MODELLING FIRE WEATHER INDEX SERIES

MODELLING FIRE WEATHER INDEX SERIES

By

SHAHZAIB BARLAS, M.Sc.

A Project
Submitted to the School of Graduate Studies
in Partial Fulfillment of the Requirements
for the Degree
Master of Science

McMaster University

MASTER OF SCIENCE (2005)	McMaster University
(Statistics)	Hamilton, Ontario

TITLE: Modelling Fire Weather Index Series

AUTHOR: Shahzaib Barlas
B.Sc (Honours), M.Sc. in Applied Statistics
University of Karachi, Pakistan

SUPERVISOR: Dr. Silvia Easterby

NUMBER OF PAGES: xii, 82

To

My late Father Mirza Abban Barlas & my Mother:

without her prayers I would not be able to finish this

&

my wonderful daughter Samrah Barlas:

who always wait me so late to take dinner with me

(Samrah: Papa love you so much)

Table of Contents

List of Tables

List of Figures

ABSTRACT

The fire weather index (FWI), useful as a measure of forest fire danger, is calculated from precipitation and other weather variables. In the present environmental study, precipitation, fuel moisture codes, and fire behavior indices were available for a reference site and 4 higher elevation sites around Smithers, British Columbia. The objective of the study was to determine whether the use of local precipitation would lead to a different FWI than obtained from precipitation at the reference site.

The features of the series of daily FWI values which needed to be taken into account were: peaks following dry periods, serial correlation, and heteroscedasticity. Two types of models were developed to characterize the record as a smooth component, for the upward and downward movements of the index, and a component of correlated error terms. The first type was a parametric Fourier series in a context of a generalized linear model (GLS) that allowed for serial correlation and heteroscedasticity. The second form was a smoothing cubic spline with a bootstrap procedure for estimation of standard errors and confidence bands. The question, of whether FWI on a particular day differed between a higher elevation station and the reference station, was addressed by adding a station effect to the GLS model and by graphical comparison of the smooth curves with confidence bands for the spline method.

The Model-3 for the combined station effect is not able to capture the sharpness of the peak and found insignificant while cubic spline smoothing curves fitted to the bootstrap behave well to capture peaks and troughs in the index but it encounter some difficulties for few lower index values.

Acknowledgements

All praises for Almighty Allah, the most beneficent and merciful, who gave me courage and provide me this opportunity that I could be able to finish this research project.

I would like to express my deepest gratitude to my supervisor, Dr. Sylvia Esterby for her constant guidance, useful comments and suggestions to improve the quality of my work. I am thankful to all professors and staffs in the Department of Mathematics and Statistics at McMaster University, particularly to my committee members Dr. Peter Macdonald, and Dr. Rong Zhu for their helpful suggestions and valuable advices.

I wish my sincere thank Dr. Judi Beck, Manager Fire Management Section, and Bradley Martin, North West Fire Center, Protection Branch, British Columbia Ministry of Forests and Range, Canada for providing the data.

I am also thankful Binod Prasad Neupane, Enayetur Raheem and Mallikarjuna Rao Rettiganti, and others that they really encouraged me a lot during my study, and proved to be very good friends.

Finally, I will give special thank to my mother, my brothers, and sister who always pray here and abroad for my success, and my wife who always encourage me and stood with me during my hard time.

Chapter 1

Introduction

1.1 Background

Precipitation plays an important role to reduce forest fire danger rating and it is one of the key ingredient to calculate daily severity rating that base on fire weather index (FWI). It is an index that represent the fire intensity of spreading fire as energy output rate per unit length of fire front (Van Wagner 1987) and based on precipitation and other weather variables known as fuel moisture codes and fire behavior indices.

Precipitation has a major influence on all human activities. It deposits on surfaces from air either in liquid or solid form. Dew, fog, frost, rain and snow are the different forms of precipitation. It is a natural unlimited gift; one of the basic needs of human life and of other living creatures.

On the other hand fire is a threat to lives, homes and businesses. Forest fire is a very dangerous form of fire because it may take lives, houses, businesses and future sources of income. Fire is one of nature's primary carbon-cycling mechanisms but human activity interferes with this, causing some 60% of forest fires (Higgins & Ramsey 1992). Canada has 420 million hectares, or 10% of the world's forest area, (Canadian Forest Service 1999), and annually experiences an average of 9,500 wildland fires that burn more than 3 million hectares of forest (Canadian Forest Service 2005a).

Canada is a participant in several international initiatives including the Kyoto Protocol to the UN Framework Convention on Climate Change, the International Convention on Biological Diversity, and the Montreal Process (Working Group on Criteria and Indicators for the Conservation and Sustainable Management of Temperate and Boreal Forests). These agreements commit Canada to reducing anthropogenic carbon emissions, maintaining biodiversity, and practicing sustainable forest management, goals which cannot be met without a thorough understanding of the impacts of forest fires and of the role of climate change on forest fires (Canadian Forest Service 2001).

Extensive resources are needed to extinguish a forest fire, even when it is small. Thus management and forest-fire fighters need the right information to be able to issue fire-danger ratings which will save human lives and precious resources. One such forest fire danger rating is the Daily Severity Rating (DSR) which is calculated from the fire weather index (FWI) on a daily basis.

1.2 The Fire Weather Index

Research on forest fire danger rating was begun in Canada by J.G. Wright in 1925. In 1970 the first Canadian forest fire weather index (FWI) was issued (Van Wagner & Pickett 1985, pp 1). Since then, revised versions were issued in 1976 and 1984, and several improvements have been made. Currently the Van Wagner & Pickett (1985) FORTRAN program is used to calculate the Fire Weather Index.

1.2.1 FWI Components

The FWI represents the fire intensity of the spreading fire (Figure 1.1) and is useful as a general measure of forest fire danger in Canada. FWI is a mixture of the fuel moisture content and fire behavior indices, with wind playing a key role in the calculations. FWI should be used for a single daily value only (Van Wagner & Pickett 1985).

The standard definitions for FWI components are as follows (Canadian Forest Service 2005b, Van Wagner 1987, Leathwick & Briggs 2001).

Fine Fuel Moisture Code (FFMC) represents the moisture content of litter and other cured fine fuels. It is an indicator of the relative ease of ignition and flammability of fine fuel.

Duff Moisture Code (DMC) represents the average moisture content of loosely compacted organic layers of moderate depth. This gives an indication of fuel consumption in moderate duff layers and medium-size woody material.

Drought Code (DC) represents a deep layer of compact organic matter. This is useful indicator of seasonal drought effects on forest fuels and the amount of smoldering in deep duff layers and large logs.

Initial Spread Index (ISI) is a combination of wind and the FFMC that represents rate of spread alone without the influence of variable quantities of fuel.

Buildup Index (BUI) is a combination of DMC and DC that represents the total amount of fuel available for the fire.

The Daily Severity Rating (DSR) is a computed from the FWI and reflects the expected efforts required for fire suppression. It is calculated as

$$\mathrm{DSR} = 0.0272\ (\mathrm{FWI})^{1.77}$$

1.2.2 Calculation of FWI

The measurements needed for the moisture content codes are daily rain, temperature, wind and humidity all of ideally be recorded at noon. Wind also plays a key role in the fire behavior indices as shown in Figure 1.1. Rain and temperature are recorded to the first decimal place, while relative humidity and wind speed are recorded as a whole number. The Van Wagner & Pickett (1985) FORTRAN program, used to calculate FWI and components, has some standard starting values for the moisture content codes; these are 85 for FFMC, 6 for DMC, and 15 for DC. Criteria for adjusting starting values for the moisture codes and for the first recording day for meteorological variables, dependent primarily upon snow cover, are discussed by Turner & Lawson (1978).

1.3 Objective of The Project

The main objective of the project was to determine whether the use of the local precipitation amount would lead to a different fire weather index (FWI) than obtained if the precipitation at a nearby weather station is used. We have a precipitation data set for a reference station, and for 4 sites at higher elevations. Throughout this report, higher elevation stations are abbreviated as ES with a suffix that shows the height above sea level. The reference station is abbreviated as RS532.

The following chapters are organized as follows. In Chapter 2, we describe the nature of the data, discuss generalized least square and fit a model in a context of a Fourier series, that allow serial correlation and heteroscedasticity in the model. In Chapter 3, we fit cubic spline smooth curve to the given data set and describe a bootstrap residuals method. In Chapter 4, we extend the idea of chapters 2 and 3 to two stations, and performed a comparison of regression curves based on both and on fitted cubic spline smooth curves using the bootstrap method. In Chapter 5, we have concluding remarks and future research proposals .

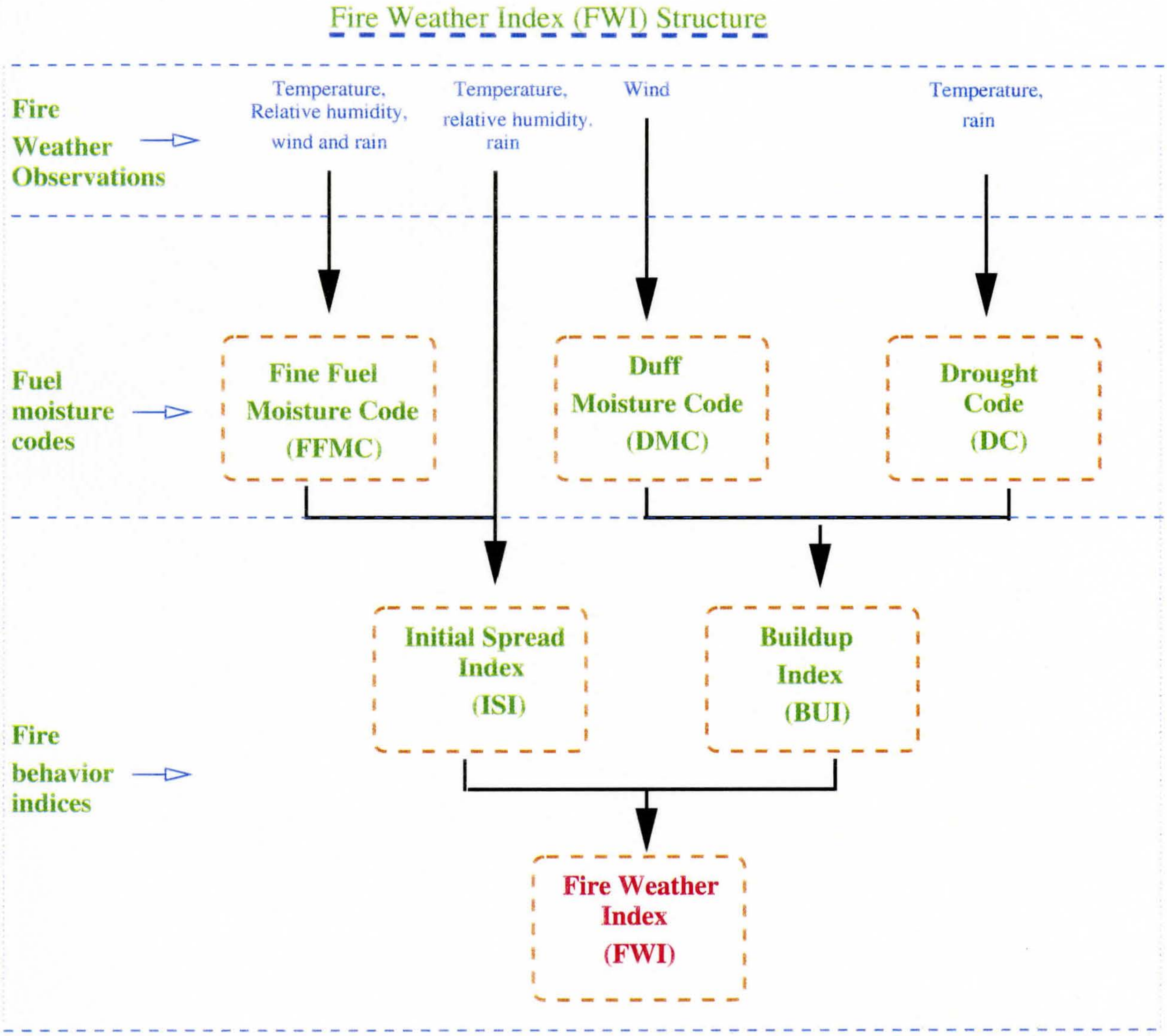


Figure 1.1: A typical Fire weather index structure (Canadian Forest Service 2005b)

Chapter 2

Data Description and Models

2.1 Data Description

The data set consists of weather measurements and calculated variables for locations in the area around Smithers, British Columbia, part of the region of responsibility of the North West Fire Centre, Protection Branch, British Columbia Ministry of Forests and Range. The reference site is the weather station at the Smithers Airport, BC., 532 m meter above sea level. The 4 other sites are at elevations of 855, 1017, 1166, and 1302 m above sea level and receive precipitation differs in amount from the reference site. The measurements were taken between May 1 and September 15, 2003, during the fire season for the high elevation sites. The four weather measurements, temperature, relative humidity, wind and 24-hour precipitation (1 PM to 1 PM), were made at the reference site, but only precipitation was measured at the high elevation sites, with the other weather measurements extrapolated on the basis of elevation using the elevation grid. The temperature adjustment is based on the United States standard Lapse Rate of $-6.5°C/km$, that is, for every kilometer gain in elevation, the temperature is assumed to drop by 6.5°C (Canadian Forest Service 2005d). The three fuel moisture codes, two fire behavior indices, and the FWI, described in Chapter 1, have been calculated using the Van Wagner & Pickett (1985) FORTRAN program. The precipitation, fuel moisture codes, fire behavior indices and FWI for the five sites were supplied by Bradley Martin of the North West Fire Centre.

Usually April 1 to September 30 is known as the fire season because this is the period in which fires occur in most areas of the country. However data collection depends on when the snow melts and when spring starts at a particular site (Turner & Lawson 1978). In the present data set, data collection started May 1, 2003, at the reference site (day 1) and ended September 15 (day 138). At the higher elevations day 1 was May 9, May 21, May 19, and June 10 for the 855 m, 1017 m, 1166 m, and 1302 m stations respectively.

The total accumulated precipitation (Table 2.1) at RS532 is 219.8 mm and which is lower than the other stations: ES855 (247.14 mm), ES1017 (222.9 mm), ES1166 (298.3 mm), and ES1302 (224.5 mm). Thus higher elevation stations have lower FWI compared to the reference site (Table 2.2). The FWI scatter plots Figure 2.1(b) for RS532 (more plots can be seen in the Appendix A (A.3)) and Figure 2.1(c) for precipitation RS532 and FWI RS532 when they plotted against day t. There is upward and downward movement of FWI in Figure 2.1(c) which is decreases when precipitation has been high and increases when there is no precipitation and a pair plot between Reference site FWI and precipitation shows exponential relationship (Figure 2.1(d)). Figure 2.2 shows precipitation, that is the main factor in FWI because all fuel moisture codes, namely FFMC, DMC, and DC that also shown there heavily depend on it along with wind, humidity and temperature. Figure 2.3 shows fire behavior indices and FWI structure. It is also important to note that, if precipitation was not measured due to the late snow melting at a particular site, then we will not have

Table 2.1: Summary Statistics, Reference and higher elevation sites Precipitation

	RS532	ES855	ES1017	ES1166	ES1302
Sum	219.8	247.14	222.9	298.3	224.5
Mean	1.59	1.9	1.89	2.49	2.29
Sdev	4.45	4.97	3.98	5.96	4.95
Q1	0.0	0.0	0.0	0.0	0.0
Q3	1.3	1.95	1.58	2.5	1.58
Min	0.0	0.0	0.0	0.0	0.0
Max	31.4	37.3	23.3	42.4	26.6

Table 2.2: Summary Statistics, Reference and higher elevation sites FWI

	RS532	ES855	ES1017	ES1166	ES1302
Sum	1040.3	622.4	435.6	317.0	178.1
Mean	7.5	4.8	3.7	2.6	1.8
Sdev	7.54	5.61	4.81	4.12	2.56
Q1	0.82	0.37	0.11	0.1	0.02
Q3	12.14	8.1	6.14	3.9	2.54
Min	0.0	0.0	0.0	0.0	0.0
Max	35.81	26.92	22.06	21.01	10.76

values for all 6-derived components.

2.2 Modelling Approaches

The precipitation at the reference site is normally used to provide the precipitation values to calculate an index for higher elevation sites. The moisture codes and fire behavior indices are calculated from deterministic equations, taking as input the weather measurements and thus it is reasonable to model all of the variables by including a random component. In addition to serial correlation that may be present in the data, the calculations include equations that assume cumulative effects. Figures 2.2 and 2.3 lead us to a model that accounts for regular peaks and troughs in the component that describe change in level over time and dependency on past values. This finally leads us to a conceptual linear model that best describes the smooth and rough components in the model

$$Y_t = S_t + R_t \tag{2.2.1}$$

where Y_t is a response variable on day t at a particular station site, S_t is a smooth component and R_t is a rough component. We obtained smooth curves in two different ways and accounted serial correlation in the rough component.

Since the interest was in FWI, and the intermediate calculations are complex, it was decided to consider only FWI in the reminder.

2.2.1 Linear Regression Models

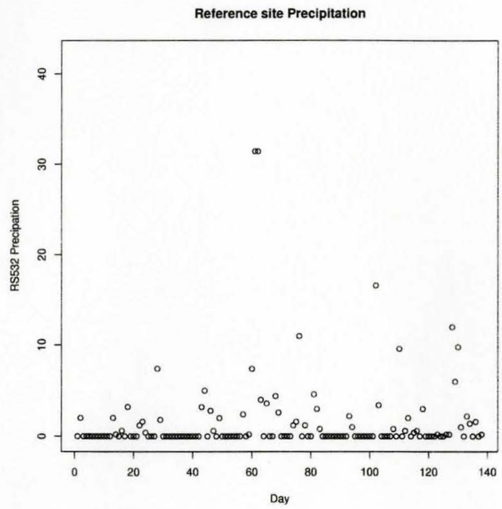
In Model-1, we adopt multiple linear regression model that is the extension of the simple linear regression model, in which we have one dependent and p independent variables, $p \geq 2$,

$$Y = \beta_0 + \sum_{i}^{p} \beta_i X_i + \epsilon \tag{2.2.2}$$

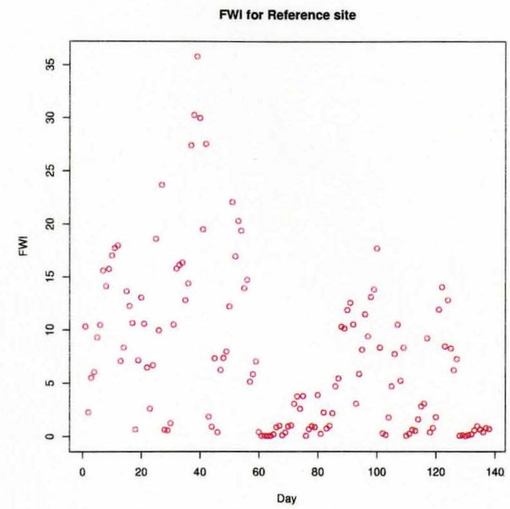
A Fourier series model is a function of sine and cosine waves added together. Sine and cosine waves have amplitude, frequency and phase angle for the horizontal phase shift. The multiple linear regression model in Fourier form can be written as

$$Y_t = \beta_0 + \sum_{s=1}^{k} \left[\beta_{1s} \sin \frac{st}{T} 2\pi + \beta_{2s} \cos \frac{st}{T} 2\pi \right] + \epsilon_t \tag{2.2.3}$$

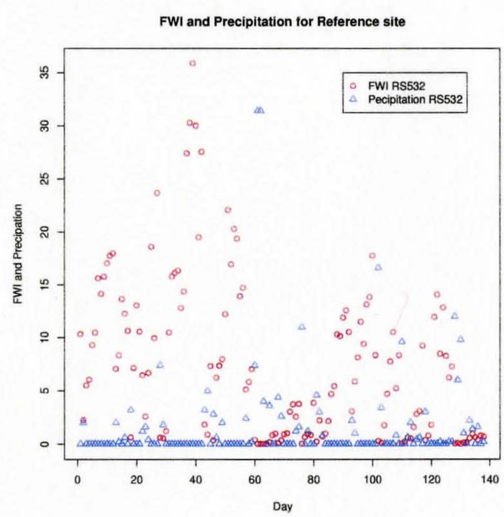
where t is the day of measurement and T is total number of days, in the series, $s = 1, 2, 3, \cdots, k$ and $t = 1, 2, \cdots, T$; $T \leq 138$ (according to the station).



