the accessible to users. The First stage involves preparing

the data for fitting; the second stage involves fitting the Markov model for the rainfall. Three procedures are written which includes; COUNT, PREPARE, FITTING, AMOUNTFITTING.

The COUNT procedure reads the raw data fed by the user and summarizes it by counting the number of rainy days for each day of the number of the years the data is given. It also calculates the amount of rain in the rainy days. The result of the summaries is then tabulated ready for calculation of the probabilities and fitting. The PREPARE procedure calculates the probability of rain, the FITTING procedure, fits the probability of rain while AMOUNTFITTING fits the amount of rainfall. This chapter provides the documentation of what the procedures that are newly created do.

5.2: COUNT Procedure.

COUNTS procedure, identifies the state condition on each day then TABULATE directive is used to summarize the frequency of each day of the year possessing a specific state over the number of years the data is given.

5.2.1: Options for COUNT Procedure

States – This is a scalar value that defines the number of states, i.e. 2, 3, 4 etc. A state is determined by the amount of the rainfall below or above a specified threshold. The number of state equals the number of limits (threshold) values plus one.

Limits – This is a scalar or a variate that is used to determine the limit of a state, the default value of limit is 0.85. If a day received rainfall amount that is less than this value (0.85), then the day is referred to as a dry day, otherwise it will be a wet day.

Order- It is a scalar value that defines the number of order the Markov model will use to model the data. The values range from zero onwards

Class – It is a factor with 366 values over which the data is to be totaled. It corresponds to the day number of the years.

High – It specify whether the Markov order is high or normal as explained in section 1.2.4: of this thesis.

Labels –Specifies whether or not the table of the *counts* will have labels or not (the default is ‘yes’)

Spread – It is a string specifying whether or not to display the result of counts and amount in a spreadsheet (the default is ‘yes’).

Initial – It is a scalar or variate that assigns initial value to days before the first data of the data. It is applicable to Markov of order one and above. The initial values are specified to avoid missing data for days before the first day.

Table – It is a string specifying whether or not to display the result of counts and amount in a table structure (the default is ‘yes’).

5.2.2: Parameters for Count Procedure

Data- It is an input vector parameter of rainfall data. The rainfall data is given as one variate for all the years.

Counts- It is an output table parameter containing the sum of the frequencies

Amounts – It is an output table containing the sum of the amounts

Categories- An output variate parameter containing the category of each observation

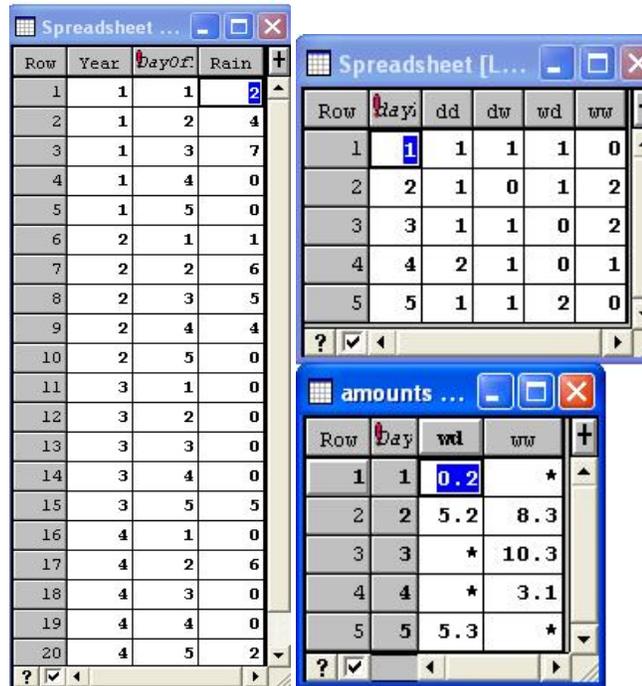
Scounts - An output pointer parameter to tables for frequencies

5.2.3: Description of Count Procedure

The COUNTS procedure counts the number of times a day of the year having a specific state-condition for the number of years and also calculates the rainfall amounts for rainy days. Amount of rainfall is defined as the actual amount of rainfall recorded minus the threshold value. That is, if the threshold is 0.85mm and in a specific day, it was recorded that the rainfall was 2mm, then the rainfall amount is 1.15mm.

The frequencies of the days with the same state conditions are organized in a table using TABULATE directive with columns representing the state condition e.g. *dry, wet, dd, dw, ddd, dww*, etc while the rows indicate factor levels for example day number of the year totaling to 366 days (see a sample in **Figure 32** (b)). The amounts of rainfall are also summed over the rainy days. This is further illustrated in the figure below. Using the sample data set in

Figure 30 (a), the resulted tables that are created when the COUNTS procedure is run with order 1 and threshold of 0.85 are shown in Figure 30 (b)



(a)

(b)

Figure 30: A sample rainfall data and results of count procedures.

Considering a state condition with wet given wet for day 2, i.e. there are only 2 days that satisfy this condition. The two days had rainfall recorded as 4mm and 6mm. the rainfall amount recorded in the amount table is 8.3 (4mm+6mm-2(0.85mm)).

The rainfall amount of a given day of the year with specific state, is calculated using the equation (3-1 and (3-2.

The results of the count procedure are given in a table structure with the number of columns determined by the order of the model.

The table of counts and amounts may be displayed in a spreadsheet by declaring SPREAD=yes. The initial record for rain can be set using the option of INITIAL. The number of values in INITIAL variate must be equal to the ORDER value.

5.3: PREPARE Procedure

The PREPARE procedure converts the input structures (*counts* and *amounts* tables) to variates by making use of VTABLE Procedure (Goedhart, 1991), this enable manipulation of the result in each columns of the table.

5.3.1: Options for PREPARE procedure

Plot-Specifies whether or not to plot the results, the default is YES

Spread- Specifies whether or not to display the results in spreadsheets, the default is YES

5.3.2: Parameter for Prepare procedure

The parameters that are used in the PREPARE procedure are;

Counts - This is an input table containing the frequencies, it correspond to the count table resulted from the COUNT procedure

Amounts – This is an input table containing the amounts of rainfall. It correspond to the amount table resulted from the COUNT procedure described above.

5.3.3: Description of Prepare Procedure

The PREPARE procedure reads data from the results created in COUNTS procedure; these are *counts* and *amounts*. The input structures are converted into variate then the probabilities are calculated by applying the expression below.

$$P_{ijk\dots} = \frac{f_{ijk\dots}}{\sum_{i=1}^C f_{ijk\dots}} \quad i, j, k = 1, 2, 3, \dots, C$$

(5-1)

Where

C= maximum number of states

$P_{ijk\dots}$ = The probability of the present state is i given that the state in the previous day is j and day before yesterday is k and so on up to the number of order. If order zero is used then we refer to P_i , for order one (P_{ij}), for order two (P_{ijk}) and so on.

$f_{ijk\dots}$ - The number of days that posses the condition that the present state is i given that the state in the previous day is j and day before yesterday is k and so on up to the number of order

$\sum_{i=1}^C f_{ijk\dots}$ - The total number of days with the previous day having state j and the day before the having state k regardless of the state in the present day.

The probabilities may be saved, displayed in the spreadsheet when SPREAD=yes or plotted by the DGRAPH directive when PLOT=yes. The high order Markov can be obtained when HIGH=yes which is only applicable when ORDER \geq 2. The order value used in the COUNTS procedure must be maintained in PREPARE procedure

5.4: FITTING command procedures.

Fitting Procedure converts the counts and amounts table into variate using Vtable procedure (Goedhart, 1991). The variates are then used to fit the probabilities of rain. The result may be displayed in a graph or saved.

5.4.1: Options for FITTING procedure

Harmonics- It is a scalar to represent the number of harmonic required for the model fitting, default is 3

Spread- Specifies whether or not to display the results in spreadsheets, the default is NO

5.4.2: Parameter for FITTING procedure

The parameters that are used in the FITTING procedure are;

Counts -Tables containing the frequencies, it correspond to the count table resulted from the count procedure

Amounts – This is an input table containing the amounts of rainfall. It correspond to the amount table resulted from the count procedure described above.

5.4.3: Description of Fitting Procedure

The `FITTING` parameter is set to read reads data structure of type table. (Counts and amounts tables from `COUNTS` procedure). The input structures are converted into variate then fitted. The variates forms the Y values while the *dayNo* factor is assigned to the X-values when plotting the fitted parameters.

5.5: AMOUNTFITTING command procedures.

`AMOUNTFITTING` Procedure converts the counts and amounts table into variate using `VTABLE` procedure (Goedhart, 1991). The variates are then used to fit the amounts of rain. The result may be displayed in a graph or not.

5.5.1: Options for AMOUNTFITTING procedure

Harmonics- It is a scalar to represent the number of harmonic required for the model fitting, default is 3

Spread- Specifies whether or not to display the results in spreadsheets, the default is NO

5.5.2: Parameter for FittingAmount procedure

The parameters that are used in the `AMOUNTFITTING` procedure are;

Counts -Tables containing the frequencies, it correspond to the count table resulted from the count procedure

Amounts – This is an input table containing the amounts of rainfall. It correspond to the amount table resulted from the count procedure described above.

5.5.3: Description of FittingAmount Procedure

The AMOUNTFITTING parameters are set to read data structure of type table. (counts and amounts tables from COUNTS procedure). The input structures are converted into variate then fitted. The variates forms the Y values while the *dayNo* factor is assigned to the X-values when plotting the fitted amounts.

Chapter 6: Results and Discussion

6.1: Application of the model to Katumani Data

This chapter discusses how the newly developed procedures are used in all the aspects of the options specified. The daily rainfall data collected from 1961- 2001 at Katumani Experiments research Station in Kenya is used to test the procedure. The data is imported from Instat library into GenStat through excel. Alternatively it can be exported to GenStat from Instat window.

6.2: Description of the data

The data is for rainfall recorded in mm for 41 years, the columns represents the years while the row represent the day of the year running from 1 to 366. Part of the screen shot showing the data set in GenStat is in the appendix I. There are 10 leap years and 31 non-leap years. The maximum of maximum annual rainfall was received in 1963 (186.9 mm) while the minimum of maximum annual rainfall was received in 1987 (26.7 mm) (see appendix II).

6.3: Organizing the data

The counts procedure reads only two variates; Rain and a factor column over which the count is to be done i.e. day of the year, months of the year, weeks of the year etc. In this text the factor DayOfYear is used. The data is stalked such that the variate Rain is in one column for all the days and DayOfYear in one column for all the years. The *Spread=> Manipulate=> Stalk* menu is used to achieve this. The ‘Stalk Column in Sheet’ dialogue box is filled as shown in Figure 31. The number of columns in this case will be the number of years (see appendix I). The stalked data is shown in Figure 32.

The factor column is changed to a variate by just right clicking the year column then choosing *Convert to variate* option. This result will now show the level values of the column. The *Spread => Calculate => Fill* menu is used to fill the year’s values as shown in Figure 33. This is important for the calculation of the *DayOfYear*. The stalked data with year variate is shown in Figure 34.

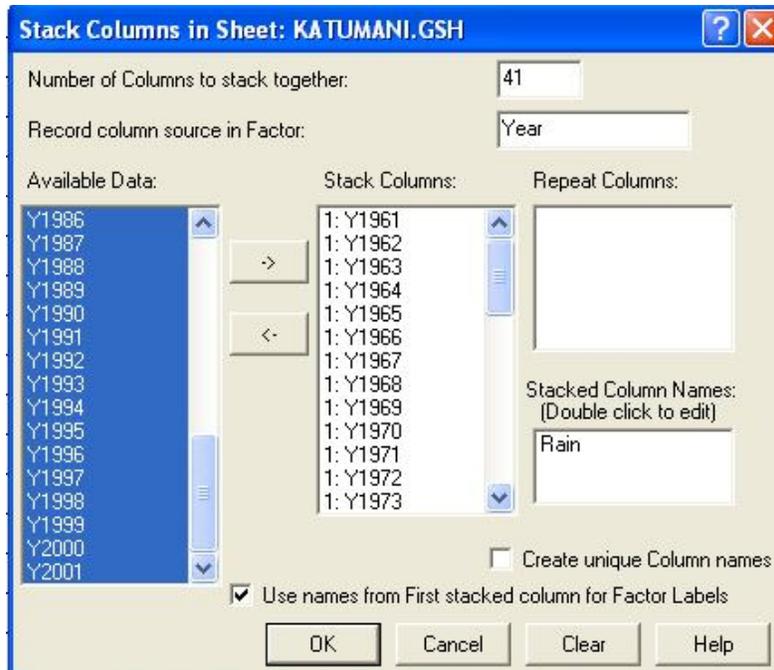


Figure 31: *Stack column dialogue box*

Row	Year	Rain
1	Y1961	0
2	Y1961	0
3	Y1961	0
4	Y1961	0
5	Y1961	0
6	Y1961	0
7	Y1961	0
...
364	Y1961	0
365	Y1961	0
366	Y1961	0
367	Y1962	0
368	Y1962	11.2
369	Y1962	31.5
...
15004	Y2001	0
15005	Y2001	0
15006	Y2001	0

Figure 32: *Stacked data*

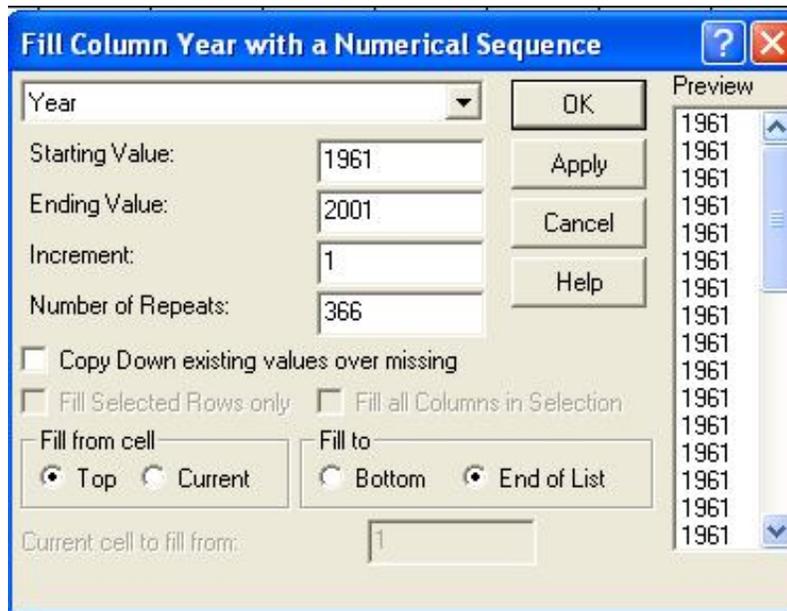


Figure 33: Fill column Year with Numerical Sequence dialogue box

Row	Year	Rain
14988	2001	2.1
14989	2001	4.2
14990	2001	2.4
14991	2001	0
14992	2001	1.6
14993	2001	1.5
14994	2001	1.8
14995	2001	0
14996	2001	0
14997	2001	0
14998	2001	3
14999	2001	0
15000	2001	0

Figure 34: Year in variate

Create two columns, day of the months and month of the year as shown in the Figure 35. The three columns (day, month and year) are used to calculate the date unless data is entered with date format. Using *Spread=>Calculate=> column*, (Figure 36) the date column is calculated (See Figure 37).

Row	DayOfMonth	Month	Year	Rain
1	1	Jan	1961	0
2	2	Jan	1961	0
3	3	Jan	1961	0
4	4	Jan	1961	0
5	5	Jan	1961	0
6	6	Jan	1961	0
7	7	Jan	1961	0
8	8	Jan	1961	0
9	9	Jan	1961	0
10	10	Jan	1961	0
11	11	Jan	1961	0
12	12	Jan	1961	0
13	13	Jan	1961	0

Figure 35: Data with day, month and year

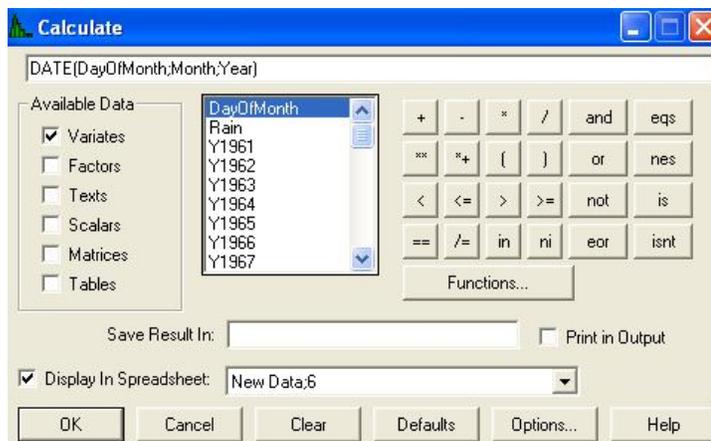


Figure 36: Calculate dialogue box for date

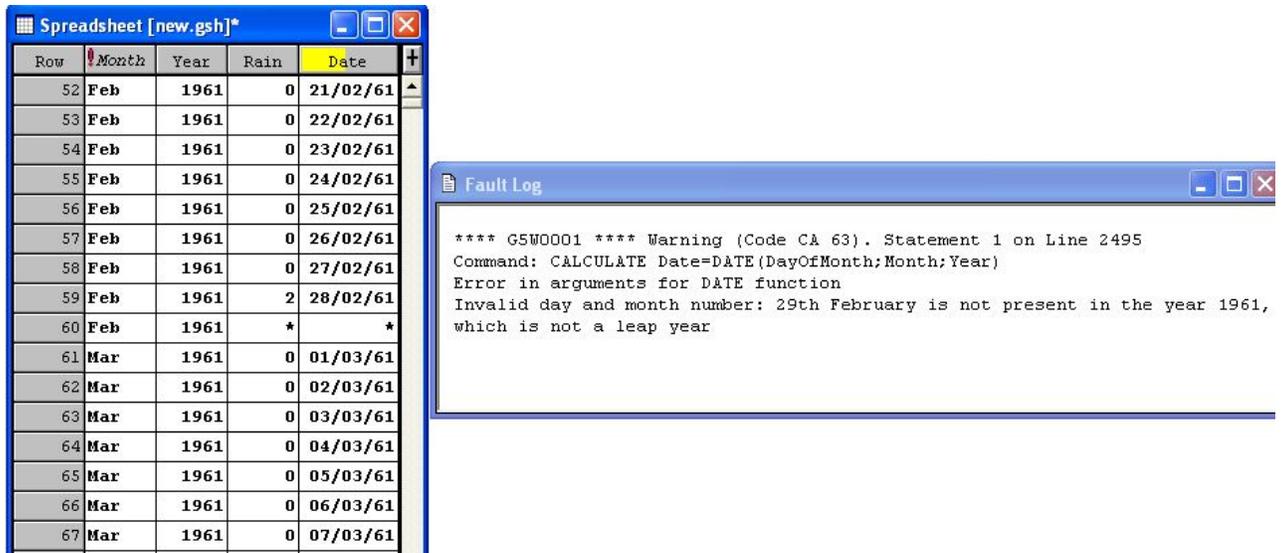


Figure 37: *Date column calculated but with a fault message*

The date column is calculated though a fault message is shown, which states that 29th Feb of a non-leap year is invalid, hence no date is created on the 29th Feb. To avoid this fault, the invalid day is eliminated before calculating the dates. To eliminate the invalid data (29th Feb of a non-leap year), you can just delete the rows which are invalid.

Column for *DayOfYear* is calculated using MFRACTION (Date; 1; 1). This always gives 1st March as day 61 in the year instead of NDAYINYEAR (Date; 1) which gives 1st March as the 61st day in a leap year, but 60th day in other years.

The data set ready for the analysis is shown in Figure 38. The *dayofyear* with the 1st march (61st day) highlighted.

