An Improved Meta-analysis for Analyzing Cylindrical-type Time Series Data with Applications to Forecasting Problem in Environmental Study

by

Shuo Wang

A Thesis

Submitted to the Faculty

of the

WORCESTER POLYTECHNIC INSTITUTE

In partial fulfillment of the requirements for the

Degree of Master of Science

in

Applied Statistics

by

_____

May 2015

APPROVED:

_____
Professor Sungsu Kim, Thesis Co-advisor

_____
Professor Thelge Buddika Peiris, Thesis Co-advisor

Abstract

This thesis provides a case study on how the wind direction plays an important role in the amount of rainfall, in the village of Somió. The primary goal is to illustrate how a meta-analysis, together with circular data analytic methods, helps in analyzing certain environmental issues. The existing GLS meta-analysis combines the merits of usual meta-analysis that yields a better precision and also accounts for covariance among coefficients. But, it is quite limited since information about the covariance among coefficients is not utilized. Hence, in my proposed meta-analysis, I take the correlations between adjacent studies into account when employing the GLS meta-analysis. Besides, I also fit a time series linear-circular regression as a comparable model. By comparing the confidence intervals of parameter estimates, covariance matrix, AIC, BIC and p-values, I discuss an improvement on the GLS meta analysis model in its application to forecasting problem in Environmental study.

Keywords: Covariance matrix, Generalized least square, Linear-circular regression, Meta-analysis, Time series

Acknowledgements

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Climate can be influenced by many variables like wind direction, humidity, SO_2 or temperature, but most studies investigated by climatology focus on wind direction and related variables[4] such as rainfall amount. A non-parametric method introduced in [3] summarizes that rainfall amount is particularly influenced by wind direction. Motivated by the paper, my goal of this study is to step further and try to model it with a parametric approach in order to determine an association between wind direction and amount of rainfall.

Differently from linear variables, it worth noting that wind direction is a circular variable, which arises when we measure it in the form of angle. Since it is presented as a point on the circumference of the unit circle, it should be dealt quite differently from usual statistical analysis, called Circular Statistics[8]. Also bearing in mind the nature of a time series, some backgrounds on time series[11]is needed, as long as some knowledge about linear-circular regression[7][6], where a linear-circular regression refers to a regression with a linear dependent variable and a circular independent variable.

In this thesis, I use a meta-analysis on a linear-circular regression, whose main

goal is to provide an estimated overall effect by combining the results from related small studies. A prevalent practice in recent years in a meta-analysis is the rationale for summarizing regression coefficients. A meta-analysis yields a better precision of regression coefficient estimates than the usual linear-regression coefficient estimates in each small study. In [1], the authors briefly review several existing approaches in meta-analysis and point out the complexities and potential problems in synthesizing coefficients from regression models.

The synthesis of regression coefficients is very difficult for several reasons including nonequivalence of the matrix for the predictors and outcomes across studies, very diverse models across studies and lack of information of raw data. First, considering that Y should be measured similarly across studies is very important, because the raw regression coefficients in each study depends on the scales of that predictor and the outcome. For two-scale coefficients to be comparable across studies, the scales of X and Y must be the same (or proportional). Another problematic assumption is that focal X should be measured similarly across studies. When the index of study results is an elasticity and represents proportional change in X and Y (commonly in economics, such as inflation), the scale of X may not be critical.

Apart from strict measurements, there also exists limitations for existing methods, the article[5] used direct and simple summaries of slopes. By using ordinary least square (OLS) regression analysis for a dummy variable that represented union membership to predict low wage between union and non-union workers, they didn't acknowledge that the errors in their model were likely to be heteroscedastic. This method focuses on a single focal coefficient, while ignores dependence and precision of coefficients. In addition, the weighted least squares (WLS) approach was proposed in [2]. Its method combines slopes by using weights after estimating slopes in each study using method like ordinary least square. The strength of weighted least

square approach is that it is relatively simple and has ability to handle regression slopes in which the coefficients are varying. But the weakness is the same as the simple summaries of slopes that it ignores dependence of coefficients.

In [1], the authors presented a new approach based on generalized least square (GLS) estimation of the synthesis of regression coefficients and outcomes, since previous existing methods ignore dependence among coefficients and the inherent precision of coefficients across studies. It combines the merits in previous meta-analysis and also accounts for covariance among coefficients. An overview of the use of the GLS is given in Chapter 3.

In my proposed model, I employ 12 months as small studies in 240 monthly time series. After confirming a significant association between rainfall amount and wind direction, I combine the results of 12 month studies consisted of 20 years data, in order to determine an overall linear relationship between rainfall and wind direction. When calculating the overall effect of wind direction on rainfall amount, the meta-analysis proposed in [1] did not consider a possible correlation between adjacent studies, and treated them as independent studies which may threaten the validity of the resulting conclusions. In this thesis, I claim that a meta-analysis that assumes independence of studies has a limitation due to an apparent dependent structure present among related small studies. Hence, in the proposed meta-analysis, a new form of covariance matrix is utilized, where the correlations between adjacent studies are taken into account and GLS method is employed.

Additionally, given the nature that the rainfall data form a monthly time series of successive measurements made over twenty years, I also fit a time series linear-circular regression of rainfall amount (linear) on wind direction (circular), and include in a comparison study as a comparable model with the proposed and existing meta-analysis models. Autoregressive moving average (ARMA) model is

considered for the dataset.

My thesis is organized as follows. In Chapter 2, data preparation is given along with data description. In Chapter 3, I provide a review of the existing methods in meta-analysis. In Chapter 4, I propose an improved model from the existing model. In Chapter 5, I build the regression model using both existing and proposed models. The results show some differences and similarities. In Chapter 6 and Chapter 7, I discuss the results from the three models: proposed GLS model, existing GLS model, and time series linear-circular regression model, where I also summarize their main strengths and limitations.

Chapter 2

Data Description and Preparation

2.1 Data Description

The dataset contains 20×240 daily observations of rainfall amount and wind direction from 1995 to 2014 obtained from the Davis station in the village of Somió, about 4 km from Gijón. The observatory coordinates are 43°32'17" N, 5°37'26" W and 30 m above sea level. The dataset is available from `http://infomet.am.ub.es/clima/gijon/`. Wind directions are measured in degrees from 0° to 359° and rainfall amounts are measured on a grid of 0.2 liters per square meter.

2.2 Data Preparation

Figure 2.1 shows histograms of wind direction conditioning on the weather being dry days and rainy days. The difference between two graphs is evidently shown in the figure.

Due to what I am interested in is the wind direction when orographic rain is produced, I select the data from the rainy days only and calculate the monthly average of the amount of rainfall, as well as the circular mean of wind direction in each
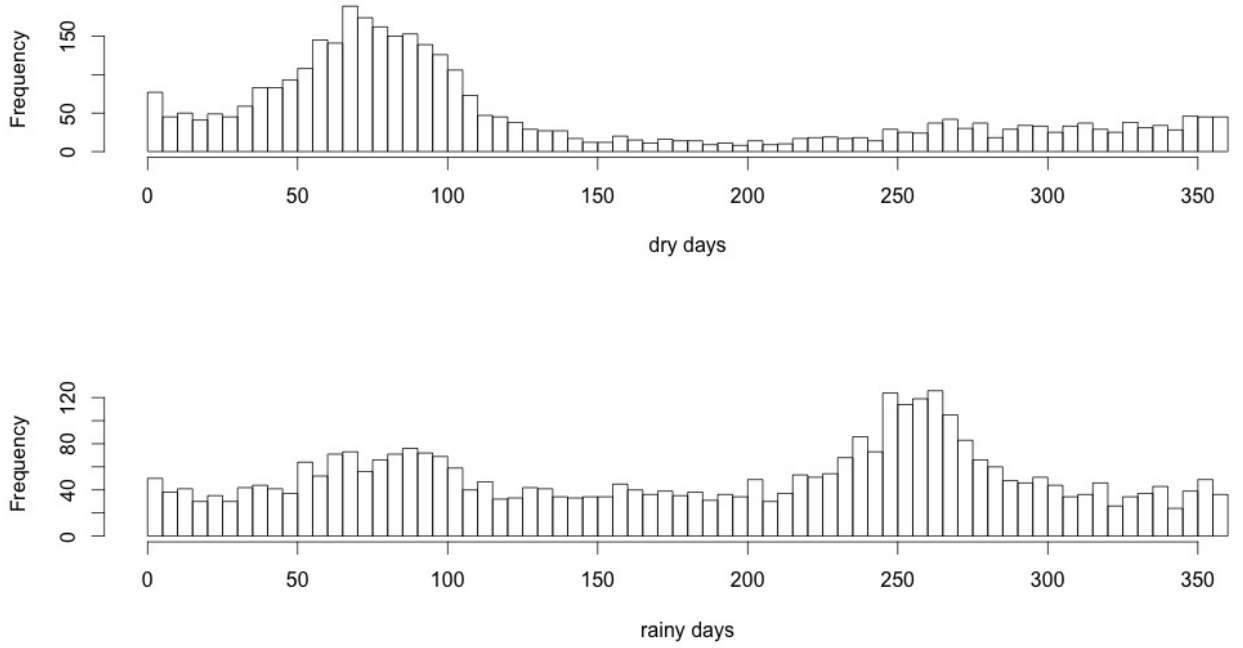
Figure 2.1: Histograms of wind direction on dry days(top) and rainy days(bottom)