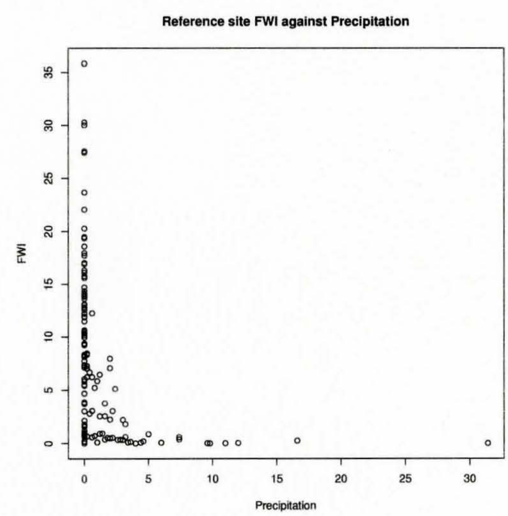
(a) Precipitation for Reference Site



(b) FWI for Reference Site (RS532)

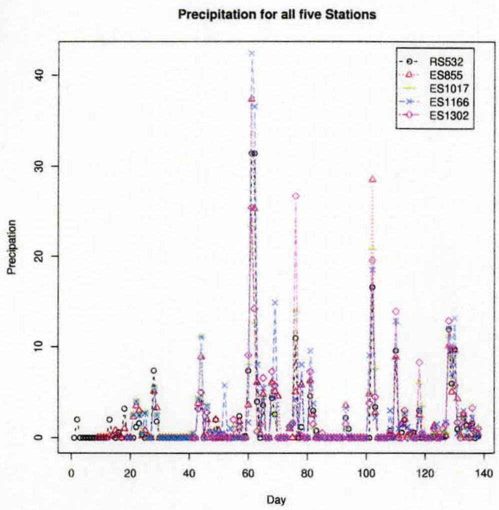


(c) Precipitation and FWI for Reference Site

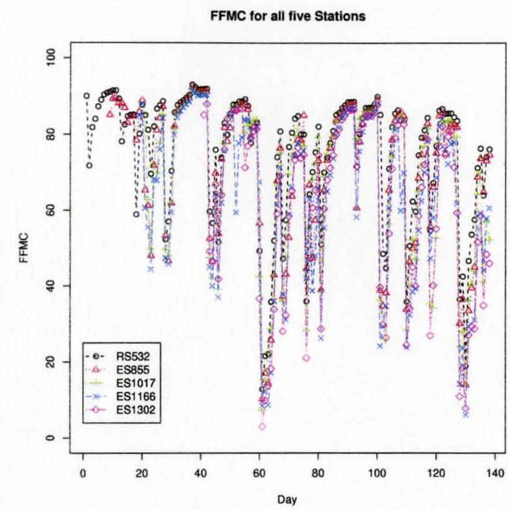


(d) Reference Site FWI against Precipitation

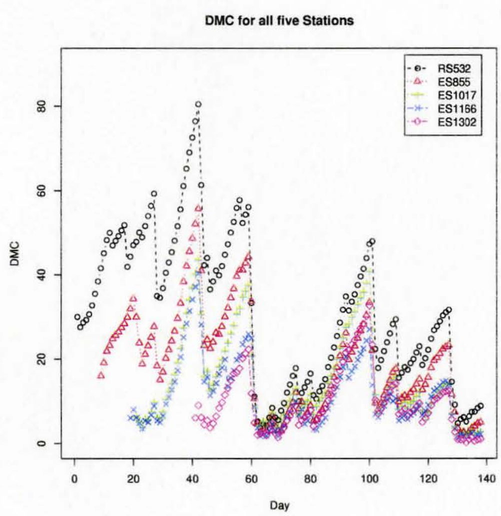
Figure 2.1: FWI and Precipitation plots for Reference Site



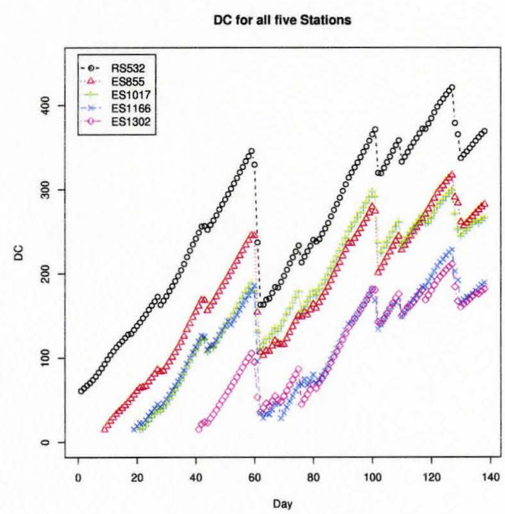
(a) Precipitation (24 hour, in mm) points join by the line



(b) FFMC points join by the line

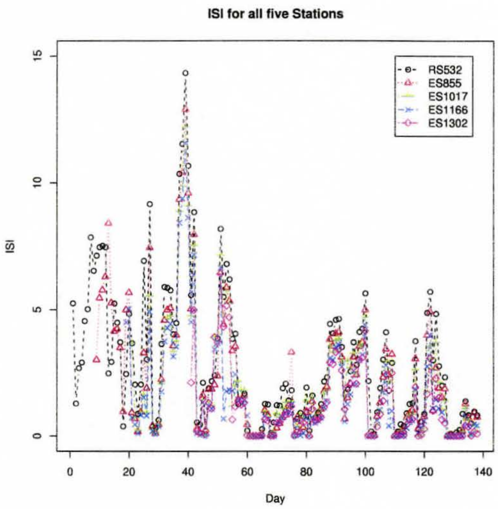


(c) DMC points join by the line

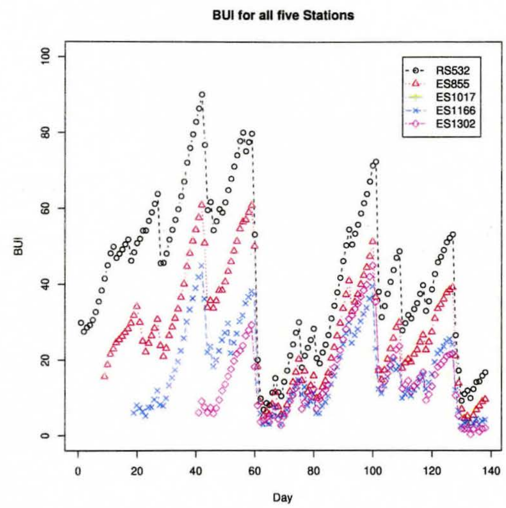


(d) DC points join by the line

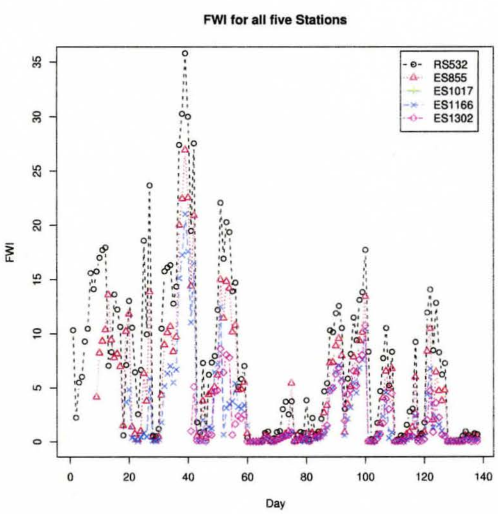
Figure 2.2: Precipitation and Fuel Moisture codes points join by line for all Stations



(a) ISI points join by the line



(b) BUI points join by the line



(c) FWI points join by the line

(d) FWI structure with 6-derived components

Figure 2.3: FWI Structure and Fire behavior Indices points join by the line for all Stations

If we take

$$X_{1s} = \sin\frac{st}{T}2\pi \tag{2.2.4}$$

$$X_{2s} = \cos\frac{st}{T}2\pi \tag{2.2.5}$$

then Equation 2.2.3 can be rewritten in the multiple linear regression form

$$Y_t = \beta_0 + \sum_{s=1}^{k} [\beta_{1s}X_{1s} + \beta_{2s}X_{2s}] + \epsilon_t \tag{2.2.6}$$

where Y_t is the response variable on day t, X_{is} is the regressor Fourier variable for the ith level and ϵ_t is the residual on day t. In a linear regression model it is common to assume that

$$\epsilon_t \overset{iid}{\sim} N(0, \sigma^2)$$

which means errors are independently and identically normally distributed with mean zero and constant variance σ^2.

2.2.2 Serial Correlation

In Model-2, we account serial correlation that is most common when we are considering a data set that is changing over time and depending on its past value (Buonaccorsi 2002, Greene 1993). In such data sets, residuals are correlated across time and give rise to autocorrelation. Hence we can not assume the residuals are identically and independently distributed.

Here we will focus only on the first order autoregressive $AR(1)$ model, where, the error ϵ_t is correlated with its own previous term ϵ_{t-1}.

$$\epsilon_t = \rho\epsilon_{t-1} + u_t, \qquad -1 < \rho < 1 \tag{2.2.7}$$

where

$$\begin{aligned} u_t &\overset{iid}{\sim} N(0, \sigma_u^2),\ \forall\, t = 1, 2, \cdots, T \\ \text{Cov}(u_t, u_k) &= 0,\ \text{if}\ \ t \neq k. \end{aligned}$$

Equation 2.2.7 is the first order autoregressive model $AR(1)$ because ϵ_t is linearly related its own past value ϵ_{t-1} for order one. In the same fashion we may define the $AR(r)$ model.

$$\epsilon_t = \sum_{k=1}^{r} (\rho_k \epsilon_{t-k}) + u_t \tag{2.2.8}$$

Since, $-1 < \rho < 1$ and $\rho \to 0$ as $k \to \infty$, thus $\lim_{k \to \infty} \rho^k \to 0$. Therefore equation 2.2.7 and 2.2.8 will yield

$$\begin{aligned} E(\epsilon_t) &= E(\rho\epsilon_{t-1} + u_t) \\ E(\epsilon_t) &= 0 \\ \text{Var}(\epsilon_t) &= \rho^2 \text{Var}(\epsilon_{t-1}) + \sigma_u^2 \\ &= \sigma_u^2/(1 - \rho^2) \\ \text{Cov}(\epsilon_t, \epsilon_{t-1}) &= E(\epsilon_t, \epsilon_{t-1}) \\ &= \rho \text{Var}(\epsilon_{t-1}) \\ &= \rho \frac{\sigma_u^2}{1 - \rho^2} \\ \text{Corr}(\epsilon_t, \epsilon_{t-1}) &= \frac{\text{Cov}(\epsilon_t, \epsilon_{t-1})}{\sqrt{\text{Var}(\epsilon_t)\text{Var}(\epsilon_{t-1})}} \\ &= \rho \end{aligned}$$

Thus, autocorrelation for the rth order autoregressive model will be

$$\text{Corr}(\epsilon_t, \epsilon_{t-k}) \Rightarrow \rho_k = \rho^k \tag{2.2.9}$$

and the covariance matrix $\mathrm{Cov}(\epsilon) = \sigma_u^2 \boldsymbol{\Omega}$ is

$$\frac{\sigma_u^2}{1-\rho^2} \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{T-1} \\ \rho & 1 & \rho & \dots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{T-3} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho^{T-1} & \rho^{T-2} & \dots & \dots & 1 \end{pmatrix}$$

2.2.3 Heteroscedasticity

In Model-3, we test heterosecdasticity in the model. It arises when residuals are not constant. For a data set that is collected over time at different locations, heteroscedasticity is common (Buonaccorsi 2002, Greene 1993). This leads us higher standard error in the residuals. If errors are pairwise uncorrelated then the generalized least square estimator $\hat{\beta}$ is

$$\begin{aligned} \hat{\beta} &= (X'\Omega X)^{-1} X' \Omega^{-1} y \\ E(\epsilon_t) &= 0 \\ \mathrm{Var}(\epsilon_t) &= \sigma_t^2 \boldsymbol{\Omega} \end{aligned}$$

$$\boldsymbol{\Omega} = \begin{pmatrix} \omega_1 & 0 & 0 & \dots & 0 \\ 0 & \omega_2 & 0 & \dots & 0 \\ 0 & 0 & \omega_3 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & \omega_T \end{pmatrix}$$

Thus Ω^{-1} is diagonal matrix whose ith element is $1/\omega_t$. Hence the GLS estimator $\hat{\beta}$ is now the weighted generalized least square estimator

$$\hat{\beta} = \left[\sum_t \omega_t \mathbf{x_t} \mathbf{x_t'} \right]^{-1} \left[\sum_t \omega_t \mathbf{x_t} y_t \right]$$

and weights, $w_t = 1/\omega_t, \quad t = 1, 2, \cdots, T$.

2.3 Data Analysis

We are using a multiple linear regression model to analyze the data since scatter plots in Figure 2.3 suggests that we may model the data using linear combination of sine and cosine waves. It also suggests that we need to accommodate peaks and troughs by smoothing the model as we stated earlier in Equation 2.2.1.

In this Chapter we fit multiple linear regression model to the FWI, and we will consider smoothing in the next chapter. We fit a linear model to FWI for different values of s. However, due to many zeros in FWI we are not able to fit a model for two stations separately for the value of $s < 5$. Since algorithm of the program did not allow us to converge the model for all stations so we are not able to fit the same model for all stations except two stations, RS532 and ES1017, where we able to fit the model for $s = 5$. Hence in this chapter we discuss the results for $s = 5$. We have presented the results for $s = 2$, and 4 only for RS532 in the Appendix A (A.1).

We used the equations 2.2.4 and 2.2.5 to calculate sine and cosine terms and a restricted maximum likelihood method in generalized least square (gls) function, by default, to estimate the parameters in Splus. It also allow correlated error structure and heteroscedasticity (Splus Development Core Team. 2002) in the model.

Table 2.3 below shows the result of a linear model after using the gls function for $s = 5$. It is obvious from Table 2.3 the alternate sin or cos parameters are significant at 5% level except for sin8 and cos8 terms. The plot of the standardized residuals against the fitted values which determines the adequacy of the distribution for the fitted model is shown in Figure 2.4 and it indicates non consistency of variance. It also revel a diagonal shape that can be observed here and other residuals plots that is occurring due to minimum or zero FWI in the model (Figure 2.1(b)).

Also, serial dependence has been considered since FWI is calculated in a way that assumes a cumulative effect. To check this we need to calculate serial correlation in residuals. Figure 2.5 also shows autocorrelation existing in the model, and suggest an autoregressive progressive process of order 1 AR(1), may be good to model the serial correlation in the residuals also see Figure 2.6 for partial autocorrelation function plot for response variable with 5% asymptotic boundary.

Table 2.3: FWI RS532 Estimated Parameters in Model-1

Parameters	Value	SE	t-value	p-value
(Intercept)	7.532875	0.4965967	15.16900	<0.0001
$\sin 2t\pi/T$	4.203463	0.7012940	5.99387	<0.0001
$\cos 2t\pi/T$	0.718448	0.7023373	1.02294	0.3083
$\sin 4t\pi/T$	0.274679	0.7012931	0.39167	0.6960
$\cos 4t\pi/T$	-3.883595	0.7023379	-5.52953	<0.0001
$\sin 6t\pi/T$	0.547333	0.7016092	0.78011	0.4368
$\cos 6t\pi/T$	2.620609	0.7027078	3.72930	0.0003
$\sin 8t\pi/T$	0.899074	0.7012931	1.28202	0.2022
$\cos 8t\pi/T$	-0.824473	0.7023379	-1.17390	0.2426
$\sin 10\pi/T$	1.898425	0.7012940	2.70703	0.0077
$\cos 10t\pi/T$	-2.162301	0.7023374	-3.07872	0.0025

AR1 Model

We can use the first lag correlation value in our model, obtained from Table 2.8 [1] column two, and that is 0.55. Once we model the data with correlation structure AR1(0.55) (Table 2.4) we do not have any significant parameter at 5% level except the intercept. Autocorrelation is still same for this model see Model-2 in table 2.8. Moreover, it increase residual standard error from 5.83 to 7.34, and yielded almost

[1] Where Model-1 represent OLS, Model-2 is GLS with AR1 and Model-3 is GLS with AR1 and heteroscedasticity

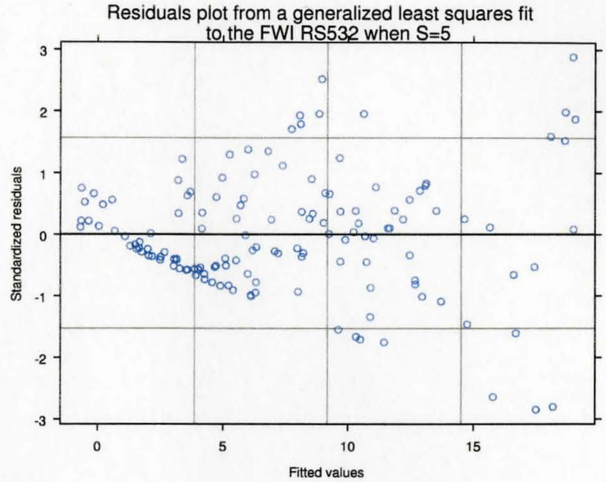


Figure 2.4: FWI RS532 Residual plot from a GLS Fourier Model-1

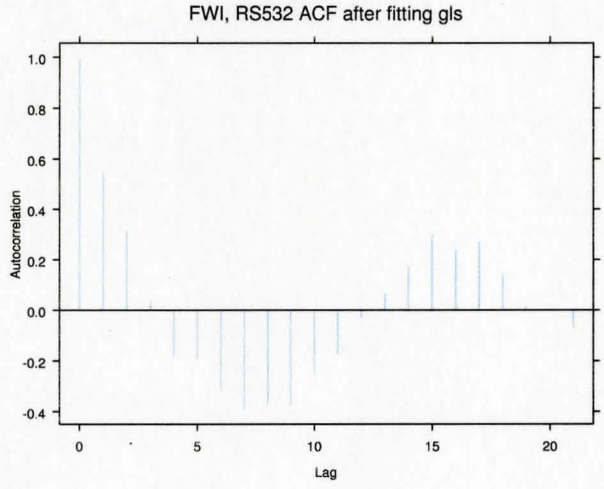


Figure 2.5: FWI RS532 ACF after fitting GLS Fourier Model-1

same lag values. Plot of residuals (Figure 2.7) shows very little improvement and still has extra variation in residuals in same form. Therefore the overall model with AR1(.55) does not show better fit.

Table 2.4: FWI RS532 Estimated Parameters with AR1 in Model-2

Parameters	Value	SE	t-value	p-value
(Intercept)	7.560747	1.563000	4.837331	<0.0001
$\sin 2t\pi/T$	4.045576	2.216941	1.824847	0.0704
$\cos 2t\pi/T$	0.808086	2.156049	0.374800	0.7084
$\sin 4t\pi/T$	0.168843	2.152466	0.078442	0.9376
$\cos 4t\pi/T$	-3.816051	2.096331	-1.820348	0.0711
$\sin 6t\pi/T$	0.444791	2.057649	0.216164	0.8292
$\cos 6t\pi/T$	2.733858	2.008477	1.361160	0.1759
$\sin 8t\pi/T$	0.735058	1.941644	0.378575	0.7056
$\cos 8t\pi/T$	-0.800277	1.900142	-0.421167	0.6743
$\sin 10t\pi/T$	1.740789	1.819437	0.956773	0.3405
$\cos 10t\pi/T$	-2.118384	1.784663	-1.186994	0.2374

Heteroscedastic Model

Since overall model with AR1(0.55) does not show any improvement that lead us to further investigate the model. The Residual plot also shows that the variance is not constant across the predicted values (Figure 2.4 and 2.7) since it is not spread equally. This indicate heterogeneity in the residuals and to accommodate this in the model, we can use the weights option [2] in our gls function with the AR(1) correlation structure. Hence we introduce weighted generalized least square to account heteroscedasticity in our model.

