



A Bayesian Hierarchical Model for Daily Average Temperature

Snæbjörn Helgi Emilsson



Faculty of Industrial Engineering, Mechanical Engineering
and Computer Science
University of Iceland
2012

A BAYESIAN HIERARCHICAL MODEL FOR DAILY AVERAGE TEMPERATURE

Snæbjörn Helgi Emilsson

30 ECTS thesis submitted in partial fulfillment of a
Magister Scientiarum degree in Industrial Engineering

Advisor

Birgir Hrafnkelsson

Ólafur Pétur Pálsson

Faculty Representative

Kristján Jónasson

Faculty of Industrial Engineering, Mechanical Engineering
and Computer Science

School of Engineering and Natural Sciences

University of Iceland

Reykjavik, May 2012

A Bayesian Hierarchical Model for Daily Average Temperature

A Bayesian Model for Temperature

30 ECTS thesis submitted in partial fulfillment of a M.Sc. degree in Industrial Engineering

Copyright © 2012 Snæbjörn Helgi Emilsson

All rights reserved

Faculty of Industrial Engineering, Mechanical Engineering
and Computer Science

School of Engineering and Natural Sciences

University of Iceland

VRII, Hjarðarhagi 2-6

107, Reykjavík

Iceland

Telephone: 525 4700

Bibliographic information:

Snæbjörn Helgi Emilsson, 2012, A Bayesian Hierarchical Model for Daily Average Temperature, M.Sc. thesis, Faculty of Industrial Engineering, Mechanical Engineering and Computer Science, University of Iceland.

Printing: Háskólaupprent, Fálkagata 2, 107 Reykjavík
Reykjavík, Iceland, May 2012

Abstract

In this thesis, a Bayesian hierarchical model for daily average temperature is presented. A multivariate normal distribution is selected as the data distribution due to its flexibility and theoretical basis. The linear fit is assumed to be governed by a seasonal effect parameter vector, a linear trend parameter, a long term fluctuation parameter vector and a model constant. The seasonal effect and fluctuations are modeled as independent Gaussian processes which are governed by Gaussian Markov random fields. The covariance matrix of the multivariate normal distribution describes temporal correlation and a seasonally changing variance of the data. A periodic autoregressive (PAR) process is used to model the temporal correlation and regression is used to estimate the parameters. An iterative process is used to update the regression parameters and the Bayesian parameters, since they are dependent on each other. This model allows for future predictions, but is limited to predicting one year ahead.

A program based on the model was developed in the R programming language. The program uses the Gibbs sampler, a Markov chain Monte Carlo algorithm, to estimate the parameters of the model by sampling from their conditional distributions. Using the R program the model is applied to observed data from four locations in Iceland over the years 1949 to 2010. These locations are Reykjavík, Akureyri, Dalatangi and Stórhöfði. Based on the model the estimated increase in average temperature over the period is from 0.05 to 0.46°C, depending on location. A prediction was made for the year 2011, which was not a part of the training set. Of the actual temperature values of 2011, only 2.5 to 4.7% of the observations were outside the 95% posterior prediction interval.

Útdráttur

Í þessari ritgerð er Bayesískt stigskipt líkan fyrir daglegan meðalhita sett fram. Margvíða normaldreifingin er valin sem gagnadreifing vegna sveigjanleika hennar. Væntigildi gagnadreifingarinnar er stjórnað af stikavigri fyrir árstíðasveiflur, stika fyrir línulega leitni, stikavigri fyrir langtíma sveiflur í meðalhita og fasta líkansins. Árstíðabundna sveiflan og meðaltalssveiflan eru metin sem óháð Gaussian ferli sem er stjórnað af Gaussian Markov slemmbiferli. Samfylgnifylkið í margvíðu normaldreifingunni inniheldur tímaháða fylgni og árstíðabundnar sveiflur í dreifni. Lotubundið eiginadhvarfs líkan (PAR) er notað til að lýsa tímaháðu fylgninni og aðhvarfsgreiningu er beitt til að meta stika líkansins. Stikanir í Bayesíska líkaninu og í PAR líkaninu eru háðir hvor öðrum svo nauðsynlegt er að nota ítrunarferli til að uppfæra stikana. Hægt er að nota líkanið til að spá eitt ár fram í tímann.

Forrit byggt á líkaninu var þróað í forritunarmálinu R. Forritið notar Gibbs hermun, Markov keðju Monte Carlo reiknirit til að meta stikana í líkaninu með því að draga úr skilyrtu líkindadreifingunum þeirra. Í gegnum R forritið er líkaninu beitt á gögn frá tímabilinu 1949 til 2010, frá fjórum stöðum á Íslandi. Þessir staðir eru, Reykjavík, Akureyri, Dalatangi og Stórhöfði. Samkvæmt líkaninu var hækkun meðalhitastigs frá 0.05 til 0.46°C á áratug yfir tímabilið, lægst á Akureyri og hæst í Reykjavík. Spá var gerð fyrir árið 2011 og hún borin saman við raunverulegan hita frá sama ári. Það kom í ljós að ekki nema 2.5 til 4.7% af gildunum voru fyrir utan 95% spábil líkansins.

Preface

This M.Sc. project was carried out at the Faculty of Industrial Engineering, Mechanical Engineering and Computer Science at the University of Iceland.

Special thanks go to my supervisor, Assistant Professor of Statistics Birgir Hrafnkels-son at the Department of Mathematics of University of Iceland, for his opinions and contributions throughout the process of making the model presented in this thesis, for the patient guidance and for all the feedback. I would also like to thank my co-advisor Professor Ólafur Pétur Pálsson at the Faculty of Industrial Engineering, Mechanical Engineering and Computer Science for his help with the thesis.

Furthermore I would like to thank my fellow students Helgi Sigurðarson and Óli Páll Geirsson for their help and support. I also thank my girlfriend Eyrún for her moral support and encouragement and my lovely daughter Emilía for her inspiration.

Reykjavik, May 2012.

Snæbjörn Helgi Emilsson

Contents

List of Figures	xi
List of Tables	xv
1. Introduction	1
1.1. Goals of the project	1
1.2. Possible applications	1
1.3. The data	2
1.4. Literature review	2
1.5. Structure of the Thesis	3
2. Theory	5
2.1. Bayesian Inference	5
2.1.1. Hierarchical Models	6
2.1.2. The Multivariate Normal distribution	6
2.1.3. Gaussian Markov Random Fields	7
2.2. The Gibbs sampler and MCMC	8
2.3. Model information criterion	9
2.3.1. AIC	9
2.3.2. DIC	10
2.4. B-Splines	11
2.5. Periodic autoregression	12
3. Data	13
3.1. Seasonal behavior	16
3.2. Long term fluctuations	18
3.3. Correlation	20
4. The Model	21
4.1. A linear model	21
4.1.1. Seasonal behavior	22
4.1.2. Long term behavior	22
4.1.3. Autocorrelation	23
4.1.4. Heteroscedasticity (non-constant variance)	24
4.2. Bayesian Inference	24
4.2.1. Prior distributions	25

4.2.2. Posterior distribution	26
4.2.3. MCMC	27
4.3. Computer implementation	27
5. Results	29
5.1. Main results	29
5.1.1. Convergence	34
5.2. Comparison to the data	35
5.3. Prediction	37
5.4. The residuals	39
6. Discussion and conclusions	45
7. Future studies	47
Bibliography	49
A. Formulations	51
A.1. Posterior Formulation	51
A.2. DIC formulation	56
A.3. The error term ϵ	57
B. Figures	59

List of Figures

2.1. Cubic B-spline basis using seven equally spaced knots.	11
3.1. The daily temperature averages in Reykjavik the first three years, 1949-1951.	13
3.2. Frequency distribution of daily mean temperature for the whole period 1949-2010.	15
3.3. The densities of the daily mean temperature for the whole period 1949-2010, estimated by kernel smoothing.	16
3.4. The mean annual temperature cycle based on daily temperature averages (solid line), together with 2.5% and 97.5% percentiles (dashed lines).	17
3.5. The sample standard deviation of the daily average temperature, as a function of Julian day.	18
3.6. An estimation of the long term fluctuation in the data, calculated by a kernel smoothing, with bandwidth 1500, of \vec{e}_1 in Equation (3.3). . .	19
3.7. ACF and PACF for residuals \vec{e}_2 from Equation (3.4).	20
4.1. Deviance information criterion for different numbers of parameters in parameter vector $\vec{\alpha}$	28
5.1. Posterior means for the four locations of $\vec{\alpha}$ (solid line) and 95% credible intervals (dashed lines).	31
5.2. Posterior means of $(\vec{\beta} + \beta_0)$ (solid line) and 95% credible intervals (dashed lines), for the four locations.	32

LIST OF FIGURES

5.3.	Posterior means of the long term temperature fluctuation as a function of time, estimated by the model, i.e. $(\beta_0\vec{1}+\gamma_0\vec{T}+\mathbf{Z}\vec{\alpha})$ (solid line), along with 95% credible interval (dashed lines), for the four locations.	33
5.4.	Histograms and MCMC trace plots after burn-in for the parameters β_0 and γ_0 at location Reykjavik.	34
5.5.	Posterior means of $(\vec{\beta}+\beta_0)$ (black line) compared to the sample means of Julian calendar day temperature from Figure 3.4 in Chapter 3 (blue line).	35
5.6.	Posterior means for the fluctuation $(\gamma_0\vec{T}+\mathbf{Z}\vec{\alpha})$ (solid line) compared to the fluctuation estimate based on smoothing from Figure 3.6 (dashed line).	36
5.7.	Prediction for the year 2011 (solid line) along with 2.5% and 97.5% percentiles (dashed line) and the actual temperature (red line), for the four locations.	38
5.8.	The standard deviation of the v_t residuals from Equation (4.8) and a trigonometric fit from Equation (4.9), for the four locations.	40
5.9.	Three different residuals for Reykjavik. Top panel, the residuals ϵ_t from the Bayesian model, middle panel, the residuals v_t from the PAR model, and at the bottom panel, the standardized residuals $\frac{v_t}{\sigma_v}$	41
5.10.	The estimated densities of the standardized residuals $\frac{v_t}{\sigma_v}$, for the four locations.	42
5.11.	The normal Q-Q plot of the standardized residuals $\frac{v_t}{\sigma_v}$, for the four locations.	43
5.12.	The autocorrelation function of the standardized residuals $\frac{v_t}{\sigma_v}$, for the four locations (the value of ACF at lag 0 is 1).	44
B.1.	Histograms and MCMC trace plots after burn-in for the parameter β_0 .	59
B.2.	Histograms and MCMC trace plots after burn-in for the parameter γ_0 .	60
B.3.	Histograms and MCMC trace plots after burn-in for the parameter σ^2 .	61
B.4.	Histograms and MCMC trace plots after burn-in for the parameter σ_β^2 .	62

B.5. Histograms and MCMC trace plots after burn-in for the parameter σ_{α}^2 .	63
B.6. Histograms and MCMC trace plots after burn-in for the parameter β_{10} .	64
B.7. Histograms and MCMC trace plots after burn-in for the parameter β_{150} .	65
B.8. Histograms and MCMC trace plots after burn-in for the parameter α_1 .	66
B.9. Histograms and MCMC trace plots after burn-in for the parameter α_{15} .	67

List of Tables

- 2.1. Hierarchical model structure. 6

- 3.1. Summary statistics of the daily average temperature series. 14

- 5.1. Posterior means of β_0 , γ_0 and σ^2 along with 95% credible intervals in brackets. 30
- 5.2. Posterior means of σ_β^2 and σ_α^2 along with 95% credible intervals in brackets. 30
- 5.3. Autoregressive parameter estimates. 30

1. Introduction

1.1. Goals of the project

The main goal of this thesis is to estimate and model the seasonal effect and long term fluctuations in average daily temperature and give reliable uncertainty estimates for these quantities. This is done by setting up a Bayesian hierarchical model which adequately describes univariate series of daily average temperature. Along with the seasonal effect and the long term fluctuation, this model takes into account temporal correlation and heteroscedasticity in the data. The programming language R is used to apply the model to the given data and to predict the weather one year ahead. These predictions are validated with data from that same year.

1.2. Possible applications

Air temperature has an impact in various subjects. Accurate and reliable information about future temperature predictions as well as information about the past are important to correctly estimate temperature effect in different subjects. Subjects that are affected by temperature include, but are not limited to, ecology, where temperature affects a wide variety of biological processes and the development of ecosystems (e.g. the growth of plants, the population of animals), district heating, where the consumption of hot water depends on the outside air temperature, tourism and in climate research. In the model presented in this thesis the uncertainty of the predictions is reduced by including a temporal correlation and seasonal variance in the model, resulting in a reliable model that can be used in any of the fields mentioned above.

1.3. The data

The weather data used in the development of the model presented here are time series of average daily air temperature at four locations in Iceland spanning over 62 years, from the beginning of 1949 to the end of 2010. The locations are Reykjavik, Akureyri, Dalatangi and Storhofdi.

The raw data from which the daily average is calculated, are from the Icelandic Meteorological Office Database. They consist of standard daily readings of air temperature taken by meteorological observers every three hours, beginning at 3 am and ending at midnight the same day.

All years are assumed to contain 365 days with February 29th being removed from all datasets. Missing values in the data are replaced using linear interpolation. Each site is modeled separately but the general form of the model is assumed to be the same for each site.

1.4. Literature review

Most meteorological time series data (e.g. temperature) contain a continuous slowly varying change over long scales (long term fluctuations). In a recent paper, Craigmile et al. (2004) propose using discrete wavelet transform (DWT) to extract polynomial fluctuations from long memory processes. This method is demonstrated to work well on annual temperature data. Király and János (2002) claim that detrended fluctuation analysis (DFA) can effectively filter out slow fluctuations in daily average temperature data. In DFA the time series analyzed is divided into non-overlapping segments of equal length. The local linear trend is then fitted in each segment by a polynomial and the data detrended by subtracting this local fit.

Das (2009) analyzes average daily temperature values for a 12 year period at some cities around the world, including Reykjavik. He attempts to remove the seasonality from the data by fitting adjusted sine functions, with time period 365 days, to the data. He then subtracts the fitted sine functions from the data, getting residuals that should be clear of all seasonality. He concludes however, that the sine functions are not flexible enough to remove the seasonal behavior completely, with the residuals containing traces of seasonality in them.

In his research, Tol (1996) presents a complete model for average daily temperature, based on 30 years of observations from De Bilt in the Netherlands. He proposes using a generalized autoregressive conditional heteroscedastic (GARCH) model to capture

regular but heteroscedastic changes between summer and winter. This model gives questionable results.

Short term correlations in meteorological records are usually explained by low order autoregressive processes (Storch and Zwiers, 2002). In most of the literature cited above first or second order autoregressive processes are used to describe the dynamics of the data.

Some notable studies have been done on the statistical analysis of temperature in Iceland. Crochet and Jóhannesson (2010) created a spatial data set of simulated gridded daily air temperature in Iceland for the period 1949 to 2010. Jónasson (2005) made predictions for annual temperature in Reykjavik using regression models connected to temperature studies. According to the model the average temperature will be 5.4°C for the period 2005 to 2015 and then increase significantly over the next century and will be 7.7°C in 2100. Hrafnkelsson et al. (2012) used Bayesian hierarchical modeling to analyze annual minimum and maximum temperatures over Iceland. One of their conclusion is that the average annual temperature increased 0.24°C per decade over the period 1961 to 2009. They also conclude that the minimum and maximum temperatures over the same period increased 0.71 and 0.47°C per decade, respectively.

In climatology, Bayesian hierarchical models are increasingly used to describe data that is highly multivariate, with many response variables or autocorrelated (Banerjee et al., 2004; Wikle and Berliner, 1998). In this thesis a flexible, three-level Bayesian hierarchical model for daily average temperature is presented. The first level, which is the data level, contains an autoregressive process with seasonal variance. Parameters of this autoregressive process are estimated within the Bayesian approach using classical inference. At level two the parameters for a seasonal effect, and parameters that describe the long term fluctuation are presented through Gaussian prior distributions based on Markov random fields. The third level contains prior distributions for the parameters at the second level, so-called hyperparameters.

1.5. Structure of the Thesis

The structure of the thesis is as follows. In Chapter 2 an overview of some of the theory used in the thesis is given. In Chapter 3 the data which are analysed here are presented. In Chapter 4 all aspects of the model are given along with the Bayesian setup, prior distributions and computational methods. The main results are presented in Chapter 5 along with a forecast for the year 2011. Chapter 6 contains some discussion and conclusions and in Chapter 7 some ideas on future projects are given.

2. Theory

In this chapter the theory used in this thesis is described. First a short introduction to Bayesian inference and hierarchical models is given, followed by an introduction to the multivariate normal distribution and Gaussian Markov random fields. Then an introduction to the posterior simulation tools used, the Gibbs sampler and to Markov chain Monte Carlo (MCMC). Further, the model criterias AIC and DIC, which are used for the model comparison, are described. Finally a brief introduction is given to B-splines, which are used to model the long term fluctuations, and periodic auto regression, which is used to model temporal correlation in the data.