month. The circular mean of wind direction is calculated in the following way[8]. Consider a random sample of circular observations of size n, denote $\theta_1, \ldots, \theta_n$. Then, we obtain

$$Y = \frac{\sum_{i=1}^{n} \sin \theta_i}{n}, \quad X = \frac{\sum_{i=1}^{n} \cos \theta_i}{n},$$

$$r = \sqrt{X^2 + Y^2}$$

$$\cos \bar{\theta} = \frac{X}{r}, \quad \sin \bar{\theta} = \frac{Y}{r}, \quad \bar{\theta} = \arctan^* \left(\frac{\sin \bar{\theta}}{\cos \bar{\theta}} \right),$$

6

where arctan* is quadrant specific as shown below, and $\bar{\theta}$ denote the circular mean.

$$\bar{\theta} = \begin{cases} \arctan(\frac{\sin\bar{\theta}}{\cos\bar{\theta}}), & \text{if } \cos\bar{\theta} > 0 \text{ and } \sin\bar{\theta} > 0. \\ \pi - \arctan(\frac{\sin\bar{\theta}}{\cos\bar{\theta}}), & \text{if } \cos\bar{\theta} < 0 \text{ and } \sin\bar{\theta} > 0. \\ \arctan(\frac{\sin\bar{\theta}}{\cos\bar{\theta}}) + \pi, & \text{if } \cos\bar{\theta} < 0 \text{ and } \sin\bar{\theta} < 0. \\ 2\pi - \arctan(\frac{\sin\bar{\theta}}{\cos\bar{\theta}}), & \text{if } \cos\bar{\theta} > 0 \text{ and } \sin\bar{\theta} < 0. \end{cases}$$

Figure 2.2 shows a rose diagram (circular histogram) of the monthly circular mean of wind directions for rainy days only. Areas of the sectors in the rose diagram indicate the frequencies in 12 divided sub-intervals of 360 degrees. It is shown that predominant wind direction is towards $\frac{17\pi}{12}$ in radians.
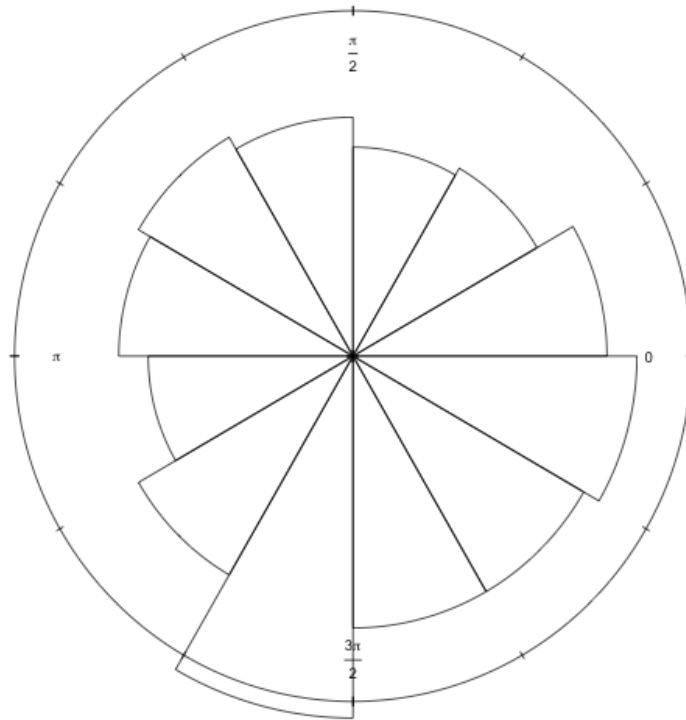
Figure 2.2: Rose diagram for the monthly circular mean of wind directions.

In Figure 2.3, it is shown that the density plot of the monthly average of rainfall amount is skewed to the right. In order to meet the assumption that the errors

follow a normal distribution, a $\log_{10}$-transform is applied to the monthly average of rainfall amount. A more normal looking of the density plot of rainfall amount after $\log_{10}$-transform is shown in Figure 2.4.
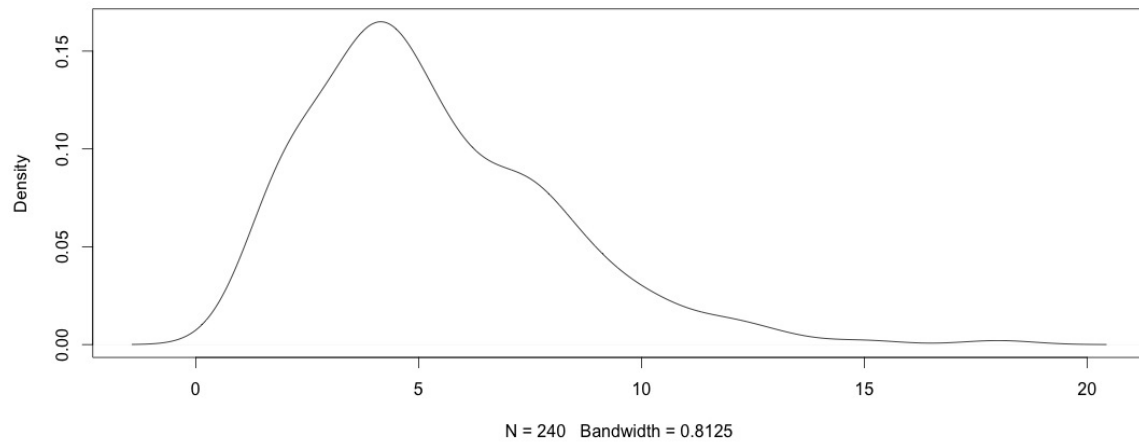
Figure 2.3: Density plot of the monthly average of rainfall amount
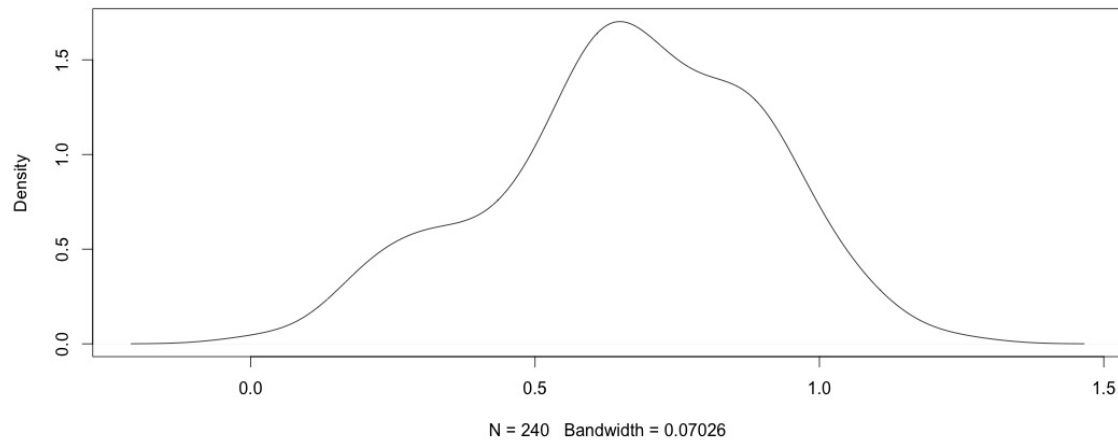
Figure 2.4: Density plot of the $\log_{10}$-transformed monthly average of rainfall amount

2.3 Linear-Circular Association

Following the logic of Spearman's rank correlation coefficient[10], Mardia proposed a rank-based analogue[9], where $x'_j s$ and $\theta'_j s$ are replaced by their ranks and uniform scores, respectively. Although ranks are not well defined on a circle, the rank correlation measure is origin-invariant and provides a useful measure.