[2] We used varPower function in Splus that represent a power variance covariate $s2(v) = |v|^{(2\times t)}$; where v denote the variance co-variate and $s2(v)$ is a variance function evaluated at v and t is the variance function coefficient (Splus Development Core Team. 2002, Variance Function Classes)

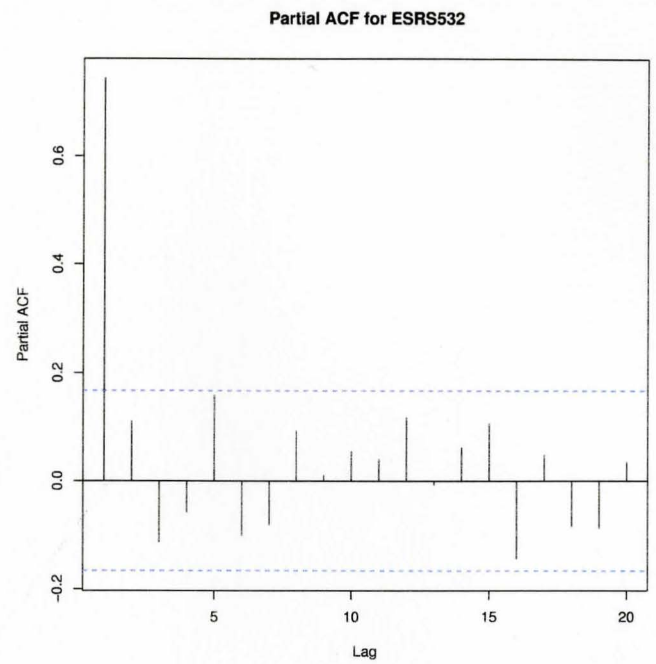


Figure 2.6: FWI RS532 Partial ACF

Table 2.5: FWI RS532 Correlation Matrix with AR1 in Model-2

	(Intr)	sin2π	cos2π	sin4π	cos4π	sn6π	cos6π	sin8π	cos8π	sin10π
sin2π	-.001									
cos2π	-.042	-.001								
sin4π	-.002	.000	-.003							
cos4π	-.039	-.001	-.058	-.002						
sin6π	-.003	.000	-.004	.001	-.004					
cos6π	-.039	-.001	-.055	-.002	-.053	-.004				
sin8π	-.003	.001	-.005	.000	-.004	.002	-.004			
cos8π	-.035	-.001	-.052	-.002	-.049	-.004	-.047	-.004		
sin10π	-.004	.003	-.005	.002	-.005	.002	-.005	.003	-.005	
cos10π	-.034	-.001	-.047	-.002	-.046	-.003	-.045	-.004	-.041	-.005

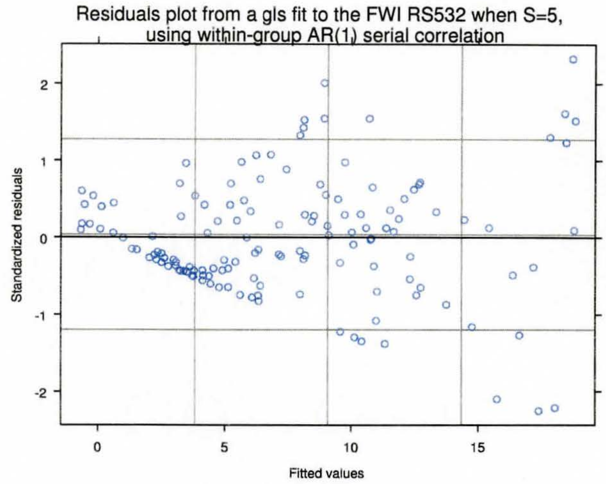


Figure 2.7: FWI RS532 Residual plot after AR(1) in Model-2

After fitting model-3 there is much improvement (Table 2.6). The alternate sin or cos parmeters are significant at 5% level including the sin8 parameter that was not significant before. The correlation matrix in (Table 2.7) also shows noticeable variance-covariance structure exists in the model as compared to the Table 2.5, where we introduced serial correlation. After introducing heteroscedasticity in the model the residual standard error also went down to 1.60 as compared to the 5.83 and 7.34, respectively. The smaller residual standard error also shows that heteroscedasticity was existing in the previous model before. This also demonstrates the earlier statement, that due to unequal variances in residuals, we get the large standard error if we use ordinary least squares. Residual plot (Figure 2.9) has also improved a lot relative to the previous two Figures 2.4 and 2.7. But still we see some departure from the ideal band of constant width. We also plotted residuals against day t (Figure 2.10).

Autocorrelation also went down from 0.55 to 0.45 (Table 2.8) and autocorrelation plot (Figure 2.11) shows much smaller values at higher lags as compare to the previous

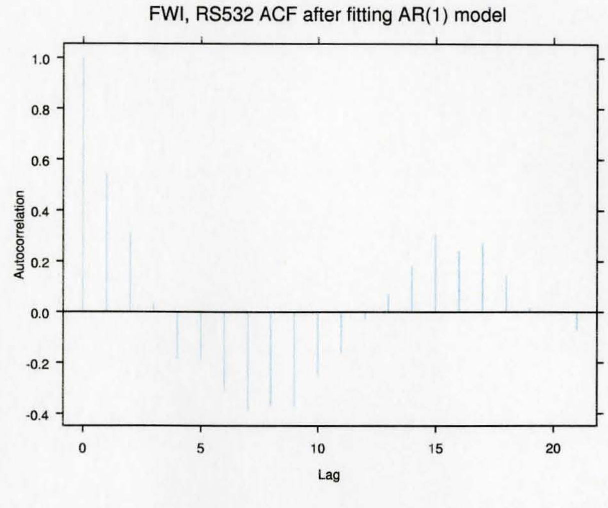


Figure 2.8: FWI RS532 ACF after fitting AR(1) in Model-2

two models. Above 3-Models are not nested but Table 2.9 shows a likelihood ratio test that clearly indicate the fitted Model-3 with AR1 and heteroscedasticity is good as compare to the other models since it has minimum AIC and BIC criteria. These criteria provide alternatives to measure the goodness of fit and show that the model fits adequately.

Table 2.6: FWI RS532 Parameters with AR1 and Heteroscedasticity in Model-3

Parameters	Value	SE	t-value	p-value
(Intercept)	7.836735	1.094279	7.161549	<0.0001
sin2tπ/T	5.347672	1.668822	3.204459	0.0017
cos2tπ/T	0.679831	1.091037	0.623105	0.5343
sin4tπ/T	0.261597	1.434091	0.182413	0.8555
cos4tπ/T	-4.650954	1.321436	-3.519621	0.0006
sin6tπ/T	0.750334	1.337100	0.561165	0.5757
cos6tπ/T	3.595814	1.362416	2.639292	0.0093
sin8tπ/T	2.664587	1.113808	2.392321	0.0182
cos8tπ/T	-1.462143	1.046406	-1.397300	0.1648
sin10tπ/T	-0.304928	0.946292	-0.322235	0.7478
cos10tπ/T	-3.598100	0.942107	-3.819207	0.0002

Table 2.7: FWI RS532 Corr. Matrix with AR1 and Heteroscedasticity in Model-3

	(Intr)	sin2π	cos2π	sin4π	cos4π	sn6π	cos6π	sin8π	cos8π	sin10π
sin2π	.541									
cos2π	.041	-.002								
sin4π	-.013	-.234	.542							
cos4π	-.553	-.366	.446	.264						
sin6π	.061	-.236	.410	.422	.412					
cos6π	.418	.192	-.588	-.430	-.352	-.080				
sin8π	.359	.460	-.138	-.498	.019	.092	.478			
cos8π	-.170	.035	-.446	-.290	-.526	-.473	.197	-.069		
sin10π	-.028	-.004	.039	-.008	-.121	-.494	.200	.183	.263	
cos10π	-.594	-.248	.044	.168	.176	-.292	-.623	-.369	.430	.076

Table 2.8: FWI RS532, Numerical Comparison of ACF in all 3-fitted Models

lag	Model-1	Model-2	Model-3
1	0.54529072	0.546483705	0.444429303
2	0.31377058	0.315071073	0.270875067
3	0.03135022	0.033027479	0.006298133
4	-0.18385005	-0.181978260	-0.135265531
5	-0.19442393	-0.192623008	-0.073049219

Table 2.9: FWI RS532, Comparison of 3-fitted Models

Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
1	12	879.6567	913.7870	-427.8284			
2	13	817.1797	854.1541	-395.5898	1 vs 2	64.47704	<.0001
3	14	754.7417	794.5603	-363.3708	2 vs 3	64.43801	<.0001

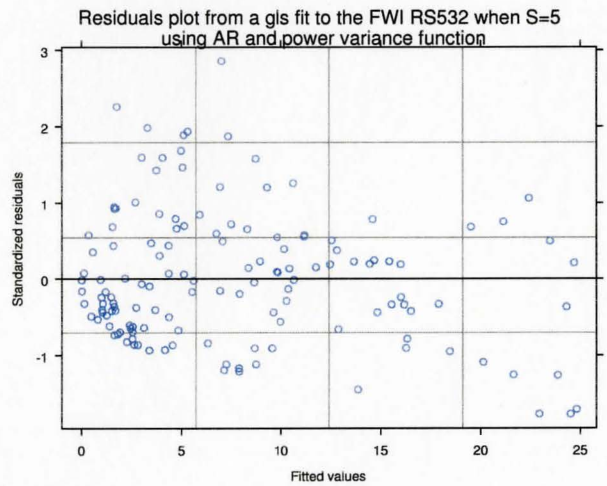


Figure 2.9: FWI RS532 Residual plot after fitting AR1 and Heteroscedasticity in Model-3

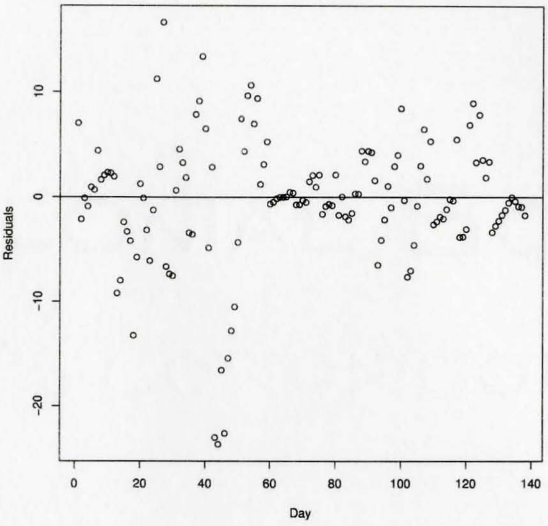


Figure 2.10: FWI RS532 Residual plot against Day after fitting AR1 and Heteroscedasticity in Model-3

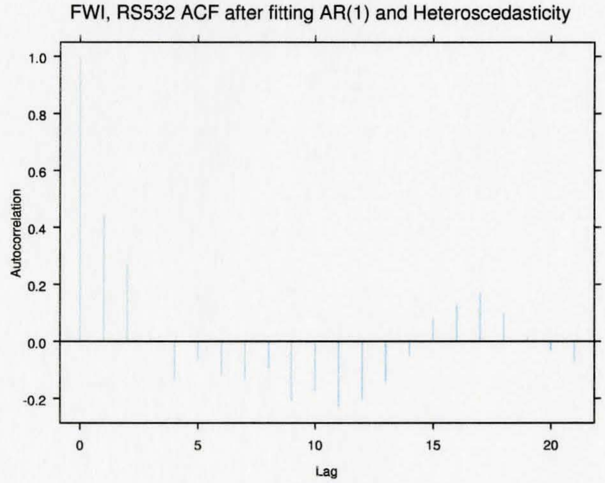


Figure 2.11: FWI RS532 ACF after fitting AR1 and Heteroscedasticity in Model-3

2.4 Linear Regression Model For ES1017

In this section we will discuss the same models again but in context of ES1017 instead of RS532. Station ES1017 which is located 1017 m above sea level has fewer days of temperature measurement so $T = 107$ days rather than $T = 138$ days as it was for RS532.

Table 2.10 shows the simple GLS model output where most parameters are significant at 5% level for the alternate sin or cos term except sin10 and cos10, while sin4 and cos4 terms are highly significant. Residual standard error is 3.63 with 96 degrees of freedom. The Residual plot in Figure 2.12 shows almost the same behavior as we saw in Figure 2.4.

Table 2.10: FWI ES1017 Estimated Parameters in Model-1

Parameters	Value	SE	t-value	p-value
(Intercept)	3.688989	0.3346565	11.02321	<0.0001
$\sin 2t\pi/T$	1.399292	0.4732481	2.95678	0.0038
$\cos 2t\pi/T$	-0.000242	0.4735387	-0.00051	0.9996
$\sin 4t\pi/T$	3.004787	0.4732481	6.34928	<0.0001
$\cos 4t\pi/T$	-2.272202	0.4735392	-4.79834	<0.0001
$\sin 6t\pi/T$	-1.487080	0.4732483	-3.14228	0.0022
$\cos 6t\pi/T$	-0.455121	0.4735388	-0.96111	0.3387
$\sin 8t\pi/T$	-0.495113	0.4732481	-1.04620	0.2978
$\cos 8t\pi/T$	-1.719303	0.4735392	-3.63075	0.0004
$\sin 10t\pi/T$	-0.039286	0.4732483	-0.08301	0.9340
$\cos 10t\pi/T$	-0.378188	0.4735388	-0.79864	0.4263

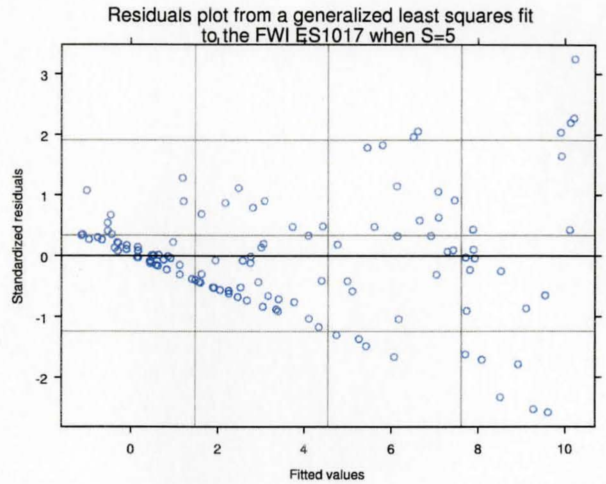


Figure 2.12: FWI ES1017 Residual plot in a GLS Fourier Model-1

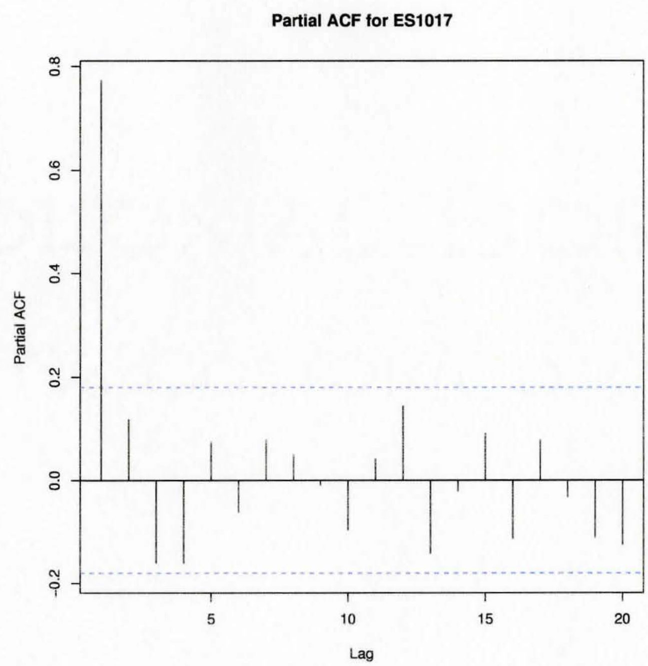


Figure 2.13: FWI ES1017 Partial ACF

AR1 Model

To investigate if there is any serial correlation among the residuals, the autocorrelation (ACF) was calculated (Table 2.15)[3] and found to be 0.58, in our model. The results in Table 2.11 indicate we do not have any significant parameters in the model at 5% except intercept. There is some variance-covariance structure (Table 2.12) but residual plot still shows much variability existing in residual structure. The correlation parameter $\hat{\rho} = 0.82$ and residual standard error went up from 3.63 to 5.41. Residual plot in Figure 2.14 almost have the same pattern as we already observed in Figure 2.7.

Table 2.11: FWI ES1017 Estimated Parameters with AR1 in Model-2

Parameters	Value	SE	t-value	p-value
(Intercept)	3.823929	1.543617	2.477253	0.0148
$\sin 2t\pi/T$	1.302706	2.163324	0.602178	0.5483
$\cos 2t\pi/T$	0.195894	2.061242	0.095037	0.9245
$\sin 4t\pi/T$	2.940860	1.977327	1.487291	0.1399
$\cos 4t\pi/T$	-2.145342	1.898234	-1.130178	0.2609
$\sin 6t\pi/T$	-1.494387	1.751442	-0.853232	0.3954
$\cos 6t\pi/T$	-0.247382	1.694571	-0.145985	0.8842
$\sin 8t\pi/T$	-0.609918	1.535628	-0.397179	0.6920
$\cos 8t\pi/T$	-1.632949	1.495666	-1.091787	0.2774
$\sin 10t\pi/T$	-0.116436	1.348797	-0.086326	0.9314
$\cos 10t\pi/T$	-0.280997	1.320549	-0.212788	0.8319

[3]Where Model-1 represent OLS, Model-2 is GLS with AR1 and Model-3 is GLS with AR1 and heteroscedasticity

Table 2.12: FWI ES1017 Corr. Matrix with AR1 in Model-2

	(Intr)	sin2π	cos2π	sin4π	cos4π	sn6π	cos6π	sin8π	cos8π	sin10π
sin2π	-.002									
cos2π	-.073	-.002								
sin4π	-.003	.002	-.005							
cos4π	-.065	-.002	-.092	-.005						
sin6π	-.004	.003	-.006	.005	-.006					
cos6π	-.057	-.002	-.080	-.004	-.072	-.005				
sin8π	-.005	.003	-.007	.005	-.007	.007	-.006			
cos8π	-.049	-.002	-.069	-.004	-.063	-.005	-.055	-.006		
sin10π	-.006	.004	-.008	.006	-.008	.006	-.007	.008	-.006	
cos10π	-.043	-.002	-.061	-.004	-.054	-.005	-.048	-.006	-.041	-.006

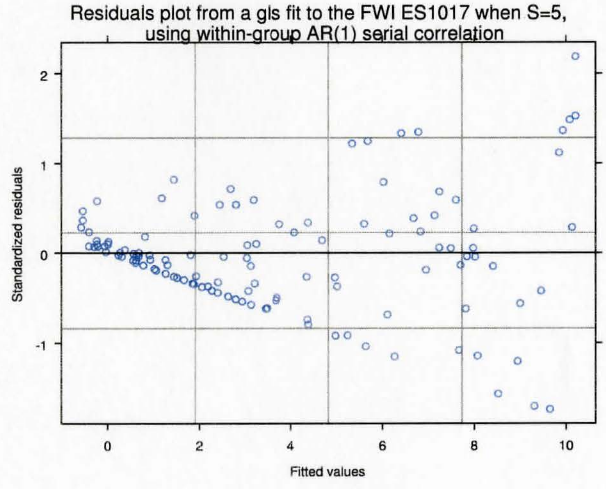


Figure 2.14: FWI ES1017 Residual plot after AR(1) in Model-2

Heteroscedastic Model

To take care the unequal variability in the residuals, we introduce heteroscedasticity into the model and the results (Table 2.13) show most of the parameters are significant at 5% level for alternate sin or cos term, while sin4 and cos4 both are significant and sin2 and cos2 are not significant at all.

It also shows the improvement in correlation structure since $\hat{\rho}$ decreases from 0.82 to 0.58. The correlation matrix (Table 2.14) also shows noticeable variance-covariance structure exists in the model as compared to the Table 2.12. After introducing heteroscedasticity in the model, the residual standard error went down to 1.53, as compared to the 3.63 and 5.41 respectively. The Residual plot (Figure2.15) also shows much improvement compared to the previous two plots (Figure 2.14 and 2.12). We also plotted residuals against day t (Figure 2.16). Autocorrelation is reduces 0.58 to 0.44 after introducing heteroscedasticity in the model (Table 2.15). AIC and BIC criteria also went down as well (Table 2.16).

2.5 Comparison of models fitted for Stations RS532 and ES1017

For the comparison purpose we plotted FWI RS532 and fitted RS532 from the Model-3 (for Reference site) and FWI ES1017 and Fitted ES1017 from the Model-3 (for higher elevation site). The fitted models (Figure 2.17) visually look different and the peaks occur differ by a constant amount the same time. This difference is wider in early June and later is quite parallel (we will discuss this in detail in Chapter 5).