2.1. Bayesian Inference

Bayesian inference is a method for statistical inference in which Bayes' theorem is used to make a statistical conclusion about a set of model parameters θ in terms of probability statements that are conditional on the given data y . Parameters θ are treated as random variables and the data are used to update prior knowledge about the parameters. The essential ingredient in Bayesian statistics is that the information about θ is summarized in the likelihood function. Based on Bayes' rule,

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} \quad (2.1)$$

where $p(\theta)$ is the prior density of θ (that expresses the knowledge about θ before observing the data), $p(\theta|y)$ is the posterior density of θ and $p(y|\theta)$ is the data distribution (the likelihood). If θ is continuous then the probability of y is

$$p(y) = \int_{\theta} p(\theta) p(y|\theta) d\theta \quad (2.2)$$

which is sometimes referred to as the normalizing constant and does not depend on θ . If θ is discrete then the probability of y is

$$p(y) = \sum_{\theta} p(\theta) p(y|\theta). \quad (2.3)$$

2. Theory

Bayes' rule can therefore be rewritten as the posterior density proportional to the product of the prior distribution and the likelihood

$$p(\theta|y) \propto p(\theta)p(y|\theta). \quad (2.4)$$

Bayesian inference about θ follows from inspection of the posterior density (where as a frequentist would look at the likelihood).

2.1.1. Hierarchical Models

For many of the complex data sets which Bayesian statistics can be applied to, the model parameters can be regarded as related or connected to each other in some way according to the structure of the problem. In that case a Bayesian hierarchical model structure can be used to ensure that the joint probability model for these parameters reflects the dependencies among them (Gelman (2004)). A general hierarchical model with three levels has the form shown in Table 2.1.

Table 2.1: Hierarchical model structure.

Level	Variables	Density
1 (Data level)	Observations Processes, Parameters	$p(y \theta, \varphi)$
2 (Latent level)	Processes Parameters	$p(\theta \varphi)$
3 (Hyperparameters)	Parameters	$p(\varphi)$

Let y denote the data as before, θ a set of parameters in the data distribution, referred to as latent parameters and φ are the parameters in a distribution describing the latent parameters, referred to as hyperparameters. In most real-world problems, there is some knowledge about how to constrain the hyperparameters φ , although the exact values of them are not known. The hyperparameters have the prior distribution $p(\varphi)$. The joint distribution of the latent- and hyperparameters is

$$p(\theta, \varphi) = p(\theta|\varphi)p(\varphi) \quad (2.5)$$

and the posterior density for the hierarchical model is

$$p(\theta, \varphi|y) \propto p(y|\theta, \varphi)p(\theta|\varphi)p(\varphi) \quad (2.6)$$

2.1.2. The Multivariate Normal distribution

In probability theory, the univariate normal (or Gaussian) distribution is a commonly known continuous probability distribution that has a probability density

function

$$f(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R} \quad (2.7)$$

where parameter μ is the mean or expected value and σ^2 is the variance. The general notation is $x \sim \mathcal{N}(\mu, \sigma^2)$. The multivariate normal distribution is a generalization of the univariate normal distribution to higher dimensions and has a probability density function

$$f_X(x_1, \dots, x_k) = \frac{1}{(2\pi)^{k/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)^T \Sigma^{-1}(\mathbf{x} - \mu)\right) \quad (2.8)$$

where $\mu \in \mathbb{R}^k$ is the mean vector, $\Sigma \in \mathbb{R}^{k \times k}$ the covariance matrix and $|\Sigma|$ the determinant of the covariance matrix. The covariance matrix Σ is a matrix which contains the covariance between the i -th and the j -th element of a vector in its (i, j) position. It is symmetric and positive definite with $\text{Cov}(x) = E((x - \mu)(x - \mu)^T)$ and $\text{Cov}(x)_{i,j} = \text{Cov}(x_i, x_j)$.

2.1.3. Gaussian Markov Random Fields

Gaussian Markov random fields (GMRF), also known as conditionally specified Gaussian fields, are useful to describe the temporal (and spatial) dependencies between neighbors. GMRFs are usually specified by their precision matrix \mathbf{P} meaning that the covariance matrix, $\Sigma = \mathbf{P}^{-1}$ is only known through its inverse.

In Bayesian hierarchical models the precision matrix can be used on the second level in the hierarchical model to generate dependence structure between the latent parameters and therefore the dependence between the observed data (Rue and Martino, 2007). In this thesis a result in Lindgren et al. (2011), referred to as *main result 1*, is used to describe the latent relationship and create the precision matrix.

Since our data is one dimensional with regular spacing the form of the matrix is relatively simple. It is done by examining the solution (on \mathbb{R}) to the linear fractional stochastic partial differential equation

$$(\kappa^2 - \nabla \cdot H \nabla)^{\alpha/2} x(\mathbf{u}) = W(\mathbf{u}), \quad \mathbf{u} \in \mathbb{R} \quad (2.9)$$

where H is a positive element. For any $\mathbf{u} = u_1, \dots, u_n \in \mathbb{R}$ let $\gamma_i = u_i - u_{i-1}$, $\delta_i = u_{i+1} - u_i$ and $s_i = (\gamma_i + \delta_i)/2$. Then the elements of row i around the diagonal of the precision matrix are given by,

$$\mathbf{P}_1 : s_i \cdot [-a_i \quad c_i \quad -b_i], \quad \text{if } \alpha = 1$$

$$\mathbf{P}_2 : s_i \cdot [a_i a_{i-1} \quad -a_i(c_{i-1} + c_i) \quad a_i b_{i-1} + c_i^2 + b_i a_{i+1} \quad -b_i(c_i + c_{i+1}) \quad b_i b_{i+1}], \quad \text{if } \alpha = 2$$

2. Theory

where $a_i = H/\gamma_i s_i$, $b_i = H/\delta_i s_i$ and $c_i = \kappa^2 + a_i + b_i$. Since the data used in this thesis has regular even spacing between all observations we can write, $s = \delta = \gamma = 1$ and $a = a_i = b_i = H/\delta = 1$ and the lines in the precision matrix as,

$$\mathbf{P}_1 : \begin{bmatrix} -1 & (\kappa^2 + 2) & -1 \end{bmatrix} \quad (2.10)$$

$$\mathbf{P}_2 : \begin{bmatrix} 1 & -(2\kappa^2 + 4) & (\kappa^4 + 4\kappa^2 + 6) & -(2\kappa^2 + 4) & 1 \end{bmatrix} \quad (2.11)$$

In some cases when sampling from a GMRF under an additional constraint $\mathbf{A}\mathbf{x} = \mathbf{e}$, using the precision matrix \mathbf{P} (i.e. $p(\mathbf{x}|\mathbf{A}\mathbf{x}) = \mathbf{e}$ where $x \sim \mathcal{N}(\mu, \mathbf{P}^{-1})$) it is necessary to use a correction algorithm. The following correction algorithm is from Rue and Held (2005).

Algorithm 1 Sampling $x|Ax = e$ where $x \sim N(\mu, Q^{-1})$.

- 1: Compute the Cholesky factorization, $Q = LL^T$
 - 2: Sample $z \sim N(0, I)$
 - 3: Solve $L^T v = z$
 - 4: Compute $x = \mu + v$
 - 5: Compute $V_{n \times k} = Q^{-1} A^T$
 - 6: Compute $W_{k \times k} = AV$
 - 7: Compute $U_{k \times n} = W^{-1} V^T$
 - 8: Compute $c = Ax - e$
 - 9: Compute $x^* = x - U^T c$
 - 10: Return x^*
-

2.2. The Gibbs sampler and MCMC

Markov chain Monte Carlo (MCMC) methods are a class of algorithms used to sample from probability distributions based on constructing a Markov chain that has the desired distribution as its equilibrium distribution. The state of the chain after a large number of iterations is used as a sample of the desired distribution and the quality of the sample improves as a function of the number of iterations. The Gibbs sampler is one of the most commonly used MCMC methods in Bayesian statistics since it is useful for multidimensional posterior densities (Gelman, 2004).

The Gibbs sampler draws samples from a set of conditional distributions which are based on the prior distributions and the posterior distribution of the parameters.

Suppose that the parameter vector of interest is $\theta = (\theta_1, \dots, \theta_K)$. The parameters are given initial values and then the Gibbs sampler is used to iterate the

parameters by drawing samples from their conditional distributions. The conditional distribution of parameter θ_j at iteration t is $p(\theta_j^{(t+1)}|\theta_{-j}^{(t)}, \text{data})$, where $\theta_{-j} = (\theta_1, \dots, \theta_{j-1}, \theta_{j+1}, \dots, \theta_K)$ is the set of all parameter excluding θ_j . For a joint distribution with full conditionals $p(\theta_1^{(t+1)}|\theta_{-1}^{(t)}, \text{data}), \dots, p(\theta_K^{(t+1)}|\theta_{-K}^{(t)}, \text{data})$, the Gibbs sampler simulates successively from all conditionals, modifying one component of θ at a time. The corresponding algorithmic representation is given in Algorithm 2.

Algorithm 2 The Gibbs sampler.

```

Specify an initial value  $\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_K^{(0)})$ 
for  $t = 1, 2, \dots, M$  do
    Draw  $\theta_1^{(t+1)}$  from  $p(\theta_1^{(t+1)}|\theta_{-1}^{(t)}, \text{data})$ 
    Draw  $\theta_2^{(t+1)}$  from  $p(\theta_2^{(t+1)}|\theta_{-2}^{(t)}, \text{data})$ 
     $\vdots$ 
    Draw  $\theta_K^{(t+1)}$  from  $p(\theta_K^{(t+1)}|\theta_{-K}^{(t)}, \text{data})$ 
end for
return the values  $(\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(M)})$ 

```

Under general conditions, draws from this simulation algorithm will converge to the target distribution of interest (the joint posterior of θ).

2.3. Model information criterion

An information criterion is a measure of the relative goodness of fit of a statistical model. The information criteria used in this thesis are Akaike's information criterion (AIC) and Deviance information criterion (DIC).

2.3.1. AIC

Akaike's information criterion is probably the most commonly used procedure for the model selection in classical statistics. It is defined as

$$\text{AIC} = -2\log(L) + 2n_i \quad (2.12)$$

where n_i is the number of parameters in the statistical model, and L is the maximized value of the likelihood function for the estimated model, see Akaike (1974) for further information.

2. Theory

2.3.2. DIC

The Deviance information criterion is a generalization of the AIC that was developed by Spiegelhalter et al. (2002). It is the commonly used criterion for model comparison within a Bayesian framework. The advantage of DIC over other criteria in Bayesian model selection is that the DIC can be fairly easily calculated from the samples generated by a Markov chain Monte Carlo simulation. Define the deviance as

$$D(y, \theta) = -2 \log \{p(y|\theta)\} \quad (2.13)$$

where y is the data, θ the unknown parameters of the model and $p(y|\theta)$ the likelihood function. A summary of $D(y, \theta)$ based on $\hat{\theta}$ is given by

$$D_{\hat{\theta}}(y) = D \{y, \hat{\theta}(y)\} \quad (2.14)$$

In MCMC simulations $\hat{\theta}$ is usually the mean of the posterior simulations. The average of $D(y, \theta)$ is given by

$$D_{avg}(y) = E \{D(y, \theta)|y\}$$

which can be estimated with

$$\hat{D}_{avg}(y) = \frac{1}{L} \sum_{l=1}^L D(y, \theta^l) \quad (2.15)$$

where θ^l is the l -th draw from $p(\theta|y)$. The summary $\hat{D}_{avg}(y)$ is a measure of how well the model fits the data. The smaller this summary is, the better the fit of the model. The effective number of parameters of the model, are commonly estimated by

$$p_D = \hat{D}_{avg}(y) - D_{\hat{\theta}}(y) \quad (2.16)$$

The higher this value is, the easier it is for the model to fit the data. The deviance information criterion is defined as

$$\text{DIC} = 2\hat{D}_{avg}(y) - D_{\hat{\theta}}(y) = \hat{D}_{avg}(y) + p_D \quad (2.17)$$

To calculate DIC, compute $D(y, \theta^l)$ for every iteration l , compute $\hat{D}_{avg}(y)$ as the average of $D(y, \theta)$ over the samples of θ and $D_{\hat{\theta}}(y)$. Then the DIC follows directly from these approximations. The DIC is used in Section 4.2.3 to determine how many parameters should be used to model the long term fluctuation in the data.

2.4. B-Splines

A B-spline basis is a set of natural splines that is particularly well suited for computation (Wasserman, 2006). An M -order B-spline is a piecewise $M - 1$ degree polynomial with $M - 2$ continuous derivatives. In this thesis only cubic B-splines are used (i.e. $M = 4$). The general definition of B-splines is as follows.

Let $x_1 < x_2 < \dots < x_k$ be a set of knot points on the interval $(0, 1)$, and $x_0 = 0$ and $x_{k+1} = 1$. Lets define new knots t_i such that $t_1 \leq t_2 \leq \dots \leq t_M \leq x_0$, $t_{j+M} = x_j$ for $j = 1, \dots, k$ and $x_{k+1} \leq t_{k+M+1} \leq \dots \leq t_{k+2M}$. Usually the extra points are chosen to be $t_1 = \dots = t_M = x_0$ and $x_{k+1} = t_{k+M+1} = \dots = t_{k+2M}$. The basis functions are then defined recursively as

$$b_{j,0}(t) := \begin{cases} 1 & \text{if } t_j \leq t < t_{j+1} \\ 0 & \text{otherwise} \end{cases}, \quad j = 1, \dots, k+2M-1 \quad (2.18)$$

$$b_{j,n}(t) := \frac{t - t_j}{t_{j+n} - t_j} b_{j,n-1}(t) + \frac{t_{j+n+1} - t}{t_{j+n+1} - t_{j+1}} b_{j+1,n-1}(t), \quad j = 1, \dots, k+2M-m \quad (2.19)$$

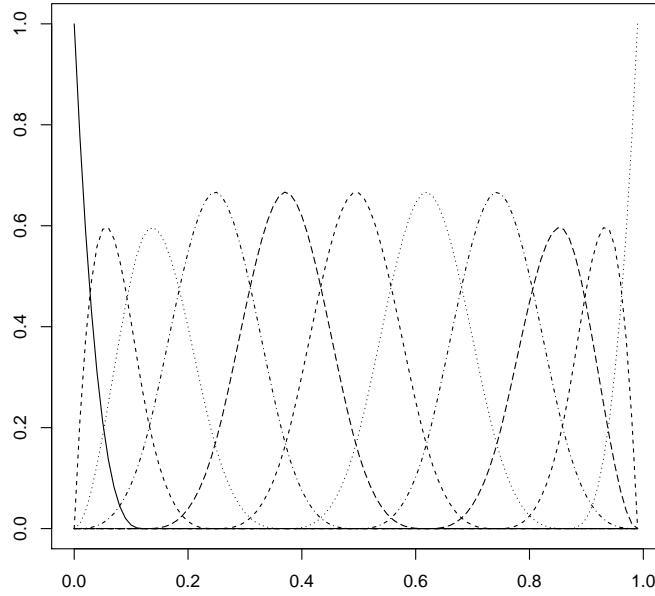


Figure 2.1: Cubic B-spline basis using seven equally spaced knots.

The B-splines will be used in Chapter 4 to estimate the long term fluctuation in the data.

2.5. Periodic autoregression

Periodic autoregressive models (PAR) are time series models which allow the value and number of AR parameters to vary with the seasons, and allows a non-constant standard deviation.

A classical AR(p) process is given by

$$Y_t = \sum_{i=1}^p \phi_i Y_{t-i} + \varepsilon_t \quad (2.20)$$

where ε_t represents an independent and identically distributed with zero mean ($E(\varepsilon_t) = 0$) and constant variance ($\text{var}(\varepsilon_t) = \sigma^2 > 0$), i.e. white noise. A PAR(p(t)) process that allows the parameters and the order to be a function of t is given by

$$Y_t = \sum_{i=1}^{p(t)} \phi_i(t) Y_{t-i} + \varepsilon_t \quad (2.21)$$

where ε_t is again independent with zero mean ($E(\varepsilon_t) = 0$), but now the variance is $\text{var}(\varepsilon_t) = \sigma^2(t) > 0$.

The periodic autoregressive process will be used in Section 4.1.3 to describe the temporal correlation in the residuals of the model. For additional information on PAR see Jones and Brelsford (1967); Troutman (1979).