To calculate the correlation coefficient, the original data vector, $x_1, \ldots, x_n$, are first reordered by $x_{(1)} \leq \cdots \leq x_{(n)}$, where $x_{(i)}$ is the i_{th} largest value among $x_1, \ldots, x_n$, $i = 1, \ldots, n$. Next, the uniform scores of the associated θ_i are computed, which is given by $\frac{2\pi r_i}{n}$, where r_i is the linear rank relative to an arbitrary origin of the angle θ_i paired with x_i in which $x'_i s$ are replaced by their ranks. The Mardia's rank correlation coefficient is obtained by,

$$U_n = \frac{24(T_c^2 + T_s^2)}{n^2(n+1)}$$

where

$$T_c = \sum_{j=1}^{n} i \cos\left(\frac{2\pi r_i}{n}\right), T_s = \sum_{i=1}^{n} i \sin\left(\frac{2\pi r_i}{n}\right)$$

A test of independence based on U_n rejects the independence of linear and circular variables, if the observed value of U_n is large compared to the percentiles of its sampling distribution under the independence. For large n, the sampling distribution of U_n under the independence is approximately chi-squared with two degrees of freedom.

Chapter 3

Existing Methods

3.1 Meta-Analysis

For a simple linear regression, consider a model in study i relating predictor X to an outcome Y for observation j, i.e

$$Y_{ij} = \beta_{i0} + \beta_{i1}X_{ij} + e_{ij}$$

for $i = 1, \ldots, k$ studies and $j = 1, 2, \ldots, n_i$ cases. The usual assumptions of normality and homoscedasticity of errors apply such that $e_{ij} \sim N(0, \sigma_i^2)$.

The generalized least square (GLS) method is used to combine ordinary least square (OLS) estimates of slopes and intercept from each of the k studies. The OLS estimator of $\beta_i = (\beta_{i0}, \beta_{i1})$ is given by $b_i = (b_{i0}, b_{i1}) = (X_i'X_i)^{-1}X_i'Y_i$ with $\Sigma_i = \text{Cov}(b_i) = (X_i'X_i)^{-1}\sigma_i^2$,where X_i is given by

$$\begin{pmatrix} 1 & x_{i1} \\ \vdots & \vdots \\ 1 & x_{in_i} \end{pmatrix}.$$

Usually σ_i^2 is unknown, so it is estimated by s_i^2, the mean square error (MSE) of the regression in study i. After stacking k coefficients. The GLS estimation is given in the following.

First, the coefficient estimates vector is given by

$$b_{2k \times 1} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix} = \begin{bmatrix} b_{10} \\ b_{11} \\ b_{20} \\ b_{21} \\ \vdots \\ b_{k0} \\ b_{k1} \end{bmatrix},$$

and Σ, the $2k \times 2k$ block diagonal matrix, is given by

$$\Sigma = \begin{bmatrix} \mathrm{cov}(b_1) & 0 & 0 & 0 \\ 0 & \mathrm{cov}(b_2) & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \mathrm{cov}(b_k) \end{bmatrix}. \tag{3.1}$$

The following model is used in order to get the OLS estimates of β_0 and β_1,

$$b = W * \beta + e$$

$$= \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \end{bmatrix} * \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + e. \tag{3.2}$$

A design matrix W is composed of 0's and 1's that identify intercept and slope estimates in each sample. The GLS estimates of β_0 and β_1 and their covariances are given by

$$\hat{\beta} = (W'\Sigma^{-1}W)^{-1}W\Sigma^{-1}b \tag{3.3}$$

and

$$\text{Cov}(\hat{\beta}) = (W'\Sigma^{-1}W)^{-1}. \tag{3.4}$$

With a large sample and under typical regularity conditions,

$$\hat{\beta} \sim N(\beta, \text{Cov}(\hat{\beta})).$$

Thus confidence intervals for each element of β are available, using $\hat{\beta}_p \pm Z_{\frac{1-\alpha}{2}} \sqrt{C_{pp}}$ where $Z_{\frac{1-\alpha}{2}}$ is the upper tail $1 - \frac{\alpha}{2}$ critical value of the standard normal distribution and C_{pp} is the $(p+1)_{th}$ diagonal element of the $\text{Cov}(\hat{\beta})$ matrix, the variance of $\hat{\beta}$, for $p = 0, 1$.

3.2 Time Series Linear-Circular Regression

A time series model known as ARMA model may include autoregressive (AR) terms and moving average (MA) terms in [11]. An autoregressive model specifies that the output variable depends linearly on its own previous values. The notation $\text{AR}(p)$ indicates an autoregressive model of order p, indicating p previous values are used to predict the present time. The model is defined as

$$Y_t = c + \sum_{i=1}^{p} \varphi_i Y_{t-i} + \varepsilon_t$$

where $\varphi_1, \ldots, \varphi_p$ are linear parameters of the model, c is a constant, and ε_t is an white noise. A moving average term in a time series model is a past error. The notation $\text{MA}(q)$ refers to the moving average model of order q. The $\text{MA}(q)$ model is defined as

$$Y_t = \mu + \varepsilon_t + \sum_{i=1}^{q} \gamma_i \varepsilon_{t-i}$$

where the $\gamma_1, \ldots, \gamma_q$ are linear parameters, μ is the mean of the series (often assumed to equal 0), and the $\varepsilon_t, \varepsilon_{t-1,\ldots}$ are white noise error terms. $\text{ARMA}(p, q)$ refers to the model with p autoregressive terms and q moving-average terms. This model contains the $\text{AR}(p)$ and $\text{MA}(q)$ models,

$$Y_t = c + \varepsilon_t + \sum_{i=1}^{p} \varphi_i Y_{t-i} + \sum_{i=1}^{q} \gamma_i \varepsilon_{t-i}$$

where $\varphi_1, \ldots, \varphi_p$ and $\gamma_1, \ldots, \gamma_q$ are linear parameters, c is a constant, and $\varepsilon_t, \varepsilon_{t-1,\ldots}$ are white noise error terms.

Looking at the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots of the stationary series, we can tentatively identify the numbers of AR and MA terms that are needed. If the partial autocorrelation function (PACF) of

13

the series displays a sharp cutoff or the lag-1 autocorrelation is positive, then adding one or more AR terms to the model is desired. If the autocorrelation function (ACF) of the series displays a sharp cutoff or the lag-1 autocorrelation is negative, then we may consider adding an MA term to the model.

Combined with the significant association between rainfall amount and wind direction, we consider a regression model that involves a simple cosine function given by

$$Y_t = \beta_0 + \beta_1 \cos(\theta_t - \mu) + \sum_{i=1}^{p} \varphi_i Y_{t-i} + \sum_{i=1}^{q} \gamma_i \varepsilon_{t-i} + \varepsilon_t$$

where Y_t is the linear response variable, μ is the sample mean direction, β_0 and β_1 are the regression coefficients, θ_t is the circular independent variable subject to time t, $\varepsilon_t's$ are white noises. Choosing a good model is an important step in the analysis of a time series regression in the later chapter. The method of linear least squares is applied to the parameter estimation.

Chapter 4

Proposed Method

A meta-analysis is used to get overall parameter estimates by combining many results from related small studies. However, the problem with the existing meta-analysis described in Section 3.1 is that the off-diagonal entries of the block diagonal matrix in (3.1) are equal to zero, indicating independence of k included small studies. Synthesizing regression coefficients with block diagonal elements only can be potentially misleading the outcome of a meta-analysis due to apparent dependent structure present among small studies.