Table 2.13: FWI ES1017 Parameters with AR1 and Heteroscedasticity in Model-3

Parameters	Value	SE	t-value	p-value
(Intercept)	3.655636	0.681831	5.361501	<0.0001
$\sin 2t\pi/T$	1.496847	1.114650	1.342885	0.1822
$\cos 2t\pi/T$	0.093911	0.708892	0.132476	0.8949
$\sin 4t\pi/T$	3.168546	0.877197	3.612125	0.0005
$\cos 4t\pi/T$	-2.572754	0.840282	-3.061775	0.0028
$\sin 6t\pi/T$	-1.864934	0.806323	-2.312885	0.0226
$\cos 6t\pi/T$	-0.246723	0.779342	-0.316578	0.7522
$\sin 8t\pi/T$	-0.990304	0.489388	-2.023556	0.0455
$\cos 8t\pi/T$	-0.471778	0.576847	-0.817858	0.4153
$\sin 10t\pi/T$	1.020615	0.532595	1.916307	0.0580
$\cos 10t\pi/T$	0.091238	0.499090	0.182809	0.8553

Table 2.14: FWI ES1017 Corr. Matrix with AR1 and Heteroscedasticity in Model-3

	(Intr)	$\sin 2\pi$	$\cos 2\pi$	$\sin 4\pi$	$\cos 4\pi$	$\mathrm{sn} 6\pi$	$\cos 6\pi$	$\sin 8\pi$	$\cos 8\pi$	$\sin 10\pi$
$\sin 2\pi$	.332									
$\cos 2\pi$	.027	.568								
$\sin 4\pi$	.755	.183	-.007							
$\cos 4\pi$	-.643	-.537	-.003	-.299						
$\sin 6\pi$	-.478	-.252	.445	-.145	.576					
$\cos 6\pi$	-.125	-.733	-.650	-.112	.090	.001				
$\sin 8\pi$	-.481	-.372	-.120	-.402	.698	.237	.109			
$\cos 8\pi$	-.173	.389	-.132	-.455	-.410	-.242	-.004	-.116		
$\sin 10\pi$	.289	-.161	-.370	-.115	-.357	-.445	.597	.110	.148	
$\cos 10\pi$	.161	.343	-.199	.209	-.300	-.639	-.320	-.022	.186	-.074

Table 2.15: FWI ES1017, Numerical Comparison of ACF in all 3-fitted Models

lag	Model-1	Model-2	Model-3
1	0.57589567	0.58169448	0.43916021
2	0.36300164	0.37133562	0.19186230
3	0.06644572	0.07717076	-0.09600589
4	-0.22610675	-0.21328322	-0.26062806
5	-0.34656631	-0.33351974	-0.20443906

Table 2.16: FWI ES1017 Comparison of 3-fitted Models

Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
1	12	649.4018	681.4757	-312.7009			
2	13	586.2290	620.9758	-280.1145	1 vs 2	65.17274	<.0001
3	14	503.5539	540.9735	-237.7770	2 vs 3	84.67508	<.0001

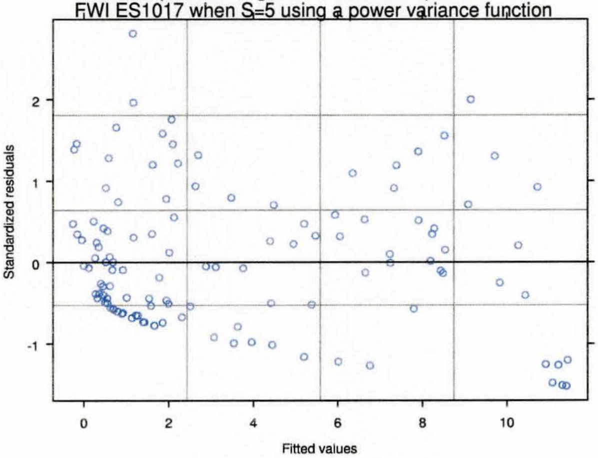


Figure 2.15: FWI ES1017 Residual plot after fitting AR1 and Heteroscedasticity in Model-3

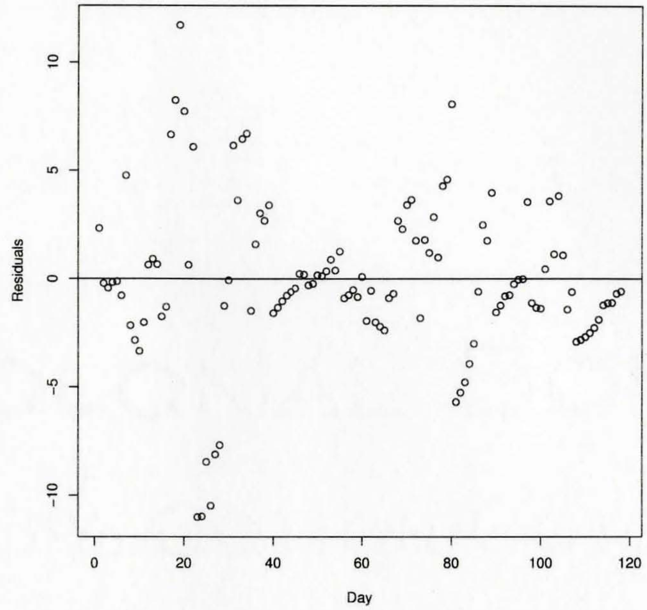


Figure 2.16: FWI ES1017 Residual plot against Day after fitting AR1 and Heteroscedasticity in Model-3

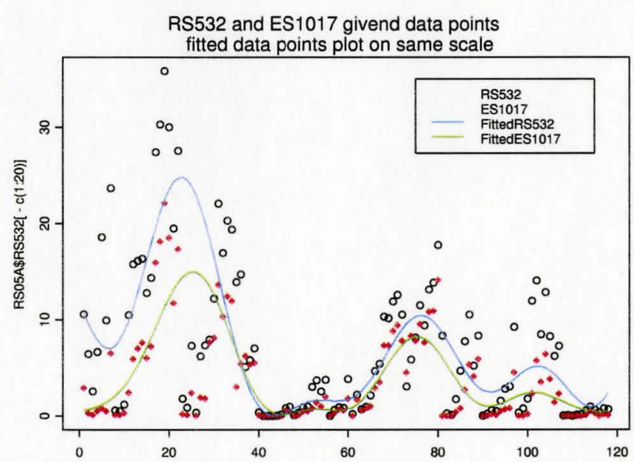


Figure 2.17: Scatter RS532 (∘) and ES1017 (•) with the $\widehat{RS532}$ and $\widehat{ES855}$ lines in an *Individual* Heteroscedastic model with AR1

Chapter 3

Smoothing the Data and Bootstrap

3.1 Cubic Spline Smoothing and Bootstrap

In this chapter we will discuss cubic spline smoothing, and smooth the given data set for the base station FWI and other higher elevation station using the cubic spline smoothing. We will use the bootstrap residuals to estimate the standard error of the smoothed values as described by Efron & Tibshirani (1998). Finally, we will graphically compare the spline fit to the data and the bootstrap smoothed results to see how well the bootstrap is performing.

3.2 Introduction

In non-parametric regression smoothing is a technique widely used in statistical analysis, because it can capture the existing structure in the data where we don't have a parametric model. There are many smoothers which a data analyst may use to find the existing structure in the data set including Kernel, Loess, Spline, Cubic spline, and Supersmoother. The idea of smoothing came from a naval draftsman who used a mechanical device called a spline to get a smooth interpolating curve by positioning the rod with a groove and weights (called ducks) (Schumaker 1993).

3.2.1 Cubic Spline Smoothing

Consider a general linear regression model for given paired observations (y_i, x_i)

$$y_i = f(x_i) + e_i$$

where $i = 1, 2, \cdots, n;\quad e_i \sim N(0, \sigma^2)$ and $x \in [a, b]$.

Our main goal is to fit a smooth curve to f within the interval $[a, b]$, where a and b are known as the boundary initial and end points. In the given interval we need to divide the data in small segments where the segments are joined by the knots, and it must be continuous at the knots to get the smoothness. Thus knots play an important role in the flexibility of the curve; more knots produce a more flexible curve. We can place more knots where data are too noisy or we need more smoothness but each knot should at least have one data point.

A spline is a function of piecewise polynomial of degree m with $m - 1$ continuous derivative. A cubic spline $(m = 3)$ has a continuous second derivative in the intervals $[a, b]$ and it produce the curve without too much rapid local variation. To quantify the local variation the integrated squared second derivative is most common to use, that is, the roughness penalty. Thus a cubic smoothing spline arises as the function f that minimizes the penalized residuals sum of square (PRSS)

$$PRSS = \sum_{i=1}^{n}[y_i - f(x_i)]^2 + \lambda \int_a^b [f''(x)]^2 dx \tag{3.2.1}$$

On the right of the equation (3.2.1), the first part is the goodness of fit for a linear regression equation and second is a roughness penalty due to smoothness. Equation 3.2.1 yields a cubic spline, i.e., a piecewise cubic polynomial with continuous second derivative.

3.2.2 Smoothing Parameter

The parameter λ is a smoothing parameter and is a trade off between the residual sum of square and the roughness penalty since it determine the influence of that. As $\lambda \to 0$, PRSS approaches a curve that interpolates the data and when $\lambda \to \infty$, then roughness becomes less important and we will get a linear regression fit. Therefore choosing λ is most important. There are certain methods to choose λ. One is cross validation proposed by Wahba (1975) and other is generalized cross validation criteria proposed by Craven & Wahba (1979). For the given data set we will use a cubic smoothing spline function in Splus with the spar argument (i.e. is equivalent to λ) and it yield a 3rd degree piecewise polynomial.

3.2.3 Graphical Output

Below we will see some graphical output from smoothing the data using the cubic spline function and plotting data and spline. In the first two graphs we plotted the given data set for FWI RS532 and ES855 and the corresponding from the spline function in Splus. This function will fit a cubic spline to the given data set with default smoothing parameter or we may specify its value. In Figure 3.1 and 3.2, we observe clearly that if we choose smoothing parameter $\lambda = 0$ (i.e. spar in Splus), then the interpolated curve will pass through the largest number of points and yield a smooth curve that provides a better fit compared to df= 5, where the curve is too smooth for our purposes. We also observe that all.knots=T and cross validation (cv=T) produce the same result as obtained with $\lambda = 0$.

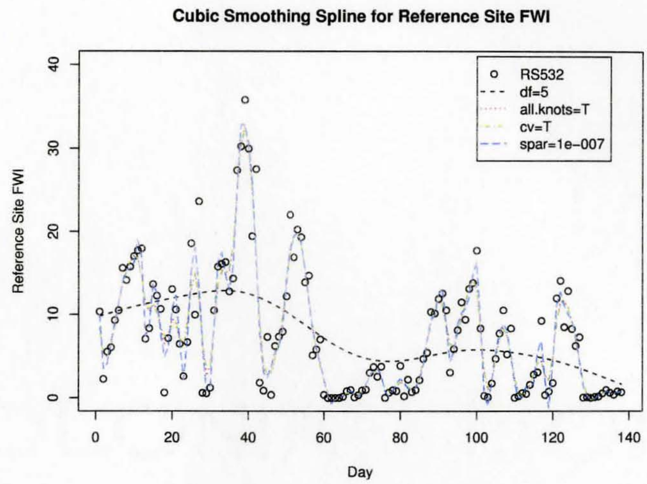


Figure 3.1: Fitted Cubic spline FWI RS532 with $\lambda = 10^{-7}$ and all knots are true

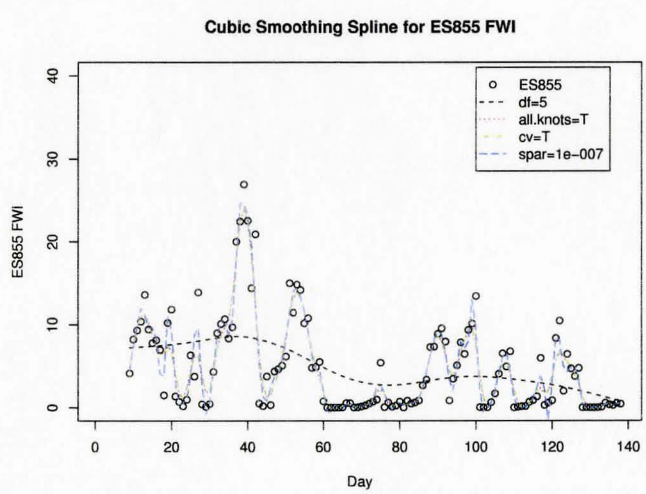


Figure 3.2: Fitted Cubic spline FWI ES855 with $\lambda = 10^{-7}$ and all knots are true

3.3 Bootstrap

Bootstrap is a method to resample statistic. It is a data-based simulation method used to estimate standard error of an estimator and provide confidence intervals which could be difficult to do in the usual way.

Its main purpose to get better results by using repeated sampling with replacement either the data set, or the residuals of the fitted model. We will use the last technique in a non-parmetric setting also known as Efron & Tibshirani (1998) approach.

3.3.1 Bootstrapping from Residuals

We will adopt a non-parametric bootstrap technique without any assumption about of the population and will use residuals from the fitted cubic spline. Because data are collected over time, we will also account for serial correlation ρ, of lag1, in the residuals and will use this value to get the data to bootstrap, in same fashion as proposed by Efron & Tibshirani (1998, page 95-96). The only difference is that instead using the slope parameter β, we are accounting for serial correlation ρ.

3.3.2 Algorithm

We have a given data set FWI RS532 (y_t) for the base station and to that we fit a cubic smoothing spline ($\hat{y}_t$). From the fitted cubic smoothing spline we will get the residuals z_t (Step 1), and account the serial correlation ρ for lag1 in the residuals z_t. Thus we would use Step 1 and the estimated $\hat{\rho}$ from residuals to get the estimated error at time t in Step 2. Now we have $T-1$ estimated error observations where $T = 138$. We can construct the bootstrap sample from this by drawing random sample of size $T-1$ with replacement, where each observation has equal chance $1/(T-1)$ to occur and ϵ_t^* denote the bootstrap data set. It is very necessary that we sample with replacement, otherwise we get the same sample every time.

From the ϵ_t^* we construct z_t^* in Step 4 and y_t^* in Step 6. In constructing z_t^* we need to treat z_1 as a fixed constant in Step 3, where z_1 is the residual on day $t = 1$ obtained from (Step 1), and calculate z_t^* recursively. By using Step 4 and smoothed values of FWI RS532 (i.e. $\hat{y}$) we can construct a new y^* vector (Step 6)

- Step 1: $z_t = y_t - \hat{y}_t$
- Step 2: $\hat{\epsilon}_t = z_t - \hat{\rho} * z_{t-1};$ where $t = 2, 3, \cdots, 138$
- Step 3: $z_2^* = \hat{\rho} * z_1 + \epsilon_{1,i}^*$
- Step 4: $z_j^* = \hat{\rho} z_{j-1}^* + \epsilon_{j-1,i}^*;$ where $i = 1, 2, \cdots, 137$ and $j = 3, 4 \cdots, 138$
- Step 5: $y_1^* = \hat{y}_1 + z_1$
- Step 6: $y_j^* = \hat{y}_j + z_j;$ where $j = 2, 3, \cdots, 138$

So first value will be same for all sets of generated y^*. To the set of y_t^*, $t = 1, 2, \cdots, 138$, the spline was fitted in the same way as done for the observed data and the fitted values are denoted by $\hat{y_t^*}$, $t = 1, 2, \cdots, 138$. These steps are repeated B times, for a total of $B = 999$ bootstrap replicates and fit a cubic smoothing spline to the bootstrap data (for codes please see Appendix A (A.2)).

3.3.3 Bootstrap Percentile Confidence Interval

To construct a confidence band for $\alpha = 0.05$ around the spline fitted to the data, the 2.5% and 97.5% percentiles were obtained from $\hat{y_t^*}$ at each t, where $t = 1, 2, \cdots, 138$. We can arrange $\hat{y^*}$ in ascending order $\hat{y_1^*}, \hat{y_2^*}, \cdots, \hat{y_B^*}$ or use a prebuilt quantile function to get order the bootstrap replicates. We can choose lower $= (B + 1)(\alpha/2)$ for 25th (i.e. 2.5%) percentile and upper $= (B + 1)(1 - \alpha/2)$ for 975th (i.e. 97.5%) percentile if $B = 999$.

We can plot different curves and numerical results to see desired results. For example we can plot the FWI RS532 data and fitted cubic spline lines that we got through $\hat{y^*}$ and construct a 95% confidence band around this.

3.3.4 Numerical Output

We chooses $B = 999$ which means we replicate the data 999 times and every time we should get a different set of observations due to sampling with replacement. Our numerical output the tables below verifies this. For illustrative purposes, we provide some numerical results for first few bootstrap data sets, including some initial rows and final rows.

Table 3.1 shows some of the data for FWI RS532 (i.e. $y1$) and Fitted spline FWI RS532 (i.e. $\hat{y}$). It also show residuals (z_t) after fitting the cubic spline to base station, and in the next column we will see estimated residuals ($\hat{\epsilon}_t$) after introducing the serial correlation ($\hat{\rho}$). Lower (2.5%) and upper (97.5%) percentile confidence interval are also given.

Table 3.2 gives $\epsilon^*_{(10\times6)}$ matrix when the bootstrap was performed 999 times. It includes the first 6 bootstraps and 10 initial rows only. Table 3.3 gives $z^*_{(10\times6)}$ matrix and Table 3.4 gives $y^*_{(10\times6)}$ matrix. Table 3.5 gives $\hat{y}^*_{(10\times6)}$ fitted cubic spline values which we got after replicating the data 999 times and every time fitted a cubic spline to y^* and stored this fitted spline value in the $\hat{y}^*$ matrix. The tables show the successive calculations from bootstrap sample of ϵ^*.

In Table 3.6 we have the corresponding Bootstrap residuals matrix, where these residuals are the differences, $y^*_t - \hat{y}^*_t$.