Row	Month	Year	Date	DayOfYear	Rain
46	Feb	1961	15/02/61	46	0
47	Feb	1961	16/02/61	47	0
48	Feb	1961	17/02/61	48	0
49	Feb	1961	18/02/61	49	0
50	Feb	1961	19/02/61	50	0
51	Feb	1961	20/02/61	51	0
52	Feb	1961	21/02/61	52	0
53	Feb	1961	22/02/61	53	0
54	Feb	1961	23/02/61	54	0
55	Feb	1961	24/02/61	55	0
56	Feb	1961	25/02/61	56	0
57	Feb	1961	26/02/61	57	0
58	Feb	1961	27/02/61	58	0
59	Feb	1961	28/02/61	59	2
60	Mar	1961	01/03/61	61	0
61	Mar	1961	02/03/61	62	0
62	Mar	1961	03/03/61	63	0

Figure 38: Data set with 'DayOfYear' and 'Rain' column ready for analysis

6.4: The User Menu

The four climatic menus corresponding to the procedures are created under the *user* menu in GenStat. Through this menu the procedures can now be accessed and used (Figure 39). Choosing *User=> Count and Totals from daily Data* will open count dialogue box (Figure 42) which correspond to the COUNT Procedure

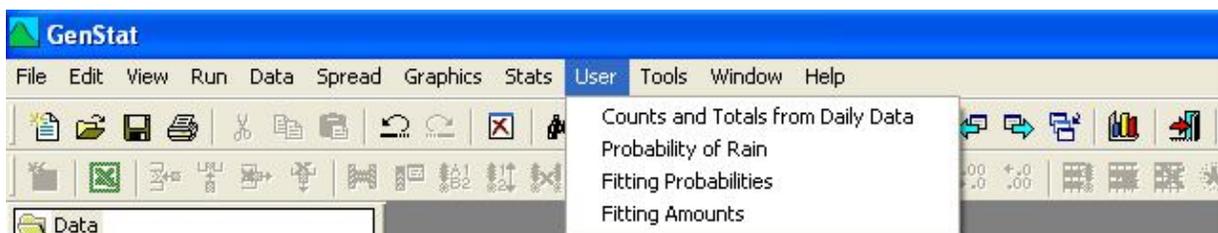


Figure 39: User Menu

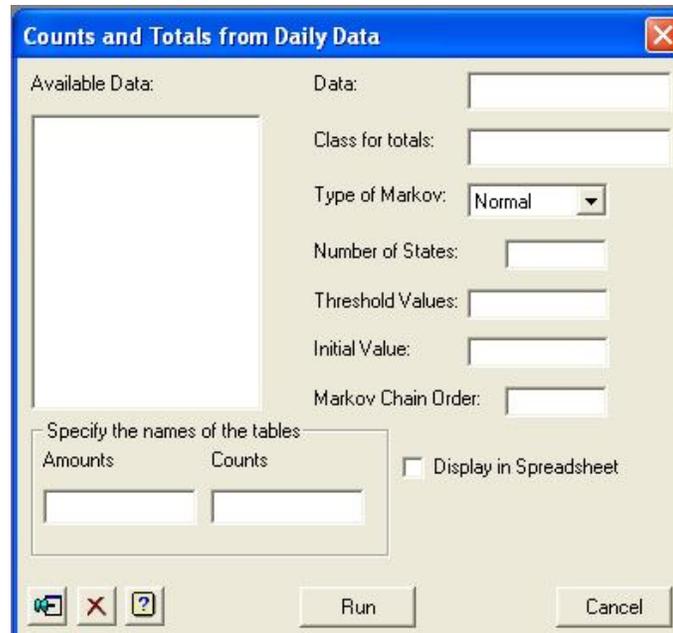


Figure 40: *Count and Totals from daily Data dialog box*

The field for ‘data’ receives the rainfall data and must be set. The ‘class for totals’ specifies the factor over which the data for counts will be totaled. This can either be the month of the year; day of year etc. ‘Type of Markov’ field specifies the type of Markov chain to model, either normal or high. The fields ‘number of states’, ‘threshold value’, ‘initial value’ and ‘Markov chain order’, when left blank will use the defaults values.

6.5: Two- State Markov chain modeling

In a two- State Markov chain, the states are dry and wet; the threshold value is only one marking the boundary of the two states. Using the default value of threshold option (0.85mm), any value of rainfall greater than 0.85mm is the state of rain while less than or equal to 0.85 is the state of dry. This section discusses Markov chain modeling of order zero and one.

6.5.1: Order zero

When modeling a normal two-state zero order Markov chain, i.e. the chance of an event occurring in day τ without considering the event that occurred in the previous day(s),

the User => *Count and Totals from daily Data* menu is used to obtain the *Count and Totals from daily Data* dialogue box and then filled as shown below.

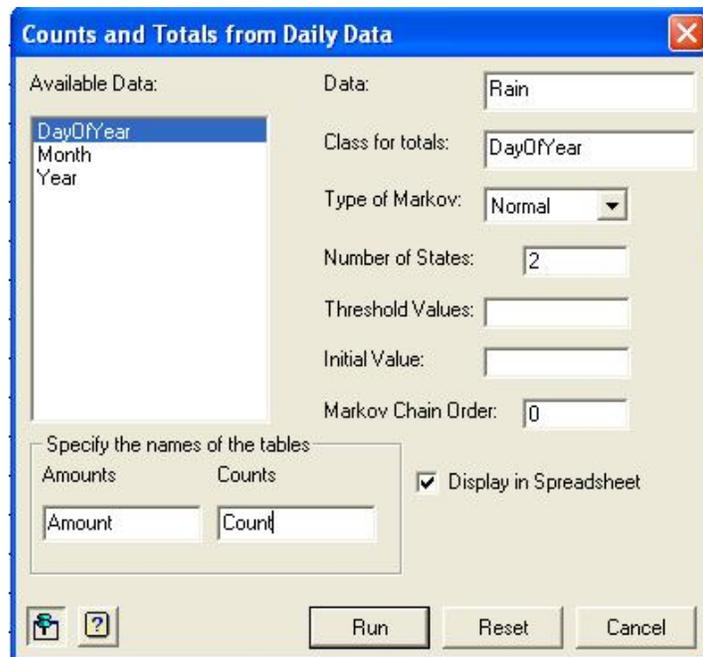


Figure 41: Dialogue box for Counts and totals from Daily Data (Order Zero)

The result gives the count and amounts based on order zero (**Figure 42** and **Figure 43**).

Row	DayOfYear	0	1
1	1	35	6
2	2	33	8
3	3	31	10
4	4	34	7
5	5	33	8
6	6	20	13

Figure 42: Counts of rainy days (order zero)

Row	DayOfYear	w
1	1	22.2
2	2	116.7
3	3	101.0
4	4	78.8
5	5	87.2
6	6	101.0
7	7	73.7

Figure 43: Rainfall amounts (order zero)

A PREPARE procedure is then used to calculate the chance of rainfall for order zero. The *User => Prepare* menu is used to load the Prepare dialog and the option and parameter set as shown in figure below.

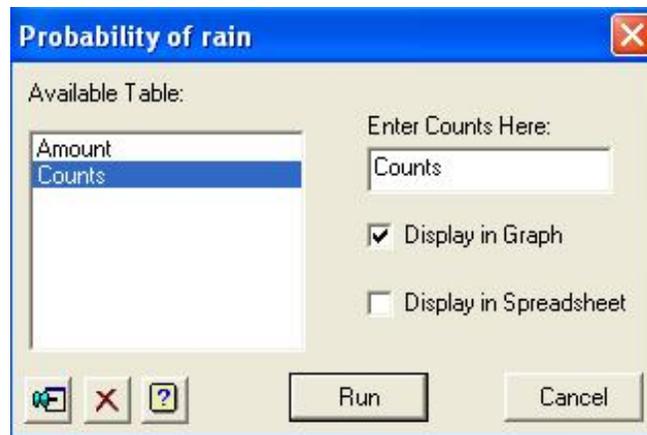
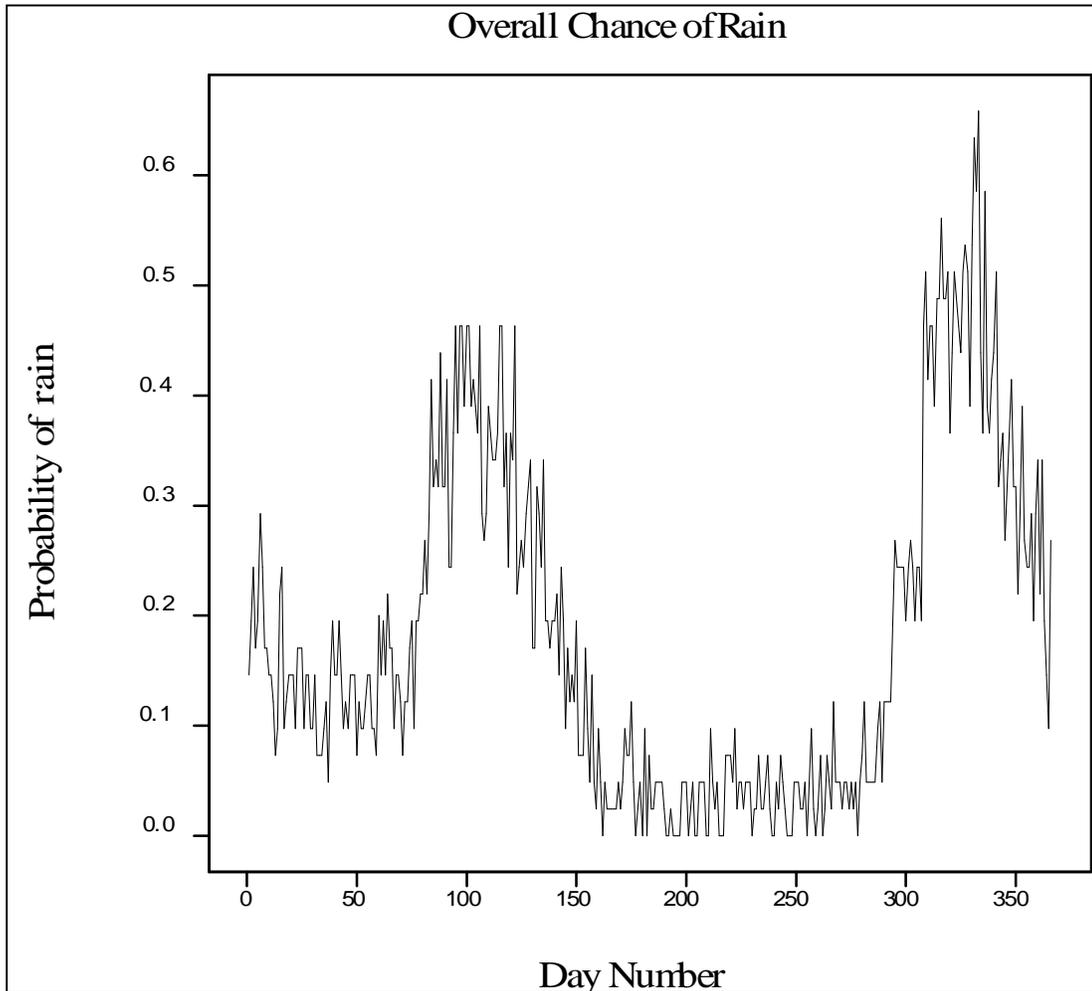


Figure 44: Dialogue box for Probability of rain (Order Zero)

The results of probability are displayed on a spreadsheet (Figure 45) and a plot of the probabilities (Graph 10) is displayed.

Row	dayNo	P_d	P_w
1	1	0.853659	0.146341
2	2	0.804878	0.195122
3	3	0.756098	0.243902
4	4	0.829268	0.170732
5	5	0.804878	0.195122
6	6	0.707317	0.292683

Figure 45: Probabilities of rainy day form GenStat (order zero)

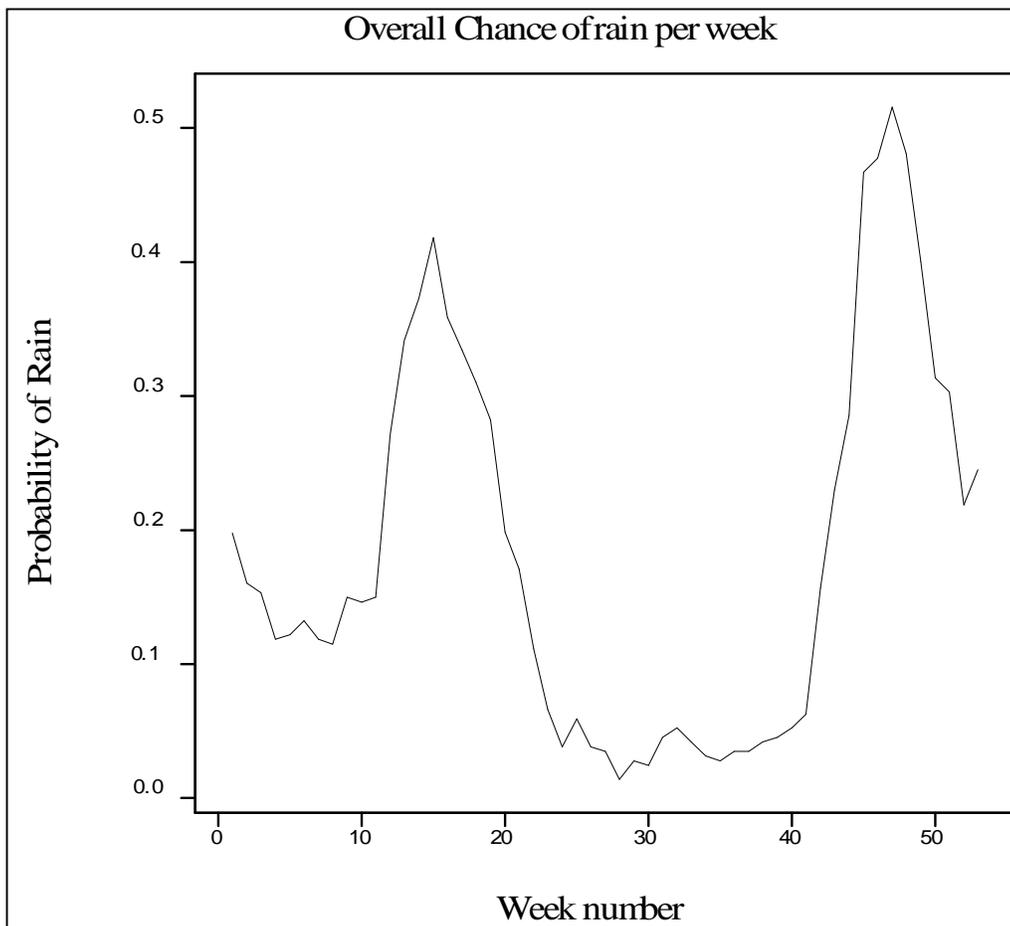


Graph 10: *Probabilities of order zero (GenStat)*

The graph indicates that Katumani has two seasons of rainfall, these are long Seasons (April-May) and short seasons (November). The chance of rain during the short rain period is higher compared to that of long rain period. The same results were depicted when Instat software was used (refer to section 3.2.2:). The day number can be grouped into 5-days totals, week's totals or monthly. To obtain weekly totals, the user will have to create a factor column of week number and use it instead of the DayOfYear (Figure 46). The count totaled in 7 days (weekly) gives a clearer graph (see Graph 11).

Row	Year	DayOfMor	Month	Date	DayOfYear	Rain	Week Number
1	1961	1	Jan	01/01/61	1	0	52
2	1961	2	Jan	02/01/61	2	0	1
3	1961	3	Jan	03/01/61	3	0	1
4	1961	4	Jan	04/01/61	4	0	1
5	1961	5	Jan	05/01/61	5	0	1
6	1961	6	Jan	06/01/61	6	0	1
7	1961	7	Jan	07/01/61	7	0	1
8	1961	8	Jan	08/01/61	8	0	1
9	1961	9	Jan	09/01/61	9	0	2
10	1961	10	Jan	10/01/61	10	0	2
11	1961	11	Jan	11/01/61	11	0	2

Figure 46: Dataset with week number created (Order Zero)



Graph 11: Overall chance of weekly rainfall of order zero (GenStat)

The fitted probability is obtained using *User=>Fitting probability of rain* menu which loads 'Fitting Probabilities of rain' dialogue box associated to the FITTING procedure. The options for the procedure are set as; Harmonics=3(Figure 46)



Figure 47: Dialogue box for fitting Probability of rain (Order Zero)

Clicking run button displays the fitted model on a graph (Graph 11) and the regression output shown below starting with harmonic of order 2.

Regression analysis

```

Response variate:  w
Binomial totals:  tot
Distribution:      Binomial
Link function:    Logit
Weight variate:   tots
Fitted terms:    Constant, half_hamonic[1], half_hamonic[2],
half_hamonic[3], half_hamonic[4]

```

Summary of analysis

			mean	deviance	approx
Source	d.f.	deviance	deviance	ratio	chi pr
Regression	4	1917.8	479.451	479.45	<.001
Residual	361	542.5	1.503		
Total	365	2460.3	6.741		
Change	-1	-119.4	119.364	119.36	<.001

Table 3: Summary of analysis of fitted probability of rain for 0-order with 2

The data have a binomial distribution with logit link function. These are the two aspects required to characterize a generalized linear model (GLM). Instead of sum of squares, the deviances are obtained when fitting a GLM.

Estimates of parameters

Parameter	estimate	s.e.	t(*)	t pr.	antilog of estimate
Constant	-1.9904	0.0336	-59.15	<.001	0.1366
half_hamonic[1] Sin[1]	0.7023	0.0390	17.99	<.001	2.018
half_hamonic[2] cos[1]	1.2007	0.0468	25.65	<.001	3.323
half_hamonic[3] Sin[2]	-1.2589	0.0411	-30.60	<.001	0.2840
half_hamonic[4] cos[2]	-0.3816	0.0355	-10.76	<.001	0.6828

Table 4: Estimate of parameters of fitted probability of rain for 0-order with 2

Accumulated analysis of deviance

Change	d.f.	deviance	mean deviance	deviance ratio	approx chi pr
+ half_hamonic[1] Sin[1]	1	93.386	93.386	93.39	<.001
+ half_hamonic[2] cos[1]	1	535.478	535.478	535.48	<.001
+ half_hamonic[3] Sin[2]	1	1169.576	1169.576	1169.58	<.001
+ half_hamonic[4] cos[2]	1	119.364	119.364	119.36	<.001
Residual	361	542.511	1.503		
Total	365	2460.315	6.741		

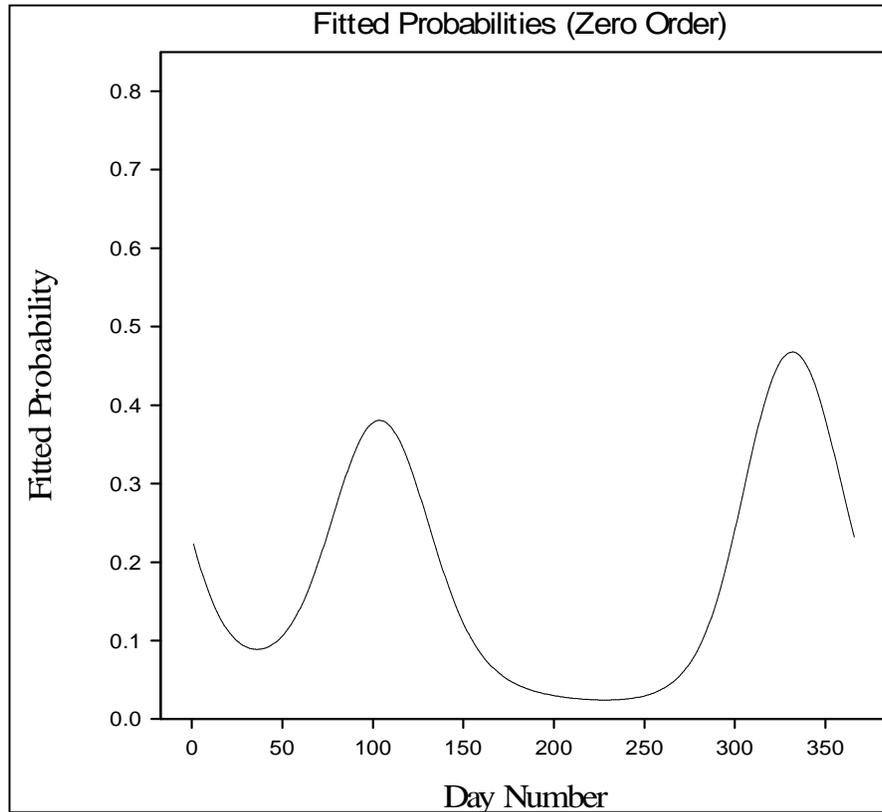
Table 5: Accumulated analysis of deviance of fitted probability of rain for 0-order with 2 harmonics

In table 3 above, shows that adding the harmonic of order 2 is significant. When harmonic of order 3 is added the results below is obtained.

Change	d.f.	deviance	mean deviance	deviance ratio	approx chi pr
+ half_hamonic[1] Sin[1]	1	93.386	93.386	93.39	<.001
+ half_hamonic[2] cos[1]	1	535.478	535.478	535.48	<.001
+ half_hamonic[3] Sin[2]	1	1169.576	1169.576	1169.58	<.001
+ half_hamonic[4] cos[2]	1	119.364	119.364	119.36	<.001
+ half_hamonic[5] Sin[3]	1	131.654	131.654	131.65	<.001
+ half_hamonic[6] cos[3]	1	1.028	1.028	1.03	0.311
Residual	359	409.829	1.142		
Total	365	2460.315	6.741		

Table 6: Accumulated analysis of deviance of fitted probability of rain for 0-order with 3 harmonics

Fitting the probability of order zero with harmonics of order three is not significant since $\cos[3] > 0.05$. Therefore harmonic of order two is applied and the fitted plot is now drawn and shown in graph 10.