In our example, we consider a model in study i relating the predictor $\cos(\theta_t - \mu)$ to an outcome Y for case j. Specifically, in study i,

$$Y_{ij} = \beta_{i0} + \beta_{i1} \cos(\theta_{ij} - \mu) + e_{ij}$$

for $i = 1, 2, \ldots, 12$ studies and $j = 1, 2, \ldots, 20$ cases. One important thing to note is that under usual assumptions of normality and homoscendasticity of errors, the covariance of error term across time periods is zero and variance of the error is constant across time. But in the presence of monthly correlation, the covariance of error term across different time periods is no longer zero. Let e_t be the error term

15

at time t. The error covariance matrix Ω is a 12×12 matrix that takes the form

$$\Omega = \begin{pmatrix} \sigma_1^2 & \sigma_{1,2} & \cdots & \sigma_{1,12} \\ \sigma_{2,1} & \sigma_2^2 & \cdots & \sigma_{2,12} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{12,1} & \sigma_{12,2} & \cdots & \sigma_{12}^2 \end{pmatrix}$$

where $\sigma_{i,j}$ means error covariance between observation i and j studies. For simplicity, I assume that only the adjacent months are related in my example, implying that

$$\mathrm{cov}(e_{t_1}, e_{t_2}) = \begin{cases} \sigma_t^2 & t_1 = t_2 = t \\ \sigma_{t_1,t_2} & |t_1 - t_2| = 1 \\ 0 & |t_1 - t_2| > 1 \end{cases}.$$

Therefore, the error covariance matrix takes the following form

$$\Omega = \begin{pmatrix} \sigma_1^2 & \sigma_{1,2} & 0 & 0 & \cdots & 0 & \sigma_{1,12} \\ \sigma_{2,1} & \sigma_2^2 & \sigma_{2,3} & 0 & \cdots & 0 & 0 \\ 0 & \sigma_{3,2} & \sigma_3^2 & \sigma_{3,4} & \cdots & 0 & 0 \\ 0 & 0 & \sigma_{4,3} & \sigma_4^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \sigma_{11}^2 & \sigma_{11,12} \\ \sigma_{12,1} & 0 & 0 & 0 & \cdots & \sigma_{12,11} & \sigma_{12}^2 \end{pmatrix}$$

where σ_{t_i,t_j} refers error covariance between adjacent month in our example, $t_1, t_2 = 1, 2, \ldots, 12$. Using (3.3) and (3.4), I obtain the vector of GLS estimates of the regression coefficients and its covariance matrix as

$$\hat{\beta}^* = (W'(\Sigma^*)^{-1}W)^{-1}W'(\Sigma^*)^{-1}b$$

and

$$\text{Cov}(\hat{\beta}^*) = (W'(\Sigma^*)^{-1}W)^{-1},$$

where b is a 24×1 vector, W is a 24×2 design matrix give in (3.2). Then employing a large sample theory, $\hat{\beta}^* \sim N(\beta, \text{Cov}(\hat{\beta}^*))$, similarly to the covariance matrix of error, the covariance structure of coefficients estimates in our example is obtained in the following. However, in the proposed model, one can assume a more complicated covariance structure.

$$\Sigma^* = \begin{pmatrix} \text{cov}(b_1, b_1) & \text{cov}(b_1, b_2) & 0 & 0 & \cdots & 0 & \text{cov}(b_1, b_{12}) \\ \text{cov}(b_2, b_1) & \text{cov}(b_2, b_2) & \text{cov}(b_2, b_3) & 0 & \cdots & 0 & 0 \\ 0 & \text{cov}(b_3, b_2) & \text{cov}(b_3, b_3) & \text{cov}(b_3, b_4) & \cdots & 0 & 0 \\ 0 & 0 & \text{cov}(b_4, b_3) & \text{cov}(b_4, b_4) & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \text{cov}(b_{11}, b_{11}) & \text{cov}(b_{11}, b_{12}) \\ \text{cov}(b_{12}, b_1) & 0 & 0 & 0 & \cdots & \text{cov}(b_{12}, b_{11}) & \text{cov}(b_{12}, b_{12}) \end{pmatrix},$$

where

$$\text{cov}(b_i, b_j) = \begin{bmatrix} \text{cov}(b_{i0}, b_{j0}) & 0 \\ 0 & \text{cov}(b_{i1}, b_{j1}) \end{bmatrix}$$

in which $\text{cov}(b_{i1}, b_{j1})$ and $\text{cov}(b_{i0}, b_{j0})$ are given by

$$\text{cov}(b_{i1}, b_{i'1}) = \text{cov}\left(\frac{\sum_{j=1}^{N}(x_{ij} - \bar{x}_i)(y_{ij} - \bar{y}_i)}{\sum_{j=1}^{N}(x_{ij} - \bar{x}_i)^2}, \frac{\sum_{j=1}^{N}(x_{i'j} - \bar{x}'_i)(y_{i'j} - \bar{y}'_i)}{\sum_{j=1}^{N}(x_{i'j} - \bar{x}_{i'})^2}\right)$$

$$= \frac{1}{(N-1)^2}\frac{1}{s_{x_i}^2 s_{x_{i'}}^2}\text{cov}\left(\sum_{j=1}^{N}(x_{ij} - \bar{x}_i)(y_{ij} - \bar{y}_i), \sum_{j=1}^{N}(x_{i'j} - \bar{x}_{i'})(y_{i'j} - \bar{y}_{i'})\right)$$

$$= \frac{1}{(N-1)^2}\frac{1}{s_{x_i}^2 s_{x_{i'}}^2}\sum_{j=1}^{N}(x_{ij} - \bar{x}_i)(x_{i'j} - \bar{x}_{i'})\text{cov}(y_{ij} - \bar{y}_i, y_{i'j} - \bar{y}_{i'})$$

$$= \frac{1}{(N-1)^2}\frac{1}{s_{x_i}^2 s_{x_{i'}}^2}\sum_{j=1}^{N}(x_{ij} - \bar{x}_i)(x_{i'j} - \bar{x}_{i'})\frac{N-1}{N}\text{cov}(y_{ij}, y_{i'j})$$

$$= \frac{1}{(N-1)^2}\frac{1}{s_{x_i}^2 s_{x_{i'}}^2}(N-1)\text{cov}(x_{ij}, x_{i'j})\frac{N-1}{N}\text{cov}(y_{ij}, y_{i'j})$$

$$= \frac{1}{N}\frac{1}{s_{x_i}^2 s_{x_{i'}}^2}\text{cov}(x_{ij}, x_{i'j})\text{cov}(y_{ij}, y_{i'j}),$$

$$\text{cov}(b_{i0}, b_{i'0}) = \text{cov}(\bar{y}_i - b_{i1}\bar{x}_i, \bar{y}_{i'} - b_{i'1}\bar{x}_{i'})$$

$$= \text{cov}(\bar{y}_i, \bar{y}_{i'}) - \text{cov}(\bar{y}_i, b_{i'1}\bar{x}_{i'}) - \text{cov}(b_{i1}\bar{x}_i, \bar{y}_{i'}) + \bar{x}_i\bar{x}_{i'}\text{cov}(b_{i1}, b_{i'1})$$

$$= \frac{1}{N}\text{cov}(y_{ij}, y_{i'j}) + \bar{x}_i\bar{x}_{i'}\text{cov}(b_{i1}, b_{i'1})$$

$$= \frac{1}{N}\text{cov}(y_{ij}, y_{i'j}) + \bar{x}_{ij}\bar{x}_{i'j}\frac{1}{N}\frac{1}{s_{x_i}^2 s_{x_{i'}}^2}\text{cov}(x_{ij}, x_{i'j})\text{cov}(y_{ij}, y_{i'j}).$$

Chapter 5

Results

In this chapter, I use 240 monthly observations of amount of rainfall and wind direction obtained in Davis station to analyze a functional relationship between rainfall amount and wind direction. By treating 12 months as small related studies. we fit the proposed meta-analysis model, the traditional meta-analysis model and a time series linear-circular regression model.

5.1 Meta-analysis

The regression model with $\cos(\theta_t - \mu)$ as predictors of Y_t was estimated within each month. The coefficient estimates are shown in Table 5.1. Inspection of the models for 12 months yields some variation in the slopes and intercepts.

5.1.1 Existing Method

In our example, the combined regression coefficients are estimated using 12 regression models of each study. The design matrix W is in (5.1), the OLS estimates of $\beta_i's$, $i = 1, \ldots, 12$ are =in Table 5.1, and the covariance matrix for 12 months are

Month	n_i	Intercept	$cos(\theta_t - \mu)$	MSE
1	20	0.682	-0.032	0.030
2	20	0.661	0.049	0.045
3	20	0.694	0.006	0.055
4	20	0.673	0.007	0.053
5	20	0.565	0.161	0.041
6	20	0.528	0.108	0.068
7	20	0.591	0.080	0.069
8	20	0.630	-0.058	0.062
9	20	0.669	0.162	0.043
10	20	0.693	0.126	0.036
11	20	0.786	0.128	0.035
12	20	0.731	0.033	0.025

Table 5.1: Fitted regression coefficients and MSE values for 12 studies

shown in (5.2).