Table 3.1: FWI RS532 (y_1), fitted Cubic Spline ($\hat{y}$), residuals (z_t), estimated residuals ($\hat{\epsilon}$) after $\hat{\rho}$, and 95% CI

t	$y1$	$\hat{y}$	z_t	$\hat{\epsilon}$	2.5%	97.5%
1	10.33682	8.16603883	2.170784642		6.35261383	9.630347
2	2.247748	5.41383588	-3.166087874	-0.718530038	4.89955279	8.356558
3	5.509882	4.90422323	0.605658308	-0.032801206	4.34192216	7.906800
4	6.038941	6.32505377	-0.286112454	0.414064712	5.16104129	8.711987
5	9.299535	8.76580584	0.533728725	-0.680173249	7.18271929	10.567550
⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮
136	0.3749921	0.60753954	-0.232547423	0.036242156	-0.89042743	2.542698
137	0.7642292	0.63072610	0.133503066	0.039297033	-1.16726651	3.070985
138	0.6682403	0.68477970	-0.016539447	-2.258176332	-2.80202343	3.977151

Table 3.2: Residuals ($\epsilon^*_{10\times 6}$) with replacement after 999 Bootstrap, few shown

	1	2	3	4	5	6
1	1.0256534	-5.33382331	-0.6151219	0.05660906	-0.2540998	-0.36173567
2	0.1434458	2.24186180	-0.6151219	-2.38220678	-2.2641185	-1.25463287
3	-1.1996549	8.63200905	1.6553581	2.24186180	-1.5812923	-1.28459072
4	0.5096793	2.24186180	-0.6151219	-2.79657424	-0.6058410	1.13091295
5	1.7019371	-0.61338190	-0.8881789	3.92367747	-0.2204663	-1.01637450
6	-0.6151219	-1.31122515	0.4976243	-0.56979689	0.3063347	-4.15912695
7	-2.2581763	1.91392239	0.3475198	0.30872285	-0.0922907	-0.56979689
8	4.2829354	-2.26411846	-1.2845907	-1.28459072	1.0127787	1.02565341
9	3.1268148	4.37281956	-1.0284353	-2.25817633	1.1309130	1.36216827
10	1.7019371	-0.10236572	-5.3338233	1.72108229	-5.7762719	-2.78191878

Table 3.3: $z^*_{10\times 6}$ after $\hat{\rho}$, 999 Bootstrap sample drawn from residuals $\hat{\epsilon}_t$, few shown

	1	2	3	4	5	6
1	-3.2901183	-2.0462971	-1.16201136	-0.8206017	-0.87166939	-6.4510304
2	2.4017164	-4.4779776	-0.12912088	0.3998185	0.11046822	2.3363509
3	-0.8610509	4.1147365	-0.56111822	-2.5494273	-2.31032082	-2.2317911
4	-0.8395280	6.9110568	1.89004085	3.3081373	-0.61502095	-0.3511638
5	0.8608045	-0.6486268	-1.40561481	5.6719982	-0.07466163	-1.5507964
7	-1.1763653	-1.1681451	0.62321918	-2.9420601	0.33756128	-3.5105200
8	-1.7661719	2.4024888	0.08686386	1.5392136	-0.23347274	0.8984472
9	5.0216213	-3.2689382	-1.32092076	-1.9283532	1.11042662	0.6498858
10	1.0265660	5.7400241	-0.47597188	-1.4516596	0.66648681	1.0903593

Table 3.4: Estimated ($y^*_{10\times6}$) from 999 Bootstrap samples from residuals, few shown

	1	2	3	4	5	6
1	2.123718	3.3675388	4.251825	4.593234	4.542166	-1.037195
2	7.305940	0.4262456	4.775102	5.304042	5.014691	7.240574
3	5.464003	10.4397902	5.763936	3.775626	4.014733	4.093263
4	7.926278	15.6768626	10.655847	12.073943	8.150785	8.414642
5	12.244553	10.7351218	9.978134	7.203575	11.035135	12.661533
6	14.861464	13.1774516	13.219258	19.191549	13.444889	11.968755
7	14.003325	14.0115448	15.802909	12.237630	15.517251	11.669170
8	14.686253	18.8549134	16.539289	17.991638	16.218952	17.350872
9	22.108960	13.8184005	15.766418	15.158986	18.197765	17.737225
10	17.710329	22.4237875	16.207791	15.232104	17.350250	17.774123

Table 3.5: Fitted bootstrap spline ($\hat{y^*_{10\times6}}$) from 999 Bootstrap samples, few shown

	1	2	3	4	5	6
1	6.037521	4.803254	6.289922	6.464023	6.202321	4.781956
2	5.654236	5.486733	5.881534	6.320520	5.292440	4.562187
3	6.633709	8.416233	6.796527	7.166708	5.839470	5.920580
4	8.717613	11.051750	8.689491	8.980012	7.839509	8.156692
5	11.179485	12.870703	11.049572	11.246369	10.570846	10.519331
6	13.360581	13.987803	13.339226	13.379905	13.223263	12.446831
7	15.341175	15.049507	15.079600	14.999015	15.350318	14.186373
8	17.182915	16.494968	15.855159	15.845540	16.573105	15.915826
9	18.133927	17.402385	15.444960	15.950856	16.419089	16.726948
10	17.498637	16.799467	13.896367	15.408636	14.675509	15.825918

Table 3.6: Fitted bootstrap spline Residuals from 999 Bootstrap samples, few shown

	1	2	3	4	5	6
1	-2.2051208	0.37489428	-1.0458966	-2.065536817	-2.3198750	2.1856724
2	-1.3455165	0.12918641	-0.3208171	-0.194894152	-2.7348450	-5.7143734
3	-0.1547040	-3.17978136	-4.7226003	-1.224723891	3.0889688	2.4559358
4	-0.4599867	1.69680174	3.1294803	-2.927866507	-1.5619209	0.5644404
5	-2.7108927	1.33803071	-2.3351549	5.193137603	-0.8928071	-1.2161194
6	8.4699105	-0.17216746	5.7054511	-1.306865672	0.2507089	-0.8560967
7	-6.9684242	1.97483707	-4.9493058	1.434270900	2.2397253	3.3260521
8	3.8300563	-0.01244875	3.1129506	-3.500593406	1.0010052	-0.7344299
9	0.9794580	-1.58189628	-1.9785704	2.389824648	-3.5960286	-2.0033394
10	-3.7006952	3.05723789	4.6753500	-0.009150813	4.5388554	3.4331919

3.3.5 Visual Output

To better show the bootstrap outputs after replicating it 999 times we plot some graphs. In Figure 3.3 we plot $\hat{y}$ as a data point and the 999 lines, obtained from bootstrap fitted cubic spline ($\hat{y}^*$). These fitted lines are quite smooth and able to capture all peaks obtained when the original data was smoothed. We also observe that some lines become negative where we have FWI zero or close to zero. Two visible upper and lower lines show the 95% confidence intervals. We will notice that lower confidence band also becomes negative where we have zero FWI. But overall 95% percentile confidence band appears reasonable, since almost all curves are inside the lower (2.5%) and upper (97.5%) confidence bound.

Because the individual resultant curves can not be distinguished due to 999 smooth lines, in Figure 3.4 we plot only first 10 curves to give better visualization for smooth curves. Almost all curves are lying within the 95% confidence band except one at initial and mid point.

How well have we done in the bootstrap? This is a question which we can better answer by comparing the curves from the given FWI RS532 data and fitted cubic spline data ($\hat{y}$). Next two figures will illustrate this. Figure 3.5 we plotted the RS532 fitted cubic spline data $\hat{y}$ and the lower (2.5%) and upper (97.5%) confidence band from 999 bootstrap sample. Here all smoothed values of the given base station data set are lying inside the 95% confidence band.

In the next, Figure 3.6, we added the data for FWI RS532 and smoothed lines $\hat{y}$ to curves plotted in Figure 3.5. Here the green line shows the RS532 cubic smooth spline.

3.3.6 Bootstrap Residuals

No analysis in statistics can be completed without the residuals discussion. Next four graphs will show the fitted spline and bootstrap residuals.

In Figure 3.7 we will see bootstrap spline residuals those we got through bootstrap cubic spline data. These residual graph we got after fitting cubic spline residuals $\hat{y^*}$ when bootstrap data replicated 999 times. The dense thickest cloud mostly oscillating between $[-5, 5]$, also we will see very visible cloud below -5 and above 5.

In next Figure 3.8 we can expect the same thing only difference is that here we plot only first 10 bootstrap residuals, now it is more visible that most of the residuals are oscillating between $[-5, 5]$ and few residuals point lying below -5 and above 5.

What do these residual plots tell us? For that we will compare them to the given data set residuals z_t and $\hat{\epsilon}$ when we fitted a cubic spline to FWI RS532 data. The Figure 3.9 and 3.10 show the cubic spline residual (z_t) and estimated residual $\hat{\epsilon}$ after introducing serial correlation $\hat{\rho}$ that we got from fitted cubic spline residuals. Here once again we will observe most of the residuals are oscillating between $[-5, 5]$ except very few below -5 and above 5.

Thus simple comparison between figure 3.8 and 3.9 show very clear that our bootstrap residuals plots are quite different to the original residuals plot in range values. We can clearly observe heteroscedastic structure of residuals from spline fitted to observed data (Figure 3.9 and 3.10) which is why we needed to account heteroscedasticity in our generalized least squares model. If we do graphical comparison among the observed data set residuals (3.9 and 3.10) and bootstrap residuals (Figure 3.7 and 3.8) we will find that the bootstrap residuals are lying between $[-10, 10]$ while original data residuals or oscillating between $[-5, 10]$.

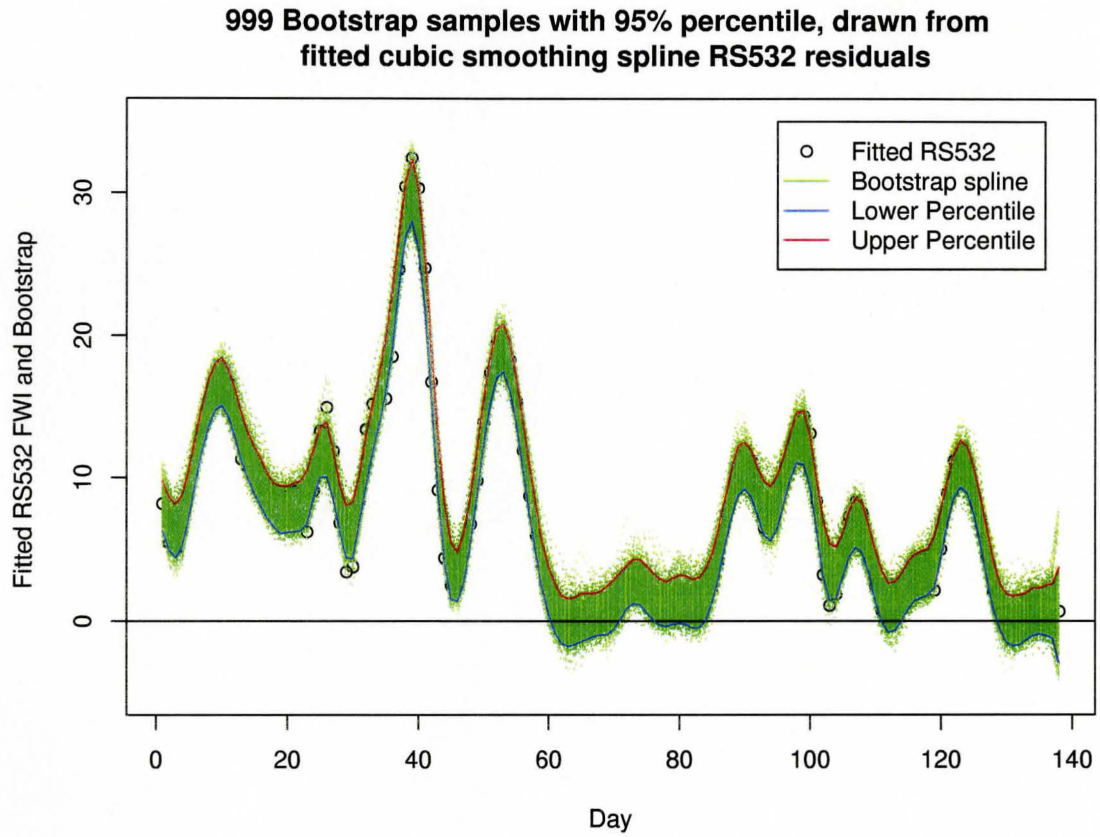


Figure 3.3: 999 Bootstrap drawn from a fitted RS532 cubic spline smooth residuals with a 95% percentile

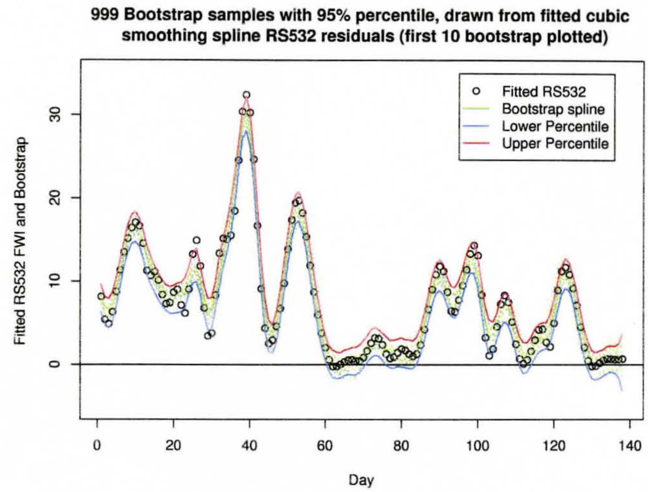


Figure 3.4: 999 Bootstrap drawn from a fitted RS532 cubic spline smooth residuals with a 95% percentile (first 10 curves plotted)

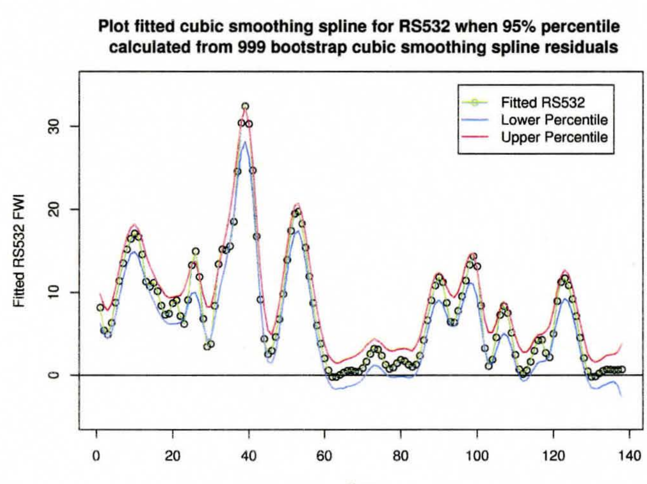


Figure 3.5: Fitted Cubic spline RS532 ($\hat{y}$) when 95% percentile calculated from 999 Bootstrap cubic spline smooth residuals

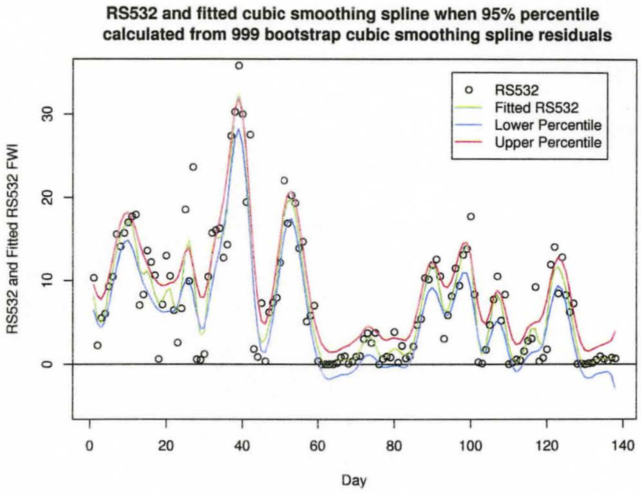


Figure 3.6: FWI RS532 and fitted cubic spline when 95% percentile calculated from 999 Bootstrap cubic spline smooth residuals

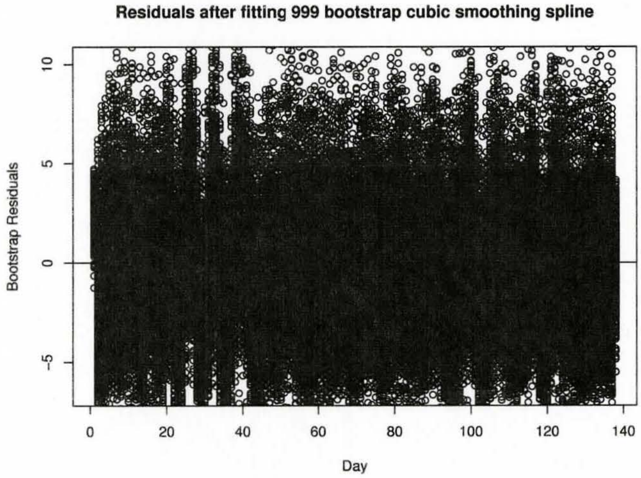


Figure 3.7: Residuals from 999 Bootstrap

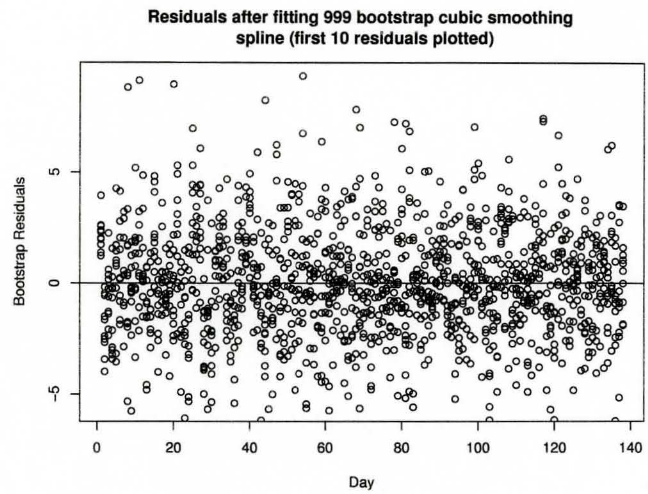


Figure 3.8: Residual from 999 Bootstrap (first 10 plotted)

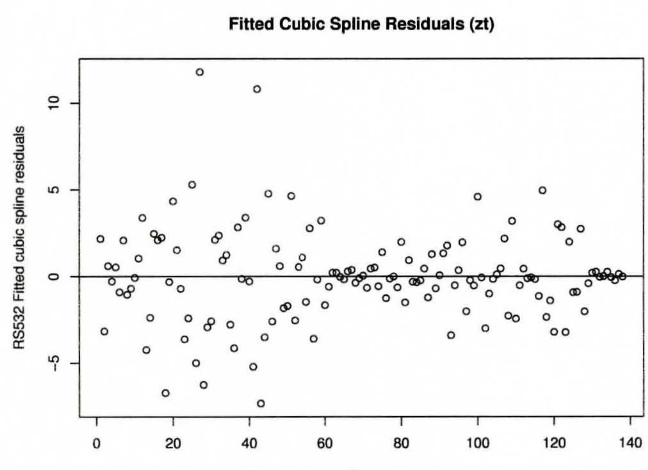


Figure 3.9: Fitted RS532 Cubic spline smoothing residuals (z_t)

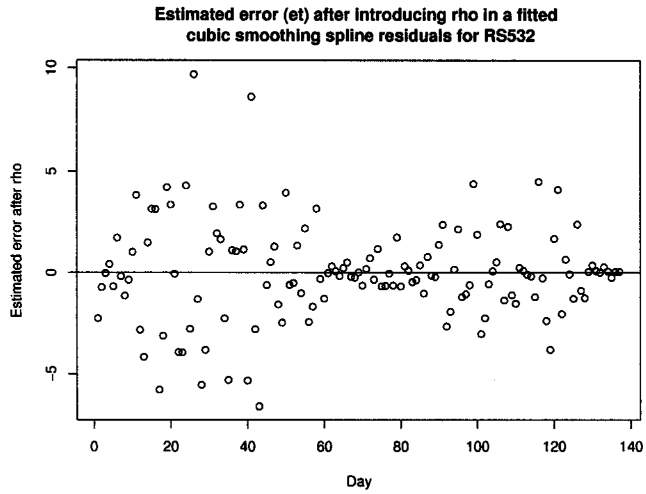


Figure 3.10: Fitted RS532 Cubic spline smoothing Residuals ($\hat{\epsilon}_t$) after introducing $\hat{\rho}$

Chapter 4

Station Effect

4.1 Combined Model to address the main question

In this Chapter we will fit a multiple linear regression model in the form of GLS to investigate the main question. To better accommodate peaks and troughs in FWI data set we will use the combined spline model and later for the same stations we will use the bootstrap residuals to estimate standard error of smoothed values.

Our main interest is to determine whether the use of the local precipitation amount would lead to a different fire weather index (FWI) than obtained if the precipitation at a nearby weather station is used. For that, we need to compare the combined model of the base station and one of the higher elevation stations. The comparison for the higher elevation station, ES855, is reported here. The first 8 observations from the base station were omitted because the data collection could not start before May 09, 2003 at ES855.

The data table we designed here, contained the 130 observations from RS532, starting May 09, 2003, first, and same number for ES855 following the RS532 observations. Thus $t = 1$, the first observation for RS532, occurs in row 1, and for ES855, in row 131. In the construction of the *sin* and *cos* terms, $T = 130$ for both stations.

We are fitting the *sin* and *cos* terms to FWI RS532 and ES855 with $k = 5$. Therefore the model under consideration is

$$Y_{t,i} = \beta_0 + \alpha_i + \sum_{s=1}^{k} \left[\beta_{1s} sin \frac{st}{T} 2\pi + \beta_{2s} cos \frac{st}{T} 2\pi \right] + \epsilon_{ti} \qquad (4.1.1)$$

where $s = 1, 2, \cdots, k$; $t = 1, 2, \cdots, T$; $T = 130$, and dummy variable α_i represent ith station effect

$$i = \begin{cases} 1 & \text{for RS532} \\ 0 & \text{for ES855} \end{cases}$$

In equation 4.1.1 *sine* and *cosine* terms are calculated using the same method as we used in Chapter 2. The model has the 11-parameters, as Chapter 2, plus one additional parameter α_i to measure the station effect.

4.1.1 Successive Models for the Combined Stations

To carry out the goal of our main interest we built a combined multiple linear regression model (Equation 4.1.1) and fitted the model to the combined stations (FWI RS532 and ES855) for $k = 5$, but, due to so many zeros in FWI and algorithm convergence problem in gls() function for higher elevation station sites we are not able to fit the combined model for the same stations (FWI RS532 and ES1017) again, as we fitted in Chapter 2 for individual stations. To test the significance of the combined stations (FWI RS532 and ES855) model we can run the ordinary least square (OLS).

After running OLS, it is evident from Table 4.1 that the alternate *sin* or *cos* parameters are significant at the 5% level except *sin*8 and *cos*8. The stations effect α is also highly significant and the residual standard error is 5.57. The residual plot (Figure 4.1) shows that the variance is not constant and shows te same diagonal shape as we alredy discussed (Chapter 2 Section 2.3).