3. Data

The data set analyzed in this thesis consists of daily average air temperatures at four locations in Iceland, which are Reykjavik, Akureyri, Dalatangi and Storhofdi, and range over 62 years from the beginning of 1949 to the end of 2010, all in all 22630 values. The daily average was calculated from air temperature measurements observed on three hour intervals by the Icelandic Meteorological Office, i.e. eight times each day. The average was calculated from 3am to midnight for each day. In the Storhofdi series linear interpolation was used to replace two years of missing observations at 3am, but apart from these missing values at Storhofdi the data sets seem to be without errors and with no apparent outliers or defects. In Figure 3.1 the first three years of the Reykjavik series are plotted, showing the general structure of all the series.

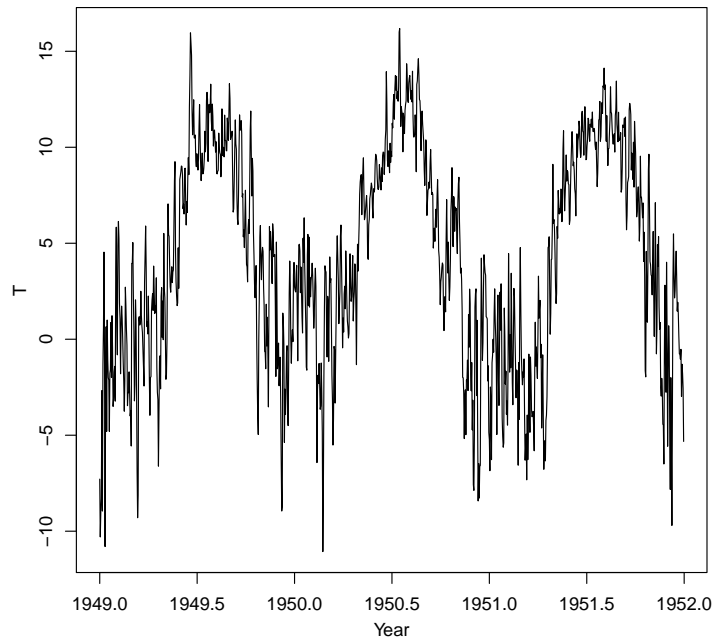


Figure 3.1: The daily temperature averages in Reykjavik the first three years, 1949-1951.

3. Data

The daily average data sets will either be referred to as y_t , where $t = 1, \dots, m \cdot n$ is an index, or by $y_{i,k}$, where $i = 1, \dots, m$, $m = 62$ stands for a year and $k = 1, \dots, n$, $n = 365$ is a Julian calendar day. A statistical summary of the data is given in Table 3.1. From the summary it is apparent that Akureyri has both highest maximum value and the lowest minimum. It also has the lowest mean temperature and the highest standard deviation, while Storhofdi has the highest mean temperature and the lowest standard deviation.

Histograms of the daily average temperature are plotted in Figure 3.2 for each location. These histograms show a negative skewness with a tail towards low temperatures. The histograms for Reykjavik and Akureyri seem to have a relatively similar shape with flat peak while Dalatangi and Storhofdi have narrower peak. This resemblance between the series is clearer in Figure 3.3 which shows the densities of the daily average temperature, estimated by a kernel smoothing function.

Table 3.1: Summary statistics of the daily average temperature series.

	Reykjavik	Akureyri	Dalatangi	Storhofdi
Min.	-15.11	-20.15	-15.91	-14.96
1st Qu.	1.05	-0.60	1.09	2.44
Median	5.08	3.94	4.23	5.61
Mean	4.68	3.63	3.84	5.12
3rd Qu.	9.03	8.45	7.10	8.34
Max.	20.08	20.90	18.91	17.41
σ	5.12	6.20	4.24	4.10
Latitude	64°07.648'	65°41.135'	65°16.090'	63°23.985'
Longitude	21°54.166'	18°06.014'	13°34.556'	20°17.299'
Altitude	52.0 m.y.s.	23.0 m.y.s.	9.0 m.y.s.	118.0 m.y.s.

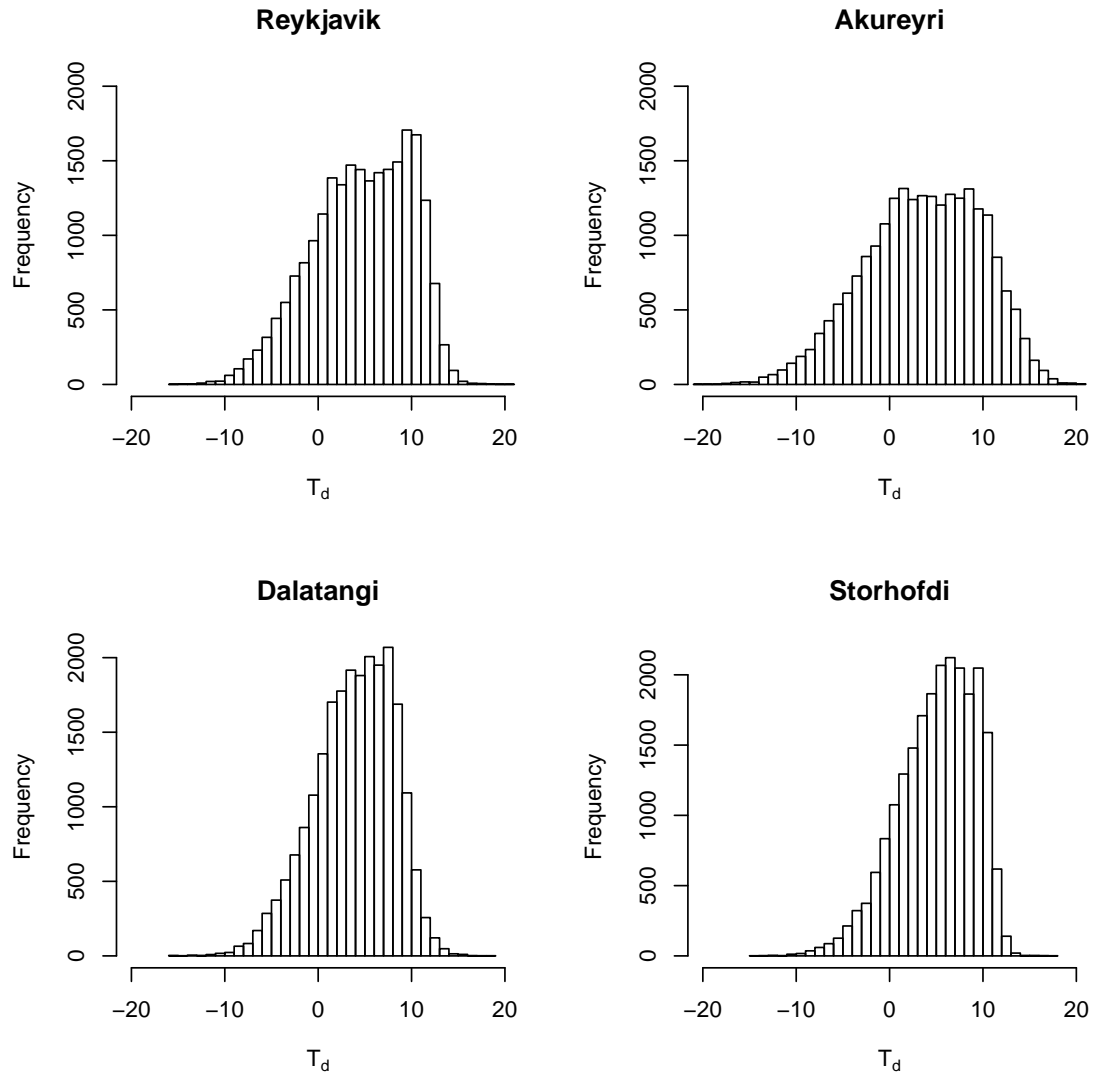


Figure 3.2: Frequency distribution of daily mean temperature for the whole period 1949-2010.

3. Data

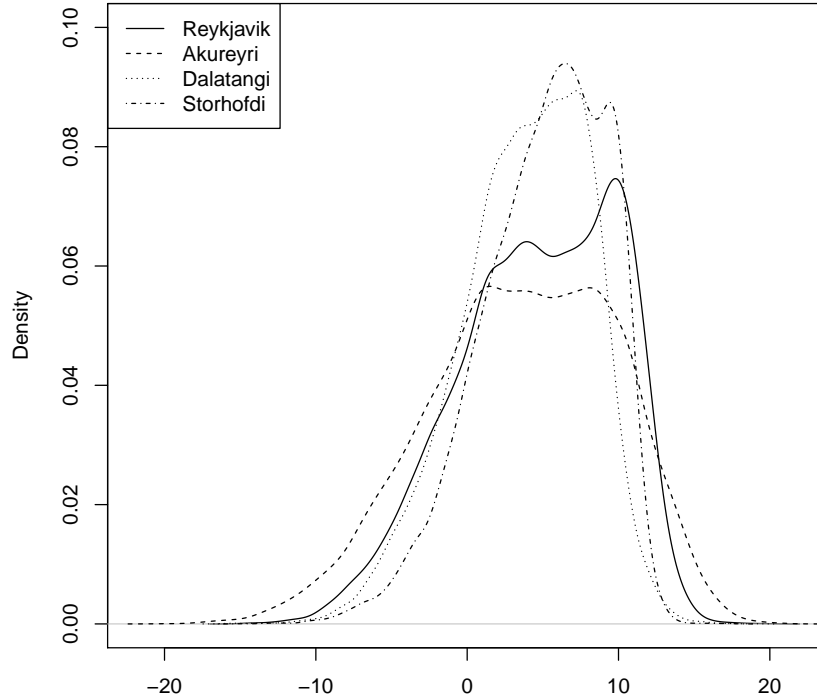


Figure 3.3: The densities of the daily mean temperature for the whole period 1949-2010, estimated by kernel smoothing.

3.1. Seasonal behavior

The data has a clear annual periodic behavior that is apparent in Figure 3.1, and should be expected in this type of seasonal data. The seasonal behavior can be examined by calculating the sample mean of each Julian calendar day by

$$\mu_d(k) = \frac{1}{m} \sum_{i=1}^m y_{i,k}. \quad (3.1)$$

The mean annual temperature cycle based on the daily average data is shown in Figure 3.4 (solid line) together with the 2.5% and 97.5% percentiles of each Julian day (dashed lines). Notice how the space between the two percentiles changes from winter to summer, indicating a difference in variance between seasons. The sample

standard deviation for each Julian calendar day is computed by

$$\sigma_d = \sqrt{\frac{1}{m-1} \sum_{i=1}^m (y_{i,k} - \mu_d(k))^2}. \quad (3.2)$$

The annual variation of the sample standard deviation of the daily mean temperature is plotted in Figure 3.5, showing a clear difference between seasons.

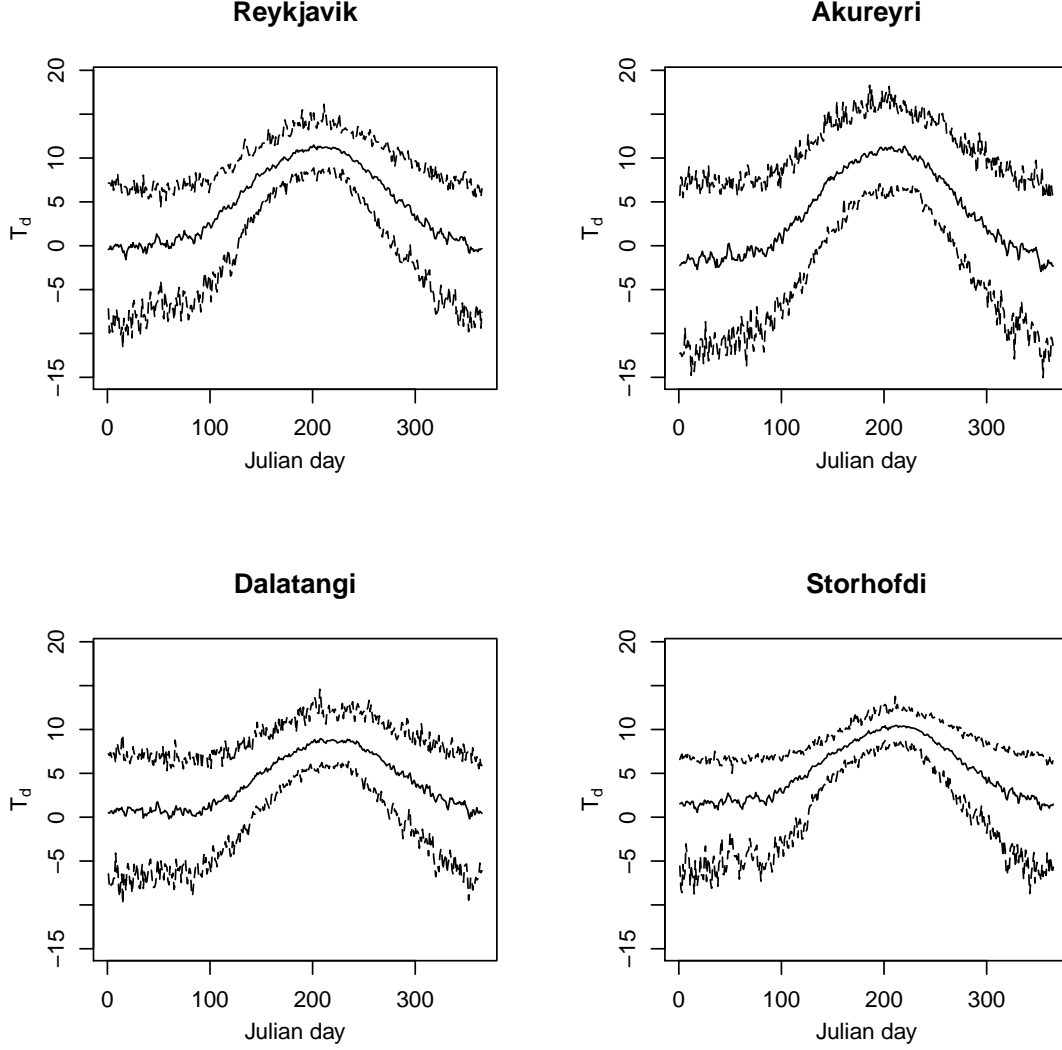


Figure 3.4: The mean annual temperature cycle based on daily temperature averages (solid line), together with 2.5% and 97.5% percentiles (dashed lines).

3. Data

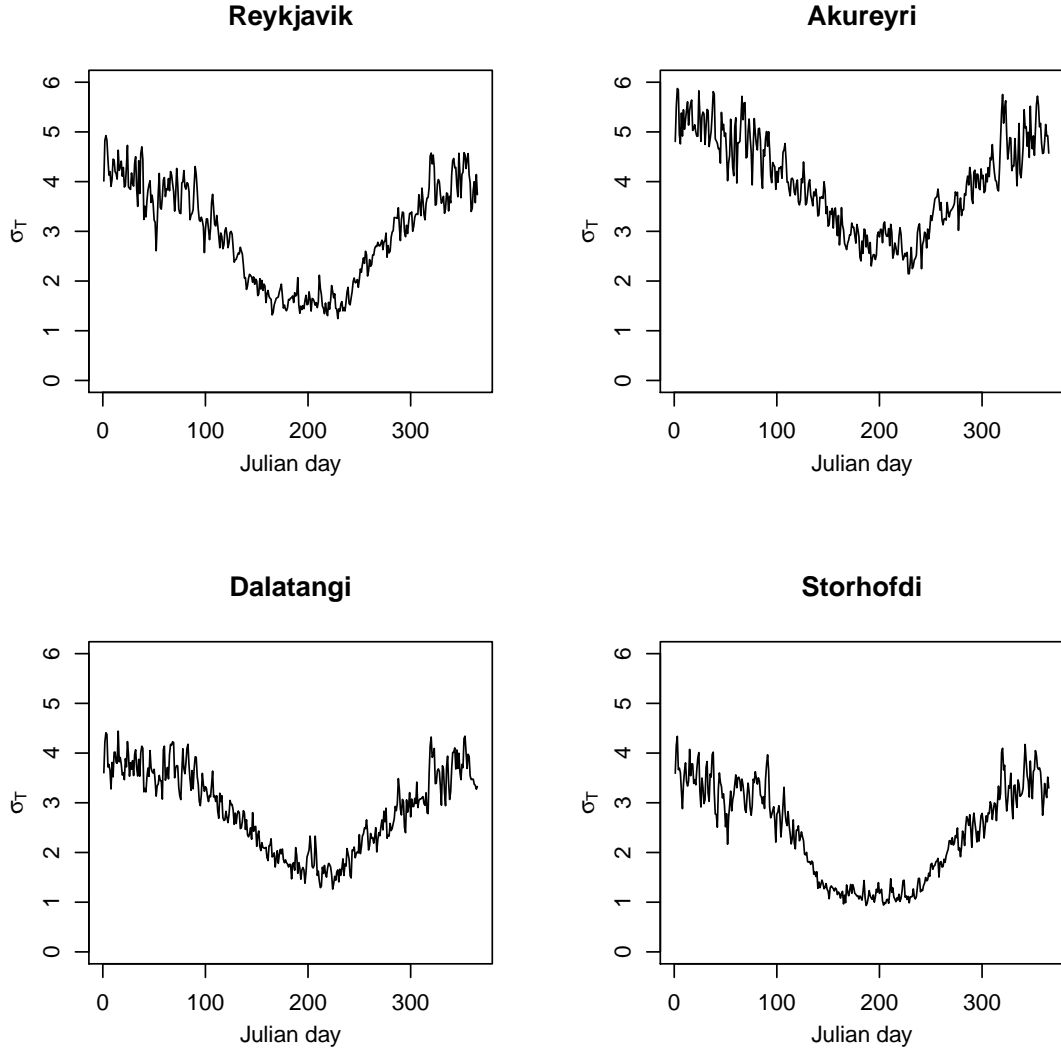


Figure 3.5: The sample standard deviation of the daily average temperature, as a function of Julian day.