$$
W = \begin{pmatrix}
1 & 0 \\
0 & 1 \\
1 & 0 \\
0 & 1 \\
1 & 0 \\
0 & 1 \\
1 & 0 \\
0 & 1 \\
1 & 0 \\
0 & 1 \\
1 & 0 \\
0 & 1 \\
1 & 0 \\
0 & 1 \\
1 & 0 \\
0 & 1 \\
1 & 0 \\
0 & 1 \\
1 & 0 \\
0 & 1 \\
1 & 0 \\
0 & 1 \\
1 & 0 \\
0 & 1
\end{pmatrix}, b = \begin{pmatrix}
0.6818 \\
-0.0317 \\
0.6608 \\
0.0485 \\
0.6940 \\
0.0061 \\
0.6731 \\
0.0072 \\
0.5647 \\
0.1612 \\
0.5276 \\
0.1079 \\
0.5911 \\
0.0798 \\
0.6302 \\
-0.0582 \\
0.6691 \\
0.1617 \\
0.6931 \\
0.1256 \\
0.7858 \\
0.1281 \\
0.7311
\end{pmatrix}.
\tag{5.1}
$$

With 12 studies in total, there are 12 within-study covariance matrix. The within study covariance matrices for study 1 through 12 are:

$$\text{Cov}(b_1) = \begin{pmatrix} 0.00178 & -0.00067 \\ -0.00067 & 0.00374 \end{pmatrix}, \text{Cov}(b_2) = \begin{pmatrix} 0.00251 & -0.00040 \\ -0.00040 & 0.00429 \end{pmatrix},$$

$$\text{Cov}(b_3) = \begin{pmatrix} 0.00328 & -0.00110 \\ -0.00110 & 0.00521 \end{pmatrix}, \text{Cov}(b_4) = \begin{pmatrix} 0.00352 & -0.00190 \\ -0.00190 & 0.00598 \end{pmatrix},$$

$$\text{Cov}(b_5) = \begin{pmatrix} 0.00232 & -0.00050 \\ -0.00050 & 0.00482 \end{pmatrix}, \text{Cov}(b_6) = \begin{pmatrix} 0.00425 & -0.00230 \\ -0.00230 & 0.01100 \end{pmatrix},$$

$$\text{Cov}(b_7) = \begin{pmatrix} 0.00502 & -0.00388 \\ -0.00388 & 0.01240 \end{pmatrix}, \text{Cov}(b_8) = \begin{pmatrix} 0.00401 & -0.00213 \\ -0.00213 & 0.00792 \end{pmatrix},$$

$$\text{Cov}(b_9) = \begin{pmatrix} 0.00239 & -0.00015 \\ -0.00015 & 0.00420 \end{pmatrix}, \text{Cov}(b_{10}) = \begin{pmatrix} 0.00199 & 0.00021 \\ 0.00021 & 0.00299 \end{pmatrix},$$

$$\text{Cov}(b_{11}) = \begin{pmatrix} 0.00202 & -0.00048 \\ -0.00048 & 0.00426 \end{pmatrix}, \text{Cov}(b_{12}) = \begin{pmatrix} 0.00146 & -0.00045 \\ -0.00045 & 0.00303 \end{pmatrix}.$$

The covariance matrix Σ for meta-analysis is a 24×24 matrix like in (3.1) as: $Diag(\text{Cov}(b_1), \ldots, \text{Cov}(b_{12}))$. After substituting an estimate of covariance matrix, we compute $\hat{\beta}^*$ and its covariance as equation (3.3) and (3.4) and get,

$$\hat{\beta} = \begin{pmatrix} 0.66760928 \\ 0.06319353 \end{pmatrix}, \quad \text{Cov}(\hat{\beta}) = \begin{pmatrix} 0.000205 & -0.0000559 \\ -0.0000559 & 0.000391 \end{pmatrix}. \qquad (5.2)$$

The data from 12 monthly studies are pooled and the overall estimated model for 12 months together is:

$$\hat{Y}_t = 0.66761 + 0.06319 \cos(\theta_t - \mu). \tag{5.3}$$

where μ is overall sample mean direction.

5.1.2 Proposed method

In the presence of serial correlation, the covariance term across different time periods is no longer zero. Our covariance matrix should look like (4.1), where any covariance between two consecutive month is calculated by

$$\text{cov}(b_i, b_{i'}) = \begin{bmatrix} \text{cov}(b_{i0}, b_{i'0}) & 0 \\ 0 & \text{cov}(b_{i1}, b_{i'1}) \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{N}\text{cov}(y_i, y_{i'}) + \bar{x}_i \bar{x}_{i'} \frac{1}{N} \frac{1}{S^2_{x_i} S^2_{x_{i'}}}\text{cov}(y_i, y_{i'}) & 0 \\ 0 & \frac{1}{N} \frac{1}{S^2_{x_i} S^2_{x_{i'}}}\text{cov}(x_i, x_{i'})\text{cov}(y_i, y_{i'}) \end{bmatrix} \tag{5.4}$$

The off-diagonal terms within covariance matrix for study 1 through 12 are:

$$\text{Cov}(b_1, b_2) = \text{Cov}(b_2, b_1) = \begin{pmatrix} 0.00035 & 0 \\ 0 & 0.0000156 \end{pmatrix},$$

$$\text{Cov}(b_2, b_3) = \text{Cov}(b_3, b_2) = \begin{pmatrix} 0.00113 & 0 \\ 0 & 0.0000816 \end{pmatrix},$$

$$\text{Cov}(b_3, b_4) = \text{Cov}(b_4, b_3) = \begin{pmatrix} 0.00091 & 0 \\ 0 & 0.0000591 \end{pmatrix},$$

$$\text{Cov}(b_4, b_5) = \text{Cov}(b_5, b_4) = \begin{pmatrix} 0.000823 & 0 \\ 0 & 0.000272 \end{pmatrix},$$

$$\text{Cov}(b_5, b_6) = \text{Cov}(b_6, b_5) = \begin{pmatrix} 0.00137 & 0 \\ 0 & 0.000516 \end{pmatrix},$$

$$\text{Cov}(b_6, b_7) = \text{Cov}(b_7, b_6) = \begin{pmatrix} -0.0011 & 0 \\ 0 & -0.00045 \end{pmatrix},$$

$$\text{Cov}(b_7, b_8) = \text{Cov}(b_8, b_7) = \begin{pmatrix} 0.0005 & 0 \\ 0 & -0.00053 \end{pmatrix},$$

$$\text{Cov}(b_8, b_9) = \text{Cov}(b_9, b_8) = \begin{pmatrix} 0.00085 & 0 \\ 0 & 0.0000582 \end{pmatrix},$$

$$\text{Cov}(b_9, b_{10}) = \text{Cov}(b_{10}, b_9) = \begin{pmatrix} 0.00065 & 0 \\ 0 & 0.00031 \end{pmatrix},$$

$$\text{Cov}(b_{10}, b_{11}) = \text{Cov}(b_{11}, b_{10}) = \begin{pmatrix} 0.00093 & 0 \\ 0 & 0.00082 \end{pmatrix},$$

$$\text{Cov}(b_{11}, b_{12}) = \text{Cov}(b_{12}, b_{11}) = \begin{pmatrix} 0.00106 & 0 \\ 0 & 0.0006 \end{pmatrix},$$

$$\text{Cov}(b_{12}, b_1) = \text{Cov}(b_1, b_{12}) = \begin{pmatrix} 0.000252 & 0 \\ 0 & 0.0000685 \end{pmatrix},$$

with the diagonal terms are same as (4.1). We compute $\hat{\beta}$ and its covariance as equation (4.4) and get

$$\hat{\beta}^* = \begin{pmatrix} 0.65136 \\ 0.05862 \end{pmatrix}, \quad \text{cov}(\hat{\beta}^*) = \begin{pmatrix} 0.000303 & -0.000065 \\ -0.000065 & 0.000430 \end{pmatrix}.$$

Hence the full estimated model using our proposed model is given by

$$\hat{Y}_t = 0.65136 + 0.05862 \cos(\theta_t - \mu)$$

where μ is overall sample mean direction.