Graph 12: *Fitted Probabilities of order zero (GenStat)*

6.5.2: Order one

A two-state Markov chain of order one is modeled for the chance of a state in a day given the state condition in the previous day. The dialogue box for counts is filled with the order option equal to one with the rest of the fields the same as in order zero discussed above. The results for counts and amounts for order 1 are displayed in a spreadsheet as shown in Figure 48. (Note: 0,1,2,3 used as column header of 'counts' table corresponds to the states *dd*, *dw*, *wd*, *ww* respectively)

Row	DayOfYear	0	1	2	3
1	1	27	7	2	4
2	2	29	4	6	2
3	3	29	2	4	6

Row	DayOfYear	wd	ww
1	1	1.8	17.4
2	2	88.1	28.6
3	3	20.3	80.7

Figure 48: *First order counts and rainfall amounts*

The number of times it was dry on 2nd January given that on 1st January was also dry are 29 out of 41. The value 29 is below *dd* on day 2. On the same row under *wd*, the 6 means

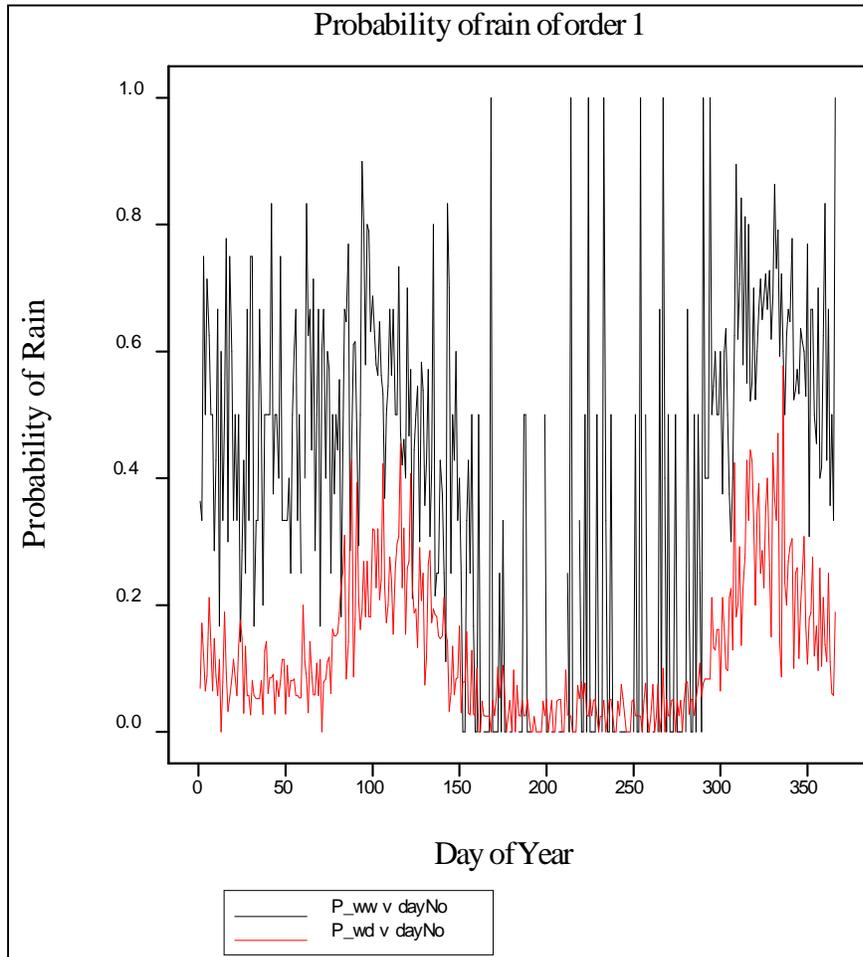
that the number of times it was wet on 2nd January given that 1st January was also dry are 6 out of 41. The amount table the columns are only two (*wd* and *ww*) which contain the amounts of rain given the previous state. On days with no rain in the present state their amounts are not calculated. i.e. on day 13th January, there was no day with rain given that the previous day was dry hence the corresponding amount is not calculated.

The probabilities of rain given the state in the previous day calculated by PREPARE procedure is shown in **Figure 49** and the corresponding plot for the probabilities (Graph 13)

Row	dayNo	P_wd	P_ww
1	1	0.0689655	0.363636
2	2	0.171429	0.333333
3	3	0.121212	0.75
4	4	0.0645161	0.5
5	5	0.0882353	0.714286
6	6	0.212121	0.625
7	7	0.137931	0.5
8	8	0.0645161	0.5
9	9	0.147059	0.285714
10	10	0.0882353	0.428571
11	11	0.0571429	0.666667
12	12	0.114286	0.166667
13	13	0	0.6
14	14	0.0789474	0.333333

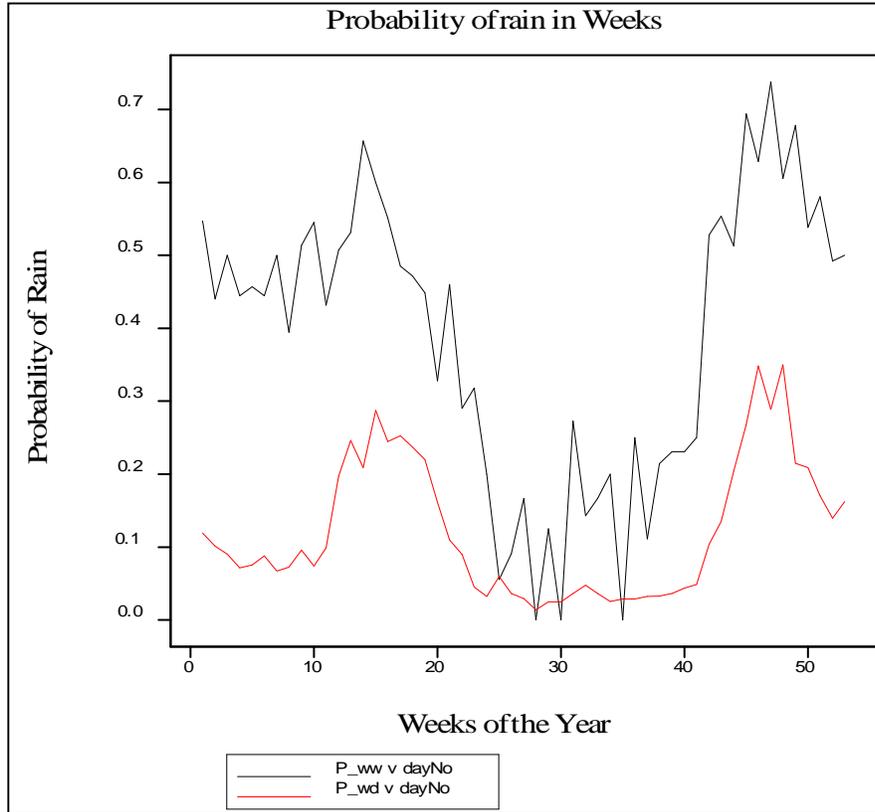
Figure 49: Chance of rain given the state in the previous day

The values in the table are probabilities of rain given the state in the previous day. For example, the probability of rain on the 1st January given that it was dry in the previous day is 0.06897

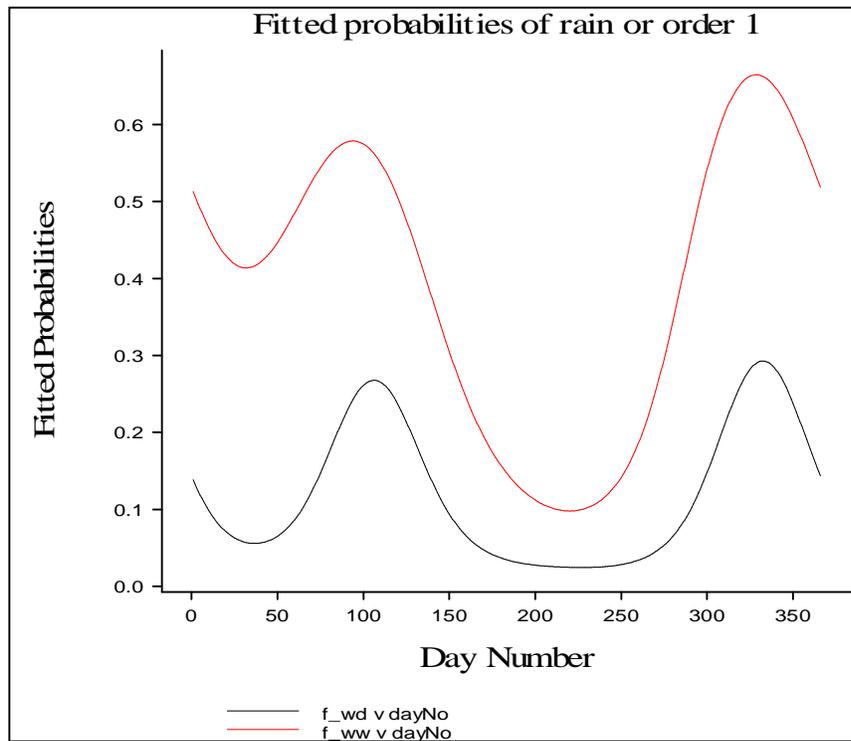


Graph 13: *Probabilities of rain for order one (GenStat)*

When the probabilities are plotted for each day, it is overcrowded and understanding the distinction of the curves can be hard. As before there are solutions for this to give better curves by first grouping the days into 5-day totals and then plot the results (Graph 14), or 10 day, weekly and monthly totals or get the fitted probabilities (Graph 15).



Graph 14: Probabilities of rain for order summarized on weekly basis



Graph 15: Fitted Probabilities of order one (GenStat)

6.6: Higher order two- state Markov Chain Modeling

The higher order two-state Markov Chain in this context would mean modeling Markov chain of order two and above with reduced number of parameters. We will discuss the results for order 2, 3, 4 and 5.

6.6.1: Order Two Normal Markov Model

The Markov chain of order 2 can be modeled for high or normal types of Markov chain. The normal order Markov is obtained when the count dialogue box is filled as shown below.

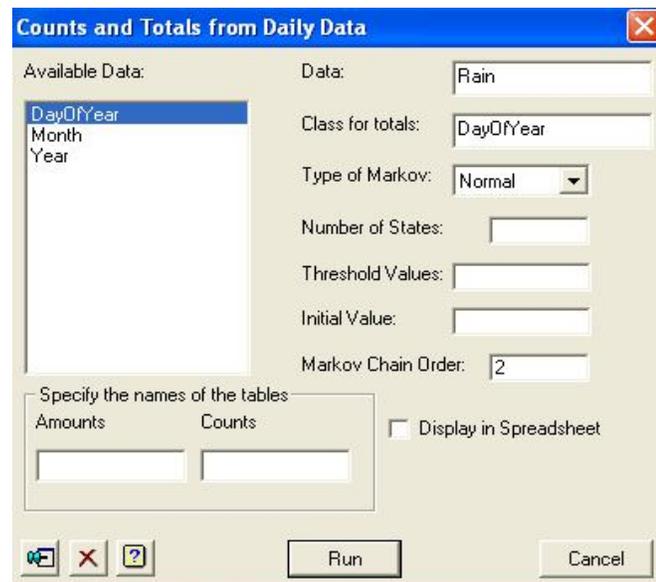
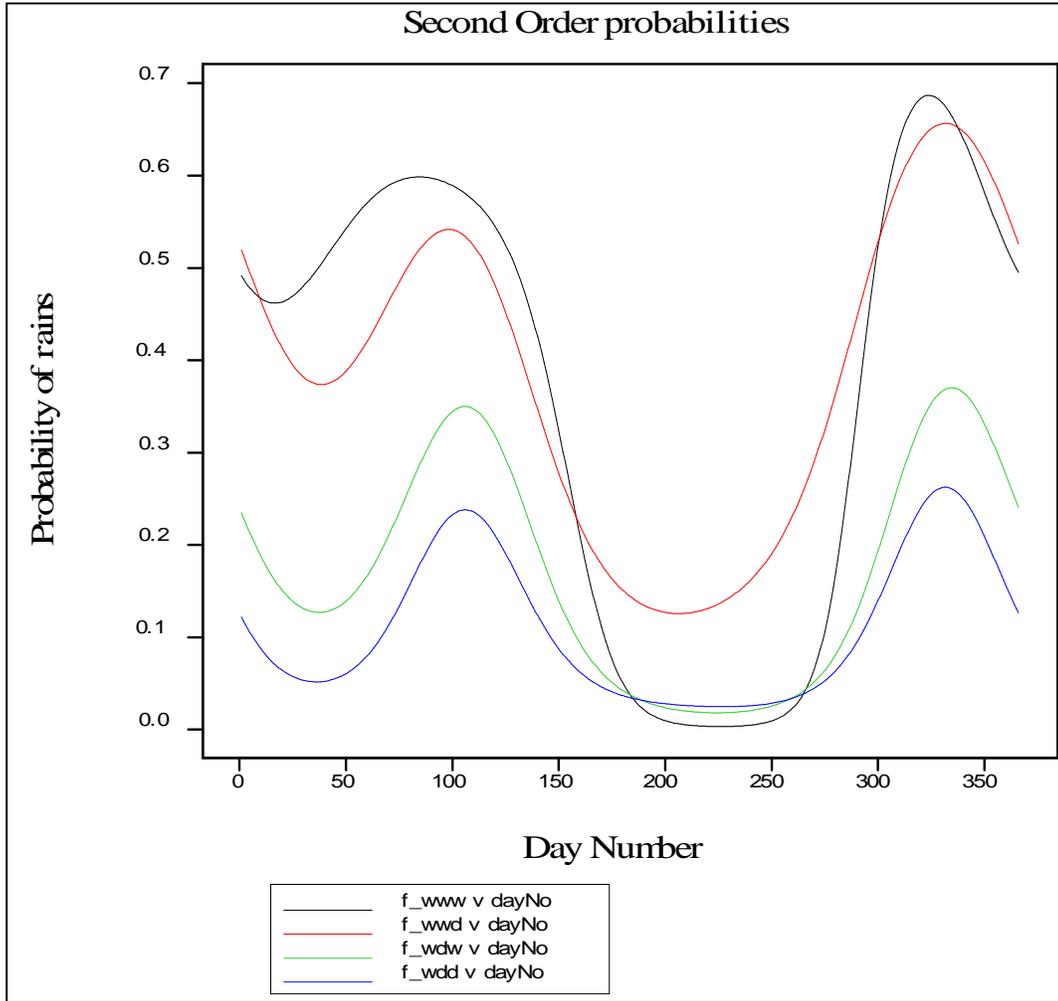


Figure 50: Dialogue box for Count (Normal Markov Chain of Order Two)

The fitted probabilities for order two are plotted in Graph 16. There is a clear distinction between the curves when it is dry on the previous day and when it is wet in the previous day during the rainy seasons. This is not the case in the dry seasons where there is no clear distinction between the curves.



Graph 16: *Fitted Probabilities of rain for order two (Normal Markov)*

6.6.2: Order Two High Markov Model

Considering a high Markov chain model for order 2, the www and wwd curves are combined to ww . The dialogue box for counts used to obtain a high Markov chain of order 2 if filled as shown in Figure 51

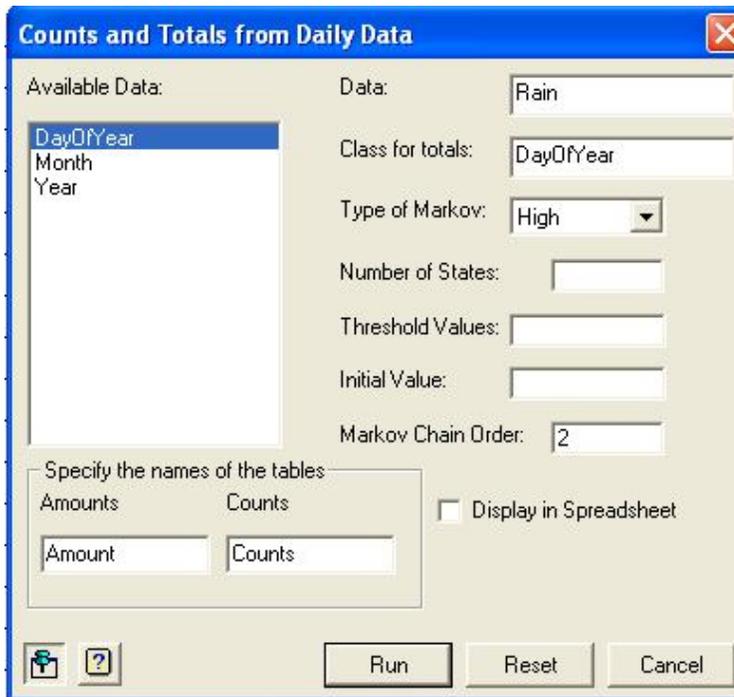


Figure 51: Dialogue box for Count (High orderMarkov Chain of Order Two)

The results for counts are displayed in a table format called count while the amounts table is displayed in the amounts table.

Row	DayOfYear	-3	-2	-1	1	2	3
1	1	2	0	4	7	0	27
2	2	5	1	2	4	6	22
3	3	3	1	6	2	3	26
4	4	2	0	5	5	2	27
5	5	2	1	5	2	4	27
6	6	7	0	5	3	2	24
7	7	2	2	6	6	1	24
8	8	1	1	5	5	5	24
9	9	2	3	2	5	2	27

Row	DayOfYear	>1dw	_wdw	_ww
1	1	4.8	*	17.4
2	2	56.9	31.1	28.6
3	3	19.9	0.4	80.7
4	4	11.0	*	67.8
5	5	13.8	11.5	61.8
6	6	45.8	*	55.1
7	7	26.6	9.9	37.2
8	8	51.9	3.4	59.6
9	9	9.5	76.6	54.3

Figure 52: Counts and amounts for 2nd-high order Markov chain(table structure)

The corresponding variate for table is shown in the figure below. The variate table is obtained when you choose 'Spreadsheet' option in the *Probability of rain dialogue* box. The column headers for counts correspond to the column headers in Figure 53.

Row	dayNo	w2d	wdw	ww	dw	_2dw	_3d
1	1	2	0	4	7	0	27
2	2	5	1	2	4	6	22
3	3	3	1	6	2	3	26

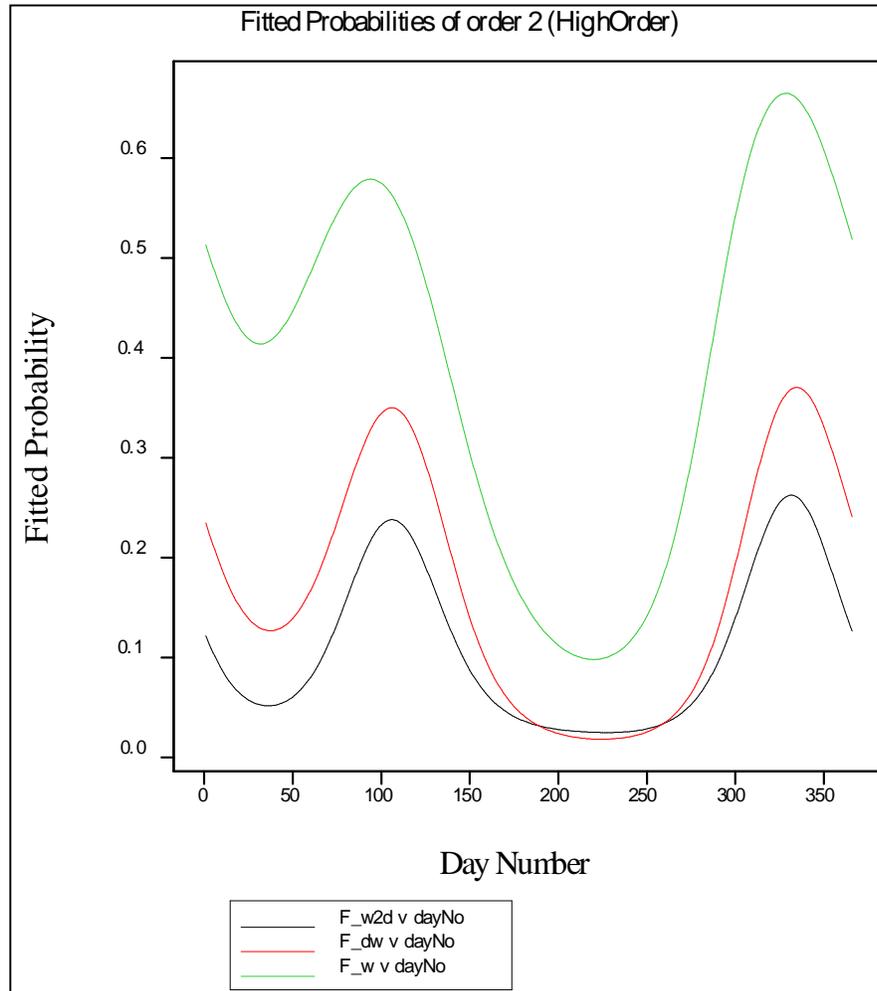
Figure 53: Counts for 2nd-high order Markov chain (variate structure)

The probability of rain for each day calculated from the counts is also displayed in the spreadsheet when the Probability of rain dialogue box is executed (see fig 51).

Row	dayNo	P_w2d	P_wdw	P_ww
1	1	0.0689655	*	0.363636
2	2	0.185185	0.142857	0.333333
3	3	0.103448	0.25	0.75
4	4	0.0689655	0	0.5

Figure 54: Probability of rain for 2nd-high order Markov chain.