3.2. Long term fluctuations

It is likely that the data contains some long term fluctuations. Since no long term fluctuation is observable visually, it has to be brought out of the data using the model, which will be done in following chapters. An estimate of the long term fluctuation can be made straight from the data. This is done by removing the seasonal

factor from the data and examining the residuals. The seasonality is removed by

$$\vec{e}_1 = \vec{y} - (\vec{1} \otimes \vec{\mu}_d) \quad (3.3)$$

where $\vec{1}$ is vector of ones of length $m = 62$ and the operator \otimes is the Kronecker product. An estimate of the fluctuations is shown in Figure 3.6, where a kernel smoothing object has been fitted to the residuals \vec{e}_1 . These figures are later be used for a comparison to the results from the model.

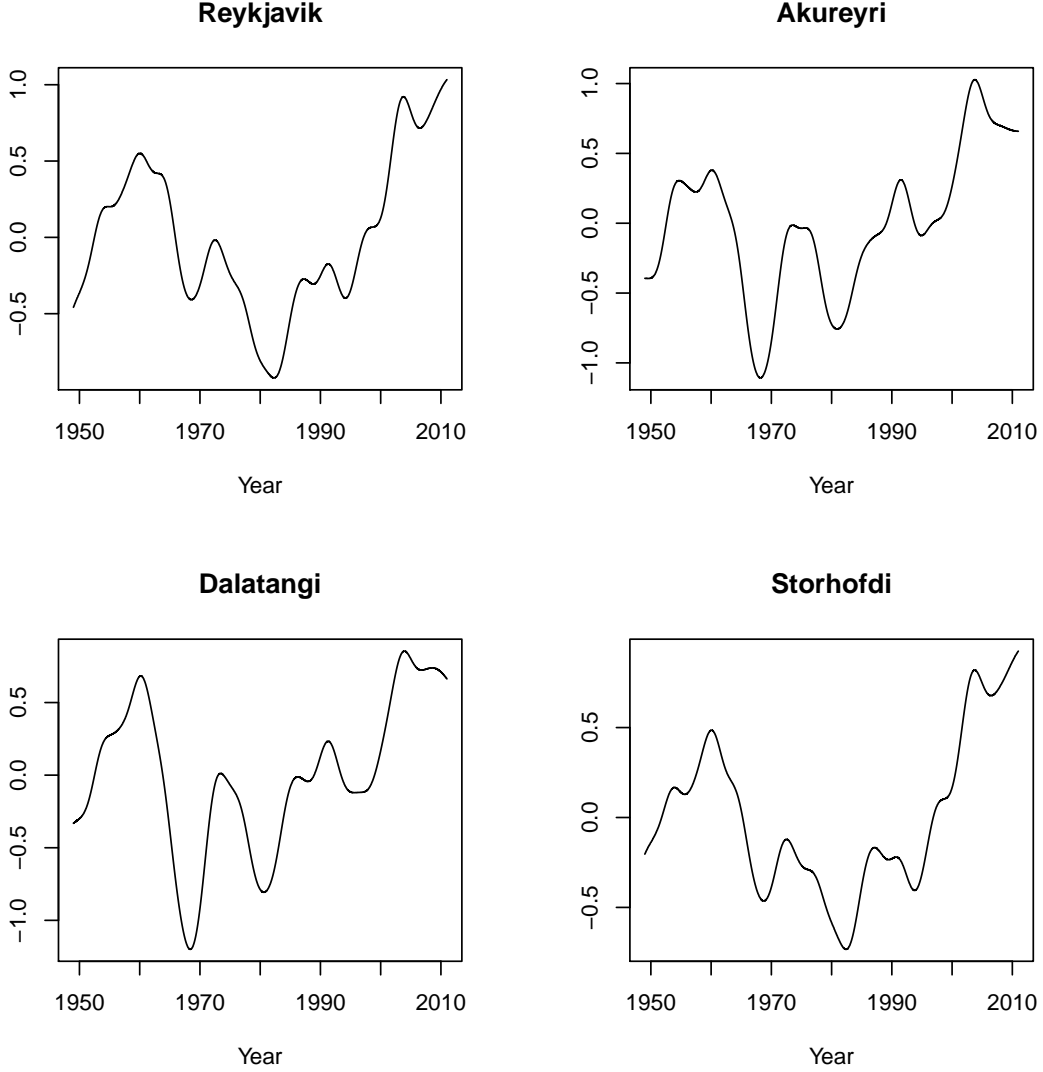


Figure 3.6: An estimation of the long term fluctuation in the data, calculated by a kernel smoothing, with bandwidth 1500, of \vec{e}_1 in Equation (3.3).

3.3. Correlation

Before examining the autocorrelation in the data, the estimated seasonal factor and long term fluctuation are subtracted from the data, resulting in the following residuals,

$$\vec{e}_2 = \vec{y} - (\vec{1} \otimes \vec{\mu}_d) - \vec{k}s \quad (3.4)$$

where $\vec{k}s$ stands for the Kernel smoothing estimate of the long term fluctuation. These residuals are similar to those of the model presented later in the thesis and should therefore give an indication of how the residuals in the model are temporally correlated. Figure 3.7 show the autocorrelation and partial autocorrelation functions of the residuals in Equation (3.4) for the Reykjavik series (the ACF and PACF figures are identical for all four locations, hence only results from one site are shown). It is apparent that there is correlation in the residuals and the PACF suggests that a low order autoregressive model (perhaps of third-order) would describe the temporal correlation in the data sufficiently.

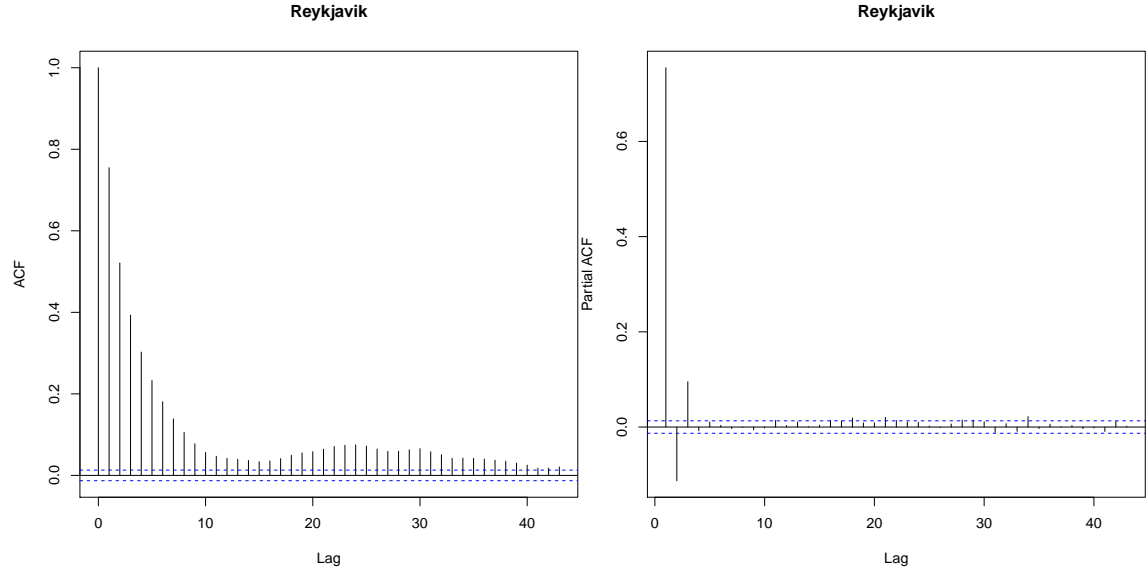


Figure 3.7: ACF and PACF for residuals \vec{e}_2 from Equation (3.4).

4. The Model

As explained in Chapter 3 there is a strong yearly seasonal behavior in the daily temperature data with temperatures being high in summer and low during winter. It is also likely that there is a long term fluctuation in the mean temperature over the period as Figure 3.6 and the discussion on global warming suggest (see for example Cox et al. (2000)). The data has a non-constant variance, which is clear from Figures 3.4 and 3.5, and its residuals after subtracting the seasonal factor and the long term fluctuation from the data are correlated (Figure 3.7). The structure of the model proposed in this chapter takes into account all the factors mentioned above. To summarize, the model should be able to capture:

- An annual change (seasonal effect).
- A linear trend.
- Long term fluctuations in the mean.
- The temporal correlation in the data.
- The seasonally changing variance.

4.1. A linear model

The model proposed in this thesis is a linear model of the form

$$\vec{y} = \mathbf{X}\vec{\beta} + \beta_0\vec{1} + \gamma_0\vec{T} + \mathbf{Z}\vec{\alpha} + \vec{\epsilon} \quad (4.1)$$

where \vec{y} represents the daily average temperature data, $\mathbf{X}\vec{\beta}$ a seasonal component, β_0 the constant of the model (which is very close to the mean of the time series), $\gamma_0\vec{T}$ a long term linear trend in the series, $\mathbf{Z}\vec{\alpha}$ a long term fluctuation term based on B-splines and an error term $\vec{\epsilon}$ that is assumed to be a normally distributed and autocorrelated with seasonally changing variance.

4. The Model

4.1.1. Seasonal behavior

The annual temperature change is represented in the model as $\mathbf{X}\vec{\beta}$, where $\vec{\beta}$ is a constrained parameter vector $\vec{\beta} = (\beta_1, \beta_2, \dots, \beta_{365})^T$ where there is one β_k parameter for every calendar day k , with February 29th being excluded (i.e. the day 1st January will get $k = 1$, 2nd January $k = 2$ up to 31th December $k = 365$). The matrix \mathbf{X} is a matrix of covariates (index variables) that ensures that every combination of day-year (i, k) gets the right β_k . It is given by

$$\mathbf{X} = \vec{1}_{m \times 1} \otimes \mathbf{I}_n$$

where m is the number of years and n number of calendar days (in this case the data spans 62 years so $m = 62$ and $n = 365$).

When multiplied with $\vec{\beta}$ the outcome will be

$$\mathbf{X} \cdot \vec{\beta} = \left(\vec{1}_{m \times 1} \otimes \mathbf{I}_n \right) \cdot \vec{\beta} = \vec{1}_{m \times 1} \otimes \vec{\beta} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_{365} \\ \beta_1 \\ \vdots \\ \beta_{365} \\ \vdots \\ \beta_1 \\ \vdots \\ \beta_{365} \end{pmatrix}$$

4.1.2. Long term behavior

The long term temperature behavior is represented in the model as $\gamma_0 \vec{T} + \mathbf{Z}\vec{\alpha}$. The term $\mathbf{Z}\vec{\alpha}$ is supposed to capture nonlinear fluctuations while $\gamma_0 \vec{T}$ is a linear trend over the whole period. In the linear trend the vector \vec{T} is a line created by

$$T_j = \frac{j}{365} - t_0 \tag{4.2}$$

where j is an index and t_0 is the center of the time series, in this case $t_0 = \frac{62 \times 365}{365} \times \frac{1}{2} = 31$. The parameter γ_0 is the slope of the line.

The matrix \mathbf{Z} is created by using B-spline Equations (2.18) and (2.19) from Section 2.4 and $\vec{\alpha}$ is a constrained parameter vector that scales the B-spline functions into their desired shape.

4.1.3. Autocorrelation

The model from Equation (4.1) can be written as

$$\vec{\epsilon} = \vec{y} - \mathbf{X}\vec{\beta} - \beta_0\vec{1} + \gamma_0\vec{T} - \mathbf{Z}\vec{\alpha} \quad (4.3)$$

which is similar to the residuals in Section 3.3 (Equation (3.4)), where an estimate of the seasonal effect and long term fluctuation have been subtracted from the data. From Figure 3.7 it is likely that an autoregressive process of order $p = 3$ will describe the serial correlation in the data adequately well.

A periodic autoregression model (PAR), presented in Section 2.5, is used to describe the serial correlation in the error term $\vec{\epsilon}$. The PAR model has the advantage over a regular AR model that it allows the variance to be non-constant.

An univariate n -period PAR model with varying order $(p(1), p(2), \dots, p(n))$ is defined by

$$\epsilon_{s \cdot n + k}^{(k)} = \varphi_1^{(k)} \epsilon_{s \cdot n + k - 1} + \dots + \varphi_{p(k)}^{(k)} \epsilon_{s \cdot n + k - p(k)} + v_{s \cdot n + k} \quad (4.4)$$

where

- k : period index, $k = 1, \dots, n$, e.g. $n = 365$
- s : year index, $s = 0, \dots, m - 1$, e.g. $m = 62$
- $v_{s \cdot n + k}$: noise term with $E(v_{s \cdot n + k}) = 0$ and $\text{var}(v_{s \cdot n + k}) = \sigma_v^2(k)$

A simple approach to the PAR model is taken by setting

$$p(1) = p(2) = \dots = p(n) = 3$$

and

$$\phi_1 = \varphi_1^{(k)}, \quad \phi_2 = \varphi_2^{(k)}, \quad \phi_3 = \varphi_3^{(k)} \quad \forall k$$

A regression model is used to estimate $v_{s \cdot n + k}$ and the auto regressive parameters ϕ_1 , ϕ_2 and ϕ_3 . Define

$$\vec{\epsilon}_t = \begin{bmatrix} \epsilon_4 \\ \vdots \\ \epsilon_{n \cdot m} \end{bmatrix}, \quad \vec{\epsilon}_{t-1} = \begin{bmatrix} \epsilon_3 \\ \vdots \\ \epsilon_{n \cdot m - 1} \end{bmatrix}, \quad \vec{\epsilon}_{t-2} = \begin{bmatrix} \epsilon_2 \\ \vdots \\ \epsilon_{n \cdot m - 2} \end{bmatrix}, \quad \vec{\epsilon}_{t-3} = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_{n \cdot m - 3} \end{bmatrix}$$

4. The Model

then Equation (4.4) can be written on the form

$$\vec{\varepsilon}_t = \begin{bmatrix} \vec{\varepsilon}_{t-1} & \vec{\varepsilon}_{t-2} & \vec{\varepsilon}_{t-3} \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \end{bmatrix} + \vec{v}_t = \boldsymbol{\epsilon} \vec{\phi} + \vec{v} \quad (4.5)$$

where $\boldsymbol{\epsilon} = \begin{bmatrix} \vec{\varepsilon}_{t-1} & \vec{\varepsilon}_{t-2} & \vec{\varepsilon}_{t-3} \end{bmatrix}$ and $\vec{\epsilon} = \begin{bmatrix} \phi_1 & \phi_2 & \phi_3 \end{bmatrix}$. The parameters of the PAR model can be estimated by

$$\hat{\vec{\phi}} = (\boldsymbol{\epsilon}^T \mathbf{W}_*^{-1} \boldsymbol{\epsilon})^{-1} \mathbf{W}_*^{-1} \boldsymbol{\epsilon}^T \vec{\epsilon}_t \quad (4.6)$$

where $\mathbf{W} = \text{diag}(\vec{1} \otimes \vec{\sigma}_v^2)$ is created using σ_v from Equation (4.9) and a vector of ones of length m . An estimate of the noise term $v_{s \cdot n + k}$ can be computed by

$$\hat{v} = \vec{\epsilon}_t - \boldsymbol{\epsilon} \hat{\vec{\phi}} \quad (4.7)$$

Finally the standard deviation of the noise term $v_{s \cdot n + k}$ can be computed by

$$\sigma_v(k) = \sqrt{\frac{1}{m} \sum_{s=0}^{m-1} v_{s \cdot n + k}^2} \quad (4.8)$$

4.1.4. Heteroscedasticity (non-constant variance)

It is important to include the seasonally changing variance in the error term ε_t in the model to ensure that the uncertainty of the model parameters is as correct as possible. The standard deviation of the noise term $v_{s \cdot n + k}$ is approximated using trigonometric functions described in the following equation,

$$\hat{\sigma}_v(k) = a_0 + \sum_{j=1}^3 \left\{ a_j \cos\left(\frac{2\pi j k}{n}\right) + b_j \sin\left(\frac{2\pi j k}{n}\right) \right\} \quad (4.9)$$

where a_j and b_j are unrestricted parameters and $n = 365$. The vector σ_v will be of length n . Parameters $a_0, a_1, a_2, a_3, b_1, b_2$ and b_3 are estimated by fitting the trigonometric function in Equation (4.9) to the calculated standard deviation from Equation (4.8).