5.2 Time Series Linear-Circular Regression

After $\log_{10}$-transformation, the time series of rainfall amount (See Figure 5.1) appears to be stationary. In order to identify important features on it, first, I obtain the autocorrelation function (ACF) and the partial autocorrelation function (PACF) plots to investigate any lagged forms for rainfall amount or white noise.
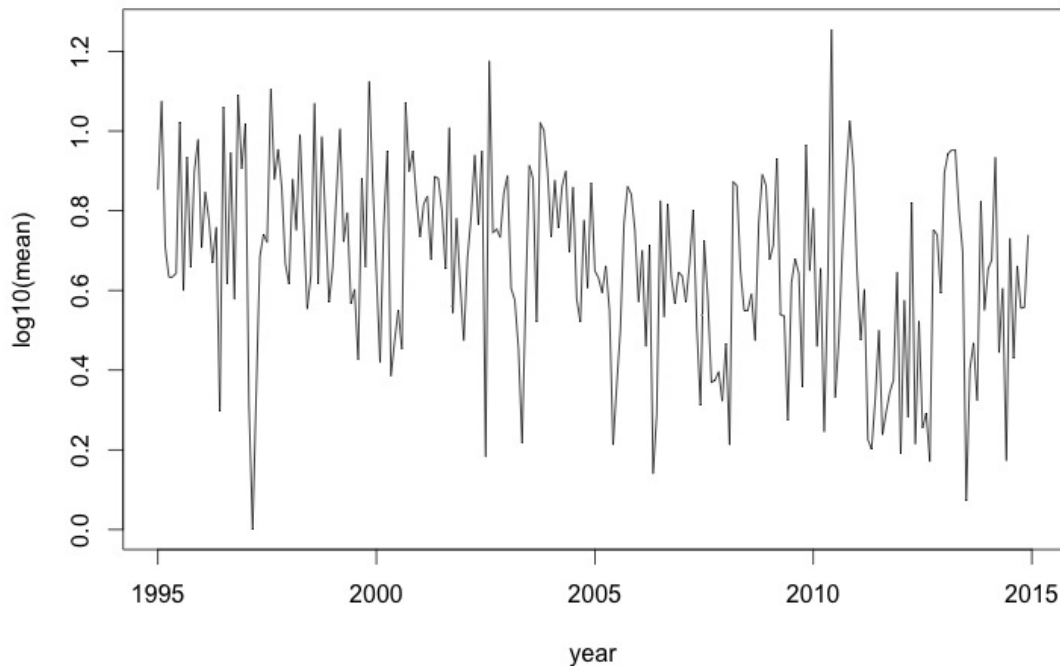

Figure 5.1: A time series plot of the $\log_{10}$-transformed monthly average rainfall amount.

Figure 5.2 and Figure 5.3 show the ACF and PACF plots. The dotted lines represent the 95% confidence intervals. Note that the PACF plot has a very large positive spike at lag 1 and no other significant spikes, indicating the lag-1 autocorrelation. However, the ACF does not display a sharp cutoff, and coefficients of lag-1 and lag-2 are both positive and almost equal. Therefore, I employ an AR(1) model and include Y_{t-1} in the time series linear-circular regression model.
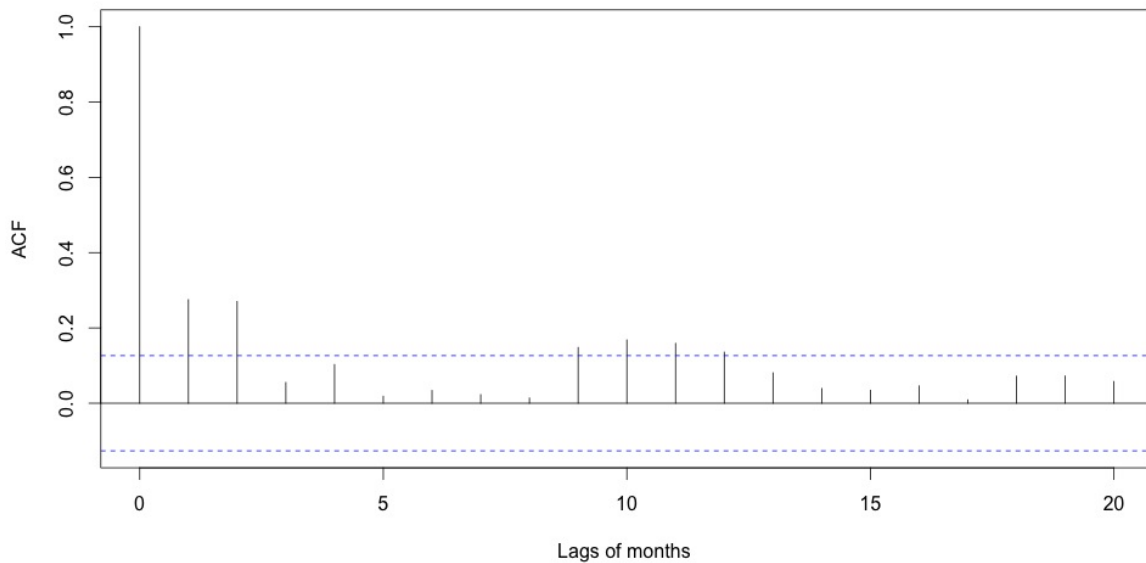
Figure 5.2: ACF of the monthly average rainfall amount after $\log_{10}$-transform.

Next, I investigate whether the dataset provides an evidence for an association between the amount of rainfall (a linear variable) and the wind direction (a circular variable), by employing the Mardia's rank correlation test [9] of independence between linear and circular variables. Since the estimated p-value is 0.0006, indicating an association between them, I reject the null hypothesis that the wind direction is independent of the rainfall amount in the dataset.

The regression model with Y_{t-1} and $\cos(\theta_t - \mu)$ as predictors of Y_t is estimated,
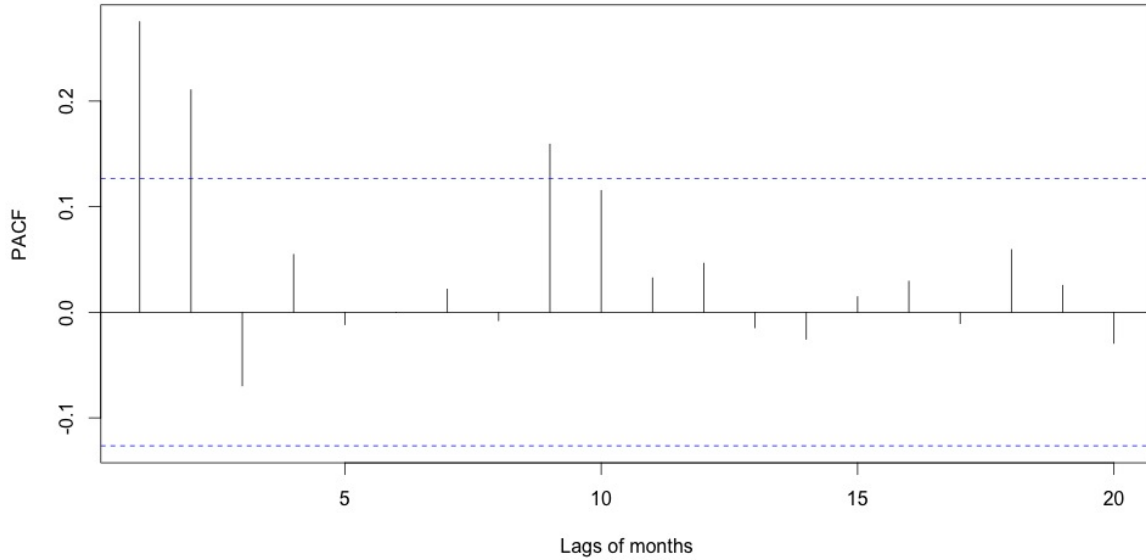
Figure 5.3: PACF of the monthly average rainfall amount after $\log_{10}$-transform.

and the estimates of coefficients and the p-values are shown in Table 5.2. The estimated model is given by $\hat{Y}_t = 0.486 + 0.256Y_{t-1} + 0.048\cos(\theta_t - \mu)$, where μ is overall sample mean direction.

Coefficients	Estimate	Standard Error	t-value	$\Pr(>|t|)$
intercept	0.486	0.044	11.144	2×10^{-16}
$\cos(\theta_t - \mu)$	0.048	0.021	2.303	0.0222
Y_{t-1}	0.256	0.062	4.102	5.65×10^{-5}

Table 5.2: Coefficient estimates and p-values of time series linear-circular regression

Chapter 6

Discussion

Table 6.1 presents the coefficients estimated using two meta-analyses. The first set of results was obtained by eliminating all off-diagonal elements from $\text{Cov}(b)$ matrix. The second one was obtained by the proposed method which utilizes correlation between adjacent studies, by including corresponding off-diagonal elements in the covariance matrix. The 95% confidence intervals (95% CI) are provided in the table.