The probability of rain in January 1st given that it was dry in the last two days is 0.06896, while the probability that it rained on January 1st given that it rained in the previous day is 0.3636. The fitted probability model is shown in the graph

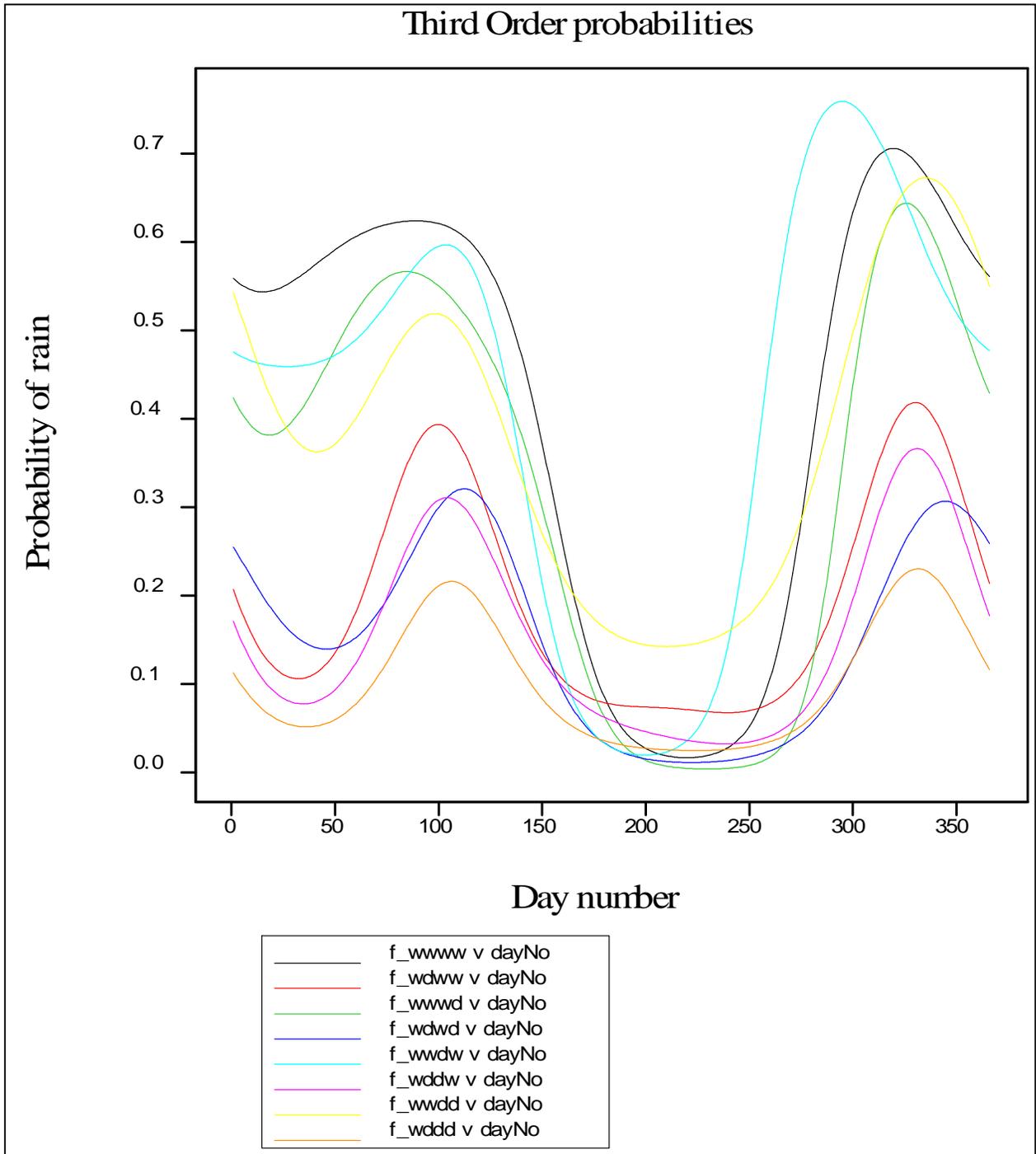


Graph 17: *Fitted Probabilities of order two-High Order (GenStat)*

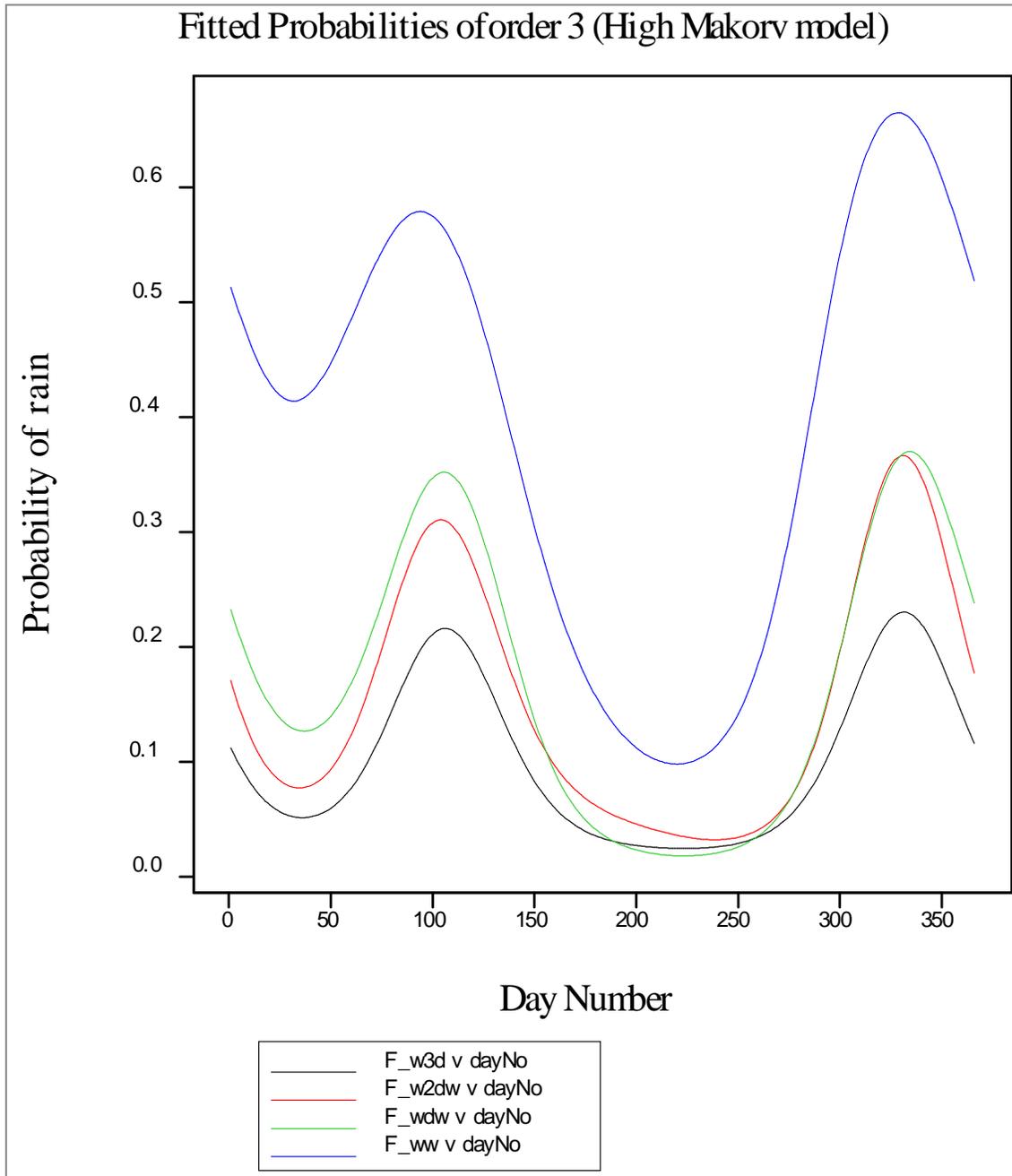
When considering the wet in the previous day only regardless of what happened the day before, then a clear distinction between the curves when it was dry in the previous day and the curves when it was wet in the previous day during dry seasons.

6.6.3: Normal and High Markov models for order three and four

The fitted normal and high Markov chain model for order three are shown in the Graph 18. The fitted model indicates that there is no distinction between the groups of curves when the previous day was dry. The same information is conveyed when the previous day was dry

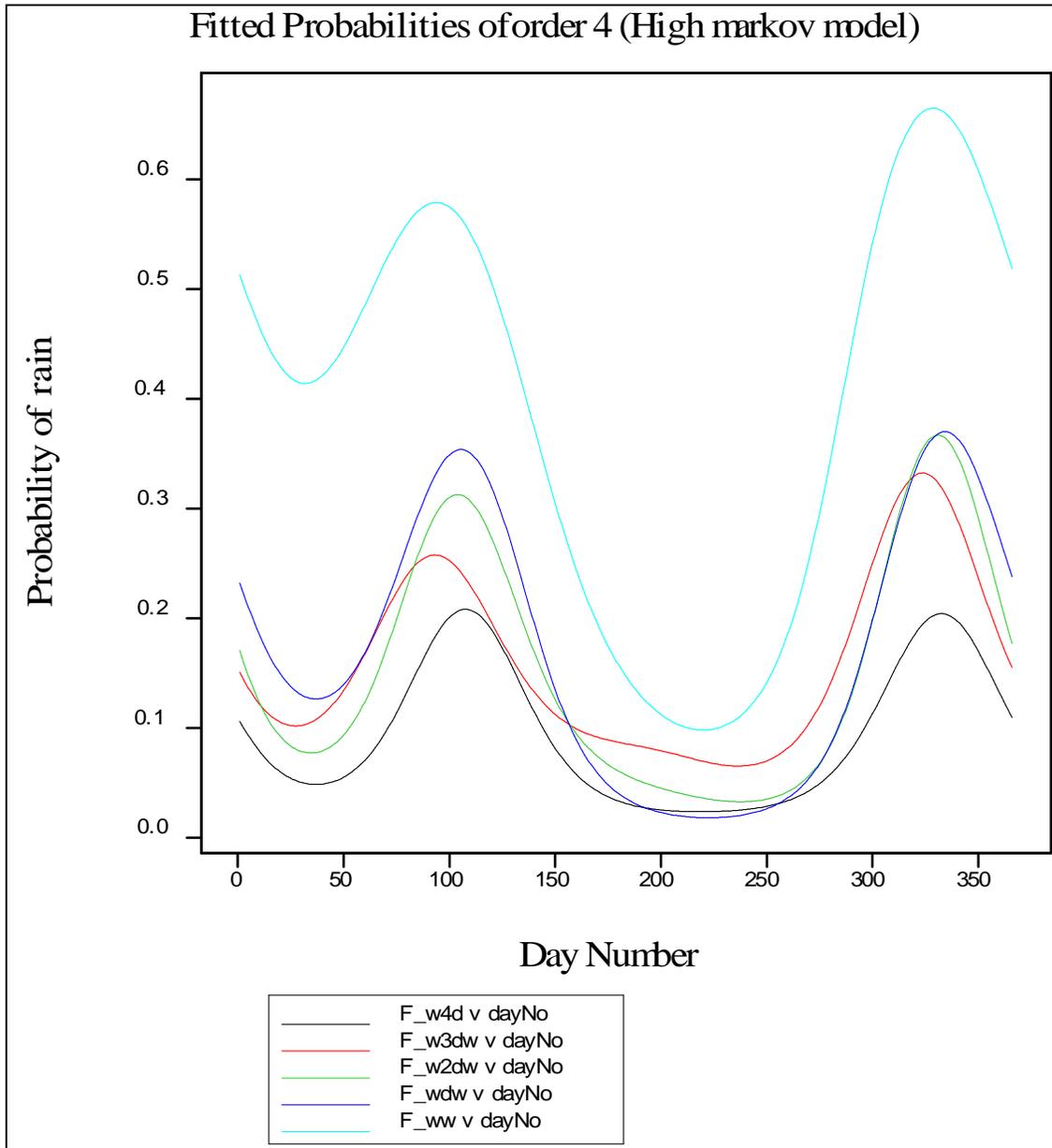


Graph 18: *Fitted Probabilities of normal Markov chain of order three.*



Graph 19: *Fitted Probabilities of high Markov chain of order three.*

All the curves in Graph 19 are distinct to each other during the rainy seasons. Using higher orders breaks the distinctions as shown for order 4 (Graph 20) and order 5 (Graph 21)



Graph 20: *Fitted Probabilities of high Markov chain of order four.*

6.6.4: Order five

Order five for a normal Markov will result in so many parameters (64) which might not be significant in Markov modeling. A high order Markov will reduce this number to 10 parameters only. The result obtained for Katumani data is as shown below.

Row	dayNo	w5d	w4d	w3d	w2d	wdw	ww	dw	_2dw	_3dw	_4dw	_5dw	_6d
1	1	2	0	0	0	0	4	7	0	3	4	7	13
2	2	3	2	0	0	1	2	4	6	0	3	2	17
3	3	1	1	0	1	1	6	2	3	5	0	2	18
4	4	2	0	0	0	0	5	5	2	3	5	0	18
5	5	1	0	1	0	1	5	2	4	2	2	5	17
6	6	5	0	1	1	0	5	3	2	3	1	2	18
7	7	1	0	1	0	2	6	6	1	2	2	1	19

Figure 55: Counts of rainy days for High Markov Chain of order 5

Row	DayOfYear	>4dw	_w4dw	_w3dw	_w2dw	_wdw	_ww
1	1	4.8	*	*	*	*	17.4
2	2	41.4	15.5	*	*	31.1	28.6
3	3	6.6	0.7	*	12.7	0.4	80.7
4	4	11.0	*	*	*	*	67.8
5	5	6.5	*	7.2	*	11.5	61.8

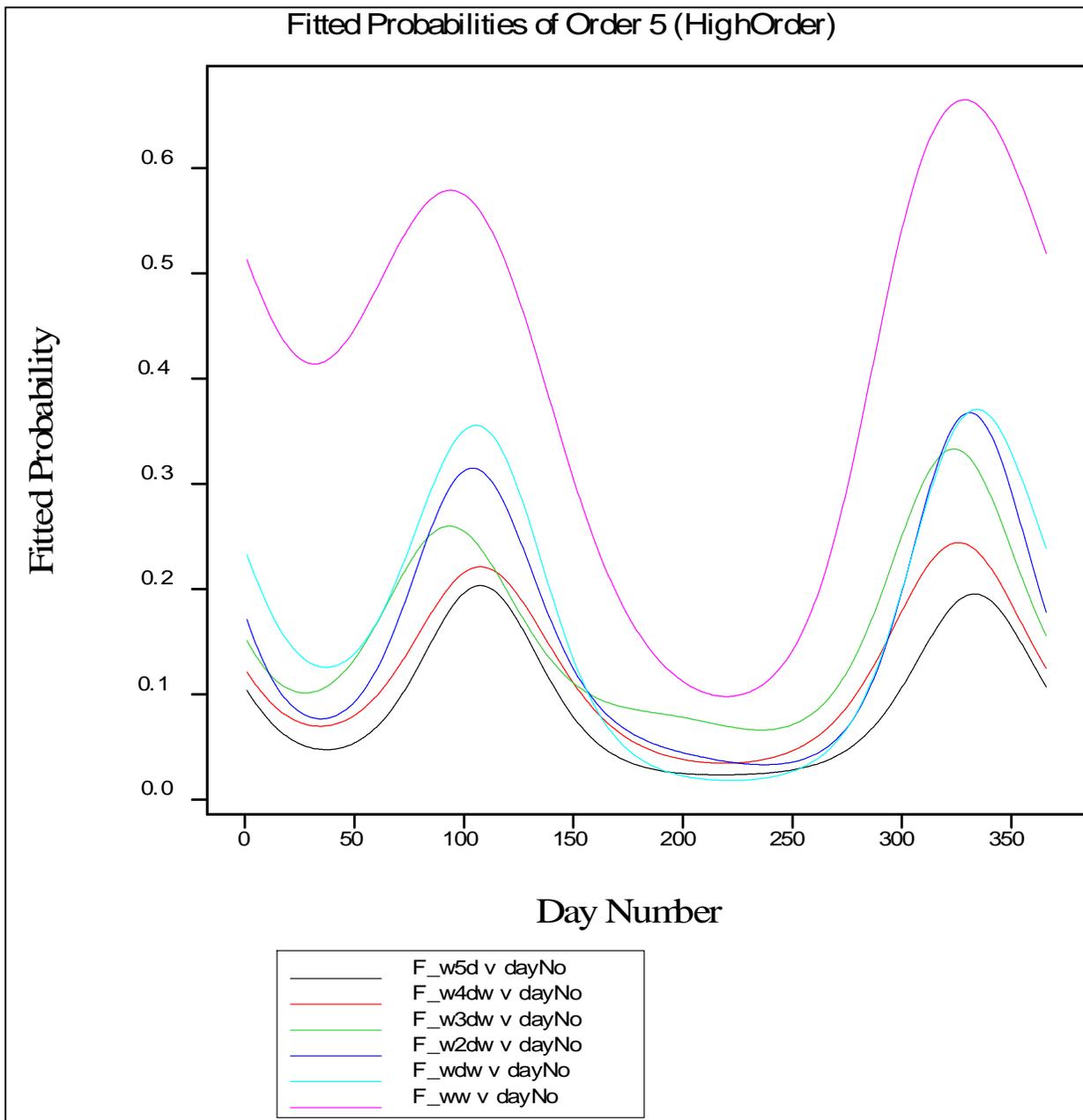
Figure 56: Amount of rainfall for High Markov Chain of order 5

The number of times that January 1st received rainfall given that the 5 previous days were dry was 2 out of 40 with amount being 4.8mm. The probability

Row	dayNo	P_w5d	P_wd4w	P_wd3w	P_wd2w	P_wdw	P_ww
1	1	0.133333	0	0	0	*	0.363636
2	2	0.15	0.5	0	*	0.142857	0.333333
3	3	0.0526316	0.333333	*	0.166667	0.25	0.75

Figure 57: Probability of rain for High Markov Chain of order 5

The probability of rain given five previous days dry is 0.1333. The fitted probability



Graph 21: *Fitted Probabilities of high Markov chain of order five.*

In graph 19, the chance of rain given rain is distinct and high from the other models while the probability of rain given longer dry spell is very low compared to other chains. The fitted model for rain given that the previous day being dry, has a lot of interactions during dry seasons implying that there is less variations in the number of wet days given dry and less interaction during the wet seasons implying that there is high variation the number of wet days given dry in the previous day. In a nutshell, during the wet season, there seems to be a

clear difference between the numbers of wet days in a month depending on whether it was wet or dry the previous day.

6.7: Markov Chain with Initial data

When the initial data is missing, the first day count will be missing if the model requires the condition of the previous state. The initial values are set to give a memory of the previous day before the first day the data starts. The number of values of initial must be equal to the Markov order. A two-state Markov of order one and two is used to demonstrate this.

6.7.1: Order one

In a 1st order Markov chain, only one initial value is needed. The dialogue box for count is filled as shown below.

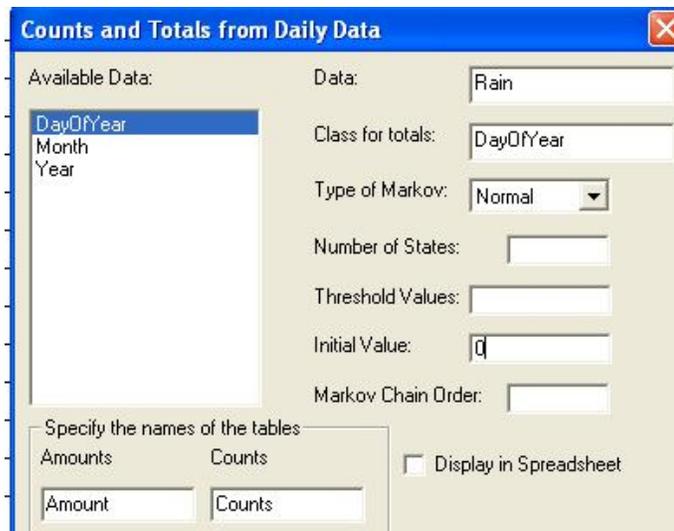


Figure 58: Counts dialogue box with initial value set of a 1st order Markov chain

There is a difference in result on 1st January for a first order Markov chain with ‘initial value’ set and ‘initial value’ not set. The total counts is more (41) in Figure 59 when initial value is set than the total counts (40) in figure 46 when the initial is not set.

Row	DayOfYear	0	1	2	3
1	1	28	7	2	4
2	2	29	4	6	2

Figure 59: Counts for Two-state 1st order Markov chain with initial

6.7.2: Order Two

When a Markov chain is of order two it will require two initial values. When the initial values are more than one, they are set in a variate structure. i.e. instead of 0,0 we use !(0,0) . The dialogue box for this is filled as shown in figure 56 below.

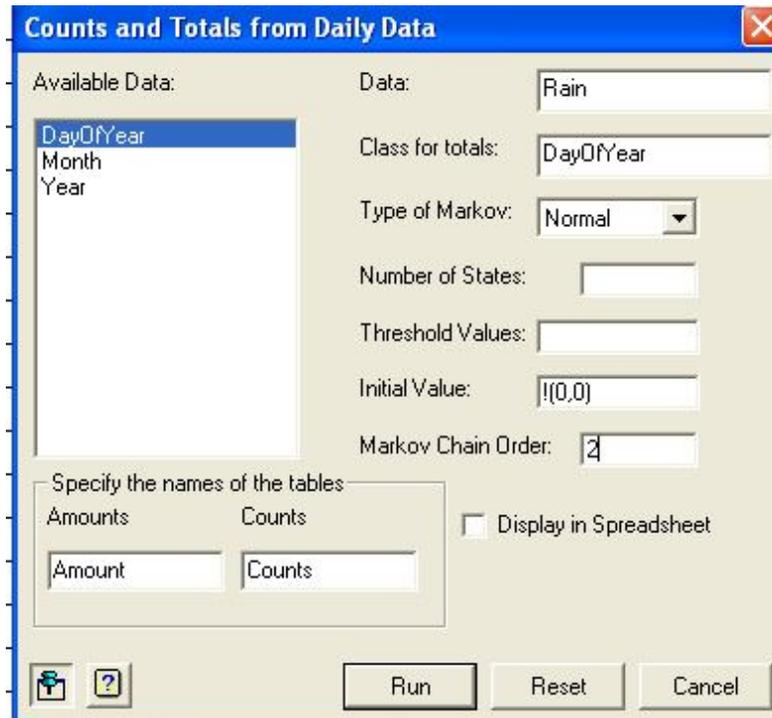


Figure 60: Counts dialogue box with initial value set of a 2nd order Markov chain

The results for this dialogue are shown in Figure 61

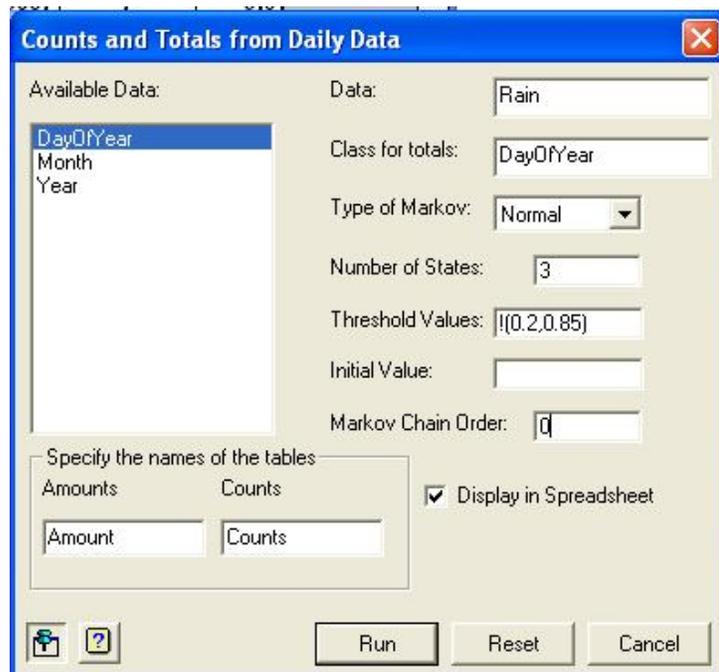
Row	DayOfYear	0	1	2	3	4	5	6	7
1	1	28	0	6	1	2	0	1	3
2	2	23	6	1	3	5	1	1	1
3	3	26	3	1	1	3	1	5	1
4	4	27	2	2	3	2	0	2	3
5	5	27	4	1	1	2	1	1	4
6	6	24	2	1	2	7	0	2	3

Row	DayOfYear	wld	wdw	wwd	www
1	1	2.4	*	7.7	3.2
2	2	11.4	31.1	3.2	25.3
3	3	6.7	0.4	14.2	9.4
4	4	5.5	*	14.9	12.6
5	5	6.9	11.5	16.5	11.3
6	6	6.5	*	11.7	10.6

Figure 61: Counts and amount table of 2nd order Markov chain with initial given

6.8: Three-state Markov chain Modeling

In certain occasions one may want to group the rain fall data into three states i.e. rain, trace and wet or even more number of states. The model has the capability of modeling Markov chains with more than two states. A three state Markov model will have two threshold values which determine the boundaries of each state. This section describes a zero order and a one order state Markov model. To obtain a zero order Markov model the dialogue box for counts can be filled as shown in the figure below.