4.2. Bayesian Inference

In this chapter assumptions are made about the prior distributions of the parameters in order to derive the posterior probability of the model and the conditional posterior distributions for each model parameter.

It is assumed that the data follows a multivariate normal distribution, as described in Section 2.1.2. Then the model in Equation (4.1) will become

$$\vec{y}|\vec{\beta}, \beta_0, \gamma_0, \vec{\alpha}, \sigma^2 \sim N(\mathbf{X}\vec{\beta} + \beta_0\vec{1} + \gamma_0\vec{T} + \mathbf{Z}\vec{\alpha}, \sigma^2\mathbf{L}^{-1}\mathbf{W}\mathbf{L}^{-T}) \quad (4.10)$$

where σ^2 is a scaling parameter that describes the variance of the model and \mathbf{L} is a matrix containing the autoregressive parameters described in Section 4.1.3,

$$L_{i,j} = \begin{cases} 1 & \text{for } (i, j = i) \\ -\phi_1 & \text{for } (i, j = i + 1) \\ -\phi_2 & \text{for } (i, j = i + 2) \\ -\phi_3 & \text{for } (i, j = i + 3) \\ 0 & \text{Otherwise.} \end{cases} \quad (4.11)$$

The matrix \mathbf{W} is a diagonal matrix containing the variance from Equation (4.9), such that $\mathbf{W} = \text{diag}(\vec{1} \otimes \vec{\sigma}_v^2)$. Both matrices are derived using the assumption that the residuals $\vec{\epsilon}$ of the model can be described by the multivariate normal distribution as $\vec{\epsilon} \sim N(0, \sigma^2\mathbf{L}^{-1}\mathbf{W}\mathbf{L}^{-T})$, see Appendix A.3.

4.2.1. Prior distributions

An important part of Bayesian inference is the selection of appropriate prior distributions for the unknown parameters. The following prior distributions are assumed for the unknown parameters in Equation (4.10) (the latent parameters of the model)

$$\begin{aligned} \vec{\beta}|\sigma_\beta^2 &\sim N(\vec{0}, \sigma_\beta^2\mathbf{P}_2^{-1}) \\ \beta_0|\mu_{\beta_0}, \sigma_{\beta_0}^2 &\sim N(\mu_{\beta_0}, \sigma_{\beta_0}^2), & \mu_{\beta_0} = 0, \sigma_{\beta_0}^2 = 10^4 \\ \gamma_0|\mu_\gamma, \sigma_\gamma^2 &\sim N(\mu_\gamma, \sigma_\gamma^2), & \mu_\gamma = 0, \sigma_\gamma^2 = 10^4 \\ \vec{\alpha}|\sigma_\alpha^2 &\sim N(\vec{0}, \sigma_\alpha^2\mathbf{P}_1^{-1}) \\ \sigma^2|V_\sigma, S_\sigma^2 &\sim \text{Inv-}\chi^2(V_\sigma, S_\sigma^2), & V_\sigma = 10^{-12}, S_\sigma^2 = 1 \end{aligned}$$

where $\mu_{\beta_0}, \sigma_{\beta_0}^2, \mu_\gamma, V_\sigma, S_\sigma^2$ are all constants selected based the knowledge of the data and $\sigma_\beta^2, \sigma_\alpha^2$ are hyperparameters. The precision matrixes \mathbf{P}_1 and \mathbf{P}_2 are constructed using Equations (2.10) and (2.11), respectively. These matrices generate a dependency structure in the parameter vectors $\vec{\beta}$ and $\vec{\alpha}$. The prior distributions selected for the hyperparameters are

$$\begin{aligned} \sigma_\beta^2|V_{\sigma\beta}, S_{\sigma\beta}^2 &\sim \text{Inv-}\chi^2(V_{\sigma\beta}, S_{\sigma\beta}^2), & V_{\sigma\beta} = 10^{-12}, S_{\sigma\beta}^2 = 1 \\ \sigma_\alpha^2|V_{\sigma\alpha}, S_{\sigma\alpha}^2 &\sim \text{Inv-}\chi^2(V_{\sigma\alpha}, S_{\sigma\alpha}^2), & V_{\sigma\alpha} = 10^{-12}, S_{\sigma\alpha}^2 = 1 \end{aligned}$$

4.2.2. Posterior distribution

The posterior distribution is a product of the prior distribution and the likelihood (recall Equations (2.4) and (2.6)). The posterior distribution for the model is

$$\begin{aligned}
 p(\vec{\beta}, \beta_0, \vec{\alpha}, \gamma_0, \sigma^2, \sigma_\beta^2, \sigma_\alpha^2 | \vec{y}) &\propto p(\vec{y} | \vec{\beta}, \beta_0, \vec{\alpha}, \gamma_0, \sigma^2) p(\vec{\beta}, \beta_0, \vec{\alpha}, \gamma_0, \sigma^2, \sigma_\beta^2, \sigma_\alpha^2) \\
 &\propto p(\vec{y} | \vec{\beta}, \beta_0, \vec{\alpha}, \gamma_0, \sigma^2) p(\vec{\beta} | \sigma_\beta^2) p(\beta_0) p(\vec{\alpha} | \sigma_\alpha^2) p(\gamma_0) \\
 &\quad p(\sigma^2) p(\sigma_\beta^2) p(\sigma_\alpha^2)
 \end{aligned}$$

The conditional posterior distributions for each parameter can be calculated using the posterior distribution above and the prior distributions from Section 4.2.1 (see Appendix A.1 for derivation). The conditional posterior distributions for the parameters are:

$$\begin{aligned}
 p(\vec{\beta} | \text{rest}) &\sim N(\vec{\mu}_{\beta, \text{post}}, \Sigma_{\beta, \text{post}}) \\
 p(\beta_0 | \text{rest}) &\sim N(\mu_{\beta_0, \text{post}}, \sigma_{\beta_0, \text{post}}^2) \\
 p(\gamma_0 | \text{rest}) &\sim N(\mu_{\gamma_0, \text{post}}, \sigma_{\gamma_0, \text{post}}^2) \\
 p(\vec{\alpha} | \text{rest}) &\sim N(\vec{\mu}_{\alpha, \text{post}}, \Sigma_{\alpha, \text{post}}) \\
 p(\sigma^2 | \text{rest}) &\sim \text{Inv-}\chi^2(V_{\sigma, \text{post}}, S_{\sigma, \text{post}}^2) \\
 p(\sigma_\beta^2 | \text{rest}) &\sim \text{Inv-}\chi^2(V_{\sigma_\beta, \text{post}}, S_{\sigma_\beta, \text{post}}^2) \\
 p(\sigma_\alpha^2 | \text{rest}) &\sim \text{Inv-}\chi^2(V_{\sigma_\alpha, \text{post}}, S_{\sigma_\alpha, \text{post}}^2)
 \end{aligned} \tag{4.12}$$

The mean and variance or covariance matrices in the normal conditional posterior distributions for $\vec{\beta}, \beta_0, \vec{\alpha}$ and γ_0 are

$$\begin{aligned}
 \Sigma_{\beta, \text{post}} &= (\sigma_\beta^{-2} P + \sigma^{-2} X^T \Sigma^{-1} X)^{-1} & \mu_{\beta, \text{post}} &= \Sigma_{\beta, \text{post}} \cdot \sigma^{-2} X^T \Sigma^{-1} \vec{\epsilon} \\
 \sigma_{\beta_0, \text{post}}^2 &= (\sigma_{\beta_0}^{-2} + \mathbf{1}^T \sigma^{-2} \Sigma^{-1} \mathbf{1})^{-1} & \mu_{\beta_0, \text{post}} &= \sigma_{\beta_0, \text{post}}^2 \cdot (\sigma_{\beta_0}^{-2} \mu_{\beta_0} + \sigma^{-2} \mathbf{1}^T \Sigma^{-1} \vec{\epsilon}) \\
 \Sigma_{\alpha, \text{post}} &= (\sigma_\alpha^{-2} P_2 + \sigma^{-2} Z^T \Sigma^{-1} Z)^{-1} & \mu_{\alpha, \text{post}} &= \Sigma_{\alpha, \text{post}} \cdot \sigma^{-2} Z^T \Sigma^{-1} \vec{\epsilon} \\
 \sigma_{\gamma_0, \text{post}}^2 &= (\sigma_\gamma^{-2} + \sigma^{-2} T^T \Sigma^{-1} T)^{-1} & \mu_{\gamma_0, \text{post}} &= \sigma_{\gamma_0, \text{post}}^2 \cdot (\sigma_\gamma^{-2} \mu_\gamma + \sigma^{-2} T^T \Sigma^{-1} \vec{\epsilon})
 \end{aligned}$$

The degrees of freedom and scale in the scaled inverse- χ^2 conditional posterior distributions of σ^2, σ_β^2 and σ_α^2 are

$$\begin{aligned}
 V_{\sigma, \text{post}} &= V_\sigma + n & S_{\sigma, \text{post}}^2 &= \frac{V_\sigma S_\sigma^2 + \vec{\epsilon}^T \Sigma^{-1} \vec{\epsilon}}{V_\sigma + n} \\
 V_{\sigma_\beta, \text{post}} &= V_{\sigma_\beta} + k & S_{\sigma_\beta, \text{post}}^2 &= \frac{V_{\sigma_\beta} S_{\sigma_\beta}^2 + \beta^T P \beta}{V_{\sigma_\beta} + k} \\
 V_{\sigma_\alpha, \text{post}} &= V_{\sigma_\alpha} + m & S_{\sigma_\alpha, \text{post}}^2 &= \frac{V_{\sigma_\alpha} S_{\sigma_\alpha}^2 + \alpha^T P_2 \alpha}{V_{\sigma_\alpha} + m}
 \end{aligned}$$

where $\vec{\epsilon} = \vec{y} - \mathbf{X}\vec{\beta} - \beta_0\vec{1} - \gamma_0\vec{T} - \mathbf{Z}\vec{\alpha}$ and $\Sigma = \mathbf{L}^{-T} \mathbf{W} \mathbf{L}^{-1}$.

4.2.3. MCMC

Inference for the parameters $\vec{\beta}, \beta_t, \gamma_0, \vec{\alpha}, \sigma^2$ will be based on the Bayesian approach but the parameters $a_0, a_1, a_2, a_3, b_1, b_2, b_3, \phi_1, \phi_2$ and ϕ_3 will be estimated with classical estimators, see Sections 4.1.3 and 4.1.4. Since the autoregressive parameters are used to estimate the Bayesian parameters and vice versa, an iterative process is used. The following iterative inference algorithm is used to update the autoregressive parameters in the model

Algorithm 3 Iterative inference

```

Assume  $\vec{\epsilon}_t \sim N(\vec{0}, \sigma^2 \mathbf{I})$ 
Compute  $\vec{\beta}, \beta_0, \gamma_0, \vec{\alpha}$  and  $\sigma^2$ 
for  $i$  in  $1, \dots, h$  do
    Now assume  $\vec{\epsilon}_t \sim N(\vec{0}, \sigma^2 \mathbf{L}^{-1} \mathbf{W} \mathbf{L}^{-T})$ 
    Compute  $\vec{\epsilon} = \vec{y} - \mathbf{X}\vec{\beta} - \beta_0\vec{1} - \gamma_0\vec{T} - \mathbf{Z}\vec{\alpha}$ 
    Get estimates for  $\phi_1, \phi_2, \phi_3, a_0, \dots, a_4, b_1, \dots, b_4$ 
    Compute  $\vec{\beta}, \beta_0, \gamma_0, \vec{\alpha}$  and  $\sigma^2$ 
end for
return  $\vec{\beta}, \beta_0, \gamma_0, \vec{\alpha}$  and  $\sigma^2$ 

```

4.3. Computer implementation

A program based on the model was developed in the statistical programming language R. The program is based on Algorithm 3, from Section 4.2.3, and uses the Gibbs sampler, described in Section 2.2, to create samples from the posterior distribution of the model, using the conditional posterior distributions of the parameters. The starting values for the Gibbs sampler are determined with experiments, the number of chains and the number of iterations are chosen to be 4 and 5000 respectively. A burn-in of 1000 is found to be sufficient by examining the MCMC trace plots and the Gelman-Rubin statistics for the chains, see Section 5.1.1. The number of iterations in Algorithm 3 is chosen to be $h = 4$.

One of the biggest obstacle in creating a computationally efficient program for the model was the size of the data. The length of the data vector \vec{y} is $m \cdot n$ so the correlation matrices of the data distribution will be $m \cdot n \times m \cdot n$. A full matrix of this size is too big for most personal computers on the market to handle (at least

4. The Model

when using R). Therefore the covariance matrices had to be estimated through their inverse as GMRF matrices and by using the PAR structure (Equation (4.11)) to create sparse band matrices.

To ensure that the parameter vectors β and α have zero mean, Algorithm 1 from Section 2.1.3 is used. This ensures that all model parameters are identifiable. The numbers of parameters in parameter vector $\vec{\alpha}$ is chosen by examining the DIC for various numbers of parameters in parameter vector $\vec{\alpha}$. The number of α parameters in the model is an important factor to create a smooth process and to avoid over-fitting. According to the DIC criteria the lowest value of DIC gives the best fit. In Figure 4.1 the lowest DIC is when the number of α parameters is 31, which means one α per every two years.

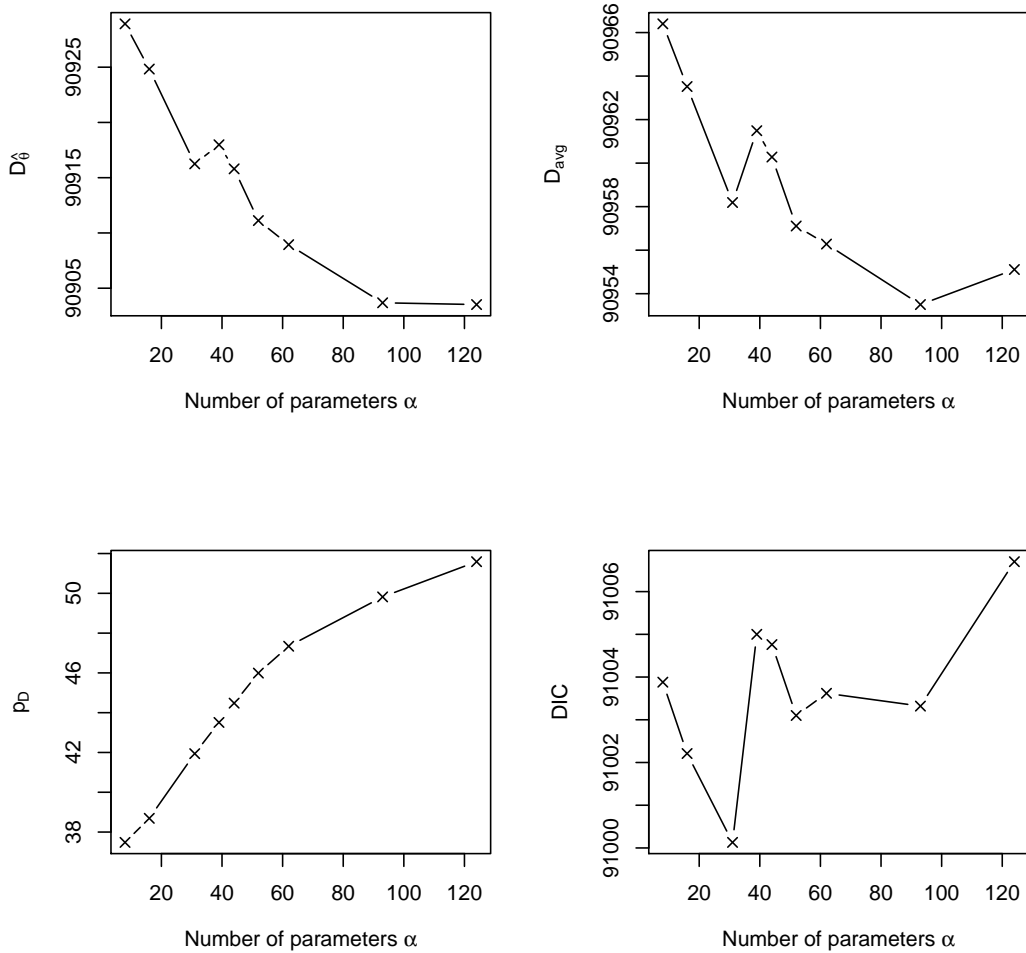


Figure 4.1: Deviance information criterion for different numbers of parameters in parameter vector $\vec{\alpha}$.