Methods	β_0	95% CI of β_0	β_1	95% CI of β_1
Existing GLS	0.66761	[0.63955, 0.69567]	0.06319	[0.02444, 0.10195]
Proposed GLS	0.65136	[0.61724, 0.68547]	0.05862	[0.01798, 0.09926]

Table 6.1: Estimated coefficients and corresponding 95% confidence intervals given by the existing meta-analysis and the proposed meta-analysis

It is somewhat problematic that the length of 95% CI of the reported coefficients from the proposed model is a little bit larger than the values from existing method. $(L_{existing}(\beta_0) = 0.056, L_{proposed}(\beta_0) = 0.068; L_{existing}(\beta_1) = 0.078, L_{proposed}(\beta_1) = 0.081)$ However, it wouldn't be generally true, simply because from one application it is not possible to determine whether this is a result of the particular nature of the example data.

Hence, different approaches to assessing the fit of a model and for comparing competing models are provided in the following. First, I use the penalized-likelihood information criteria, such as Akaike's Information Criterion (AIC) and the Bayesian Information Criterion (BIC) for model selection. AIC is an estimate of the relative distance between the unknown true likelihood function of the data and the fitted likelihood function of the model, so that a low AIC indicates that a model considered is closer to a truth model. BIC is an estimate of a function of the posterior probability of a model being a true model, under a certain Bayesian setup, so that a low BIC means that a model is considered to be more likely to be the true model. The AIC and BIC values are given in Table 6.2, from which we can learn that the proposed model is a better-fit to the dataset.

Method	AIC	BIC
Existing Method	81.283	85.995
Proposed Method	80.790	85.502

Table 6.2: Akaike information criterion (AIC) and Bayesian information criterion (BIC) for the existing method and the proposed method

Since a model selection should depend not only on the goodness-of-fit of a model, but also on the objective of the analysis, I perform an in-sample prediction to select a better model. The dataset is split into two groups. One group will be used to train the model, and the second group will be used to investigate the resulting model's error. In our example, we use 19 years of samples (1995~2013) to build the model and the remaining 1 year of samples (2014) to evaluate that model's error. Resulting in-sample prediction plots are shown in Figure 6.1.
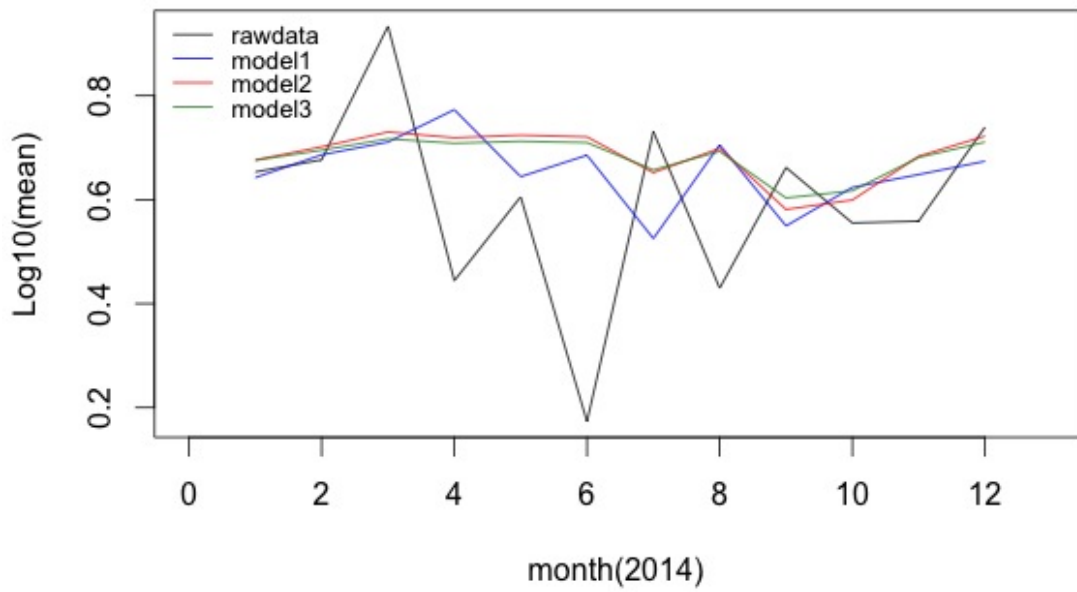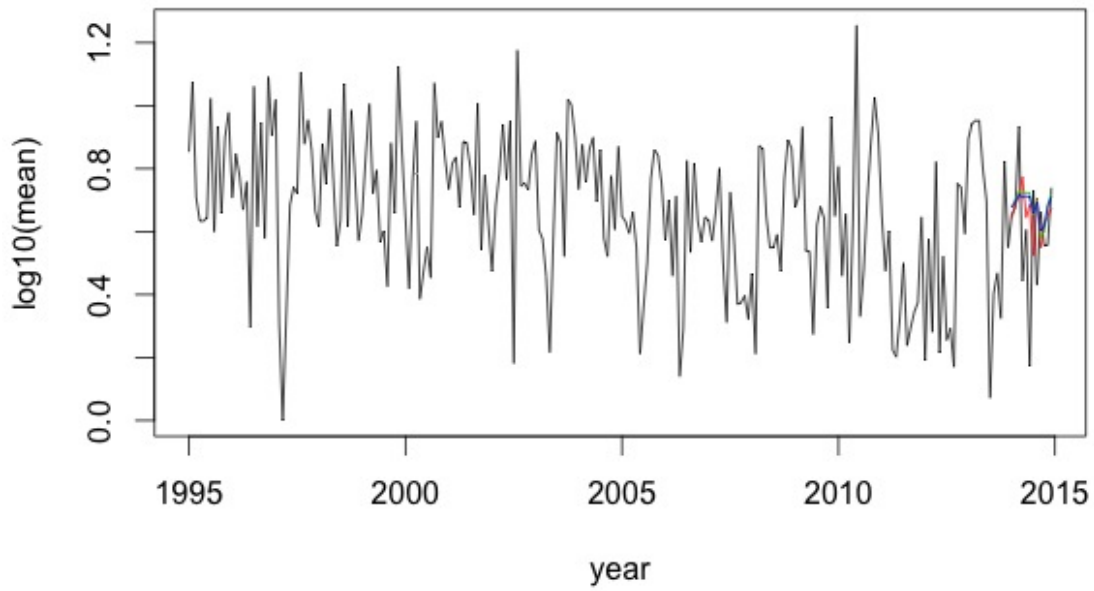
Figure 6.1: In-sample prediction plots of the monthly rainfall amount in 2014, where the top plot shows the entire time series, and the bottom plot shows only 2014 time series.

29

Table 6.3 shows the predicted values of the last year using the three models based on the 19 years samples. The in-sample mean squared prediction error of an estimator is calculated as:

$$MSPE = E\left[\left(Y_t - \widehat{Y_t}\right)^2\right]$$

where Y_t and $\widehat{Y_t}$ are actual and forecast values, respectively. In-sample mean square prediction error (MSPE) is not usually of direct interest since future values of the features are not likely to coincide with their training set values. But for comparison between models, in-sample error is convenient and often leads to effective model selection. Comparing the MSPE of the three models, Table 6.3 shows that the proposed GLS meta-analysis has the least MSPE, which indicates that the prediction bias is smallest for the proposed method.

2014	Original	Time Series	Existing GLS	Proposed GLS
Jan	0.6532	0.6424	0.6762	0.6758
Feb	0.6767	0.6863	0.7016	0.6951
Mar	0.9335	0.7108	0.7301	0.7169
Apr	0.4440	0.7730	0.7191	0.7084
May	0.6053	0.6441	0.7240	0.7122
Jun	0.1732	0.6855	0.7206	0.7096
Jul	0.7308	0.5252	0.6519	0.6572
Aug	0.4298	0.7054	0.6988	0.6930
Sep	0.6618	0.5494	0.5811	0.6032
Oct	0.5551	0.6235	0.5993	0.6171
Nov	0.5587	0.6483	0.6835	0.6813
Dec	0.7396	0.6738	0.7221	0.7107
MSPE		0.04749	0.04457	0.04289

Table 6.3: Predicted values and MSEPs from the time-series linear-circular regression, the existing GLS meta-analysis and the proposed GLS meta-analysis

Table 6.4 summarizes the strengths and weaknesses of all of the three methods. A time series linear-circular regression is relatively easy to fit but it requires stationarity of thine series. The raw data is often transformed to meet the stationary requirement. The existing GLS meta-analysis provides an alternative method to model a cylindrical time series data. But it ignores an apparent dependence structures among related studies. Yet, the proposed method utilizes some or all covariances to improve the precision of synthesized coefficients.

Method	Strength	Weakness
Existing GLS Method	Accounts for covariation. Only needs coefficients and Cov(b).	Ignore dependence structures among studies.
Proposed GLS Method	Accounts for covariation. Incorporate correlation structures among related studies.	Computationally involved.
Time-Series Regression	Relatively easy to fit	Requires stationary data. Require large number of parameters usually.