The image shows a software dialog box titled "Counts and Totals from Daily Data". It contains the following fields and controls:

- Available Data:** A list box containing "DayOfYear", "Month", and "Year". "DayOfYear" is selected.
- Data:** A text box containing "Rain".
- Class for totals:** A text box containing "DayOfYear".
- Type of Markov:** A dropdown menu set to "Normal".
- Number of States:** A text box containing "3".
- Threshold Values:** A text box containing "(0.2,0.85)".
- Initial Value:** An empty text box.
- Markov Chain Order:** A text box containing "0".
- Specify the names of the tables:** A section with two columns: "Amounts" and "Counts". Below "Amounts" is a text box containing "Amount". Below "Counts" is a text box containing "Counts".
- Display in Spreadsheet:** A checked checkbox.
- Buttons:** "Run", "Reset", and "Cancel" buttons are located at the bottom right. There are also help icons (a question mark and a document icon) at the bottom left.

Figure 62: Counts dialogue zero order 3-state Markov chain

6.9: Modeling Rainfall Amount

When fitting rainfall amount, The *User=>Fitting amount* is used, and the dialogue box for fitting amounts is as shown in the figure below.

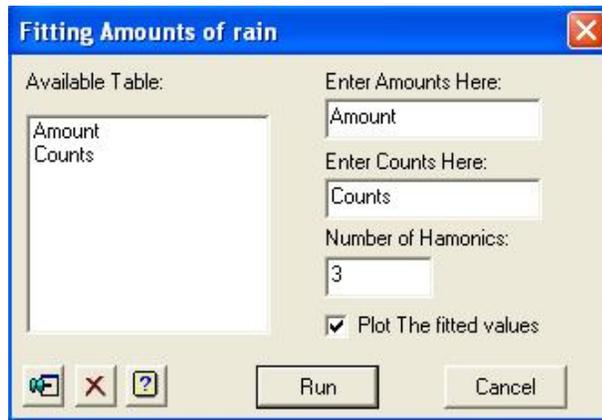
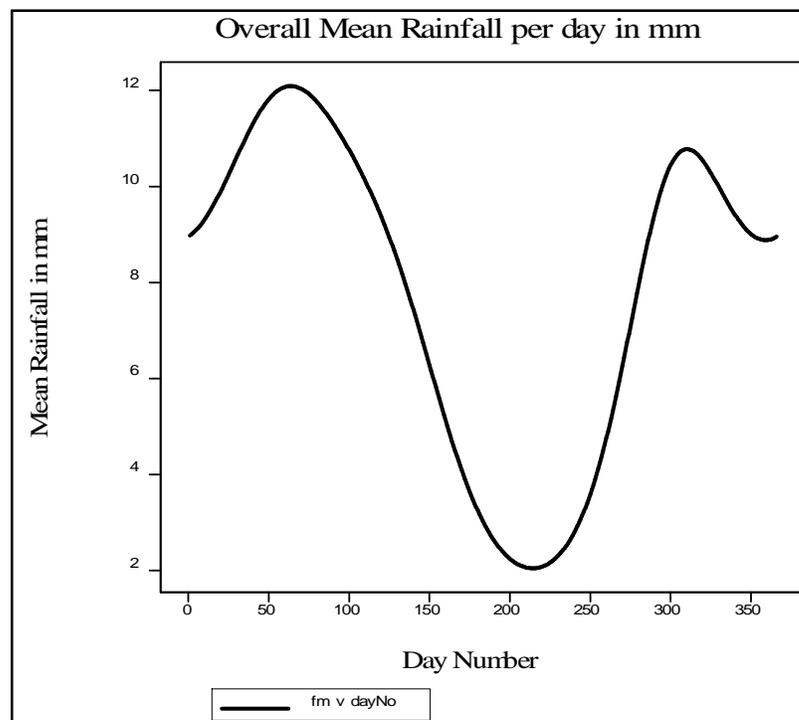


Figure 63: Dialogue box for fitting Amount

6.9.1: Modeling rainfall Amount of order zero

The model for rainfall amount of order zero is fitted in graph shown in graph 20

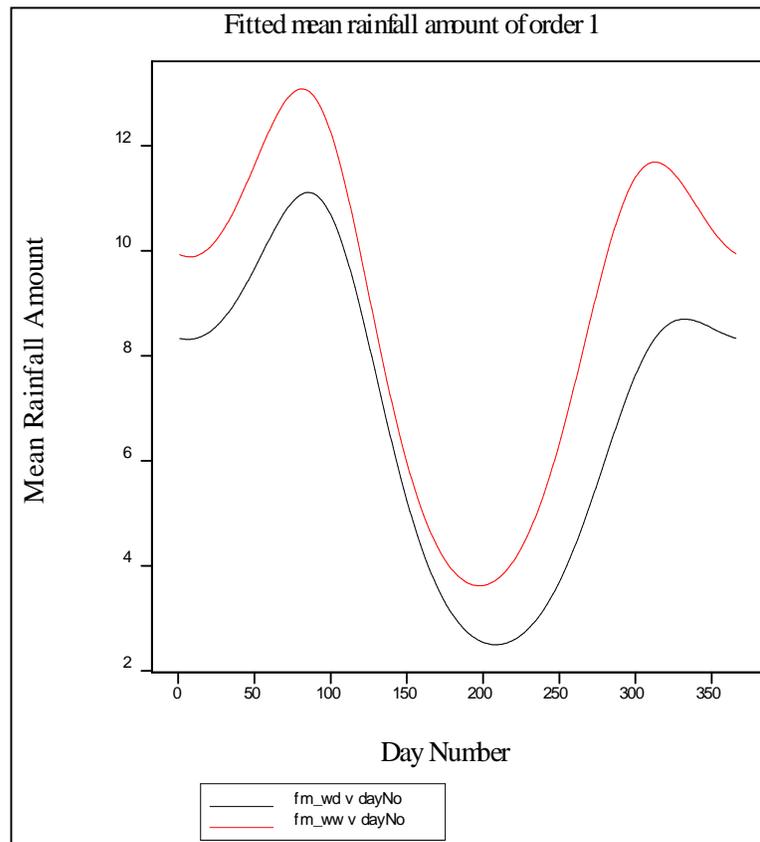


Graph 22: Fitted Mean rainfall of order zero (GenStat)

The result in Graph 22 indicates that the overall mean rainfall in the long season is higher than in the short season. The mean annual rainfall, result (Graph 26) is realistic based on the conditions that Katumani experienced during the period when the data was collected.

6.9.2: Modeling rainfall Amount of order one

When the Markov model or order one for rainfall amount is fitted starting with harmonic of order 2 followed by harmonic of order 3, the result is shown in the graph below.



Graph 23: Fitted Mean rainfall of order one with harmonic of order 2

Estimates of parameters

Response variate: w_d

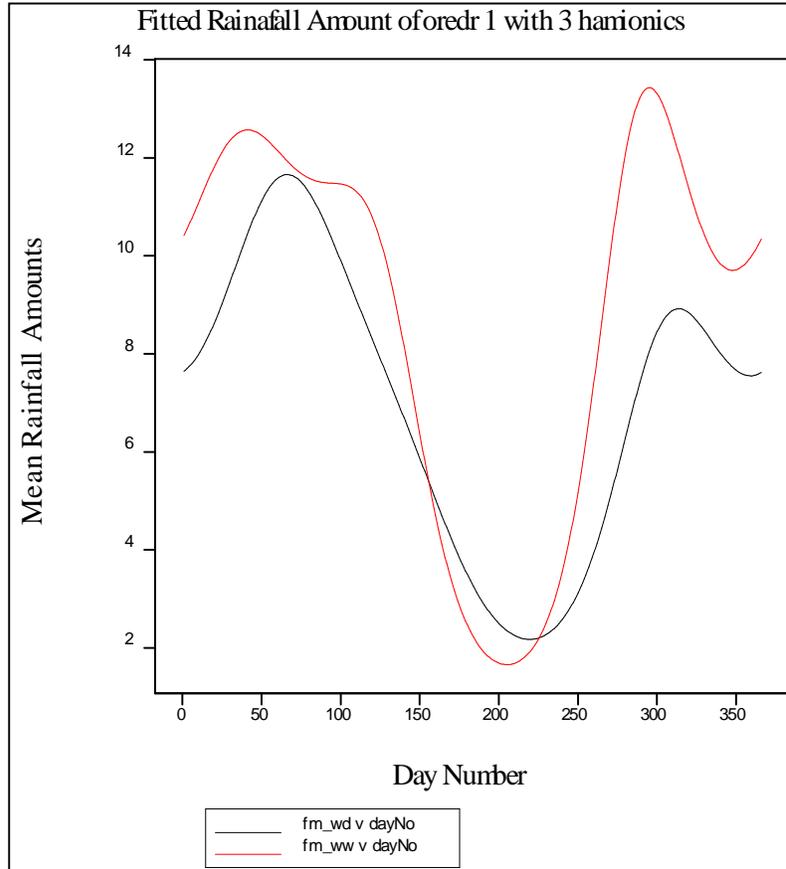
Parameter		estimate	s.e.	t(325)	t pr.
Constant		1.8291	0.0472	38.79	<.001
half_hamonic[1]	Sin[1]	0.3452	0.0601	5.74	<.001
half_hamonic[2]	Cos[1]	0.5191	0.0662	7.85	<.001
half_hamonic[3]	Sin[2]	-0.1965	0.0601	-3.27	0.001
half_hamonic[4]	Cos[2]	-0.2273	0.0608	-3.74	<.001

Table 7: Estimate of parameter for rainfall amount model (w given d) with 2 harmonics

Response variate: w_w

Parameter		estimate	s.e.	t(256)	t pr.
Constant		2.1022	0.0805	26.13	<.001
half_hamonic[1]	Sin[1]	0.1678	0.0818	2.05	0.041
half_hamonic[2]	Cos[1]	0.478	0.105	4.54	<.001
half_hamonic[3]	Sin[2]	-0.1301	0.0787	-1.65	0.099
half_hamonic[4]	Cos[2]	-0.2827	0.0674	-4.19	<.001

Table 8: Estimate of parameter for rainfall amount model (w given w) with 2 harmonics



Graph 24: Fitted Mean rainfall of order one with harmonic of order 3

Response variate: w_d

Parameter		estimate	s.e.	t(323)	t pr.	estimate
Constant		1.8267	0.0469	38.91	<.001	6.214
half_hamonic[1]	Sin[1]	0.3939	0.0639	6.17	<.001	1.483
half_hamonic[2]	Cos[1]	0.5287	0.0689	7.67	<.001	1.697
half_hamonic[3]	Sin[2]	-0.2301	0.0615	-3.74	<.001	0.7944
half_hamonic[4]	Cos[2]	-0.2121	0.0648	-3.27	0.001	0.8089
half_hamonic[5]	Sin[3]	0.0748	0.0597	1.25	0.211	1.078
half_hamonic[6]	Cos[3]	-0.1121	0.0609	-1.84	0.067	0.8940

Table 9: Estimate of parameter for rainfall amount model (w given d) with 3 harmonics

Response variate: w_w

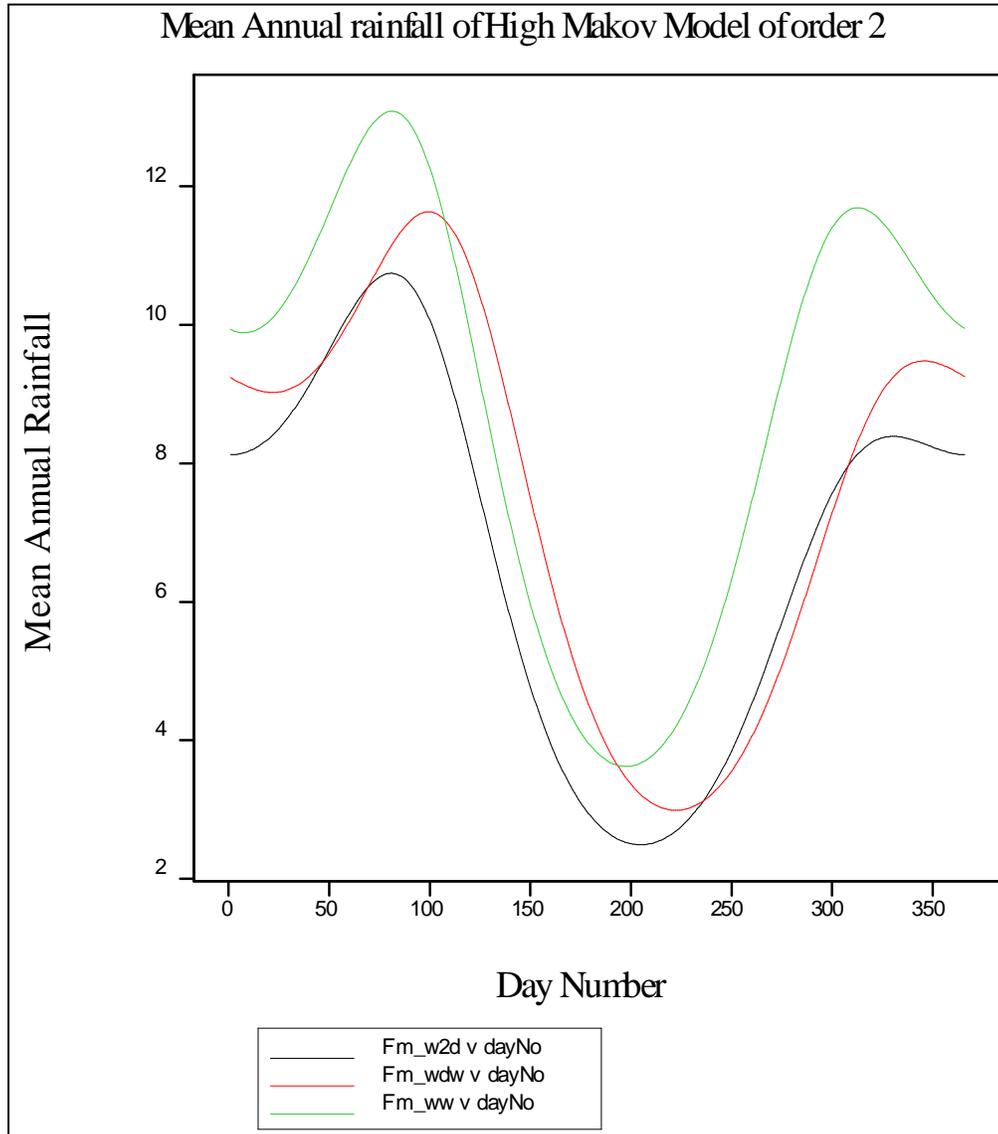
Parameter		estimate	s.e.	t(254)	t pr.	estimate
Constant		1.9994	0.0842	23.75	<.001	7.385
half_hamonic[1]	Sin[1]	0.254	0.102	2.48	0.014	1.289
half_hamonic[2]	Cos[1]	0.717	0.129	5.58	<.001	2.049
half_hamonic[3]	Sin[2]	-0.2933	0.0942	-3.11	0.002	0.7458
half_hamonic[4]	Cos[2]	-0.4191	0.0946	-4.43	<.001	0.6576
half_hamonic[5]	Sin[3]	0.2311	0.0700	3.30	0.001	1.260
half_hamonic[6]	Cos[3]	0.0390	0.0757	0.51	0.607	1.040

Table 10: Estimate of parameter for rainfall amount model (w given w) with 3 harmonics

Table 7 and Table 8 indicates that the harmonic or order 2 is significant while of order 3 is not significant (Table 9 and Table 10)

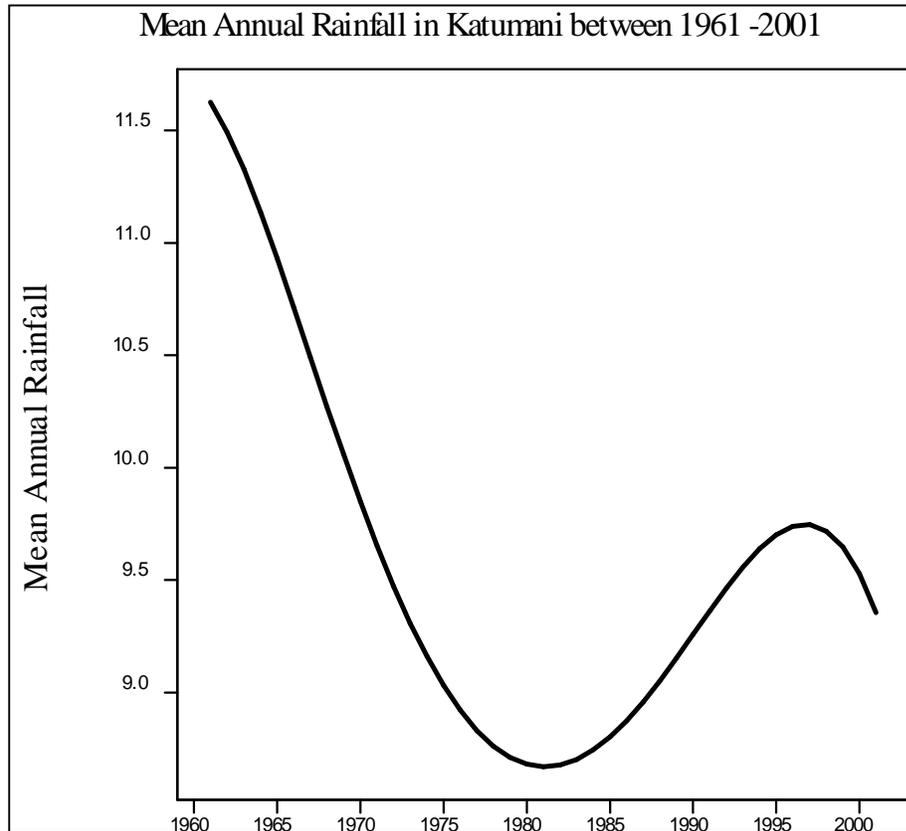
6.9.3: Modeling High Markov model of order two for Rainfall Amount

Graph 23 shows fitted mean rainfall amount of high Markov chain model of order 2



Graph 25: *Fitted Mean rainfall of High Markov of order two*

Order two normal Markov fits better than order highs for Katumani data.



Graph 26: *Fitted Mean rainfall of order zero (GenStat)*

During 1961-1963 Kenya received the highest rainfall called *uhuru*^Y. The lowest rainfall was received in 1980 -1983 when the greatest drought was experienced. The second peak was during El-Niño rains in 1998.

^Y Uhuru means Independence (The heavy rains were experience during the period of Kenyan independence.)

Chapter 7: Conclusion and Further Work

A comprehensive analysis of rainfall data is a fundamental component in the planning phase of agricultural research since rainfall is the key element in the tropical regions amongst other climatic constrains. The step taken in developing the climatic routine is therefore meant to help users make full use of their rainfall data. It will support agricultural research and many other fields that need an analysis of rainfall data as part of their work.

The successful implementation of the four menus encourages the potential statistical programmers that it is possible to have a full menu for climatic data analysis. This study has just jumpstarted the process by tackling one of the most challenging statistical techniques, “*Markov Chain Modeling*”. I strongly believe it is possible to implement the rest of the work involved in climatic analysis into GenStat. The challenges met and knowledge gained in this work highly motivates me to achieve the dream of full climatic menu as further work beyond this study

The analyses in GenStat are now more flexible compared to what can be achieved by the current version of InStat. Using the menus in GenStat, it is now possible to model a three-state Markov chain.

The area of immediate address include modeling climatic events, crop performance index analysis, summaries of climatic data, time series analysis, and temperature analysis etc. For a full utilization of the package in handling climatic data, it is important to look forward in implementing all these aspects of climatic analysis.