5. Results

In this chapter the Bayesian model described in Chapter 4 is applied to the four data sets using the programming language R. The resulting parameter estimates and other results are presented in Section 5.1. The convergence is discussed in Subsection 5.1.1. In Section 5.2 the outcome from the model is compared to the data analysis done in Chapter 3. In Section 5.3 a forecast for the year 2011 is made, using the model and fitted parameters and compared to the actual data for the year 2011. In Section 5.4 the residuals of the model are analyzed and it is checked whether a filtered version of the residuals has reduced to white noise.

5.1. Main results

The results from fitting the model to the data are presented in the following tables and figures. The expected values and 95% posterior intervals for the single latent parameters are shown in Table 5.1 and for the hyperparameters in Table 5.2. It is apparent from Table 5.1 that the parameter β_0 is very similar to the mean of the data shown in Table 3.1, as expected and β_0 is statistically significant for all locations. Based on the parameter γ_0 it can be seen that the linear trend increases by about 0.05 to 0.46°C per 10 years over the time period. The γ_0 parameters are all statistically significant except the parameter for trend at Dalatangi (zero in the 95% posterior interval for γ_0). The parameters σ^2 , σ_β^2 and σ_γ^2 are difficult to interpret since they are not the marginal variances (since the inverses of the precision matrices they are multiplied to are not correlation matrices). Parameter estimates for the PAR model are shown in Table 5.3 and the results for the parameter vector $\vec{\alpha}$ along with 95% posterior intervals is plotted in Figure 5.1.

The outcome for seasonal component of the model ($\beta + \beta_0$) is plotted in Figure 5.2 along with 95% posterior intervals, showing a curve similar to the mean Julian calendar temperature from Chapter 3. The posterior intervals are narrow indicating a small uncertainty in the estimate, this is due to the large data set and neighbor structure. The posterior intervals also have a similar varying behavior as is apparent in Figure 3.4, suggesting that the model is capturing the changing uncertainty over Julian days. The long term temperature fluctuation is shown in Figure 5.3. From

5. Results

Figure 5.3 it is easy to imagine a linear trend for Dalatangi for the period 1970 to 2010 but it is difficult to see a linear trend when looking at the whole period. That suggest that perhaps the γ_0 parameter for Dalatangi is insignificant.

Table 5.1: Posterior means of β_0 , γ_0 and σ^2 along with 95% credible intervals in brackets.

	β_0	γ_0 [°C/10 years]	σ^2
Reykjavik	4.72 (4.66,4.78)	0.29 (0.15,0.46)	1.01 (0.99,1.03)
Akureyri	3.65 (3.60,3.71)	0.20 (0.05,0.35)	1.01 (0.99,1.03)
Dalatangi	3.86 (3.82,3.91)	0.17 (-0.03,0.36)	1.01 (0.99,1.03)
Storhöfði	5.17 (5.12,5.21)	0.25 (0.08,0.43)	1.02 (1.00,1.04)

Table 5.2: Posterior means of σ_β^2 and σ_α^2 along with 95% credible intervals in brackets.

	$\sigma_\beta^2 \cdot 10^{-3}$	σ_α^2
Reykjavik	1.14 (0.66,1.86)	0.12 (0.04,0.30)
Akureyri	1.52 (0.88,2.58)	0.10 (0.02,0.28)
Dalatangi	0.67 (0.38,1.12)	0.27 (0.10,0.62)
Storhöfði	0.68 (0.39,1.13)	0.18 (0.05,0.47)

Table 5.3: Autoregressive parameter estimates.

	ϕ_1	ϕ_2	ϕ_3
Reykjavik	0.844	-0.192	0.0902
Akureyri	0.819	-0.183	0.0857
Dalatangi	0.795	-0.176	0.0861
Storhofdi	0.828	-0.193	0.0929

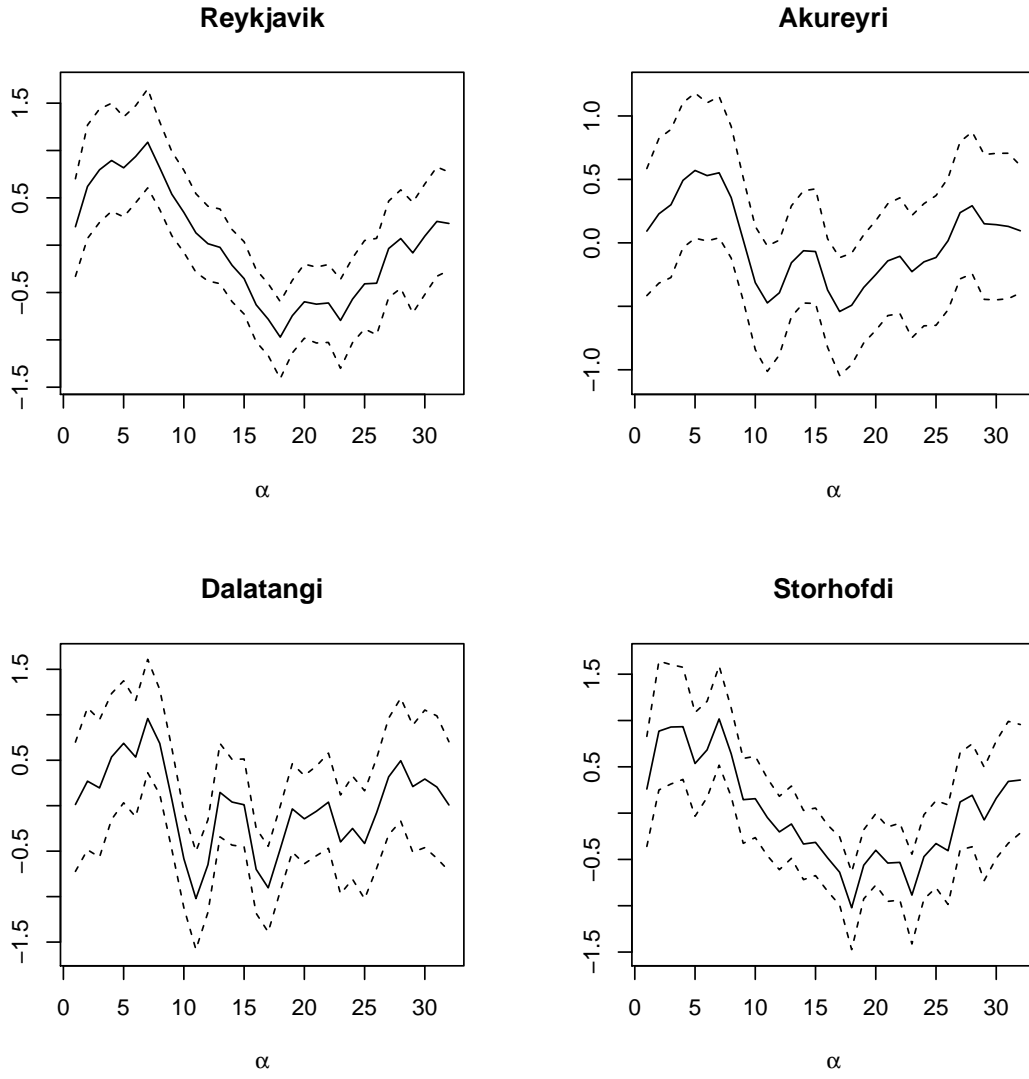


Figure 5.1: Posterior means for the four locations of $\vec{\alpha}$ (solid line) and 95% credible intervals (dashed lines).

5. Results

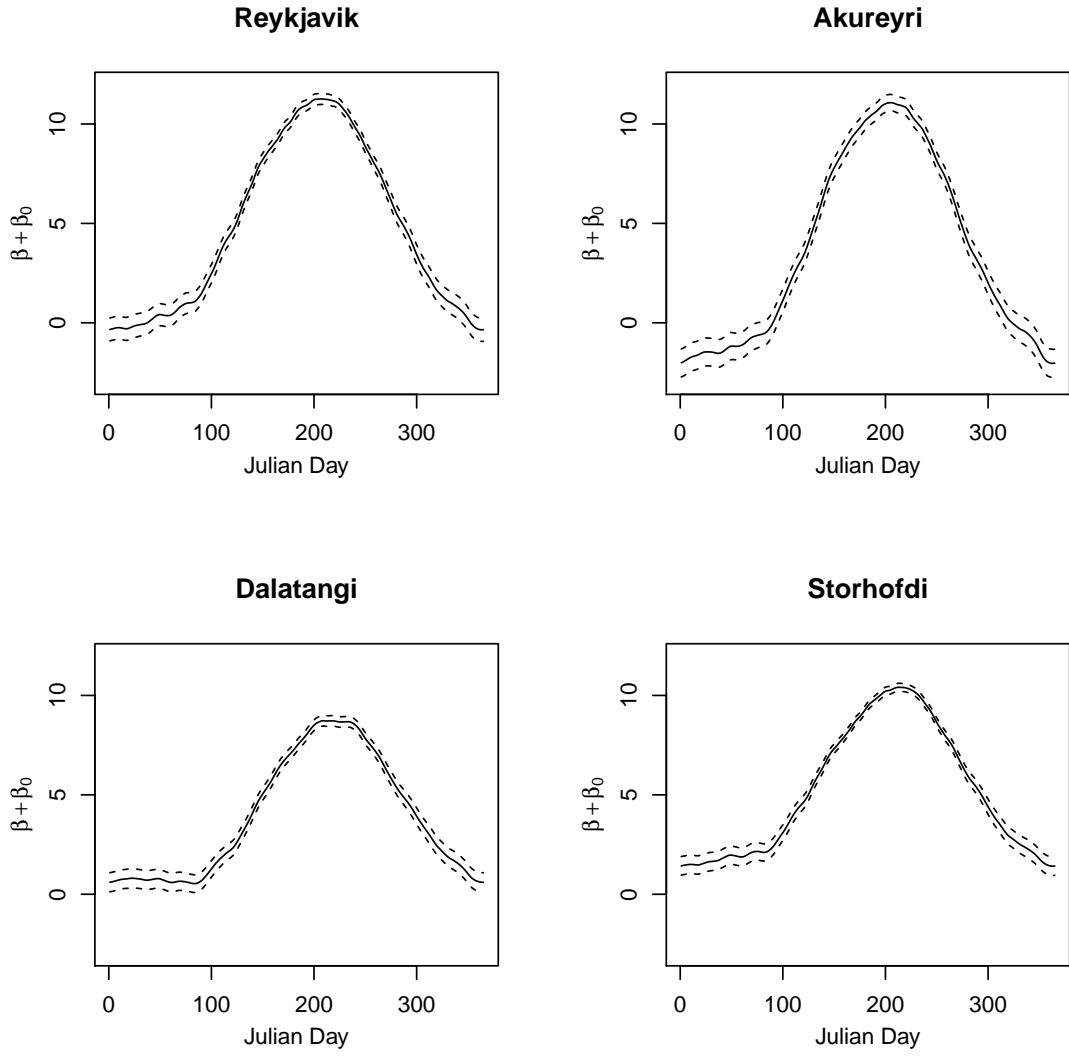


Figure 5.2: Posterior means of $(\vec{\beta} + \beta_0)$ (solid line) and 95% credible intervals (dashed lines), for the four locations.

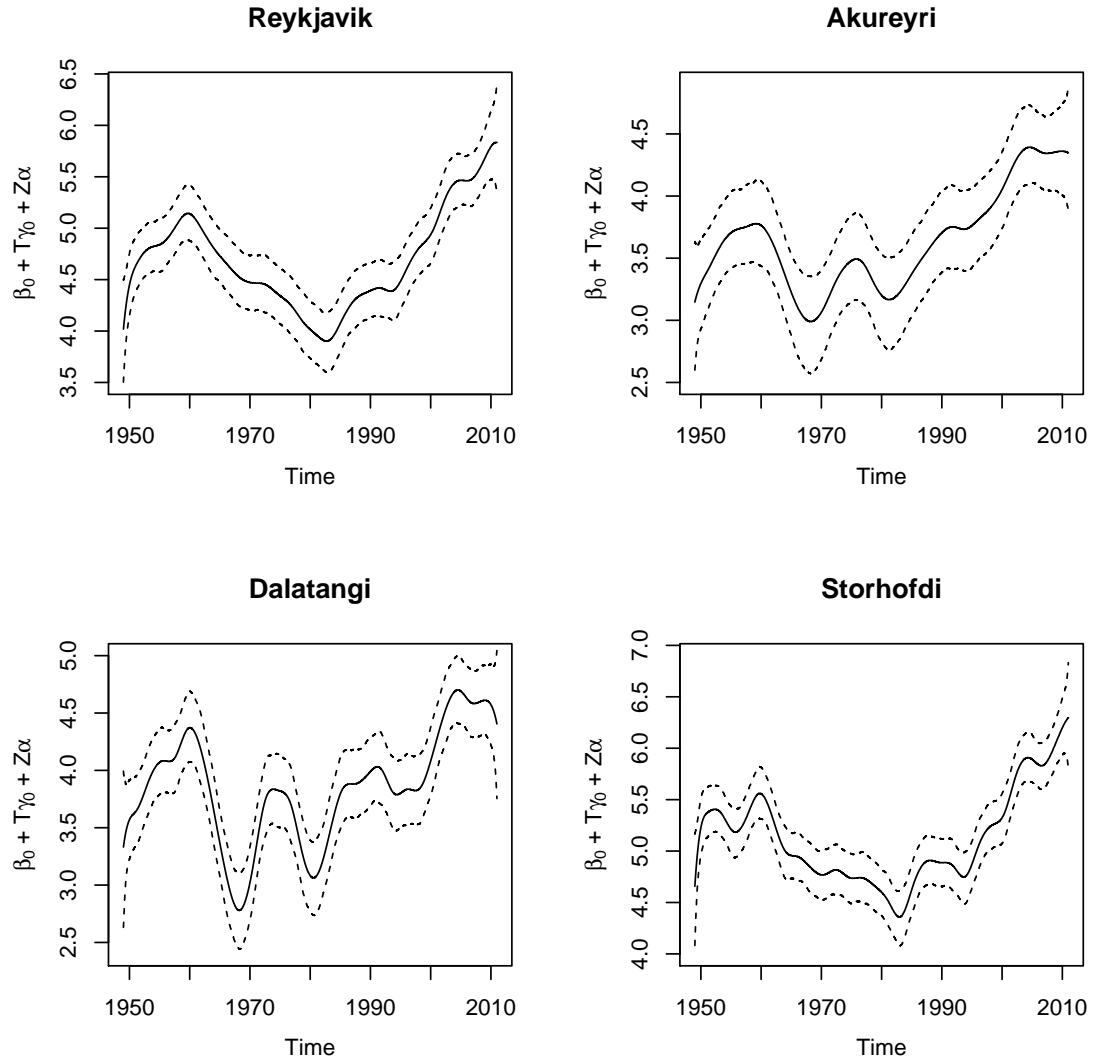


Figure 5.3: Posterior means of the long term temperature fluctuation as a function of time, estimated by the model, i.e. $(\beta_0 \vec{1} + \gamma_0 \vec{T} + \mathbf{Z}\vec{\alpha})$ (solid line), along with 95% credible interval (dashed lines), for the four locations.

5.1.1. Convergence

Convergence of iterative parameters in Bayesian statistics is usually checked by visual inspection of trace plots for each parameter and by the Gelman-Rubin statistic. The Gelman-Rubin statistic, first presented in Gelman and Rubin (1992), is an estimate of how far the current sampling distribution is from its target distribution. In most cases values of the Gelman-Rubin statistic below 1.1 are acceptable (Gelman, 2004). For all the parameters in this case the value is below 1.1 (for most of them close to 1). This and the MCMC trace plots (which are shown for two parameters at Reykjavik in Figure 5.4) confirm that 5000 iterations with a burn-in of 1000 is sufficient for the model parameters to converge to their target distribution. This fast convergence is most likely due to the way the model is set up with the normal distribution describing most of the model parameters. Histograms and MCMC trace plots for other parameters and locations are shown in Appendix B.