Table 6.4: Summary of properties of the three methods

Chapter 7

Conclusion

In this thesis, I discuss and compare the existing generalized least square (GLS) meta-analysis and the improved GLS meta-analysis. The difference between two methods is that the improved method utilizes a general form of covariance matrix which takes into account correlations between among studies. For the purpose of comparison, a time series linear-circular regression model was also fitted. All three methods have been applied to a cylindrical real dataset, to explore a functional relationship between wind direction (circular) and rainfall amount (linear). Several model diagnostic procedures have been performed. The primary objective of our improved method is to improve some limitation of the existing GLS method that assumes independence of those studies.

Firstly, I used the Mardia's rank correlation test to check if there exists a significant association between wind direction and rainfall amount in the dataset. Next, I fitted the existing and the proposed GLS meta-analyses without and with off-diagonal entries in the covariance matrix, respectively, along with the time series linear-circular regression. Then I provided the confidence intervals of coefficients, the Akaike's Information Criterion (AIC), Bayesian Information Criterion (BIC)

and an in-sample prediction plots to compare the goodness of the three models.

It was shown that the proposed model provides a better fit for the relationship between wind direction and rainfall amount. I claim that it is mainly due to taking a general covariance structure into consideration. However, obtaining the estimate of the full covariance matrix of $\mathrm{Cov}(b)$ is considered to be computationally involved. I provided the full calculation for off-diagonal entries in Chapter 4. In my example, I only included the covariances between adjacent months for simplicity after examine ACF and PACF plots.

The proposed model retains the advantages of usual meta-analysis that accounts for covariation, and also overcome the weakness of ignoring the dependence among coefficients. Future work will focus on how to apply the improved GLS meta-analysis on a multiple linear regression or on an occasion when samples do not all assume the same model (i.e., some models use fewer than the full set of p predictors).

Appendix A

R-codes

 # Mardia's rank correlation coefficient test[1] #

OrderScores <− **function**(lvar, cvar) {

  ranklvar <− **rank**(lvar, ties.method="random")

  n <− **length**(cvar) ; cvar2 <− 0

  for (j in 1:n) {cvar2[ranklvar[j]] <− cvar[j] }

  rankcvar <− **rank**(cvar2, ties.method="random")

  uscores <− rankcvar∗2∗pi/n ; **return**(uscores)}

Ustar <− **function**(uniscores) {

  n <− **length**(uniscores) ; Tc <− 0 ; Ts <− 0

  for (j in 1:n) {

  Tc <− Tc+j∗**cos**(uniscores[j]) ;

  Ts <− Ts+j∗**sin**(uniscores[j]) }

  UstarVal <− (Ts∗Ts)+(Tc∗Tc) ; **return**(UstarVal) }

MardiaRankIndTestRand <− **function**(uniscores, NR) {

  UstarObs <− Ustar(uniscores) ; nxtrm <− 1

```r
  for (r in 1:NR) {

  uniscoresRand <- sample(uniscores) ;

  UstarRand <- Ustar(uniscoresRand)

  if (UstarRand >= UstarObs) { nxtrm <- nxtrm + 1 } }

  pval <- nxtrm/(NR+1) ; return(c(UstarObs, pval))}


uniscores <- OrderScores(logmean, direction)

MardiaRankIndTestRand(uniscores, 9999)


#Proposed Method#

#off-diagonal elements#


covariance<-function(x1,x2,y1,y2)

  {v1<-var(x1);v2<-var(x2)

  cov1<-cov(x1,x2);cov2<-cov(y1,y2)

  m1<-mean(x1);m2<-mean(x2)

  d1<-(1/20+(1/20)*m1*m2*cov1/(v1*v2))*cov2;d2<-(1/20)*cov1*cov2/(v1*v2)

  return(matrix(c(d1,0,0,d2),nrow=2,ncol=2,byrow = TRUE))}

cov12<-covariance(cos1,cos2,log1,log2)

cov23<-covariance(cos2,cos3,log2,log3)

cov34<-covariance(cos3,cos4,log3,log4)

cov45<-covariance(cos4,cos5,log4,log5)

cov56<-covariance(cos5,cos6,log5,log6)

cov67<-covariance(cos6,cos7,log6,log7)

cov78<-covariance(cos7,cos8,log7,log8)

cov89<-covariance(cos8,cos9,log8,log9)

cov910<-covariance(cos9,cos10,log9,log10)
```

```
cov1011<-covariance(cos10,cos11,log10,log11)

cov1112<-covariance(cos11,cos12,log11,log12)

cov121<-covariance(cos12,cos1,log12,log1)


#Construct Sigma#

c1<-vcov(reg1)

s1<-cbind(c1,cov12,matrix(0,2,18),cov121)

c2<-vcov(reg2)

s2<-cbind(cov12,c2,cov23,matrix(0,2,18))

c3<-vcov(reg3)

s3<-cbind(matrix(0,2,2),cov23,c3,cov34,matrix(0,2,16))

c4<-vcov(reg4)

s4<-cbind(matrix(0,2,4),cov34,c4,cov45,matrix(0,2,14))

c5<-vcov(reg5)

s5<-cbind(matrix(0,2,6),cov45,c5,cov56,matrix(0,2,12))

c6<-vcov(reg6)

s6<-cbind(matrix(0,2,8),cov56,c6,cov67,matrix(0,2,10))

c7<-vcov(reg7)

s7<-cbind(matrix(0,2,10),cov67,c7,cov78,matrix(0,2,8))

c8<-vcov(reg8)

s8<-cbind(matrix(0,2,12),cov78,c8,cov89,matrix(0,2,6))

c9<-vcov(reg9)

s9<-cbind(matrix(0,2,14),cov89,c9,cov910,matrix(0,2,4))

c10<-vcov(reg10)

s10<-cbind(matrix(0,2,16),cov910,c10,cov1011,matrix(0,2,2))

c11<-vcov(reg11)

s11<-cbind(matrix(0,2,18),cov1011,c11,cov1112)
```

```
c12<−vcov(reg12)

s12<−cbind(cov121,matrix(0,2,18),cov1112,c12)

sigma<−as.matrix(rbind(s1,s2,s3,s4,s5,s6,s7,s8,s9,s10,s11,s12))

t(sigma)


#Design Matrix#

a<−diag(1, 2, 2)

w<−rbind(a,a,a,a,a,a,a,a,a,a,a,a)


#Parameter Estimates#

sigmainverse<−ginv(sigma)

wsigmaw<−t(w)%*%sigmainverse%*%w

covbeta<−ginv(wsigmaw)

beta<−covbeta%*%t(w)%*%sigmainverse%*%b
```

Bibliography

[1] B. J. Becker and M. Wu. The synthesis of regression slopes in meta-analysis. *Statistical Science*, 22(3):414–429, 2007.

[2] L. M. Bini, A. S. G. Coelho, and J.A. F. Diniz-Filho. Is the relationship between population density and body size consistent across independent studies? a meta-analytical approach. *Revista Brasileira de Biologia*, 61(1):1–6, Feb 2001.

[3] J. A. Carnicero, M. P. Wiper, and C. Ausín. Non-paramatric methods for circular-circular and circular-linear data based on bernstein copulas. *Statistics and Econometrics*, 04, Apr 2011.

[4] E. Garcia-Portugues, R. M. Crujeiras, and W. Gonzales-Manteiga. Exploring wind direction and SO_2 concentration by circular-linear density estimation. *Reports in Statistics anf Operations Research*, 2011.

[5] Stephen B. Jarrell and T. D. Stanley. A meta-analysis of the union-nonunion wage gap. *Industrial and Labor Relations Review*, 44(1):54–67, 1990.

[6] Richard A. Johnson and Thomas E. Wehrly. Some angular-linear distribution and related regression models. *Journal of the American Statistics Association*, 73(363):602–606, 1978.

[7] S. Kim and A. SenGupta. Inverse circular-linear/linear-circular regression. *Journal of Multivariate Analysis*, 2012.

[8] K. V. Mardia. Statistics of directional data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 37:349–393, 1975.

[9] K. V. Mardia. Linear-circular correlation coefficients and rhythmometry. *Biometrika*, 63(2):403–405, 1976.

[10] A. Pewsey, M. Neuhauser, and G. D. Ruxton. *Circular Statistics in R*. Oxford University Press, 2013.

[11] R. H. Shumway and D. S. Stoffer. *Time Series Analysis and Its Applications: With R Examples*. Springer, New York, 2011.