References

- Avik, G., Deepanwita, G., & Dashgupta, \. (2010). Higher order Markov Chain Model for monsoon rainfall over West Bengal, India. *Indian Journal of Radio and Space Physics* , 39, 39-44.
- Buchdahl, J. (1999). Global Climate Change Student Guide: A review of contemporary and prehistoric global climate change. Atmosphere, Climate & Environment, Information Programme, aric. *Manchester Metropolitan University*, (p. 99). Manchester.
- Chandler, R. E., & Wheater, H. S. (2002). Analysis of rainfall variability using generalized linear models: A case study from the west of Ireland. *Water Resources Research* . , 38, 1192.
- Chin, E. H. (1977). Modeling Daily Precipitation Occurrence Process with Markov Chain. *Water Resources Research* , 13(6), 949-956.
- Didkovsky, N. (1996, Feb 16). Retrieved May 13th, 2010, from <http://www.doctornerve.org/nerve/pages/interact/markhelp.htm>
- Dumont, A. G., & Boyce, D. S. (1974). The probabilistic simulation of weather variables. *Journal of Agricultural Engineering Research* , 19, 131-145.
- Dunn, P. K. (1999, July 12). A Graphical User Interface to Generalized Linear Models in matlab.
- Gabriel, K. R., & Neumann, J. (1961). A Markov Chain Model for Daily Rainfall Occurrence at Tel Aviv. *Quarterly Journal of Royal Meteorology Society* , 88, 90-95.
- Gallagher, J., & Stern, R. (2004). *Analyzing Climatic Data Using Genstat for Windows*. University of Reading.
- Gallagher, J., & Stern, R. (2009). *Analyzing Climatic Data Using Genstat for Windows*. University of Reading.
- Gates, P., & Tong, H. (1976). Markov Chain Modeling to Weather Data. *Journal of Applied Meteorology* , 15 , 1145-51.
- Goedhart, P. W. (1991). Vtable. In R. W. Payne, & G. M. Arnold, *GenStat Release 7.1 Reference Manual* (p. 464). VSN International.
- Haan, C. T., Allen, D. M., & Street, J. O. (1976). A Markov Chain Model of Daily Rainfall. *Water Resources Research* , 12, 443-449.

- Hill, S. J., & Mjelde, J. W. (2002). Challenges and Opportunities Provided by Seasonal Climate Forecasts. *Journal of Agricultural and Applied Economics* , 34(3), 603-632.
- Jones, P. G., & Thornton, P. K. (1999). Fitting a third-order Markov rainfall model to interpolated. *Agricultural and Forest Meteorology* , 97, 213-231.
- Khamsi, M. (1999). *Markov Chain*. Retrieved May 23, 2010, from S.O.S. Mathematics CyberBoard. : <http://www.sosmath.com/matrix/Markov/Markov.html>
- Kottegoda, N. T., Natale, L., & Raiteri, E. (2004). Some considerations of periodicity and persistence in daily rainfalls. *J Hydrology* , 296, 23–37.
- Longhai, L., & Radford, M. (2008). Compressing Parameters in Bayesian High-order Models with Application to Logistic Sequence Models. *Bayesian Analysis* , 3, 793-82.
- Mimikou, M. (1983). Daily precipitation occurrences modelling with Markov chain of seasonal order. *Hydrological Sciences* , 28, 2-6.
- Nelder, J., & Wedderburn, R. (1972). Generalized Linear Model. *Roy Statistics Society* , A135, 370-384.
- Payne, R., Murray, D., Harding, S., Baird, D. S., Glaser, A., I.C., C., et al. (2009). *Genstat Release 12 Reference Manual, Part 3 Procedure Library*. (R. Payne, Ed.) Hemel Hempstead, Oxford, UK: VSN International.
- Perera, H., Sonnadara, D., & Jayewardene, D. (2002). Forecasting the Occurrence of Rainfall in Selected Weather Stations in the Wet and Dry Zones of Sri Lanka. *Sri Lankan Journal of Physics* , 3, 39-52.
- Samuel, E. S. (2009). Construction Environmental Monitoring of Transport Projects and its Relevance to Climate Change. *17th Annual Conference of the Transportation Science Society of the Philippines*, (pp. 32-38).
- Siriwardena, L., Srianthan, R., & McMahan, T. A. (2002, December 14). *Evaluation of two daily rainfall data generation models*, . Retrieved from www.toolkit.net.au/cgi-bin/webobjects/toolkit.woa/wa/downloadPublications?
- Stern, R. D. (1979). Analysis of daily rainfall at Samaru, Nigeria, using a simple two-part model. *Theoretical and Applied Climatology* , Volume 28 (Numbers 1-2 / March, 1980), 123-135.
- Stern, R. D., & Coe, R. (1982). The Use of Rainfall Models in Agricultural Planning. *Agricultural Meteorology* , 26, 35-40.

- Stern, R., & Coe, R. (1984). A model fitting analysis of Daily Rainfall Data. *The journal of the Royal Statistical Society* , 147 (1), 1-34.
- Stern, R., & Coe, R. (1982). Fitting Models to Daily Data. *Journal of Applied Meteorology* , 12 (7), 1024-1031.
- Stern, R., Rijks, D., Dale, I., & Knock, J. (2006). *Instat Climatic Guide*. University of Reading.
- Vardi, Y., & Ju, W. (1999). *A Hybrid High-order Markov Chain Model for Computer Intrusion Detection*. National Institute of Statistical Sciences, Alexander.
- Woolhiser, D. A., & Pegram, G. S. (1978). . Maximum Likelihood Estimation of Fourier Coefficient to Describe Seasonal Variations of Parameters in Stochastic Daily Precipitation Models. *Journal of Applied Meteorology* , 18, 34-42.

Appendix I: Rainfall data for Katumani imported directly from Instat

Row	Y1961	Y1962	Y1963	Y1964	Y1965
1	0	0	0	5.8	5.6
2	0	11.2	0	26.2	4.1
3	0	31.5	0	0	10.3
4	0	0	0	0.3	2.5
5	0	0	0	0	0
6	0	0	0	20.8	0
7	0	6.6	0	0	0
8	0	2.5	0	4.3	52.8
9	0	0	0	0	0

Y1997	Y1998	Y1999	Y2000	Y2001
0	0	0	0	0
0	6	0	0	0
0	3.2	0.1	0	7.4
0	33.6	0	0	18.3
0	0.9	0	0	25.6
3.8	5	0	0	19.8
0	19	0	0	1.3
0	1.7	0.5	0	37.6
0	25.2	0.1	0	30.8

358	0	0	0	0	0
359	0	0	7.6	0	0
360	0	0	5.8	0	0
361	8.9	22.9	4.1	0	0
362	7.4	0	1.3	0	0
363	73.7	5.3	12.7	0	0
364	0	20.6	2.8	0	0
365	0	0	15.7	0	0
366	0	0	12.7	0	0

0	0	1.1	0	3
3.9	0	1.3	0	0
24.3	0	0	0	0
7.3	0	0.3	0	0
5.4	0	5.7	0	0
18	0	8	0	0
0	0	3.8	0	0
0.8	0	0	0	0
0	0	0	0	0

Figure 64: Rainfall data for Katumani imported directly from Instat

Appendix II: Summary Statistics for Katumani Rainfall data.

Year	NOBS	MAX rainfall (mm)
1963	365	186.9
1994	365	109.0
1981	365	95.3
1965	365	86.9
1983	365	86.0
1961	365	82.8
1998	365	81.2
1967	365	80.0
1966	365	79.0
1990	365	73.7
1985	365	73.5
1975	365	70.0
1971	365	67.0
1974	365	64.1
1989	365	63.7
1977	365	58.5
1997	365	54.0
1978	365	52.9
1993	365	50.4
1962	365	48.5
1973	365	44.0
2001	365	43.4
1979	365	43.1
1986	365	42.9
1991	365	40.6
1995	365	39.3
1970	365	37.5
1969	365	35.5
1999	365	35.4
1982	365	34.3
1987	365	26.7
1964	366	82.5
2000	366	69.7
1968	366	64.0
1972	366	59.0
1992	366	59.0
1980	366	58.6
1984	366	47.9
1996	366	46.0
1976	366	42.1
1988	366	37.1

Table 11: *Summary statistics for Katumani rainfall data*

Appendix III: COUNT Procedure

```

PROCEDURE [restore=case]'Counts'
OPTION NAME= \
'STATES' , "(I: Scalar giving the number of states of the chain)"\
'LIMITS' , "(I: Scalar or vector giving the values to determine the states)"\
'ORDER' , "(I: Order of the Markov Chain)"\
'CLASS' , "(I: Factor over which the data are to be totalled)"\
'HIGH' , "(I: YES/NO indicating if high order, default is NO)"\
'LABELS' , "(I: YES/NO on whether to label the tables of results, default is YES)"\
'SPREAD' , "(I: YES/NO on whether to display the results in spreadsheets)"\
'INITIAL'; "(I: Scalar or variate - length =ORDER with initial values)"\
MODE=4(P),3(T),P; SET=3(yes),no,3(yes),no; DECLARED=8(yes);\
PRESENT=8(yes) ;\
DEFAULT=2,! (0.85),1,*,'NO','YES','YES',*;\
TYPE=!T('scalar'),!T('scalar','variate'),!T('scalar'),!T('factor'),3(!T('text')),\
!T('scalar','variate');VALUE=*,*,*,*,3(!T('YES','NO')),*
PARAMETER NAME= \
'DATA', "(I: vector which containa the rainfall data" \
'COUNTS', "(O: tables containing the frequencies)" \
'AMOUNTS', "(O: table containing the sum of the amounts)"\
'CATEGORIES', "(O: variate containing the category of each observation)"\
'SCOUNTS'; "(O: pointer to tables for frequencies)"\
MODE=P; SET=yes,4(no); DECLARED=yes,4(no); \
TYPE=!T('variate'),2(!T('table')),!T('variate'),!T('pointer'); \
PRESENT=yes,4(no)
set [case=signif]
dummy states,order,limit,Rain, DayOfYear
assignSTATES,ORDER,LIMITS,DATA,CLASS;states,order,limit,Rain,DayOfYear
geta [type,nval]SCOUNTS;pp
calc states,order=round(states,order)
exit [con=proc;expl='**** The number of STATES must be two or more']states<2
exit [con=proc;expl='**** The ORDER of the chain must be non-negative']order<0
if (order>3).and.(HIGH.eqs.'NO')
    exit[con=proc;expl='**** The order is too high for this type of Markov']
endif
calc nlimits=states-1
exit [con=proc;expl='Number of LIMITS must be one less than the number of \
STATES'](nlim=nval(limit))<>nlimits
if .not.unset(INITIAL)
    exit [con=proc;expl='Number of elements of the INITIAL values must equal \
the ORDER of the chain'](ninit=nval(INITIAL))<>ORDER
    vari [n=ninit;val=#INITIAL]init
endif
DELETE [Redefine=yes]counts,amounts
vari [n=nlim;val=#limit]limits
sort limits
vari x[0..order]
group [l=*]Rain;limits=limits;factor=x0
calc x[0]=x0-1
scalar top
calc top=limits${nlimits}

if .not.unset(SCOUNTS):exit[con=proc;expl='**** SCOUNTS must be a pointer with \
no. of elements = number of STATES']nval(COUNTS)<>states
endif

scalar do;1
if unset(AMOUNTS)
    calc do=0
endif
if do
exit [con=proc;ex='**** There are no observations higher than the top limit']\
top>max(Rain)
rest Rain;Rain>top
calc amounts=(Rain-top)
rest Rain
endif

```

```

calc mult=states
if order>0
    calc orderm=order-1
    if unset(INITIAL)
        for i=1...order;j=0...orderm
            calc x[i]=shift(x[j])
        endfor
    else
        for i=1...order;j=0...orderm;k=order...1
            calc x[i]=shift(x[j])
            calc x[i]${!(1...#i)}=init${!(#k...#order)}
        endfor
    endif
    for i=orderm...0
        calc x[i]=mult*x[i]
        calc mult=mult*states
    endfor
endif
calc ncat=mult-1

calc sum=vsum(x) & summ=vnmv(x)

if sum(summ)>0
    rest summ/summ>0;save=v1
    rest summ
    calc sum${v1}=c('*')
    rest sum
endif
"Note Use the newlevels function for high order chains"
if HIGH.eqs.'NO'
    get [env]
    set [diag=f]
    factor [levels=!(0...#ncat)]sumf;sum
else
    exit [con=proc;explan='**** High order chains must have order at least2'](order<2)
    exit [con=proc;expl='**** The HIGH option is only available for 2 state chains']
    states>2
    calc high=order+1
    vari [n=#mult;values=#mult(#high)]ch & [val=1...#mult]pn
    scalar pos;1
    for i=1...order
        rest ch;pn>#pos+1
        calc ch=ch-1 & pos=pos*2
        rest ch
    endfor
    calc pos1=pos+1
    calc ch${!(#pos1...#mult)}=-ch${!(1...#pos)}
    calc min=-high
    factor [levels=!(0...#ncat)]sumt;sum
    factor [levels=!(#min...-1,1...#high)]sumf
    get [env]
    set [diag=f]
    calc sumf=(sum=newlevels(sumt;ch))
endif
set [diag=#env[3]]
if .not.unset(CATEGORIES)
    calc CATEGORIES=sum
endif
if unset(CLASS)
    exit[cont=proc]
endif
dummy DayOfYear
assign CLASS;DayOfYear
calc ndayno=nval(DayOfYear)
tabu [class=DayOfYear,sumf;count=freq]
restrict DayOfYear,sumf
if HIGH.eqs.'NO'
    calc newmult=round(mult/states)
    calc newm=newmult-1
    vari [n=newmult]oldlevs,newlevs;!(0...#newm);dec=0

```

```

if (states==2).and.(order<4)
  case order+1
    text [n=1;val=d]dry
    text [n=1;val=w]wet
  or
    text [n=2;val=dd,dw]dry
    text [n=2;val=wd,ww]wet
  or
    text [n=4;val=ddd,ddw,dwd,dww]dry
    text [n=4;val=wdd,wdw,wwd,www]wet
  or
    text [n=8;val=dddd,dddw,ddwd,ddww,dwdd,dwdw,dwwd,dwww]dry
    text [n=8;val=wddd,wddw,wdwd,wdww,wwdd,wwdw,wwwd,www]wet
  endcase
  if LABELS.eq. 'YES'
    fact [n=ndayno;val=#ndayno(0);lev=newlevs;lab=dry]sudry
    fact [n=ndayno;val=#ndayno(0);lev=newlevs;lab=wet]su
    comb [freq;fre[0]]sumf;sudry;oldlevs;newlevs
    calc oldlevs=oldlevs+newmult
  else
    fact [lev=newlevs;n=ndayno;val=#ndayno(0)]su
    comb [freq;fre[0]]sumf;su;oldlevs;newlevs
    calc oldlevs=oldlevs+newmult
  endif
  comb [freq;fre[1]]sumf;su;oldlevs;newlevs
else
  factor [levels=newlevs]su
  for i=0..nlimits
    combine [freq;fre[i]]sumf;su;oldlevs;newlevs
    calc oldlevs=oldlevs+newmult
  endfor
  calc oldlevs=oldlevs-newmult
endif
else
  calc order1=order-1
  print [c=ordd;ipr=*;squ=yes]'>',order,'d';dec=0
  print [c=ordw;ipr=*;squ=yes]'>',order1,'dw';dec=0
  edit [!T('g/ // :')]ordd
  edit [!T('g/ // :')]ordw
  text [n=2;val='_wdd','_wd']twod
  text [n=2;val='_wdw','_ww']twow
  if order==2
    text [n=high;value=#ordd,#twod]facnamed
    text [val=#ordw,#twow]facnamew
  else
    for i=3..order;j=order...3
      calc j1=j-1
      print [c=orddi[i];ipr=*;squ=yes]'_w',j,'d';d=0
      print [c=ordwi[i];ipr=*;squ=yes]'_w',j1,'dw';d=0
      edit [!T('g/ // :')]orddi[i] & ordwi[i]
    endfor
    text [n=high;value=#ordd,#orddi[],#twod]facnamed
    text [n=high;val=#ordw,#ordwi[],#twow]facnamew
  endif
  vari levhigh;values=(#high...1);dec=0
  if LABELS.eq. 'YES'
    fact [n=ndayno;val=#ndayno(1);lev=levhigh;lab=facnamed]su0 &
    fact [lab=facnamew]sul
    comb [freq;fre[0]]sumf;su0;levhigh;levhigh
  else
    fact [n=ndayno;val=#ndayno(1);leve=levhigh]sul
    comb [freq;fre[0]]sumf;sul;levhigh;levhigh
  endif
  comb [freq;new=fre[1]]sumf;sul;!(#min...-1);levhigh
endif

if .not.unset(COUNTS):
  calc COUNTS=freq
  table [mod=yes]COUNTS;val=freq;dec=0
endif

```

```

if do
rest sumf,amounts;Rain>top
tabu [class=DayOfYear,sumf]amounts;mean=asum
rest sumf,amounts
if HIGH.eqs.'YES'
comb [asum;asum1]sumf;su1;!(#min...-1);!(#high...1)
else
combine [asum;asum1]sumf;su;oldlevs;newlevs
endif
if .not.unset(AMOUNTS)
calc AMOUNTS=asum1
table [mod=yes]AMOUNTS;val=asum1;dec=1
calc AMOUNTS,asum1=mvreplace(AMOUNTS,asum1;0,-100)
calc AMOUNTS=mvinsert(AMOUNTS;asum1=-100)

endif
endif
if SPREAD.eqs.'YES'
fspread COUNTS
fspread AMOUNTS
endif
endproc

```

Appendix IV: PREPARE Procedure

```

PROCEDURE [restore=case] 'prepare'

OPTION NAME= \
  'SPREAD', "(I: YES/NO on whether to display the results in spreadsheets, default is YES)"\
  'PLOT'; "(I: YES/NO on whether to plot the probabilities or not, default is YES)"\
  MODE=2(T); SET=2(yes);DECLARED=2(yes);PRESENT=2(yes);DEFAULT=2('YES'); TYPE=2(!T('text'));\
  VALUE=2(!T('YES','NO'))

PARAMETER NAME= \
  'COUNTS', "(I: tables containing the frequencies)" \
  'AMOUNT'; "(I: table containing the sum of the amounts)"\
  MODE=P; SET=yes,no; DECLARED=2(yes);TYPE=2(!T('table'));PRESENT=2(yes)

set [case=signif]
dummy counts,amount
assign COUNTS,AMOUNT;counts,amount
GETATTRIBUTE [attribute=class] counts; save=classC
scal nrowC,ncolC;nlevels(classC['classification']][1,2])
calc MinVal=min(classC['classification']][2])
if (MinVal==0).and.(ncolC==16)
  calc order=3
  text[value='NO'] HIGH
elseif (MinVal==0).and.(ncolC==8)
  calc order =2
  text[value='NO'] HIGH
elseif (MinVal==0).and.(ncolC==4)
  calc order =1
  text[value='NO'] HIGH
elseif (MinVal==0).and.(ncolC==2)
  calc order =0
  text[value='NO'] HIGH
else
  calc order =(ncolC-2)/2
  text[value='YES'] HIGH
endif

XAXIS [RESET=yes] WINDOW=1; LPOSITION=outside; LDIRECTION=parallel; MPOSITION=outside;\
ARROWHEAD=omit; ACTION=display
YAXIS [RESET=yes] WINDOW=1; LPOSITION=outside;LDIRECTION=perpendicular; MPOSITION=outside;\
ARROWHEAD=omit; ACTION=display

if HIGH.eqs.'NO'

  if order==0
    PEN [RESET=yes] 1,2; METHOD=line,line; JOIN=ascending; SYMBOL=0; LIFESTYLE=1;
    vtable COUNTS; Var; !p(F1,F2)
    unstack [DATASET=F2;IDSTACKED=F1;IDUNSTACKED=dayNo] 2(Var); Data=0,1; d,w
    calc lw=log(wet)
    calc tot=d+w
    calc P_d=d/tot
    calc P_w=w/tot
    if PLOT.eqs.'YES'
      Dgraph Y=P_w; X=dayNo
    endif
    if SPREAD.eqs.'YES'
      fspread dayNo,P_w
    endif
  elseif order==1
    PEN [RESET=yes] 1,2; METHOD=line,line; JOIN=ascending; SYMBOL=0; LIFESTYLE=1
    vtable COUNTS; Var; !p(F1,F2)
    unstack [DATASET=F2;IDSTACKED=F1;IDUNSTACKED=dayNo]4(Var);Data=0,1,2,3;\ dd,dw,wd,ww
    calc _d=dd+wd & "DY-Dry yesterday" _w=dw+ww " WY-Wet yesterday"
    calc P_wd=wd/_d & P_ww=ww/_w
    if SPREAD.eqs.'YES'
      fspread dayNo,P_ww,P_wd
    endif
  endif
endif