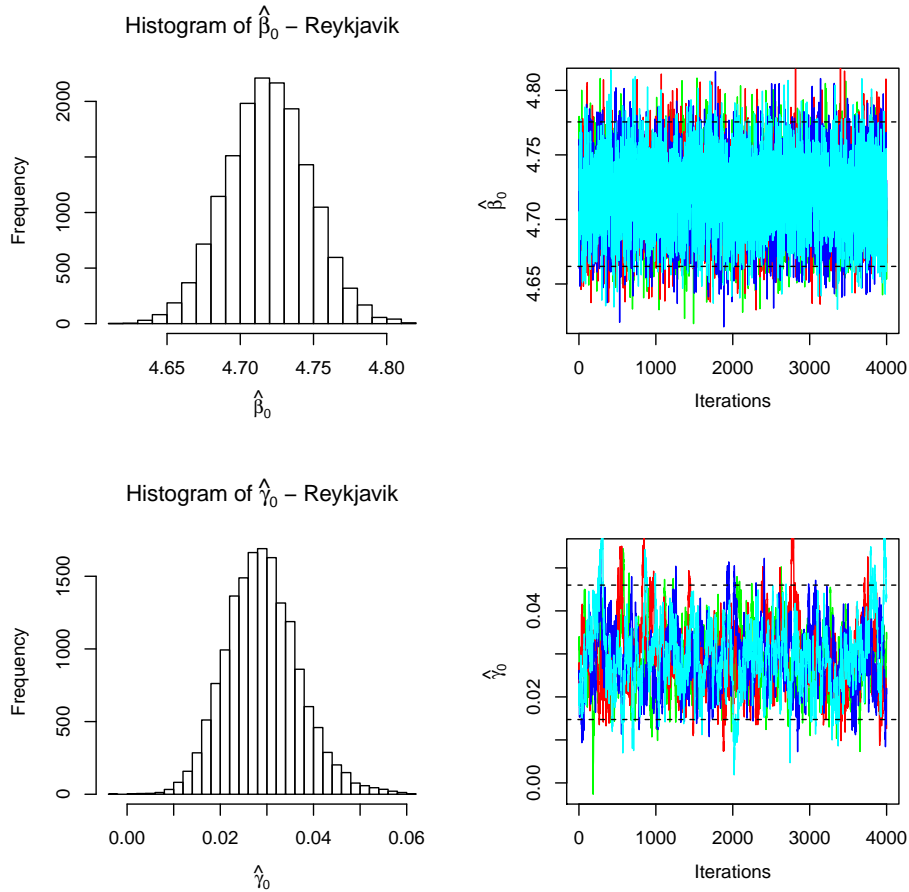


Figure 5.4: Histograms and MCMC trace plots after burn-in for the parameters β_0 and γ_0 at location Reykjavik.

5.2. Comparison to the data

To see how well the model describes the data the outcome of the model is compared to the original data described in Chapter 3. The seasonal effect from the model should be close to the mean for every calendar day from Figure 3.4, but smoother due to its prior structure. In Figure 5.5 the fitted values of the seasonal factor in the model ($\beta + \beta_0$) are plotted as a solid line along with the calendar day sample mean from Chapter 3 as a dashed line. This appears to be a perfect fit with the model creating a smooth process through the mean samples. The long term fluctuation from the model should also be similar to the fluctuation estimated using kernel smoothing in Section 3.2. Figure 5.6 shows the fitted values of the long term fluctuation in the model compared to the long term fluctuation estimate from Section 3.2, showing that the fit of these two approaches is similar.

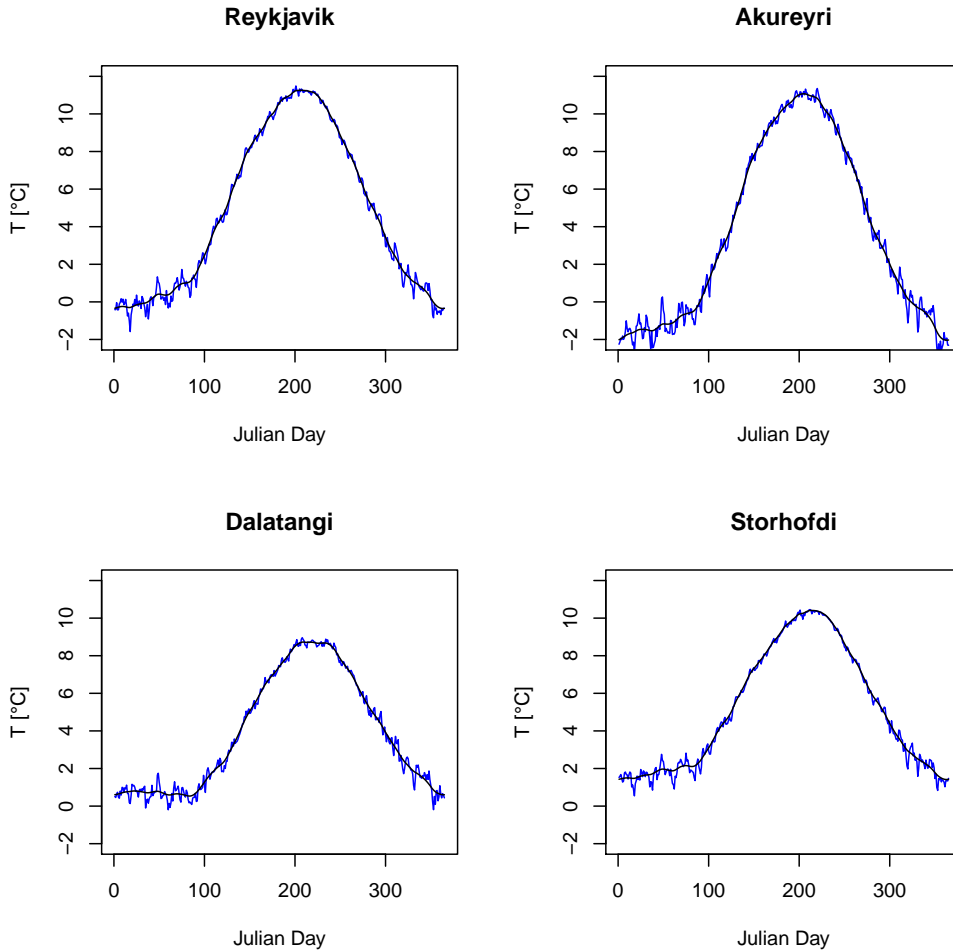


Figure 5.5: Posterior means of $(\vec{\beta} + \beta_0)$ (black line) compared to the sample means of Julian calendar day temperature from Figure 3.4 in Chapter 3 (blue line).

5. Results

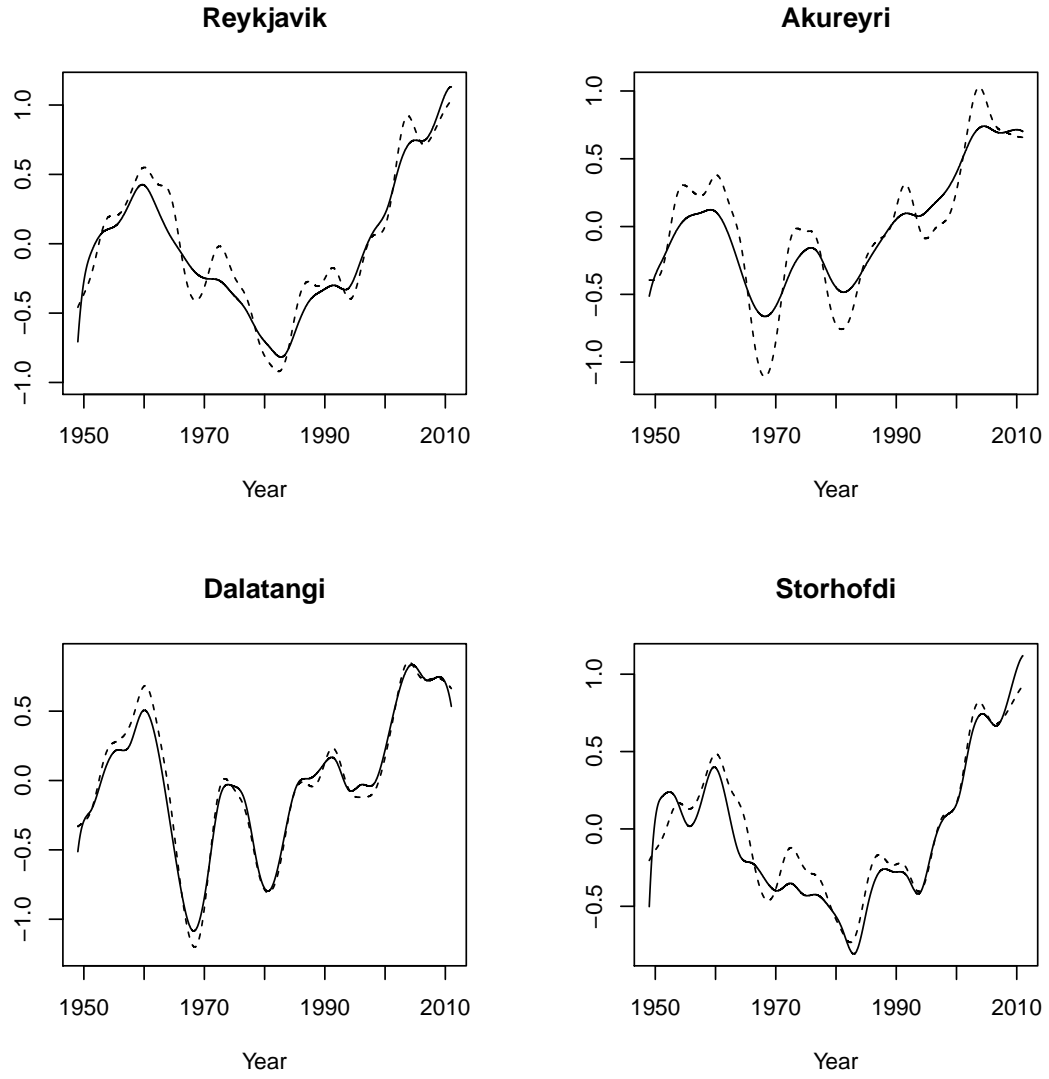


Figure 5.6: Posterior means for the fluctuation $(\gamma_0 \vec{T} + \mathbf{Z} \vec{\alpha})$ (solid line) compared to the fluctuation estimate based on smoothing from Figure 3.6 (dashed line).

5.3. Prediction

In this section the parameter estimates are used to make a prediction for one year ahead. An equation for a future y , denoted by \tilde{y} , can be written as

$$\tilde{y} \sim N(\vec{\mu}_{\text{pred}}, \Sigma_{\text{pred}}) \quad (5.1)$$

with

$$\vec{\mu}_{\text{pred}} = \vec{\beta} + \beta_0 \cdot \vec{1} + \gamma_0 \cdot \vec{T}_* + \mathbf{Z}_* \cdot \vec{\alpha} \quad (5.2)$$

and

$$\Sigma_{\text{pred}} = \sigma^2 \cdot \mathbf{L}_*^{-1} \mathbf{W}_* \mathbf{L}_*^{-T} \quad (5.3)$$

where $\vec{\beta}, \beta_0, \gamma_0, \vec{\alpha}$ and σ^2 are parameter estimates from the model and \vec{T}_{pred} and \mathbf{Z}_{pred} are extensions of \vec{T} and \mathbf{Z} , with \vec{T}_{pred} being calculated from Equation (4.2) using $j = m \cdot n + 1, \dots, m \cdot n + n$ and $t_0 = 31$ as before.

The matrices in the covariance matrix have the same structure as before, only now they contain information for just one year. That is, the matrix \mathbf{L}_* is created using Equation (4.11) and $\mathbf{W}_* = \text{diag}(\sigma_v^2)$.

In Figure 5.7 a prediction for one year ahead (solid line) is plotted along with 2.5% and 97.5% percentiles (dashed line) created by drawing a sample of 30.000 from the normal distribution in Equation (5.1). The red line shows the actual daily average temperature for 2011. The actual temperature is within the 95% confidence interval between 2.5% to 4.7% of the days for the four locations.

5. Results

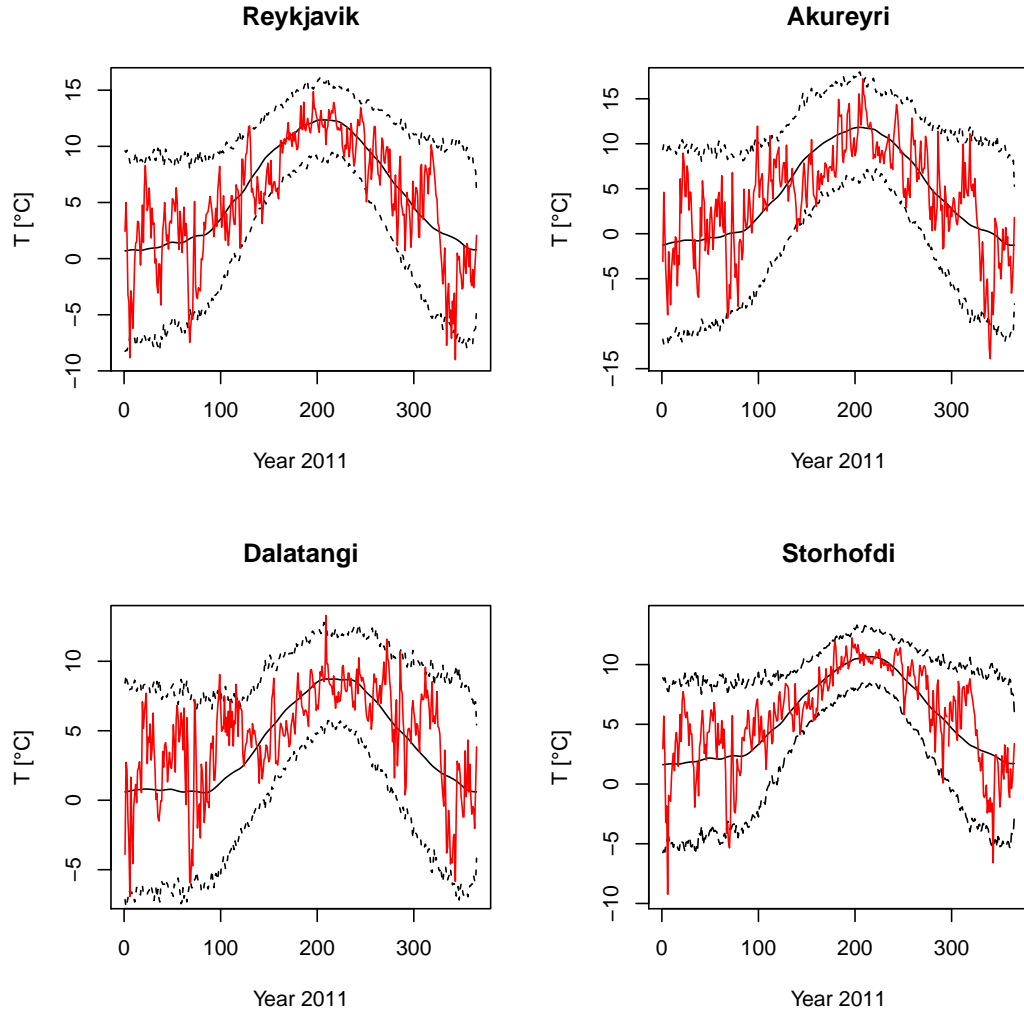


Figure 5.7: Prediction for the year 2011 (solid line) along with 2.5% and 97.5% percentiles (dashed line) and the actual temperature (red line), for the four locations.

5.4. The residuals

One way of checking if the model is describing the data adequately is by examining the residuals of the model to check whether or not they resemble white noise.

In Figure 5.8 the standard deviation of the residuals $v_{s \cdot n+k}$ from the PAR model are plotted (calculated from Equation (4.8)) along with the trigonometric fit from Equation (4.9). The trigonometric function fits the values of the standard deviation well.

Figure 5.9 shows different residuals from the model, using Reykjavik data. At the top are the correlated residuals ϵ_t that have seasonal variance. In the middle are the residuals v_t where the correlation has been removed but the variance still unequal, which is clearly visible in the figure. The bottom residuals have been standardized by dividing the uncorrelated residuals v_t with the fitted standard deviation $\hat{\sigma}_v$. The standardized residuals should be uncorrelated with unit variance, that is they should be close to being white noise.

A density estimate of the standardized residuals is plotted in Figure 5.10, showing a distribution that resembles the normal distribution closely. In Figure 5.11 a normal Q-Q plot is shown for the standardized residuals. Most of the residual quantiles are on the plotted line, implying that they are close to the theoretical quantiles. The quantiles only deviate from the line at both the ends, indicating that the distribution for the residuals has heavier tails than the normal distribution (perhaps a student-t distribution would describe it better). The autocorrelation function plot of the standardized residuals are structured, shown in Figure 5.12, also suggest that the residuals are not completely normally distributed and not white noise.