Table 4.1: Estimated Parameters in a Combined OLS Model

Parameters	Value	SE	t-value	p-value
Intercept	4.744695	0.4906943	9.669351	<.0001
α	2.648324	0.6909449	3.832903	0.0002
$\sin 2t\pi/T$	3.277964	0.4885718	6.709279	<.0001
$\cos 2t\pi/T$	1.111684	0.4885718	2.275374	0.0237
$\sin 4t\pi/T$	1.333359	0.4885718	2.729095	0.0068
$\cos 4t\pi/T$	-3.535216	0.4979115	-7.100089	<.0001
$\sin 6t\pi/T$	-1.135803	0.4885718	-2.324742	0.0209
$\cos 6t\pi/T$	0.467266	0.4978260	0.938613	0.3488
$\sin 8t\pi/T$	0.079537	0.4885718	0.162794	0.8708
$\cos 8t\pi/T$	-0.047402	0.5005359	-0.094702	0.9246
$\sin 10t\pi/T$	1.890030	0.4885718	3.868479	0.0001
$\cos 10t\pi/T$	0.127245	0.4885718	0.260444	0.7947

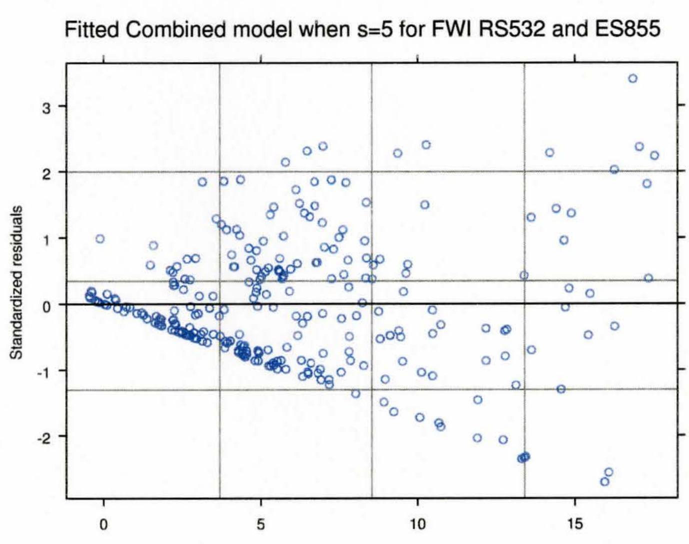


Figure 4.1: Fitted combined model

AR1 Model

The lag 1 autocorrelation estimated from the residuals of the ordinary least squares model was 0.60. Using this in the generalized least squares (GLS) analysis (Table 4.2) results in only 3 terms being significant, that of the intercept, *sin2* and *cos4*. The correlation parameter $\hat{\phi}$ increases to 0.68 and the residual standard error increases from 5.57 to 6.10. Moreover, the residual plot (Figure 4.2) has almost the same variability, and still indicates that we need to account for heteroscedasticity in the model. The station effect, parameter α also increased (Table 4.2) as compare to the previous one, but more pronounced is the increase in the standard error.

Table 4.2: Estimated Parameters in a Combined GLS Model with AR1

Parameters	Value	SE	t-value	p-value
Intercept	5.007631	1.198332	4.178833	<.0001
α	2.302654	1.650812	1.394861	0.1643
$\sin 2t\pi/T$	3.254209	1.226775	2.652654	0.0085
$\cos 2t\pi/T$	1.222526	1.208797	1.011358	0.3128
$\sin 4t\pi/T$	1.288052	1.198850	1.074407	0.2837
$\cos 4t\pi/T$	-3.355993	1.184666	-2.832861	0.0050
$\sin 6t\pi/T$	-1.198880	1.156328	-1.036800	0.3008
$\cos 6t\pi/T$	0.095121	0.386554	0.246074	0.8058
$\sin 8t\pi/T$	0.003137	1.103856	0.002842	0.9977
$\cos 8t\pi/T$	0.130505	1.094026	0.119289	0.9051
$\sin 10t\pi/T$	1.804621	1.045997	1.725263	0.0857
$\cos 10t\pi/T$	0.217654	1.034386	0.210418	0.8335

Heteroscedasticity

Non constancy of variance (Figure 4.2) in the residuals lead us to account for heteroscedasticity in the model. Taking account of it (Table 4.3) results again in most parameters being significant at the 5% level except for $s = 3$ and 4, but the station effect is not significant. The residual standard error went down to 1.55 as compared to the 5.57 and 6.10. Thus the model which has accounted for heteroscedasticity has reduced estimate of error variance which is one of the objective of the modifications to the model. The residual plot (Figure 4.3) is improved but still retain features seen for the individual stations. Autocorrelation (Table 4.4) does not change much. The comparison of all 3-fitted models (Table 4.5) shows that model-3, which accounts for serial correlation and heteroscedasticity, has minimum AIC and BIC values, as seen for individual stations.

Thus the model-3 fitted here shows, if we take both serial correlation and heteroscedasticity into account, that there is no difference between stations, since α is not significant. However, this is illustrated by the plot of the final model, model-3, with the data ($T = 130$) for RS532 and ES855 (Figure 4.4). Here the two curves are close together and parallel, it is clear (Figure 4.4) that the fitted regression curves are not able to capture the sharpness of the peaks. This causes more variability in the residuals, and leads to residuals that do not form a rough uniform band around the horizontal axis. Hence that affects the overall model fit and contributes to the station effect being insignificant.

Another way in which the data and model for combined stations, model-3 could be plotted is shown in Figure 4.6. A vertical dotted line splits the graph into two to show the fitted and scatter plot of the individual stations. The first half corresponds to RS532 FWI scatter and fitted line, and the second half to FWI ES855 scatter and fitted line. These figures clearly show that the fitted curves of the combined model, after introducing the serial correlation and heteroscedasticity are parallel.

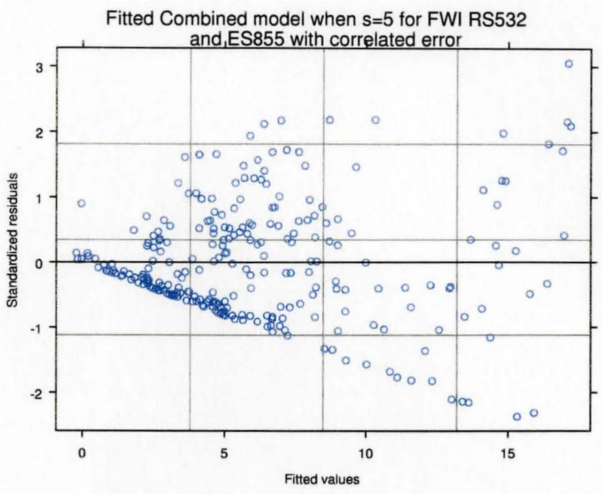


Figure 4.2: Fitted combined model with AR1

Table 4.3: Estimated Parameters in a Combined GLS Model with AR1 and Heteroscedasticity

Parameters	Value	SE	t-value	p-value
Intercept	5.963002	0.834765	7.143329	<.0001
α	0.805836	0.637260	1.264534	0.2072
sin2tπ/T	4.485287	1.294079	3.466008	0.0006
cos2tπ/T	2.036939	0.683385	2.980662	0.0032
sin4tπ/T	2.140716	0.973768	2.198385	0.0288
cos4tπ/T	-4.238541	1.099899	-3.853571	0.0001
sin6tπ/T	-0.822295	1.020905	-0.805458	0.4213
cos6tπ/T	0.162671	0.222239	0.731966	0.4649
sin8tπ/T	1.109884	0.690926	1.606372	0.1095
cos8tπ/T	-0.100857	0.685320	-0.147167	0.8831
sin10tπ/T	1.319460	0.548877	2.403926	0.0170
cos10tπ/T	-1.229745	0.510793	-2.407521	0.0168

Table 4.4: Comparison of ACF in all 3-fitted Models

lag	Model-1	Model-2	Model-3
1	0.6031644491	0.607090974	0.58267865
2	0.3875643317	0.391929878	0.38601172
3	0.1455408307	0.146089473	0.16800554

Table 4.5: Comparison of Combined fitted Models for RS532 and ES855 FWI

Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
1	14	1572.102	1621.234	-772.0511			
2	15	1426.897	1479.538	-698.4485	1 vs 2	147.2053	$<.0001$
3	16	1301.497	1357.647	-634.7485	2 vs 3	127.4000	$<.0001$

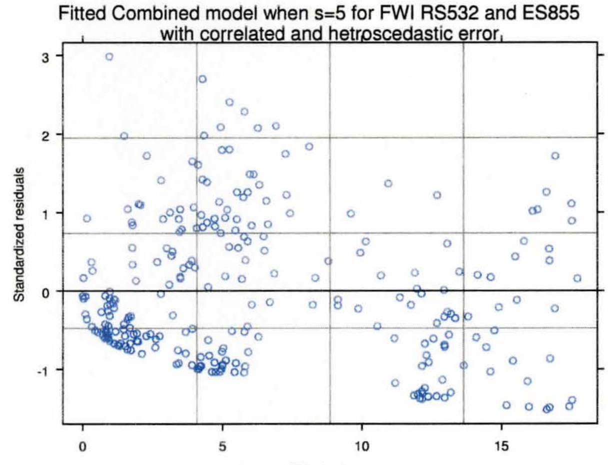


Figure 4.3: Fitted combined model with AR1 and Heteroscedasticity

For the comparison purposes we plot FWI RS532 and fitted RS532 from GLS model-3 (from Chapter 2) for $s = 5$ and on same graph we plot FWI ES1017 and Fitted ES1017. We need to ignore first 20 observation in RS532 due to ES1017 data collection starting at a later date (Figure 4.5 and Figure 4.7, where in last one, we also joined the scatter points with the line).

Figure 4.4 and 4.5 enhance our understanding of that ANCOVA-type models fitted, i.e regression curves for the individual stations should be parallel to each other see the acid rain example in Hall & Hart (1990), Bowman & Young (1996, Section 4.3). Figure 4.5 is quite different form Figure 4.4. This suggest that there is a difference between these two stations, although not tested for significance. The same peaks are present but they differ in amplitude much more in the early part of the series than do the two latter peaks. The early peak mid June, in Figure 4.4, which is missing in Figure 4.5, is missing because of the later starting date for ES855.

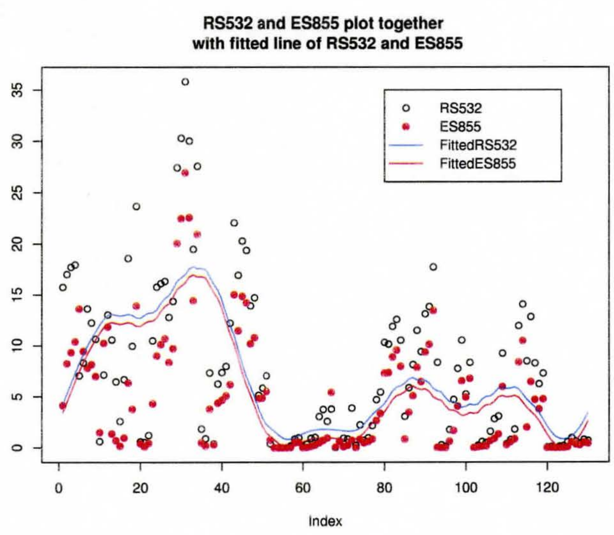


Figure 4.4: Scatter RS532 (∘) and ES855 (•) with the $\widehat{\text{RS532}}$ and $\widehat{\text{ES1017}}$ lines in a combined Heteroscedastic model with AR1

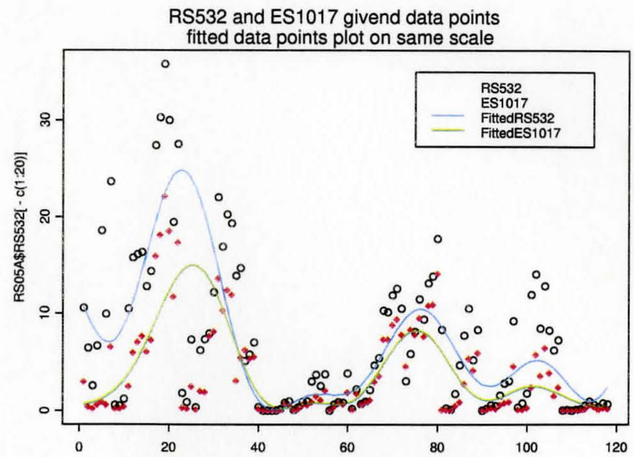


Figure 4.5: Scatter RS532 (∘) and ES1017 (•) with the $\widehat{\text{RS532}}$ and $\widehat{\text{ES1017}}$ lines in an *Individual* Heteroscedastic model with AR1

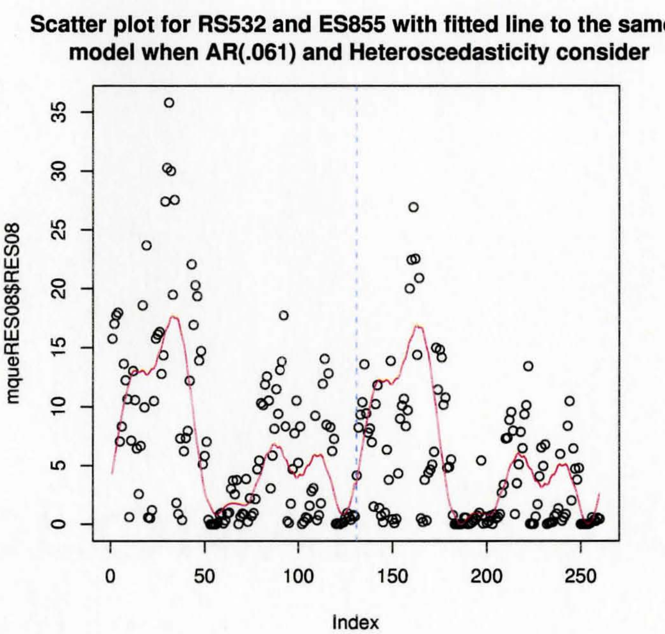


Figure 4.6: Scatter $\widehat{\text{RS532}}$ and $\widehat{\text{ES1017}}$ in a combined Heteroscedastic and AR1 model

4.2 Stations Effect in a Cubic Spline Smoothing Model

To account for peaks and troughs in our general model (Chapter 2 Equation 2.1), we smooth the data (Chapter 3) in a non-parametric setting. Here we continue the same idea and investigate the main question in a cubic spline smoothing scenario.

Thus we fitted a combined cubic spline smooth curve to FWI RS532 and ES855 separately and superimposed (Figure 4.8). Here we can clearly observe that the curves have peaks that coincide approximately in late June and early August, where FWI is zero or very close to zero (for that period we already observed diagonal structure in our residuals plots in Chapter 2 and in the previous section as well). The ES855 smooth line touches the RS532 line. This shows that fire danger rating is zero or very minimal then these lines may touch each other. When we have higher FWI, the amplitude of the peaks differ with ES855 FWI always exceeding that of RS532, but somewhere in proportion to the level. That also provide supports to our visual results from the previous section (Figure 4.4 and 4.5) after fitting the combined and individual station model.

4.3 Combined Stations Effect Using Bootstrap

To carry out our main objective we can use bootstrap residuals technique, as described in Chapter 3, and see if the spline regression curves for the combined stations are parallel to each other as proposed by Hall & Hart (1990), Bowman & Young (1996). We can boot FWI RS532 fitted cubic spline residuals and FWI ES855 fitted cubic spline residuals B times as we did in the Chapter 3. Below we will see some combined bootstrap graphical outputs.

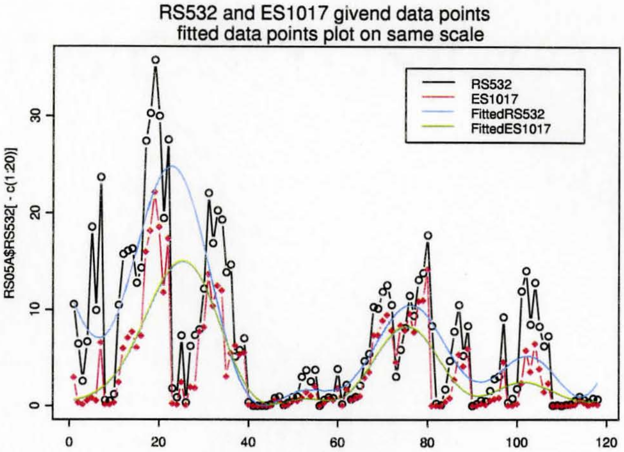


Figure 4.7: Scatter RS532 ($\circ$) and ES1017 ($\bullet$) with the $\widehat{\text{RS532}}$ and $\widehat{\text{ES1017}}$ lines in an *Individual* Heteroscedastic model with AR1 when scatter points join by a line

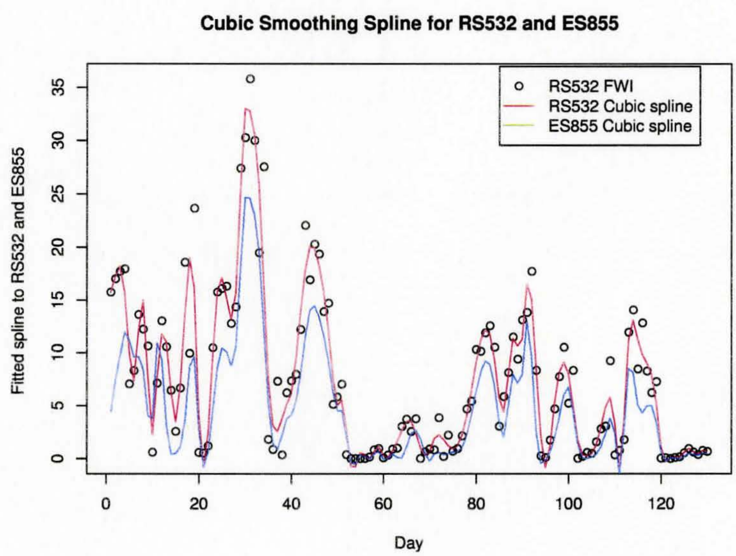


Figure 4.8: Cubic spline smooth curves for FWI RS532 and ES855 when they superimposed, $\lambda = 0$

We plot RS532 999 bootstrap lines those we got from bootstrap fitted cubic spline ($\hat{y_1^*}$) and ES855 bootstrap fitted cubic spline $\hat{y_2^*}$ (Figure 4.9). Although the individual resultant curves can not be distinguished due to 999 smooth lines for each station but it is still visible that fitted cubic spline lines to ES855 are approximately parallel to the base station lines (Figure 4.9).

We also plot the first 10 bootstrap fitted curves for both stations (Figure 4.10) to make it more visible. Here we can observe very clearly that bootstrap lines are quite parallel except in the middle and end. These figures are very similar as Figure 4.8, and this shows the ability to capture the peaks. Thus it is clear from visual output that our fitted bootstrap lines are able to capture the essence of the original cubic spline smooth data.

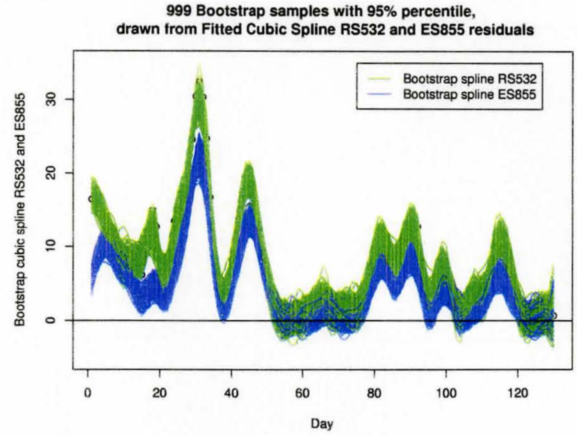


Figure 4.9: 999 Bootstrap drawn from a fitted RS532 and ES855 cubic spline smooth residuals

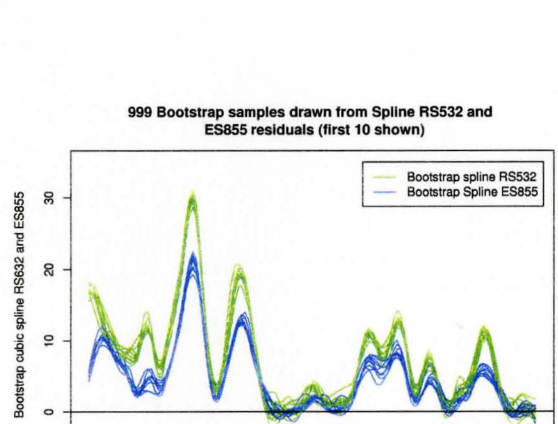


Figure 4.10: 999 Bootstrap drawn from a fitted RS532 and ES855 cubic spline smooth residuals (first 10 curves plotted)

Chapter 5

Discussion And Future Prospect

5.1 Discussion

The wildland fire is the worst kind of fire. The current project is designed to give the right information to forest fighters and management to issue the right fire-danger rating. Following we will summarize the results of this study in the light of the main objective.