```

```

        if PLOT.eq. 'YES'
Dgraph Y=P_ww,P_wd; X=dayNo
        endif

elseif order==2
PEN [RESET=yes] 1,2,3,4; METHOD=line,line; JOIN=ascending; SYMBOL=0; LINESYLE=1;
vtable COUNTS; Var; !p(F1,F2)
unstack [DATASET=F2; IDSTACKED=F1; IDUNSTACKED=dayNo] 4(Var);
        Data='wdd','wdw','wwd','www'; w_dd,w_dw,w_wd,w_ww
        calc _dd=ddd+wdd & _dw=ddw+wdw & _wd=dwd+wwd & _ww=dww+www
        calc P_www=www/_ww & P_wwd=wwd/_wd & P_wdw=wdw/_dw & P_wdd=wdd/_dd
        if SPREAD.eq. 'YES'
                fspread dayNo,P_www,P_wwd,P_wdw,P_wdd
        endif
        if PLOT.eq. 'YES'
                Dgraph Y=P_www,P_wwd,P_wdw,P_wdd; X=dayNo
        endif

elseif order==3
PEN [RESET=yes]1,2,3,4,5,6,7,8;METHOD=line,line;JOIN=ascending; SYMBOL=0; LINESYLE=1
vtable COUNTS; Var; !p(F1,F2)
unstack [DATASET=F2; IDSTACKED=F1; IDUNSTACKED=dayNo] 16(Var);
        Data=0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15;dddd,dddw,ddwd,ddww,dwdd,dwdw,dw\
        wd,dwww,wddd,wddw,wdwd,wdww,wwdd,wwdw,wwwd,wwww
        calc _ddd=ddd+wddd & _ddw=ddd+wddw & _dwd=ddwd+wdwd & _dww=ddww+wdww & \
        _wdd=dwdd+wddd
        calc _wdw=dwdw+wwdw & _wwd=dwwd+wwwd & _www=dwww+wwww
        calc P_www=www/_www & P_wwd=wwd/_wwd & P_wdw=wdw/_dw & P_wdd=wdd/_dwd & \
        P_wdw=wdw/_dww
        calc P_wdwd=wdwd/_dwd & P_wddw=wddw/_ddw & P_wddd=wddd/_ddd
        if SPREAD.eq. 'YES'
                fspread dayNo,P_www,P_wwd,P_wdw,P_wdd,P_wdww,P_wdwd,P_wddw, P_wddd
        endif
        if PLOT.eq. 'YES'
                Dgraph Y=P_www,P_wwd,P_wdw,P_wdd,P_wdww,P_wdwd,P_wddw, P_wddd; X=dayNo
        endif

endif

elseif HIGH.eq. 'YES'

        vtable COUNTS; Var; !p(F1,F2)
        calc Power= order+1
        calc high1=order*2+2 "calculating the number of variables"
        calc max1=Power
        calc negMax=-#max1

        print [c=ord[1];ipr=*;squ=yes]'w',order,'d';d=0
        edit [!T('g/ // :')]ord[1]

        for i=3...order;j=order...3
                calc j1=j-1
                calc i1=i-1
                print [c=ord[i1];ipr=*;squ=yes]'w',j1,'d';d=0
                edit [!T('g/ // :')]ord[i1]
        endfor

        print [c=ord[order];squ=yes]'wdw'
        calc order1=order+1
        print [c=ord[order1];squ=yes]'ww'
        calc order2=order+2
        print [c=ord[order2];squ=yes]'dw'

        for i=2...order;
                calc order3=order+1+i
                print [c=ord[order3];ipr=*;squ=yes]_'_',i,'dw';d=0
                edit [!T('g/ // :')] ord[order3]
        endfor

```

```

print [c=ord[high1];ipr=*;squ=yes]('_',order1,'d';d=0
edit [!T('g/ // :')]ord[high1]
text [value=#ord[]]Collumaames
pointer[nvalues=#high1]cols
read [channel=Collumaames]cols
variate cols[]
unstack [DATASET=F2;IDSTACKED=F1;IDUNSTACKED=dayNo] #high1(Var);\
  Data=#negMax...-1,1...#max1;#cols
fspread dayNo,#cols

" calculates Probabilities"
calc Power=order+1
print[c=ordP[1];ipr=*;squ=yes]'P_w',order,'d';d=0
edit [!T('g/ // :')]ordP[1]
for i=2...order;j=order...3
  calc j1=j-1
  print [c=ordP[i];ipr=*;squ=yes]'P_wd',j1,'w';d=0
  edit [!T('g/ // :')]ordP[i]
endfor
print [c=ordP[order];squ=yes]'P_wdw'
calc order1=order+1
print [c=ordP[order1];squ=yes]'P_ww'
text [value=#ordP[]]Coll
pointer[nvalues=#Power]colsP
read [channel=Coll]colsP
variate colsP[]
Calc Power1=Power+1
vari variables ;values=(1...#Power);dec=0
for i =1...Power;
  calc k=i-1
  calc j=#high1-k
  calc colsP[i]=cols[i]/(cols[i]+cols[j])
endfor
if SPREAD.eq. 'YES'
  fspread dayNo,#colsP
endif
PEN [RESET=yes]#variables; METHOD=line,line; JOIN=ascending; SYMBOL=0;
  LINSTYLE=1
if PLOT.eq. 'YES'
  Dgraph Y=#colsP; X=dayNo
endif
endif
endproc

```

Appendix V: FITTING Probability Procedure

```

PROCEDURE [restore=case] 'fitting'

OPTION NAME= \
'HAMONICS'; "(Number of hamonic required Default is 3)"\
MODE=P; SET=yes;DECLARED=yes;PRESENT=yes ;DEFAULT=2;TYPE=!T('scalar');VALUE=*
PARAMETER NAME= \
'COUNTS'," (I: tables containing the frequencies)" \
'AMOUNT'; "(I: table containing the sum of the amounts)"\
MODE=P; SET=yes,no; DECLARED=2(yes);TYPE=2(!T('table'));PRESENT=2(yes)

set [case=signif]
dummy counts,amounts,hamonics
assign COUNTS,AMOUNT,HAMONICS;counts,amount,hamonics

GETATTRIBUTE [attribute=class] counts; save=classC
scal nrowC,ncolC;nlevels(classC['classification']][1,2])
calc MinVal=min(classC['classification']][2])
if (MinVal==0).and.(ncolC==16)
    calc order=3
    text[value='NO'] HIGH
elseif (MinVal==0).and.(ncolC==8)
    calc order =2
    text[value='NO'] HIGH
elseif (MinVal==0).and.(ncolC==4)
    calc order =1
    text[value='NO'] HIGH
elseif (MinVal==0).and.(ncolC==2)
    calc order =0
    text[value='NO'] HIGH
else
    calc order =(ncolC-2)/2
    text[value='YES'] HIGH
endif

XAXIS [RESET=yes] WINDOW=1; LPOSITION=outside; LDIRECTION=parallel;
MPOSITION=outside;ARROWHEAD=omit; ACTION=display
YAXIS [RESET=yes] WINDOW=1; LPOSITION=outside; LDIRECTION=perpendicular;
MPOSITION=outside;ARROWHEAD=omit; ACTION=display

scalar const
calc const=2*c('pi')/366
if HIGH.eqs.'NO'
if order==0
PEN [RESET=yes] 1,2; METHOD=line,line; JOIN=ascending; SYMBOL=0; LIFESTYLE=1
vtable COUNTS; Var; !p(F1,F2)
unstack [DATASET=F2;IDSTACKED=F1;IDUNSTACKED=dayNo] 2(Var); Data=0,1; d,w
vtable AMOUNT; Var; !p(F1,F2)
unstack [DATASET=F2;IDSTACKED=F1;IDUNSTACKED=dayNo] 1(Var); Data='w'; wet
calc tot=d+w
calc tots=(tot>0) "This gives a weight of zero if there are no dry days"
calc tot=tot*(tot>0)+(tot.eq.0) "Now give a denominator of at least 1 always"
for i=1...hamonics
    calc sin[i]=sin(i*const*dayNo)
    calc cos[i]=cos(i*const*dayNo)
endfor

MODEL [DISTRIBUTION=binomial; LINK=logit; DISPERSION=1;wei=tots] w; NBINOMIAL=tot
FITindividually [PRINT=m,s,e,a; FPROB=yes; TPROB=yes]sin[+cos[ ]
RKEEP FITTEDVALUES=f_wt
calc f_wt=f_wt/tot

PEN [RESET=yes] 1; METHOD=line; JOIN=ascending; SYMBOL=0; LIFESTYLE=1;
DGRAPH [TITLE='Zero Order probabilities']Y=f_wt; X=dayNo

elseif order==1
vtable COUNTS; Var; !p(F1,F2)

```

```

unstack [DATASET=F2;IDSTACKED=F1;IDUNSTACKED=dayNo] 4(Var);
  Data=0,1,2,3; dd,dw,wd,ww
vtable AMOUNT; Var; !p(F1,F2)
unstack [DATASET=F2;IDSTACKED=F1;IDUNSTACKED=dayNo] 2(Var); Data='wd','ww'; w_d,w_w
calc _d=dd+wd & "DY-Dry yesterday" _w=dw+ww " WY-Wet yesterday"

calc wtdry=(_d>0) "This gives a weight of zero if there are no dry days"&
  _d=_d*( _d>0)+(_d.eq.0)"Now give a denominator of at least 1 always"
calc wtrain=( _w>0)& _w=_w*( _w>0)+(_w.eq.0)
  for i=1..hamonics
    calc sin[i]=sin(i*const*dayNo)
    calc cos[i]=cos(i*const*dayNo)
  endfor
"General Model for 2 harmonics."
MODEL [DISTRIBUTION=binomial; LINK=logit; DISPERSION=1;wei=wtdry] wd; NBINOMIAL=_d
FITindividually [PRINT=m,s,e,a; FPROB=yes; TPROB=yes]sin[ ]+cos[ ]
RKEEP FITTEDVALUES=f_wd
calc f_wd=f_wd/_d
PEN [RESET=yes] 1,2; METHOD=line,line; JOIN=ascending; SYMBOL=0; LINESSTYLE=1;

MODEL [DISTRIBUTION=binomial; LINK=logit; DISPERSION=1;wei=wtrain] ww; NBINOMIAL=_w
FITindividually [PRINT=m,s,e,a; FPROB=yes; TPROB=yes]sin[ ]+cos[ ]
RKEEP FITTEDVALUES=f_ww
calc f_ww=f_ww/_w
PEN [RESET=yes] 1,2; METHOD=line,line; JOIN=ascending; SYMBOL=0; LINESSTYLE=1
DGRAPH Y=f_wd,f_ww; X=dayNo

elseif order==2
  vtable COUNTS; Var; !p(F1,F2)
  unstack [DATASET=F2;IDSTACKED=F1;IDUNSTACKED=dayNo] 8(Var); Data=0,1,2,3,4,5,6,7;
ddd,ddw,dwd,dww,wdd,wdw,wwd,www
vtable AMOUNT; Var; !p(F1,F2)
unstack [DATASET=F2;IDSTACKED=F1;IDUNSTACKED=dayNo] 4(Var);
Data='wdd','wdw','wdd','www'; w_dd,w_dw,w_wd,w_ww
calc _dd=ddd+wdd & _dw=ddw+wdw & _wd=dwd+wwd & _ww=dww+www
calc wtdrydry=( _dd>0) & _dd=_dd*( _dd>0)+(_dd.eq.0) & wtdrywet=( _dw>0) &
_wd=_dw*( _dw>0)+(_dw.eq.0) & wtraindry=( _wd>0)& _wd=_wd*( _wd>0)+(_wd.eq.0)
calc wtwtwt=( _ww>0) & _ww=_ww*( _ww>0)+(_ww.eq.0)
  for i=1..hamonics
    calc sin[i]=sin(i*const*dayNo)
    calc cos[i]=cos(i*const*dayNo)
  endfor

MODEL [DISTRIBUTION=binomial; LINK=logit; DISPERSION=1;wei=wtdrydry] wdd;
  NBINOMIAL=_dd
FITindividually [PRINT=m,s,e,a; FPROB=yes; TPROB=yes]sin[ ]+cos[ ]
RKEEP FITTEDVALUES=f_wdd
calc f_wdd=f_wdd/_dd

MODEL [DISTRIBUTION=binomial; LINK=logit; DISPERSION=1;wei=wtdrywet] wdw; NBINOMIAL=_dw
FITindividually [PRINT=m,s,e,a; FPROB=yes; TPROB=yes]sin[ ]+cos[ ]
RKEEP FITTEDVALUES=f_wdw
calc f_wdw=f_wdw/_dw

MODEL [DISTRIBUTION=binomial; LINK=logit; DISPERSION=1;wei=wtraindry]wwd;
  NBINOMIAL=_wd
FITindividually [PRINT=m,s,e,a; FPROB=yes; TPROB=yes]sin[ ]+cos[ ]
RKEEP FITTEDVALUES=f_wwd
calc f_wwd=f_wwd/_wd

MODEL [DISTRIBUTION=binomial; LINK=logit; DISPERSION=1;wei=wtwtwt] www;NBINOMIAL=_ww
FITindividually [PRINT=m,s,e,a; FPROB=yes; TPROB=yes]sin[ ]+cos[ ]
RKEEP FITTEDVALUES=f_www
calc f_www=f_www/_ww

PEN [RESET=yes] 1,2,3,4; METHOD=line,line; JOIN=ascending; SYMBOL=0; LINESSTYLE=1
DGRAPH[TITLE='Second Order probabilities'] Y=f_www,f_wwd,f_wdw,f_wdd; X=dayNo

elseif order==3

```

```

        vtable COUNTS; Var; !p(F1,F2)
    unstack [DATASET=F2; IDSTACKED=F1; IDUNSTACKED=dayNo] 16(Var);
        Data=0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15; dddd,dddw,ddwd,ddww,dwdd,dwdw,dwwd,dwww,wd
        dd,wddw,wdwd,wdww,wddd,wwdw,wwwd,wwww
    vtable AMOUNT; Var; !p(F1,F2)
    unstack [DATASET=F2; IDSTACKED=F1; IDUNSTACKED=dayNo] 8(Var);
        Data='wddd','wddw','wdwd','wdww','wddd','wddw','wwwd','wwww';
        w_ddd,w_ddw,w_dwd,w_dww,w_wdd,w_wdw,w_wwd,w_www
    calc _ddd=dddd+wddd & _wdd=dwdd+wddd & _ddw=dddw+wddw & _wdw=dwdw+wwdw & _dwd=ddwd+wddw &
        _wwd=dwwd+wwwd & _dww=ddww+wdww & _www=dwww+wwww
    calc wddy=(_ddd>0) & _ddd=_ddd*( _ddd>0)+( _ddd.eq.0)& wddy=( _wdd>0) &
        _wdd=_wdd*( _wdd>0)+( _wdd.eq.0)& wddwt=( _ddw>0) & _ddw=_ddw*( _ddw>0)+( _ddw.eq.0)
    calc wdwwt=( _wdw>0) & _wdw=_wdw*( _wdw>0)+( _wdw.eq.0)& wdwdy=( _dwd>0) &
        _dwd=_dwd*( _dwd>0)+( _dwd.eq.0) & wwwdy=( _wwd>0) & _wwd=_wwd*( _wwd>0)+( _wwd.eq.0)
    calc wdwwt=( _dww>0) & _dww=_dww*( _dww>0)+( _dww.eq.0)& wwwwt=( _www>0) &
        _www=_www*( _www>0)+( _www.eq.0)
    for i=1..hamonics
        calc sin[i]=sin(i*const*dayNo)
        calc cos[i]=cos(i*const*dayNo)
    endfor
    MODEL [DISTRIBUTION=binomial; LINK=logit; DISPERSION=1;wei=wddy] wddd; NBINOMIAL=_ddd
    FITindividually [PRINT=m,s,e,a; FPROB=yes; TPROB=yes]sin[ ]+cos[ ]
    RKEEP FITTEDVALUES=f_wddd
    calc f_wddd=f_wddd/_ddd

    MODEL [DISTRIBUTION=binomial; LINK=logit; DISPERSION=1;wei=wddw] wwdd; NBINOMIAL=_wdd
    FITindividually [PRINT=m,s,e,a; FPROB=yes; TPROB=yes]sin[ ]+cos[ ]
    RKEEP FITTEDVALUES=f_wddw
    calc f_wddw=f_wddw/_wdd

    MODEL [DISTRIBUTION=binomial; LINK=logit; DISPERSION=1;wei=wdwt] wddw; NBINOMIAL=_ddw
    FITindividually [PRINT=m,s,e,a; FPROB=yes; TPROB=yes]sin[ ]+cos[ ]
    RKEEP FITTEDVALUES=f_wddw
    calc f_wddw=f_wddw/_ddw

    MODEL [DISTRIBUTION=binomial; LINK=logit; DISPERSION=1;wei=wdwdy] wdwd; NBINOMIAL=_dwd
    FITindividually [PRINT=m,s,e,a; FPROB=yes; TPROB=yes]sin[ ]+cos[ ]
    RKEEP FITTEDVALUES=f_wdwd
    calc f_wdwd=f_wdwd/_dwd

    MODEL [DISTRIBUTION=binomial; LINK=logit; DISPERSION=1;wei=wwdy] wwwd; NBINOMIAL=_wwd
    FITindividually [PRINT=m,s,e,a; FPROB=yes; TPROB=yes]sin[ ]+cos[ ]
    RKEEP FITTEDVALUES=f_wwwd
    calc f_wwwd=f_wwwd/_wwd

    MODEL [DISTRIBUTION=binomial; LINK=logit; DISPERSION=1;wei=wdwwt] wdww; NBINOMIAL=_dww
    FITindividually [PRINT=m,s,e,a; FPROB=yes; TPROB=yes]sin[ ]+cos[ ]
    RKEEP FITTEDVALUES=f_wdww
    calc f_wdww=f_wdww/_dww

    MODEL [DISTRIBUTION=binomial; LINK=logit; DISPERSION=1;wei=wwwwt] wwww; NBINOMIAL=_www
    FITindividually [PRINT=m,s,e,a; FPROB=yes; TPROB=yes]sin[ ]+cos[ ]
    RKEEP FITTEDVALUES=f_wwww
    calc f_wwww=f_wwww/_www

    PEN [RESET=yes] 1,2,3,4,5,6,7,8; METHOD=line; JOIN=ascending; SYMBOL=0; LIFESTYLE=1
    DGRAPH[TITLE='Third Order probabilities']
        Y=f_wwww,f_wdww,f_wwwd,f_wdwd,f_wddw,f_wddw,f_wddd; X=dayNo

endif
else
    vtable COUNTS; Var; !p(F1,F2)
    calc Power= order+1

```

```

calc high1=order*2+2 "calculating the number of variables"
calc max1=Power
calc negMax=-#max1

print[c=ord[1];ipr=*;squ=yes]'w',order,'d';d=0
edit [!T('g/ // :')]ord[1]

for i=3...order;j=order...3
  calc j1=j-1
  calc i1=i-1
  print [c=ord[i1];ipr=*;squ=yes]'w',j1,'dw';d=0
  edit [!T('g/ // :')]ord[i1]
endfor

print [c=ord[order];squ=yes]'wdw'
calc order1=order+1
print [c=ord[order1];squ=yes]'ww'
calc order2=order+2
print [c=ord[order2];squ=yes]'dw'

for i=2...order;
  calc order3=order+1+i
  print [c=ord[order3];ipr=*;squ=yes]('_',i,'dw');d=0
  edit [!T('g/ // :')] ord[order3]
endfor
print [c=ord[high1];ipr=*;squ=yes]('_',order1,'d');d=0
edit [!T('g/ // :')]ord[high1]
text [value=#ord[]]Collumaames
pointer[nvalues=#high1]cols
read [channel=Collumaames]cols
variate cols[]
unstack [DATASET=F2;IDSTACKED=F1;IDUNSTACKED=dayNo] #high1(Var); Data=#negMax...-
  1,1...#max1;#cols
calc Power=order+1
print[c=ordT[1];ipr=*;squ=yes]('_',order,'d');d=0
edit [!T('g/ // :')]ordT[1]
for i=2...order;j=order...3
  calc j1=j-1
  print [c=ordT[i];ipr=*;squ=yes]('_',j1,'dw');d=0
  edit [!T('g/ // :')]ordT[i]
endfor
print [c=ordT[order];squ=yes]_'dw'
calc order1=order+1
print [c=ordT[order1];squ=yes]_'w'
text [value=#ordT[]]CollT
pointer[nvalues=#Power]colsT
read [channel=CollT]colsT
variate colsT[]
Calc Power1=Power+1
for i =1...Power;
  calc k=i-1
  calc j=#high1-k
  calc colsT[i]=cols[i]+cols[j]
endfor

print[c=ordWT[1];ipr=*;squ=yes]'WT_',order,'d';d=0
edit [!T('g/ // :')]ordWT[1]
for i=2...order;j=order...3
  calc j1=j-1
  print [c=ordWT[i];ipr=*;squ=yes]'WT_',j1,'dw';d=0
  edit [!T('g/ // :')]ordWT[i]
endfor
print [c=ordWT[order];squ=yes]'WT_dw'
calc order1=order+1
print [c=ordWT[order1];squ=yes]'WT_w'
text [value=#ordWT[]]CollWT
pointer[nvalues=#Power]colsWT
read [channel=CollWT]colsWT
variate colsWT[]

```

```

for i =1...Power
    calc colsWT[i]=(colsT[i]>0) & colsT[i]=colsT[i]*(colsT[i]>0)+(colsT[i].eq.0)
endfor

for i=1...hamonics
    calc sin[i]=sin(i*const*dayNo)
    calc cos[i]=cos(i*const*dayNo)
endfor

print [c=ordF[1];ipr=*;squ=yes]'F_w',order,'d';d=0
edit [!T('g/ // :')]ordF[1]
for i=2...order;j=order...3
    calc j1=j-1
    print [c=ordF[i];ipr=*;squ=yes]'F_w',j1,'dw';d=0
    edit [!T('g/ // :')]ordF[i]
endfor
print [c=ordF[order];squ=yes]'F_wdw'
calc order1=order+1
print [c=ordF[order1];squ=yes]'F_ww'
text [value=#ordF[]]CollF
pointer [nvalues=#Power]colsF
read [channel=CollF]colsF
variate colsF[]
vari variables ;values=(1...#Power);dec=0
vari variables2 ;values=(#Power...1);dec=0
for i=1...Power
    MODEL [DISTRIBUTION=binomial; LINK=logit; DISPERSION=1;wei= colsWT[i]] cols[i];
        NBINOMIAL=colsT[i]
    FITindividually [PRINT=m,s,e,a; FPROB=yes; TPROB=yes]sin[+cos[]]
    RKEEP FITTEDVALUES=colsF[i]
    calc colsF[i]=colsF[i]/colsT[i]
endfor
print [c=TIT[order];ipr=*;squ=yes]'Fitted Probabilities of order ',#order,' (High
Order)';d=0
edit [!T('g/ // :')]TIT[]
text [value=#TIT[]]TITS

PEN [RESET=yes]#variables; METHOD=line,line; JOIN=ascending; SYMBOL=0; LINSTYLE=1
DGRAPH [Title=#TITS]Y=#colsF; X=dayNo

endif
endproc

```

Appendix VI: AMOUNTFITTING Procedure

```

PROCEDURE [restore=case] 'AmountFitting'

OPTION NAME= \
'HAMONICS'; "(Number of hamonic required Default is 3)"\
MODE=P; SET=yes;\
DECLARED=yes;PRESENT=yes;DEFAULT=2;TYPE=!T('scalar');VALUE=*

PARAMETER NAME= \
'COUNTS', "(I: tables containing the frequencies)" \
'AMOUNT'; "(I: table containing the sum of the amounts)"\
MODE=P; SET=yes,no; DECLARED=2(yes); TYPE=2(!T('table'));PRESENT=2(yes)
set [case=signif]
dummy counts,amounts,hamonics
assign COUNTS,AMOUNT,HAMONICS;counts,amount,hamonics
GETATTRIBUTE [attribute=class] counts; save=classC
scal nrowC,ncolC;nlevels(classC['classification'][[1,2])
calc MinVal=min(classC['classification'][[2])
if (MinVal==0).and.(ncolC==16)
calc order=3
text[value='NO'] HIGH
elseif (MinVal==0).and.(ncolC==8)
calc order =2
text[value='NO'] HIGH
elseif (MinVal==0).and.(ncolC==4)
calc order =1
text[value='NO'] HIGH
elseif (MinVal==0).and.(ncolC==2)
calc order =0
text[value='NO'] HIGH
else
calc order =(ncolC-2)/2
text[value='YES'] HIGH
endif

XAXIS [RESET=yes] WINDOW=1; LPOSITION=outside; LDIRECTION=parallel;
MPOSITION=outside;ARROWHEAD=omit; ACTION=display
YAXIS [RESET=yes] WINDOW=1; LPOSITION=outside; LDIRECTION=perpendicular;
MPOSITION=outside;ARROWHEAD=omit; ACTION=display
scalar const
calc const=2*c('pi')/366
if HIGH.eqs.'NO'
if order==0
PEN [RESET=yes] 1,2; METHOD=line,line; JOIN=ascending; SYMBOL=0; LIFESTYLE=1
vtable COUNTS; Var; !p(F1,F2)
unstack [DATASET=F2;IDSTACKED=F1;IDUNSTACKED=dayNo] 2(Var); Data=0,1; d,w
vtable AMOUNT; Var; !p(F1,F2)
unstack [DATASET=F2;IDSTACKED=F1;IDUNSTACKED=dayNo] 1(Var); Data='w'; mrain
for i=1...hamonics
calc sin[i]=sin(i*const*dayNo)
calc cos[i]=cos(i*const*dayNo)
endifor
MODEL [DISTRIBUTION=gamma; LINK=log;wei=w] mrain
FIT [PRINT=model,summmary,estimates; CONSTANT=estimate; FPROB=yes; TPROB=yes;
FACT=9]sin[ ]+cos[ ]
RKEEP FITTEDVALUES=fm
PEN [RESET=yes] 1,2; METHOD=line,line; JOIN=ascending; SYMBOL=0; LIFESTYLE=1;th=4,1
DGRAPH Y=fm; X=dayNo
elseif order==1
vtable COUNTS; Var; !p(F1,F2)
unstack [DATASET=F2;IDSTACKED=F1;IDUNSTACKED=dayNo] 4(Var); Data=0,1,2,3; dd,dw,wd,ww
vtable AMOUNT; Var; !p(F1,F2)
unstack [DATASET=F2;IDSTACKED=F1;IDUNSTACKED=dayNo] 2(Var); Data='wd','ww'; w_d,w_w
for i=1...hamonics
calc sin[i]=sin(i*const*dayNo)
calc cos[i]=cos(i*const*dayNo)
endifor
"General Model for 2 harmonics."

```

```

MODEL [DISTRIBUTION=gamma; LINK=log;wei=wd] w_d
FIT [PRINT=model,summary,estimates; CONSTANT=estimate; FPROB=yes; TPROB=yes;
FACT=9]sin[ ]+cos[ ]
RKEEP FITTEDVALUES=fm_wd

MODEL [DISTRIBUTION=gamma; LINK=log;wei=ww] w_w
FIT [PRINT=model,summary,estimates; CONSTANT=estimate; FPROB=yes; TPROB=yes;
FACT=9]sin[ ]+cos[ ]
RKEEP FITTEDVALUES=fm_ww
PEN [RESET=yes] 1,2; METHOD=line,line; JOIN=ascending; SYMBOL=0; LIFESTYLE=1
DGRAPH Y=fm_wd,fm_ww; X=dayNo

elseif order==2
vtable COUNTS; Var; !p(F1,F2)
unstack [DATASET=F2; IDSTACKED=F1; IDUNSTACKED=dayNo] 8(Var); Data=0,1,2,3,4,5,6,7;
ddd,ddw,dwd,dww,wdd,wdw,wwd,www
vtable AMOUNT; Var; !p(F1,F2)
unstack [DATASET=F2; IDSTACKED=F1; IDUNSTACKED=dayNo] 4(Var);
Data='wdd','wdw','wdd','www'; w_dd,w_dw,w_wd,w_ww
calc _dd=ddd+wdd & _dw=ddw+wdw & _wd=dwd+wwd & _ww=dww+www
calc wtdrydry=( _dd>0 ) & _dd=_dd*( _dd>0 )+( _dd.eq.0 ) & wtdrywet=( _dw>0 ) &
_dw=_dw*( _dw>0 )+( _dw.eq.0 ) & wtraindry=( _wd>0 ) & _wd=_wd*( _wd>0 )+( _wd.eq.0 )
calc wtwwt=( _ww>0 ) & _ww=_ww*( _ww>0 )+( _ww.eq.0 )
for i=1..hamonics
calc sin[i]=sin(i*const*dayNo)
calc cos[i]=cos(i*const*dayNo)
endfor

MODEL [DISTRIBUTION=gamma; LINK=log;wei=wdd] w_dd
FIT [PRINT=model,summary,estimates; CONSTANT=estimate; FPROB=yes; TPROB=yes;
FACT=9]sin[ ]+cos[ ]
RKEEP FITTEDVALUES=fm_wdd

MODEL [DISTRIBUTION=gamma; LINK=log;wei=wdw] w_dw
FIT [PRINT=model,summary,estimates; CONSTANT=estimate; FPROB=yes; TPROB=yes;
FACT=9]sin[ ]+cos[ ]
RKEEP FITTEDVALUES=fm_wdw

MODEL [DISTRIBUTION=gamma; LINK=log;wei=wwd] w_wd
FIT [PRINT=model,summary,estimates; CONSTANT=estimate; FPROB=yes; TPROB=yes;
FACT=9]sin[ ]+cos[ ]
RKEEP FITTEDVALUES=fm_wwd

MODEL [DISTRIBUTION=gamma; LINK=log;wei=www] w_ww
FIT [PRINT=model,summary,estimates; CONSTANT=estimate; FPROB=yes; TPROB=yes;
FACT=9]sin[ ]+cos[ ]
RKEEP FITTEDVALUES=fm_www

PEN [RESET=yes] 1,2,3,4; METHOD=line,line; JOIN=ascending; SYMBOL=0; LIFESTYLE=1
DGRAPH[TITLE='Second Order probabilities'] Y=fm_www,fm_wwd,fm_wdw,fm_wdd; X=dayNo

elseif order==3
vtable COUNTS; Var; !p(F1,F2)
unstack [DATASET=F2; IDSTACKED=F1; IDUNSTACKED=dayNo] 16(Var);
Data=0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15; dddd,dddw,ddwd,ddww,dwdd,dwdw,dwwd,dwww,w
ddd,wddw,wdwd,wdww,wwdd,wwdw,wwwd,wwww
vtable AMOUNT; Var; !p(F1,F2)
unstack [DATASET=F2; IDSTACKED=F1; IDUNSTACKED=dayNo] 8(Var);
Data='wddd','wddw','wddw','wddw','wddd','wddw','wddd','wddd';
w_ddd,w_ddd,w_ddd,w_ddd,w_ddd,w_ddd,w_ddd,w_ddd
calc _ddd=ddd+wddd & _ddd=dwdd+wddd & _ddw=dddw+wddw & _dwd=dwdw+wddw & _dwd=dddw+wddw &
_wwd=dwwd+wwww & _dww=dddw+wddw & _www=dwww+wwww
calc wddd=( _ddd>0 ) & _ddd=_ddd*( _ddd>0 )+( _ddd.eq.0 ) & wddd=( _ddd>0 ) &
_ddd=_ddd*( _ddd>0 )+( _ddd.eq.0 ) & wddw=( _dwd>0 ) & _dwd=_dwd*( _dwd>0 )+( _dwd.eq.0 )
calc wddw=( _dwd>0 ) & _dwd=_dwd*( _dwd>0 )+( _dwd.eq.0 ) & wddw=( _dwd>0 ) &
_dwd=_dwd*( _dwd>0 )+( _dwd.eq.0 ) & wddd=( _wdd>0 ) & _wdd=_wdd*( _wdd>0 )+( _wdd.eq.0 )

```

```

calc wdwwt=(_dww>0) & _dww=_dww*(_dww>0)+(_dww.eq.0)& wwwwt=(_www>0) &
   _www=_www*(_www>0)+(_www.eq.0)
for i=1...hamonics
  calc sin[i]=sin(i*const*dayNo)
  calc cos[i]=cos(i*const*dayNo)
endfor
MODEL [DISTRIBUTION=gamma; LINK=log;wei=wddd] w_ddd
FIT [PRINT=model,summary,estimates; CONSTANT=estimate; FPROB=yes; TPROB=yes;
    FACT=9]sin[ ]+cos[ ]
RKEEP FITTEDVALUES=fm_wddd

MODEL [DISTRIBUTION=gamma; LINK=log;wei=wddw] w_ddw
FIT [PRINT=model,summary,estimates; CONSTANT=estimate; FPROB=yes; TPROB=yes;
    FACT=9]sin[ ]+cos[ ]
RKEEP FITTEDVALUES=fm_wddw

MODEL [DISTRIBUTION=gamma; LINK=log;wei=wdwd] w_dwd
FIT [PRINT=model,summary,estimates; CONSTANT=estimate; FPROB=yes; TPROB=yes;
    FACT=9]sin[ ]+cos[ ]
RKEEP FITTEDVALUES=fm_wdwd

MODEL [DISTRIBUTION=gamma; LINK=log;wei=wdww] w_dww
FIT [PRINT=model,summary,estimates; CONSTANT=estimate; FPROB=yes; TPROB=yes;
    FACT=9]sin[ ]+cos[ ]
RKEEP FITTEDVALUES=fm_wdww

MODEL [DISTRIBUTION=gamma; LINK=log;wei=wddd] w_wdd
FIT [PRINT=model,summary,estimates; CONSTANT=estimate; FPROB=yes; TPROB=yes;
    FACT=9]sin[ ]+cos[ ]
RKEEP FITTEDVALUES=fm_wdd

MODEL [DISTRIBUTION=gamma; LINK=log;wei=wddw] w_wdw
FIT [PRINT=model,summary,estimates; CONSTANT=estimate; FPROB=yes; TPROB=yes;
    FACT=9]sin[ ]+cos[ ]
RKEEP FITTEDVALUES=fm_wddw

MODEL [DISTRIBUTION=gamma; LINK=log;wei=wwwd] w_wwd
FIT [PRINT=model,summary,estimates; CONSTANT=estimate; FPROB=yes; TPROB=yes;
    FACT=9]sin[ ]+cos[ ]
RKEEP FITTEDVALUES=fm_wwwd

MODEL [DISTRIBUTION=gamma; LINK=log;wei=www] w_www
FIT [PRINT=model,summary,estimates; CONSTANT=estimate; FPROB=yes; TPROB=yes;
    FACT=9]sin[ ]+cos[ ]
RKEEP FITTEDVALUES=fm_www

PEN [RESET=yes] 1,2,3,4,5,6,7,8; METHOD=line; JOIN=ascending; SYMBOL=0; LINESYLE=1
DGRAPH[TITLE='Third Order probabilities']
    Y=fm_wwww, fm_wdww, fm_wwwd, fm_wdwd, fm_wwdw, fm_wddw, fm_wwdd, fm_wddd; X=dayNo
Endif
else
  exit [con=proc;expl='**** Only available to normal order (Maximum Order three')]
endif
endproc

```

Appendix VII: Commands for Menus and dialogue (.grc files)

```

DIALOG dialog1, 260, 208
CAPTION "Counts and Totals from Daily Data"
PROCEDURE "COUNTS"
BEGIN
    PUSHBUTTON "&Run",IDRUN,110, 190, 45, 15,BS_RUN
    PUSHBUTTON "&Reset",IDCLEAR,160,190,45,15,BS_CLEAR
    PUSHBUTTON "&Cancel",IDCANCEL,210,190,45,15,BS_CANCEL
    "PUSHBUTTON "&Default",IDDEFAULT,210,180,30,15,BS_DEFAULTS"
    LTEXT "&Available Data:",ITEXT,5,5,55,10
    LISTBOX IDLIST,5,20,90,120,GW_AVAILABLE
    LTEXT "&Data:",ISTATIC,115,5,60,10
    EDITTEXT DATA,175,5,80,14,GW_AVAILABLE | DT_TYPPVAR | GW_PARAM | EC_SET
    LTEXT "&Class for totals:",ISTATIC,115,25,,55,10
    EDITTEXT CLASS,175,25,80,14,GW_AVAILABLE | DT_TYPPFAC | GW_OPT | EC_SET
    LTEXT "&Type of Markov:",ISTATIC,115,45,,60,10
    COMBOBOX HIGH,175,45,55,40,CB_DROPLIST | GW_OPT
    SLIST HIGH,"Normal","High"
    RLIST HIGH,"NO","YES"
    LTEXT "&Number of States:",ISTATIC,115,65,60,10
    EDITTEXT STATES,190,65,40,11, GW_OPT
    LTEXT "&Threshold Values:",ISTATIC,115,83,,60,10
    EDITTEXT LIMITS,175,83,55,11, GW_OPT
    LTEXT "&Initial Value:",ISTATIC,115,100,,60,10
    EDITTEXT INITIAL,175,100,55,11, GW_OPT
    LTEXT "&Markov Chain Order:",ISTATIC,115,118,,70,10
    EDITTEXT ORDER,190,118,40,11, GW_OPT
    GROUPBOX "Specify the names of the tables", IDSTATIC, 5, 130, 140, 50
    LTEXT "&Amounts",ISTATIC,10,140,40,10
    EDITTEXT a,10,155,60,13, GW_PARAM
    LTEXT "&Counts",ISTATIC,75,140,40,10
    EDITTEXT counts,75,155,60,13, GW_PARAM
    TOGGLEBOX "Display in Spreadsheet", SPREAD, 150, 140, 90, 14,GW_OPT,"YES", "NO"
    IMAGEBUTTON IDPIN, BS_PINDIALOG
    IMAGEBUTTON IDHELP, BS_HELP, "counthelp.html"
END

```

Figure 65: Dialogue command for Count Procedure

```

DIALOG dialog2,210 , 112
CAPTION "Probability of rain"
PROCEDURE "prepare"
BEGIN
    PUSHBUTTON "&Run",IDRUN,50, 93, 45, 15,BS_RUN
    PUSHBUTTON "&Reset",IDCLEAR,105,93,45,15,BS_CLEAR
    PUSHBUTTON "&Cancel",IDCANCEL,160,93,45,15,BS_CANCEL
    "PUSHBUTTON "&Default",IDDEFAULT,210,180,30,15,BS_DEFAULTS"
    LTEXT "&Available Table:",ITEXT,5,5,55,10
    LISTBOX IDLIST,5,20,90,80,GW_AVAILABLE
    LTEXT "&Enter Amounts Here:",ISTATIC,115,5,70,10
    EDITTEXT AMOUNT,115,15,80,14,GW_AVAILABLE | DT_TYPTAB | GW_PARAM | EC_SET
    LTEXT "&Enter Counts Here:",ISTATIC,115,30,,70,10
    EDITTEXT COUNTS,115,40,80,14,GW_AVAILABLE | DT_TYPTAB | GW_PARAM | EC_SET
    TOGGLEBOX "Display in Graph", PLOT, 115,55, 90, 14,GW_OPT,"YES", "NO"
    TOGGLEBOX "Display in Spreadsheet", SPREAD, 115, 65, 90, 14,GW_OPT,"YES", "NO"
    IMAGEBUTTON IDPIN, BS_PINDIALOG | GW_SEL
    IMAGEBUTTON IDHELP, BS_HELP, "myhelp.html"
END

```

Figure 66: *Dialogue command for Prepare Procedure*

```
DIALOG dialog3,220 , 123
CAPTION "Fitting Probabilities of rain"
PROCEDURE "fitting"
BEGIN
    PUSHBUTTON "&Run",IDRUN,50, 105, 45, 15,BS_RUN
    PUSHBUTTON "&Reset",IDCLEAR,105,105,45,15,BS_CLEAR
    PUSHBUTTON "&Cancel",IDCANCEL,159,105,45,15,BS_CANCEL
    "PUSHBUTTON "&Default",IDDEFAULT,210,180,30,15,BS_DEFAULTS"
    LTEXT "&Available Table:",ITEXT,5,5,55,10
    LISTBOX IDLIST,5,20,90,90,GW_AVAILABLE
    LTEXT "&Enter Amounts Here:",ISTATIC,115,5,70,10
    EDITTEXT AMOUNT,115,15,80,14,GW_AVAILABLE | DT_TYPTAB | GW_PARAM | EC_SET
    LTEXT "&Enter Counts Here:",ISTATIC,115,35,,70,10
    EDITTEXT COUNTS,115,45,80,14,GW_AVAILABLE | DT_TYPTAB | GW_PARAM | EC_SET
    LTEXT "Number of Hamonics:",ISTATIC,115,70,,80,10
    EDITTEXT HAMONICS,115,80,40,14,GW_AVAILABLE | DT_TYPTAB | GW_OPT
    IMAGEBUTTON IDPIN, BS_PINDIALOG
    IMAGEBUTTON IDHELP, BS_HELP, "myhelp.html"
END
```

Figure 67: *Dialogue command for Fitting probability Procedure*

```
DIALOG dialog4,220 , 123
CAPTION "Fitting Amounts of rain"
PROCEDURE "AmountFitting"
BEGIN
    PUSHBUTTON "&Run",IDRUN,50, 105, 45, 15,BS_RUN
    PUSHBUTTON "&Reset",IDCLEAR,105,105,45,15,BS_CLEAR
    PUSHBUTTON "&Cancel",IDCANCEL,159,105,45,15,BS_CANCEL
    "PUSHBUTTON "&Default",IDDEFAULT,210,180,30,15,BS_DEFAULTS"
    LTEXT "&Available Table:",ITEXT,5,5,55,10
    LISTBOX IDLIST,5,20,90,90,GW_AVAILABLE
    LTEXT "&Enter Amounts Here:",ISTATIC,115,5,70,10
    EDITTEXT AMOUNT,115,15,80,14,GW_AVAILABLE | DT_TYPTAB | GW_PARAM | EC_SET
    LTEXT "&Enter Counts Here:",ISTATIC,115,35,,70,10
    EDITTEXT COUNTS,115,45,80,14,GW_AVAILABLE | DT_TYPTAB | GW_PARAM | EC_SET
    LTEXT "Number of Hamonics:",ISTATIC,115,70,,80,10
    EDITTEXT HAMONICS,115,80,40,14,GW_AVAILABLE | DT_TYPTAB | GW_OPT
    IMAGEBUTTON IDPIN, BS_PINDIALOG
    IMAGEBUTTON IDHELP, BS_HELP, "myhelp.html"
END
```

Figure 68: *Dialogue command for Fitting Amounts Procedure*

Appendix VIII: Commands for Menus and dialogue (.gad files)



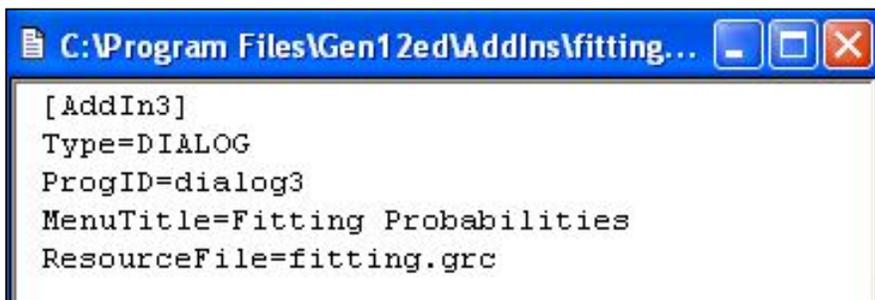
```
[AddIn1]
Type=DIALOG
ProgID=dialog1
MenuTitle=Counts and Totals from Daily Data
ResourceFile=counts.grc
```

Figure 69: .gad file for count dialogue



```
[AddIn2]
Type=DIALOG
ProgID=dialog2
MenuTitle=Probability of Rain
ResourceFile=prepare.grc
```

Figure 70: .gad file for Prepare dialogue



```
[AddIn3]
Type=DIALOG
ProgID=dialog3
MenuTitle=Fitting Probabilities
ResourceFile=fitting.grc
```

Figure 71: .gad file for Fitting probability dialogue



```
[AddIn4]
Type=DIALOG
ProgID=dialog4
MenuTitle=Fitting Amounts
ResourceFile=FittingAmount.grc
```

Figure 72: .gad file for fitting amount dialogue