In Chapter 2, the generalized least squares model for Fire Weather Index based on Fourier terms with autoregressive order1 and heteroscedastic errors, FWI model-3, showed much improved results relative to the other fitted models. In particular, this model resulted in an improved residual plot and the diagonal shape of the residuals also faded.

In the non-parametric approach (Chapter 3), the fitted cubic spline smoothing was able to capture the steepness of the peaks much better than the model with Fourier terms. In the cubic spline smoothing curves fitted to the bootstrap FWI, we observed almost the same behavior in capturing peaks, as we saw in cubic spline smoothing, but for lower FWI values some fitted curves were not able to observe that.

In Chapter 4, FWI combined model-3, the fitted curves are not able to capture the sharpness of the peaks, and as a result more of the variability remains in the residuals. This affects the overall model fit.

In FWI data set, we were interested to find if the local precipitation amount would lead to a different fire weather index than that obtained if the precipitation at a nearby station were used, so we tested the station effect, and found it to be insignificant here.

We have data for one year fire season only. With a short data series at the higher elevation stations due to later snow melt we may not have enough data to evaluate the method. Due to algorithm convergence problem in the FWI data, we were not able to use the same higher elevation station as we did for the individual station. Precipitation is a natural process, one year we get less rain and other year more rain, and higher rainfall may lead to more zeros in the FWI data, hence the algorithm convergence problem in the gls function is of concern, and the method will not work in some of the cases we will encounter. Therefore, these results can not establish the basis for a generalized result without further investigation of the model. A few years of additional data may yield a better fit.

However, the model still enhances our understanding of that ANCOVA-types models as discussed by Hall & Hart (1990) and Bowman & Young (1996). This suggests that there is a difference between two stations, although of a form more complex than that tested for significance.

The cubic smoothing spline fitted curves for individual stations when plotted separate and superimposed provided good visual summaries and comparisons. The residuals based simulation also demonstrated good smooth curves but at lower index values it encountered some difficulties. This can also be investigated with some additional years data.

There are number of model extensions that can be proposed with some additional years data, for example, the use of proper log transformation with and without dummy variable. One can also consider a model which includes different curves for each station as well as difference in levels to test whether including different curves for each station accounts for significantly more variation than current model does.

Bibliography

Bowman, A. & Young, S. (1996), *Graphical Comparison of Nonparametric Curves*, Applied Statistics **45**, 83-98.

Buonaccorsi, P. John and Elkinton, S. J. (2002), *Regression Analysis in a Spatial-Temporal Context: Least Squares, Generalized Least Squares, and the Use of Bootstrap*, Journal of Agricultural, Biological, and Enviromental Statistics **7** (1), 4-20 pp.

Canadian Forest Service (1999), *Climate Change and Forests: Context for the Canadian Forest Service's Science Program*, Science Branch, Canadian Forest Service, Natural Resources Canada, Ottawa, 13 pp.

Canadian Forest Service (2001), *Forest Fire: Context for the Canadian Forest Service's Science Program*, Science Branch, Canadian Forest Service Natural Resources Canada, Ottawa, 13 pp.

Canadian Forest Service (2005a), *Fighting Wildland Fire with Technology*, Canadian Forest Service, Natural Resources Canada, Northern Forestry Centre, Edmonton, 16 pp.
URL: *http://cwfis.cfs.nrcan.gc.ca/en/publications/fighting_fire_tech_e.php*

Canadian Forest Service (2005b), *Background Information: Canadian Forest Fire Weather Inex System (FWI)*, Canadian Forest Service, Natural Resources Canada.
URL: *http://cwfis.cfs.nrcan.gc.ca/en/publications/bi_FWI_summary_e.php*

Canadian Forest Service (2005d), *Background Information: Canadian Forest Fire Weather Inex System (FWI)*, Canadian Forest Service, Natural Resources Canada. **URL:** *http://cwfis.cfs.nrcan.gc.ca/en/background/bi_FWI_dsm_e.php*

Craven, P. & Wahba, G. (1979), *Smoothing noisy data with spline functions*, Numer. Math **31**, 377-403 pp.

Efron, B. & Tibshirani, R. J. (1998), *An Introduction to the Bootstrap*, New York: Chapman and Hall.

Greene, H. W. (1993), *Econometric Analysis*, New York, Macmillan, 397, 415-416 pp.

Hall, P. & Hart, J. D. (1990), *Bootstrap Test for Difference Between Means in Nonparametric Regression*, JASA **85** (412), 1039-1049 pp.

Higgins, D. G. & Ramsey, G. S. (1992), *Canadian forest fire statistics: 1988-1990*, Forestry Canada, Petawawa National Forestry Institute, Chalk River, ON. Info., Rep. PI-X-107, 76 pp.

Leathwick, J. R. & Briggs, C. M. (2001), *Spatial Prediction of Wildfire Hazard across New Zealand*, Prepared for New Zealand Fire Service Commission, Landcare Research, Contract Report, LCR 0001/081, 13 pp.

Schumaker, L. L. (1993), *Spline Functions: Basic Theory*, Rieger Publishing Co.

Splus Development Core Team. (2002), *S-Plus 6 for Windows Guide to Statistics, Volume 1, chapter 10*, S-PLUS 6.1 Insightful Corp. 298 pp.

Turner, J. A. & Lawson, B. B. (1978), *Weather in the Canadian Forest Fire Danger Rating System. A user guide to national standards and parctices*, Environment Canada, Forestry Service, Pacific Forest Reseaarch Centre, Victoria, BC., BC-X-177, 40 pp.

Van Wagner, C. E. (1987), *Development and Structure of the Canadian Forest Fire Weather Index System*, Canadian Forestry Service, Ottawa, Forestry Technical Report No. 35, 3, 35 pp.

Van Wagner, C. E. & Pickett, T. L. (1985), *Equations and FORTRAN Program for the Canadian Forest Fire Weather Index System*, Canadian Forestry Service, Ottawa, Technical Report No. 33, 18 pp.

Wahba, G. and Wold, S. (1975), *A completely automatic french curve*, Communications in Statistics **4**, 1-17 pp.

Appendix A

A.1 Base Station Parameters for S=2 and 4

Follwing we will see the output for the reference site when we use S=2 and S=4 in the Fourier model. Table A.1 to A.6 will show you output of a Fourier model when S=2 without AR1, with AR1 and AR1 plus heteroscedasticity, while rest tables will show for S=4. Following to that we will see bootstrap codes.
Here we will see line plots for all indices including precipitation, FWI, BUI, DC, DMC, FFMC and ISI, see figures A.1 to A.5.

Table A.1: Estimated Parameters and their p-values in RS532 FWI in GLS Fourier Model when s=2

Parameters	Value	SE	t-value	p-value
(Intercept)	7.533831	0.5469377	13.77457	< 0.0001
sin2tpi.T	4.200018	0.7723850	5.43773	< 0.0001
cos2tpi.T	0.720006	0.7735350	0.93080	0.3536
sin4tpi.T	0.275812	0.7723850	0.35709	0.7216
cos4tpi.T	-3.883010	0.7735360	-5.01982	< 0.0001

Table A.2: Estimated Parameters and p-values When AR(0.64) was introduced for RS532

Parameters	Value	SE	t-value	p-value
(Intercept)	7.563320	1.422834	5.315672	< 0.0001
sin2tpi.T	4.068520	2.022955	2.011176	0.0463
cos2tpi.T	0.806951	1.965143	0.410632	0.6820
sin4tpi.T	0.183986	1.974275	0.093192	0.9259
cos4tpi.T	-3.814440	1.920168	-1.986514	0.0490

Table A.3: Estimated Parameters and p-values with AR1 and Heteroscedasticity in RS532

Parameters	Value	SE	t-value	p-value
(Intercept)	7.634210	1.620709	4.710414	< 0.0001
sin2tpi.T	3.830139	2.393403	1.600290	0.1119
cos2tpi.T	1.535354	1.206210	1.272875	0.2053
sin4tpi.T	1.270297	1.634666	0.777099	0.4385
cos4tpi.T	-4.959893	1.779411	-2.787380	0.0061

Table A.4: Correlation Matrix When AR1 was introduced for RS532

	sin2tpi	cos2tpi	sin4tpi	cos4tpi
sin2tpi	- 0.001			
cos2tpi	- 0.044	-0.002		
sin4tpi	- 0.002	0.000	-0.003	
cos4tpi	- 0.041	-0.001	-0.061	-0.003

Table A.5: Correlation Matrix after AR1 and Heteroscedasticity for RS532

	sn2tpi	cs2tpi	sn4tpi	cos4tpi
sin2tpi	0.475			
cos2tpi	0.221	0.173		
sin4tpi	0.214	0.285	0.335	
cos4tpi	-0.737	-0.309	0.193	-0.072

Table A.6: Comparison of 3-fitted Models for RS532

Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
1	6	906.1213	923.4634	-447.0607			
2	7	828.2879	848.5204	-407.1440	1 vs 2	79.83341	< 0.0001
3	8	764.9921	788.1149	-374.4960	2 vs 3	65.29585	< 0.0001

Table A.7: Estimated Parameters and their p-values in RS532 in GLS Fourier Model when s=4

Parameters	Value	SE	t-value	p-value
Intercept	7.532875	0.5243223	14.36688	< 0.0001
sin2tpi.T	4.199720	0.7404466	5.67187	< 0.0001
cos2tpi.T	0.721130	0.7415491	0.97246	0.3326
sin4tpi.T	0.274679	0.7404471	0.37096	0.7113
cos4tpi.T	-3.883595	0.7415502	-5.23713	< 0.0001
sin6tpi.T	0.548433	0.7407807	0.74034	0.4604
cos6tpi.T	2.619217	0.7419406	3.53022	0.0006
sin8tpi.T	0.899074	0.7404471	1.21423	0.2269
cos8tpi.T	-0.824473	0.7415502	-1.11182	0.2683

Table A.8: Estimated Parameters and their p-values when AR(0.6) was introduced for RS532

Parameters	Value	SE	t-value	p-value
Intercept	7.502145	1.643756	4.564027	<.0001
sin2tpi.T	4.017179	2.331912	1.722698	0.0873
cos2tpi.T	0.700945	2.261704	0.309919	0.7571
sin4tpi.T	0.145919	2.256558	0.064664	0.9485
cos4tpi.T	-3.921663	2.192355	-1.788790	0.0760
sin6tpi.T	0.419683	2.146968	0.195477	0.8453
cos6tpi.T	2.642465	2.091426	1.263475	0.2087
sin8tpi.T	0.702699	2.015109	0.348715	0.7279
cos8tpi.T	-0.886099	1.969045	-0.450015	0.6535

Table A.9: Correlation Matrix when s=4 after AR1(0.64)

	sin2tpi	cos2tpi	sin4tpi	cos4tpi	sn6tpi	cs6tpi	sn8tpi	cos8tpi
sin2tpi	0.001							
cos2tpi	-0.047	-0.002						
sin4tpi	-0.002	0.001	-0.003					
cos4tpi	-0.043	-0.002	-0.064	-0.003				
sin6tpi	-0.003	0.000	-0.004	0.001	-0.004			
cos6tpi	-0.042	-0.002	-0.060	-0.003	-0.058	-0.004		
sin8tpi	-0.003	0.001	-0.005	0.000	-0.005	0.002	-0.005	
cos8tpi	-0.038	-0.001	-0.056	-0.003	-0.053	-0.004	-0.051	-0.005

Table A.10: Comparison of 2-fitted Models for RS532 when s=4

Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
7	10	893.9328	922.5310	-436.9664			
8	11	821.5069	852.9649	-399.7535	7 vs 8	74.42588	< 0.0001

A.2 Bootstrap Codes

```
function(B,fwi)
{
        y1 <- RS
        n <- length(y1)
        s.smooth <- smooth.spline(y1)
        zt <- s.smooth$yin - s.smooth$y
        zt.1 <- zt[-138]
        zt.1 <- c(NA, zt.1)
        yhat <- s.smooth$y
        rho <- cor(zt[-1], zt[-138])
        et <- zt - rho * zt.1
        et <- et[-1]
        z <- rep(NA, n)
        y <- rep(NA, n)
        estar <- matrix(rep(NA, B), nrow = n - 1, ncol = B)
        zstar <- matrix(rep(NA, B), nrow = n, ncol = B)
        ystar <- matrix(rep(NA, B), nrow = n, ncol = B)
        yhatstar <- matrix(rep(NA, B), nrow = n, ncol = B)
```

```
        perclo <- rep(NA, n)
        percup <- rep(NA, n)
        for(i in 1:B) {
                estar[, i] <- sample(et, replace = T)
                z[1] <- zt[1]
                z[2] <- rho * z[1] + estar[1, i]
                for(j in 3:n) {
                        z[j] <- rho * z[j - 1] + estar[j - 1, i]
                        z <- z
                }
                y[1] <- yhat[1] + z[1]
                for(h in 2:n) {
                        y[h] <- yhat[h] + z[h]
                        y <- y
                }
                ystar[, i] <- y
                zstar[, i] <- z
                fit <- smooth.spline(ystar[, i])
                yhatstar[, i] <- fit$y
        }
        for(k in 1:n) {
                perclo[k] <- quantile(yhatstar[k,  ], 0.025)
                percup[k] <- quantile(yhatstar[k,  ], 0.975)
        }
        plot((yhat), ylim = c(-5, 35), main =
                "1000 Bootstrap samples with 95% percentile
drawn from the fitted Cubic Spline RS532 residuals"
                )
        # For fig-1
for(i in 1:B) {
                lines(yhatstar[, i], spar = 1e-007),
                        lty = 1, col=3)
                abline(h = 0)
                lines(perclo, col = 4, lty = 1)
                lines(percup, col = 2, lty = 1)
legend(80,35, c("Fitted Spline RS532", "Bootstrap Cubic Spline ",
"Lower Percentile", "Upper Percentile"),
lty=c(-1,1,1,1), col=c(1,3,4,2), pch=c(21,-1,-1,-1))
        }
       list(RS532=y1[1:10] , CubicSpline = yhat[1:10],
                yhatstar = yhatstar[1:3, 1:3], LowerPerc = perclo[
```

```
                1:10], UpperPerc = percup[1:10])
}
```

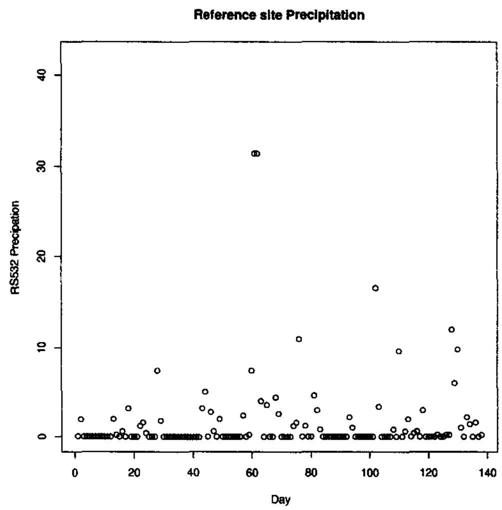
Note:
For additional figures we may do change in the plotting and lines forloop. For example:

```
# For Fig-2
#In last forloop just replace B from 10
for(i in 1:10) {
                lines(yhatstar[, i], spar = 1e-007),
                        lty = 1, col=3)................
}

# For Fig-3
# Replace  ystarhat in the lines command  by yhat
        plot((yhat), ylim = c(-5, 35), main =
                "Cubic Spline to FWI RS532 when 95% percentile
calculated from 1000 Bootstrap cubic Spline residuals")
for(i in 1:B) {
                lines(yhat,lty = 1, col=3) ...............
}

# For Fig-4
```

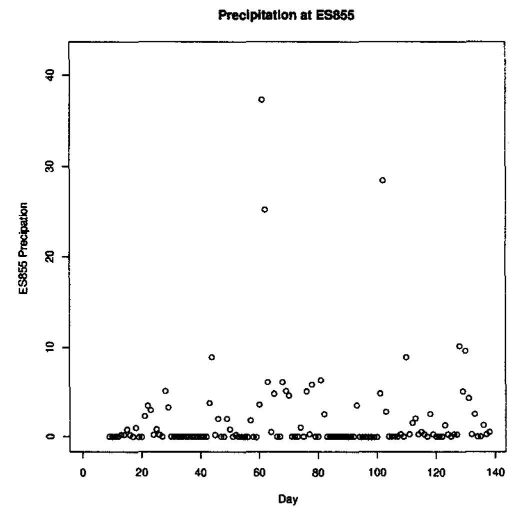
Replace yhat by y1 in the plot and ystarhat by yhat in the lines command.

```
        plot((y1), ylim = c(-5, 35), main =
                "Cubic Spline to FWI RS532 when 95% percentile
calculated from 1000 Bootstrap cubic Spline residuals")
for(i in 1:B) {
                lines(yhat,lty = 1, col=3) ..................
}
```

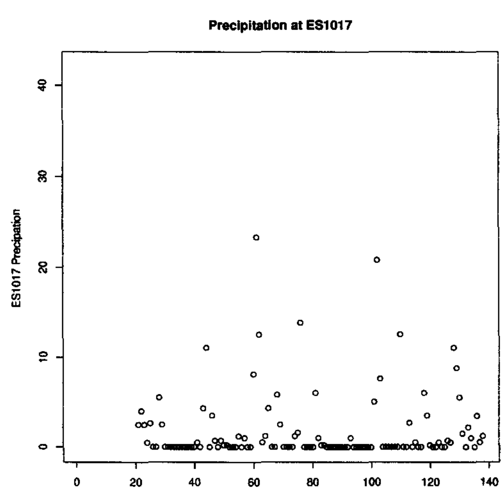
A.3 Graphical Output for Precipitation, Fuel Moistures Codes and Fire Behavior Indices for Reference and Elevation Sites



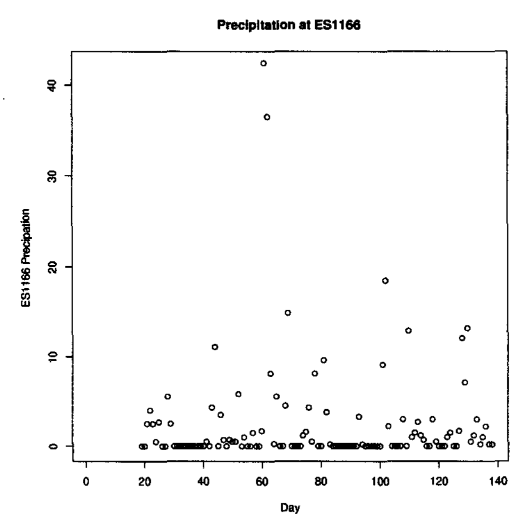
(a) Precipitation for Reference Site



(b) Precipitation for ES855

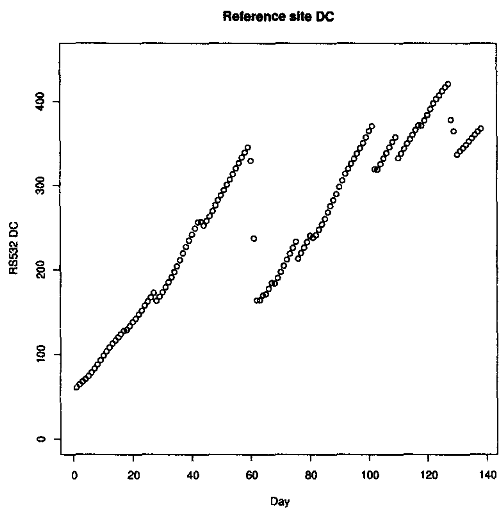


(c) Precipitation for ES1017

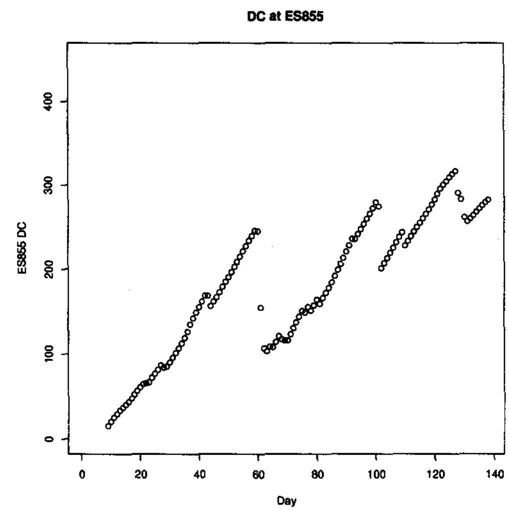


(d) Precipitation for ES1166

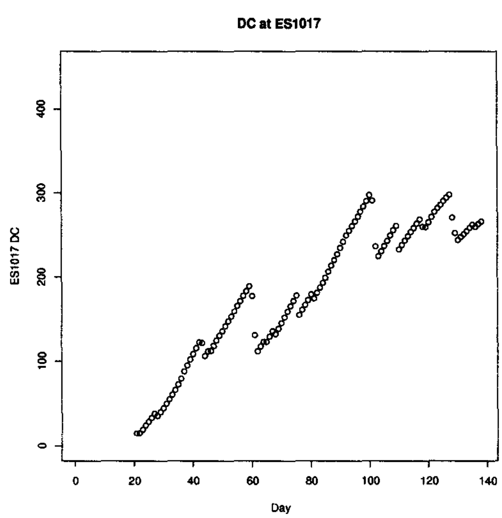
Figure A.1: 24 hour Precipitation in mm



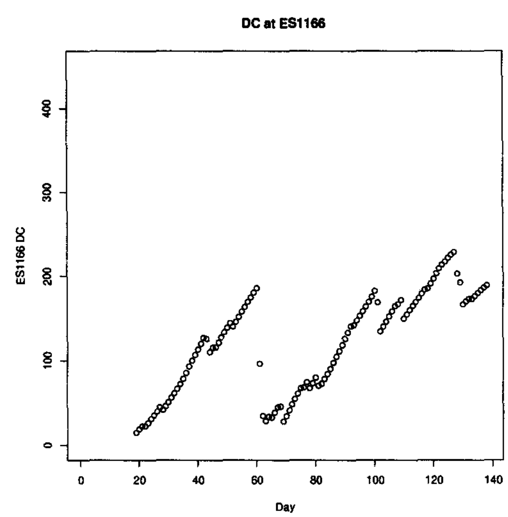
(a) DC for Reference Site



(b) DC for ES855

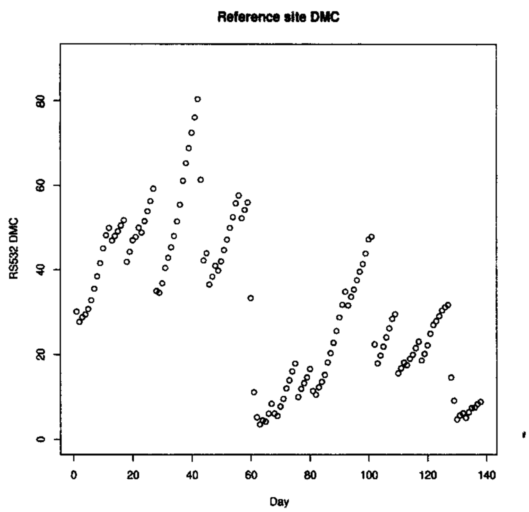


(c) DC for ES1017

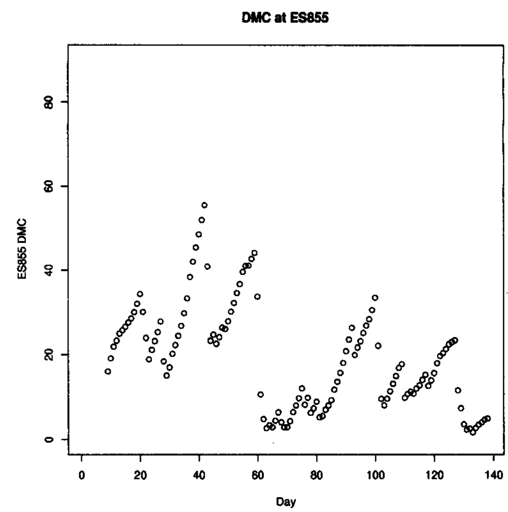


(d) DC for ES1166

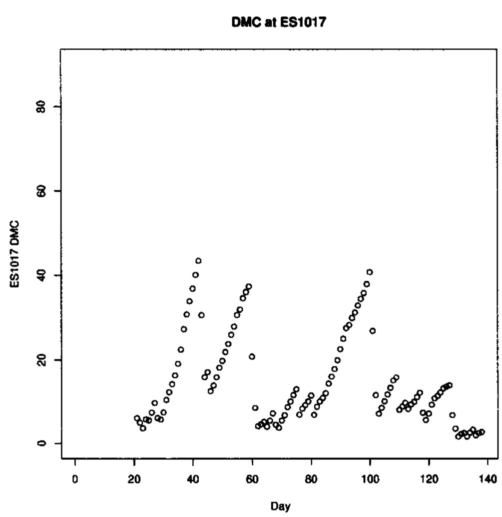
Figure A.2: Drought Code (DC) for Reference and Elevation Sites



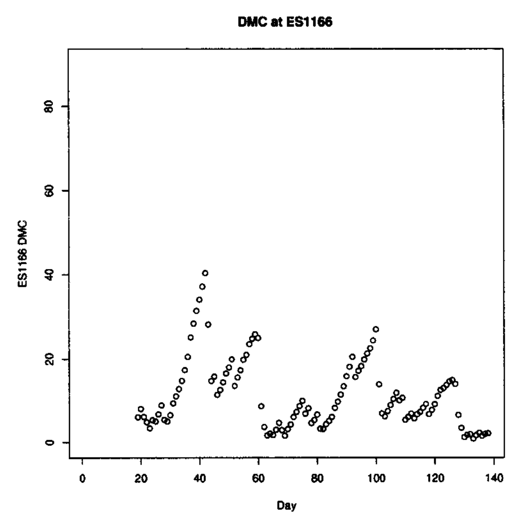
(a) DMC for Reference site



(b) DMC for ES855

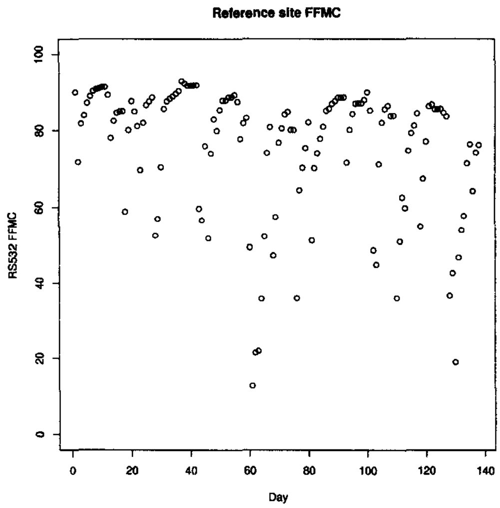


(c) DMC for ES1017

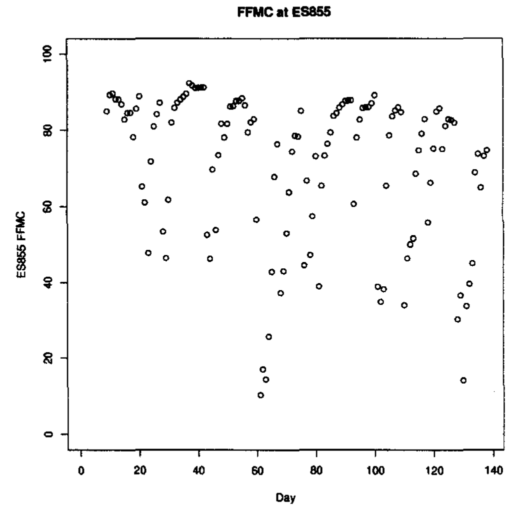


(d) DMC for ES1166

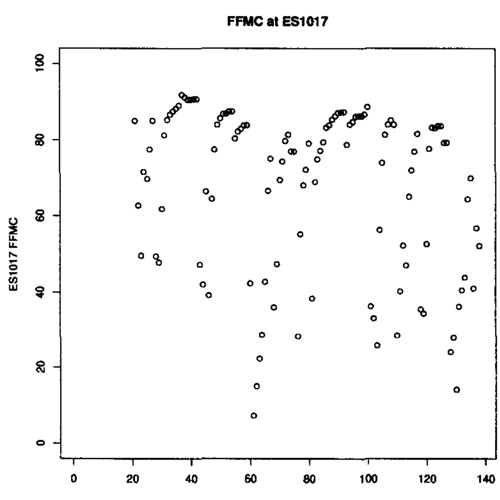
Figure A.3: Duff Moisture Code (DMC) for Reference and Elevation Sites



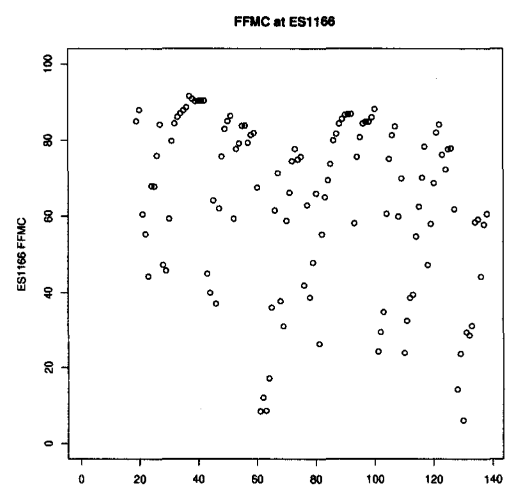
(a) FFMC for Reference site



(b) FFMC for ES855

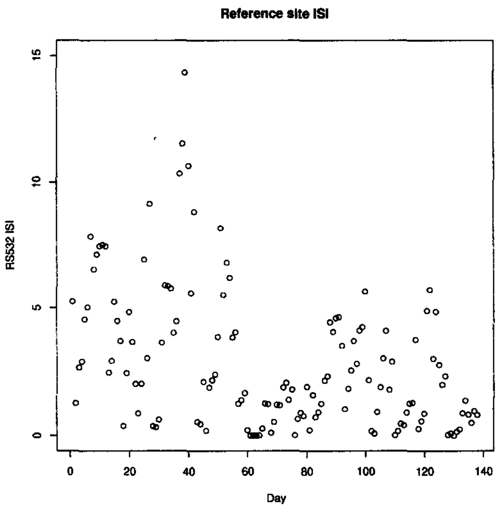


(c) FFMC for ES1017

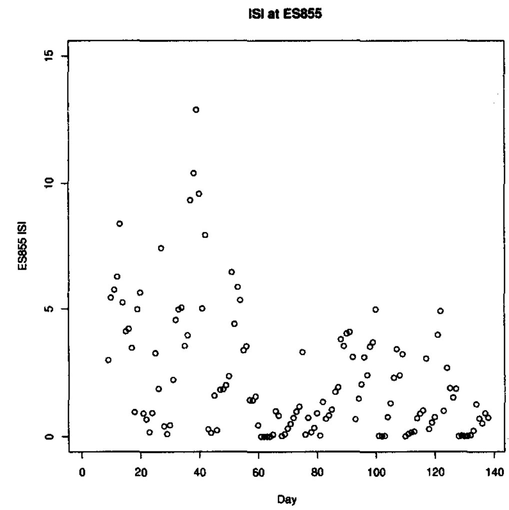


(d) FFMC for ES1166

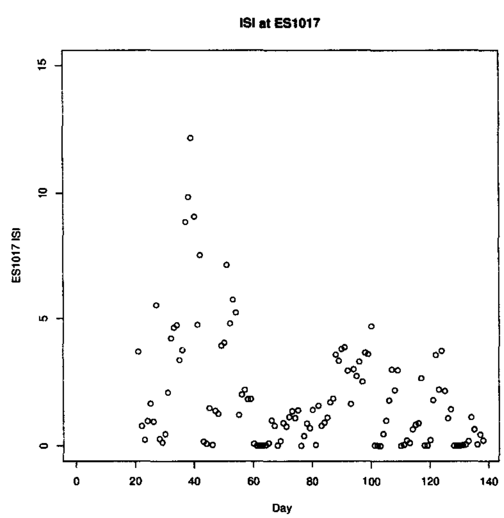
Figure A.4: Fine Fuel Moisture Code (FFMC) for Reference and Elevation Sites



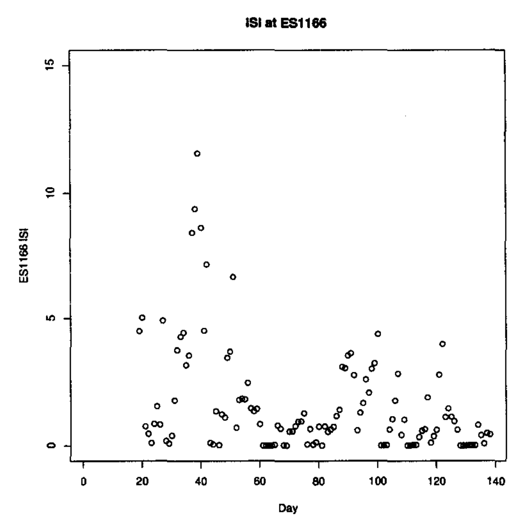
(a) ISI for Reference site



(b) ISI for ES855

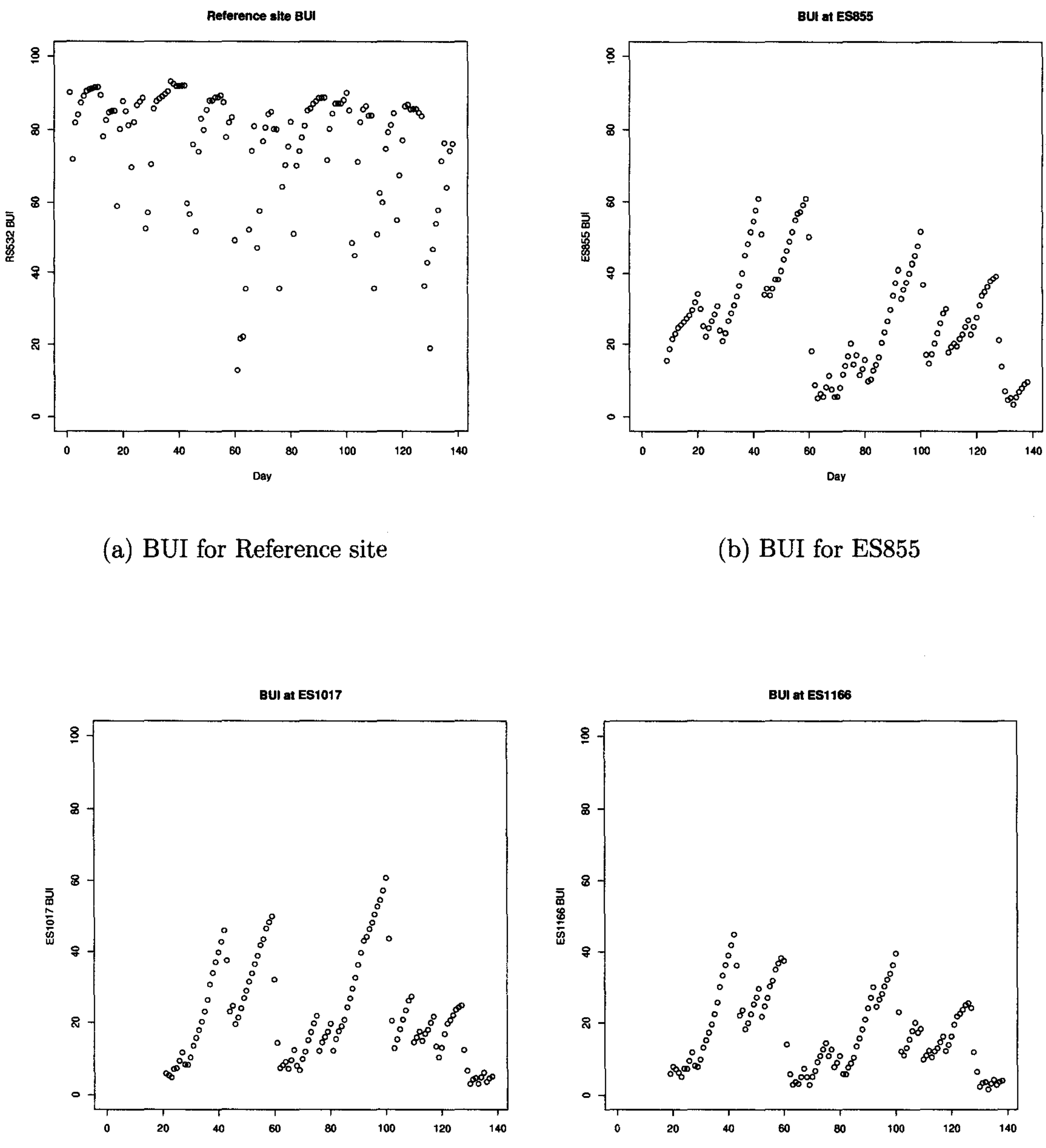


(c) ISI for ES1017



(d) ISI for ES1166

Figure A.5: Initial Spread Index (ISI) for Reference and Elevation Sites



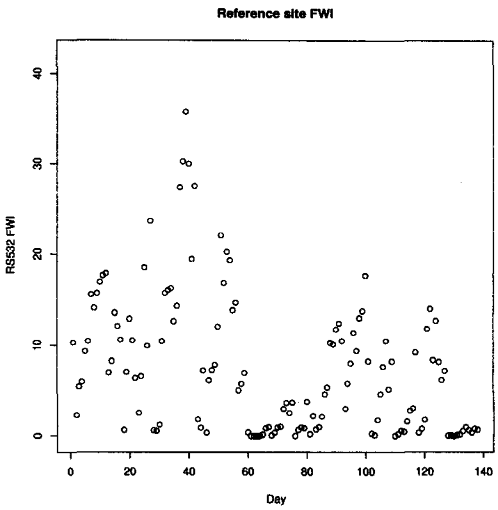
(a) BUI for Reference site

(b) BUI for ES855

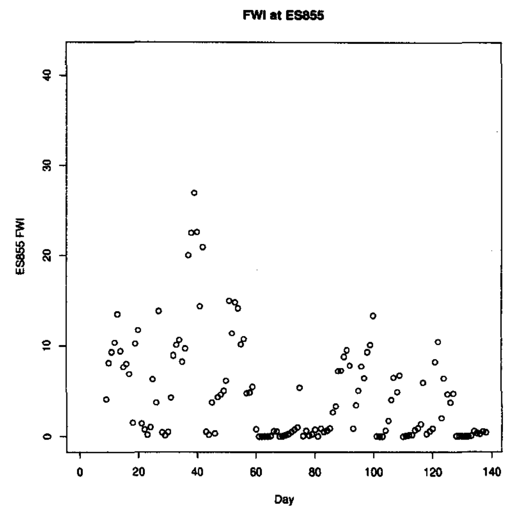
(c) BUI for ES1017

(d) BUI for ES1166

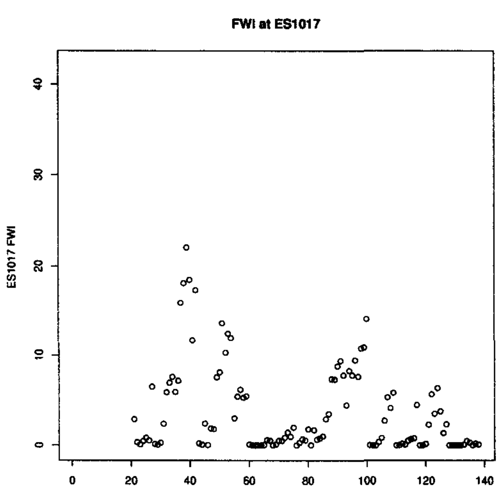
Figure A.6: Building Index (BUI) for Reference and Elevation Sites



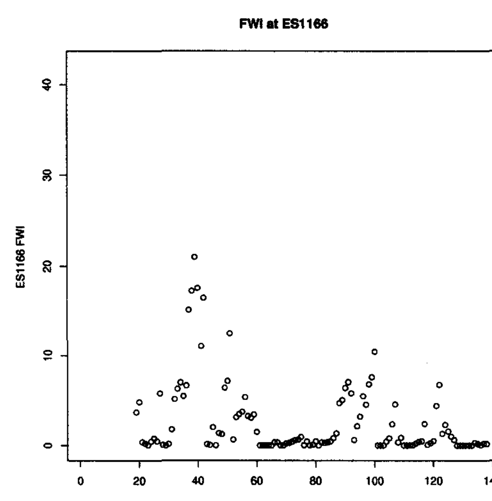
(a) FWI for Reference site



(b) FWI for ES855



(c) FWI for ES1017



(d) FWI for ES1166

Figure A.7: Fire Weather Index (FWI) for Reference and Elevation Sites