5. Results

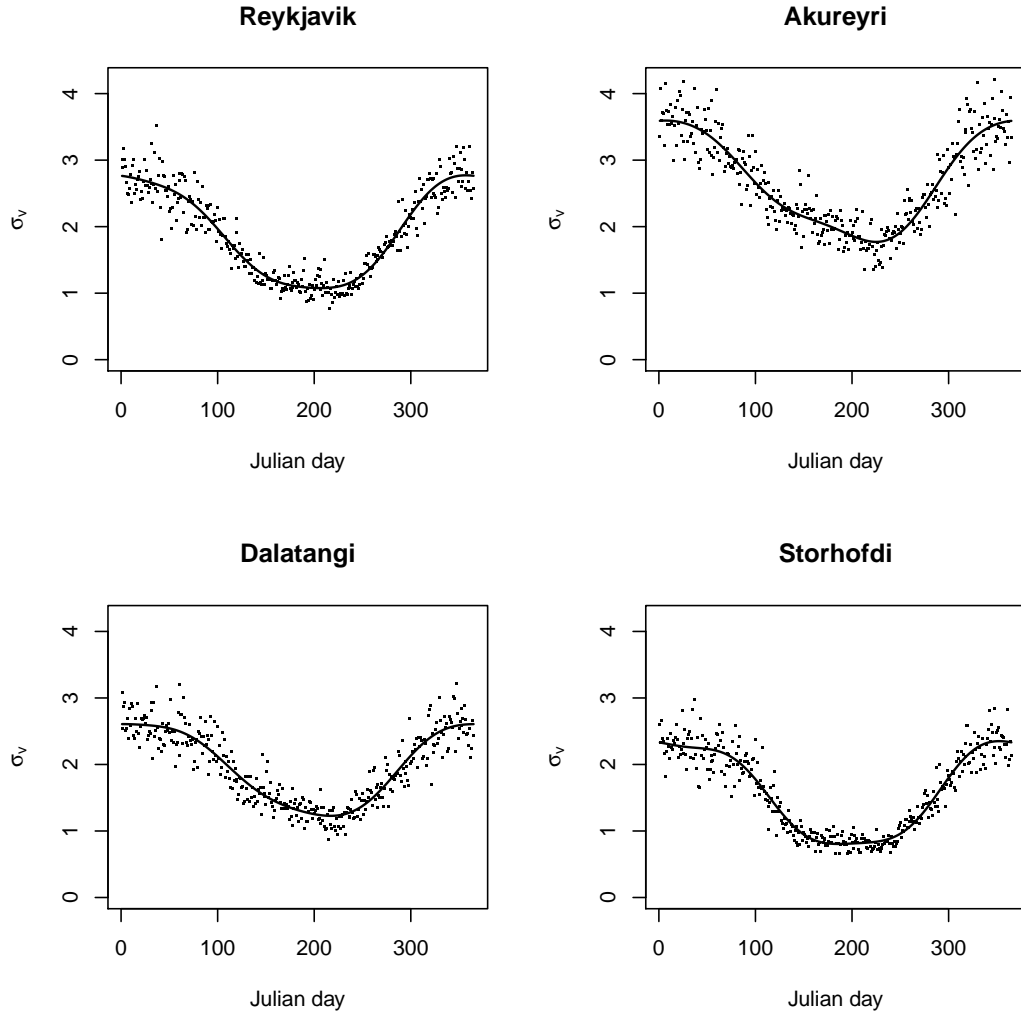


Figure 5.8: The standard deviation of the v_t residuals from Equation (4.8) and a trigonometric fit from Equation (4.9), for the four locations.

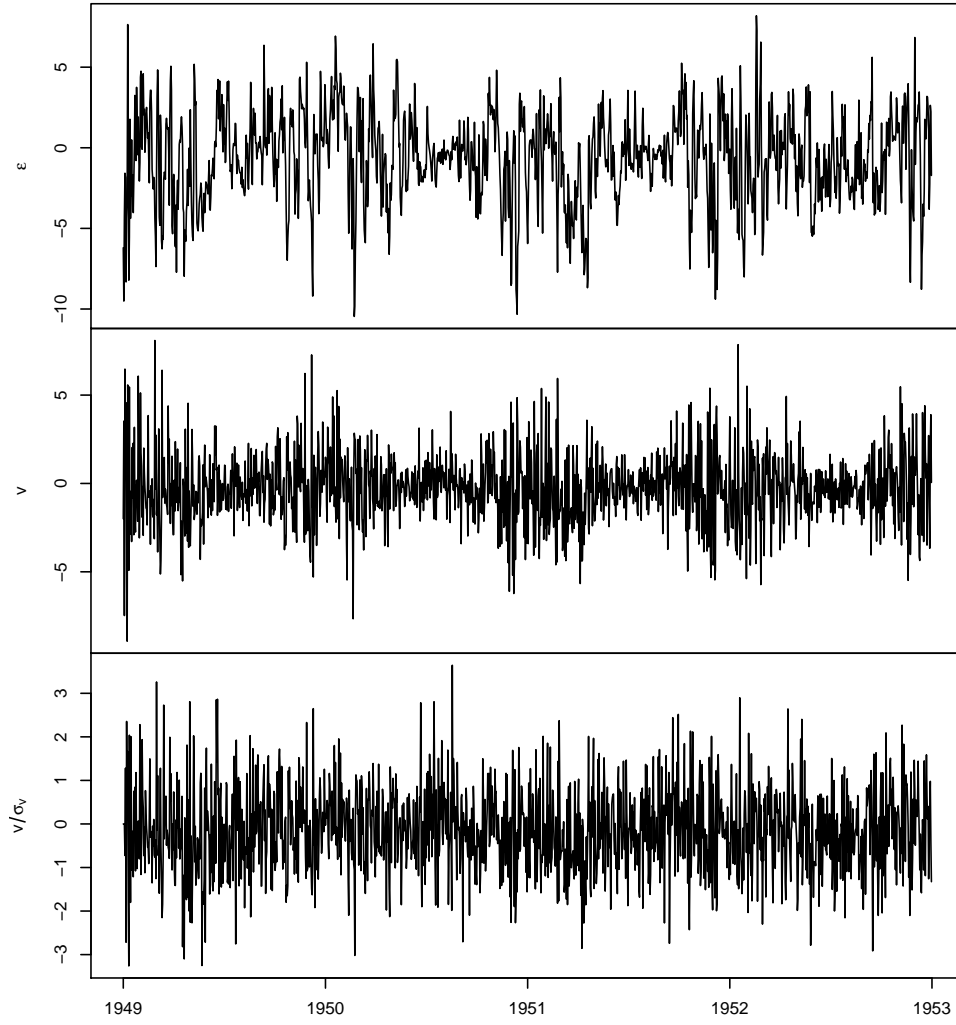


Figure 5.9: Three different residuals for Reykjavik. Top panel, the residuals ϵ_t from the Bayesian model, middle panel, the residuals v_t from the PAR model, and at the bottom panel, the standardized residuals $\frac{v_t}{\sigma_v}$.

5. Results

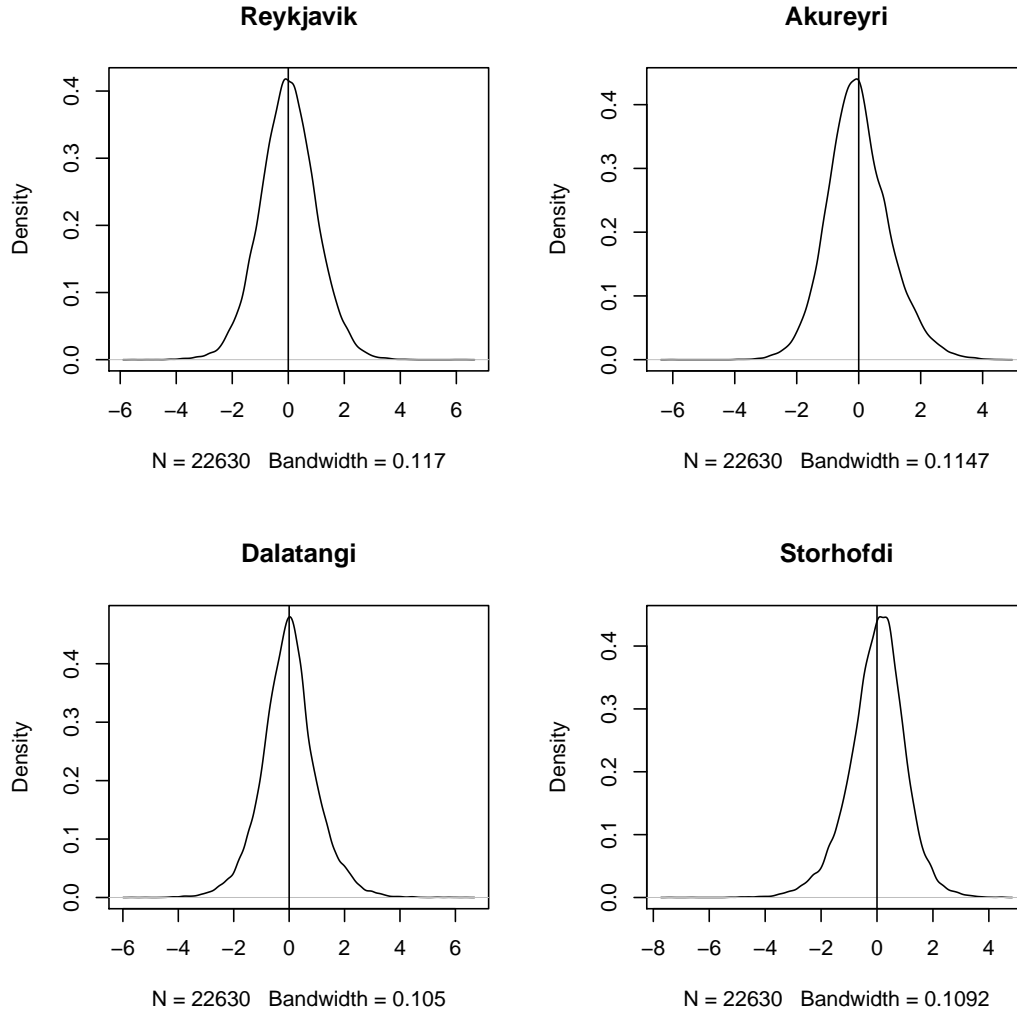


Figure 5.10: The estimated densities of the standardized residuals $\frac{v_t}{\sigma_v}$, for the four locations.

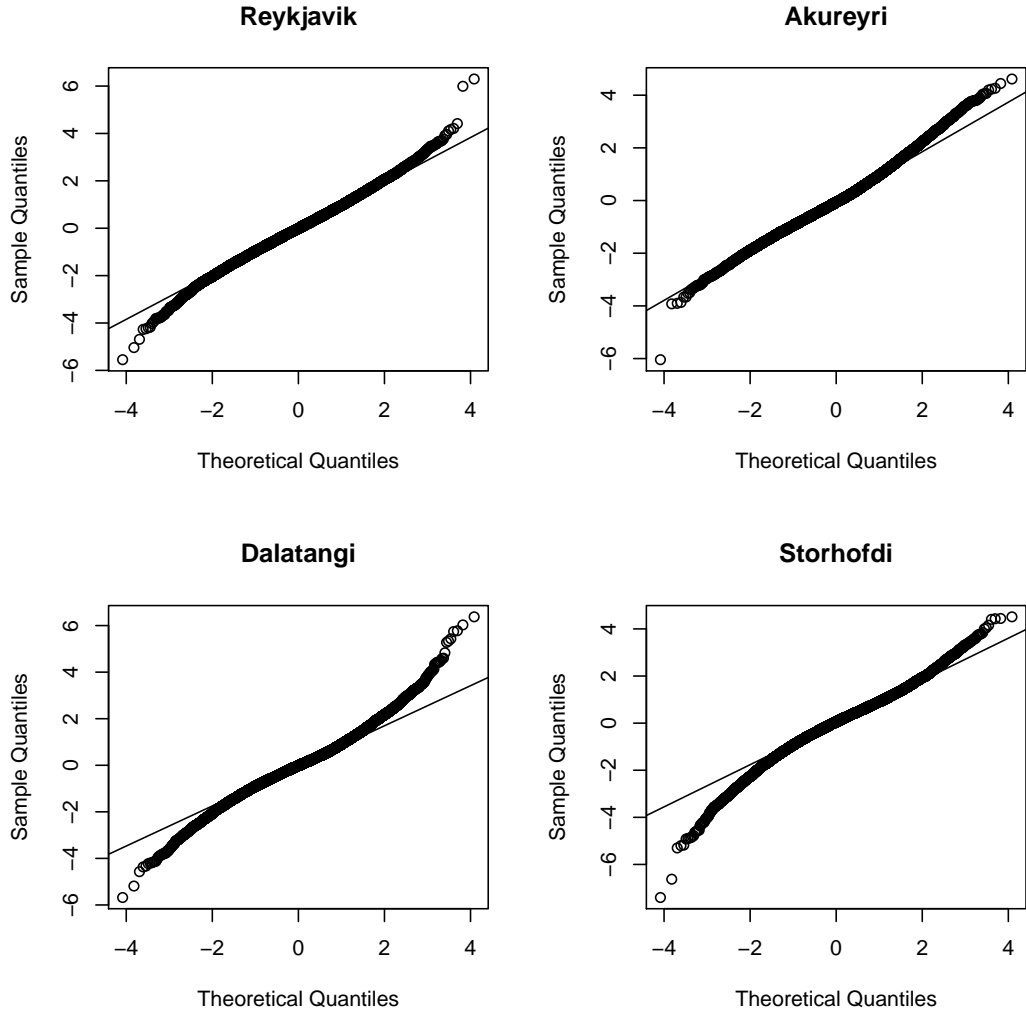


Figure 5.11: The normal Q-Q plot of the standardized residuals $\frac{v_t}{\sigma_v}$, for the four locations.

5. Results

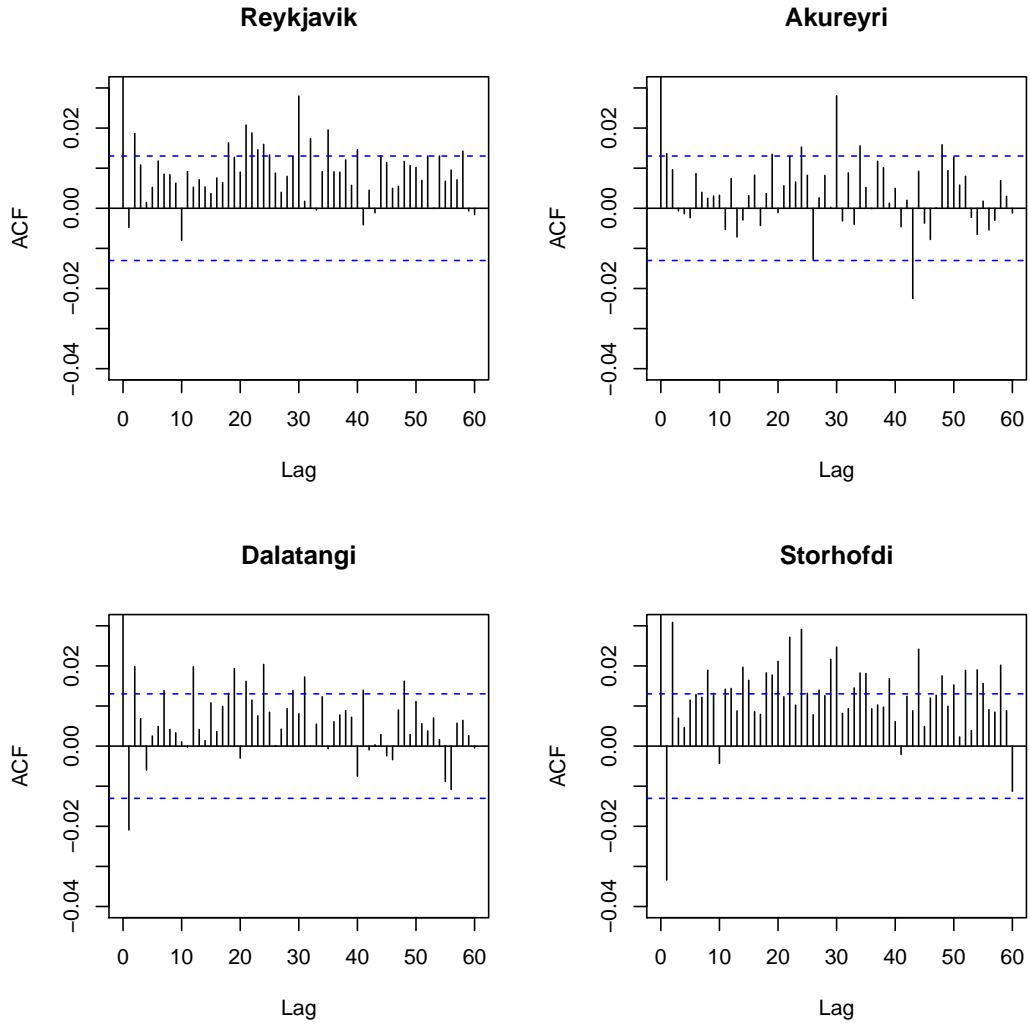


Figure 5.12: The autocorrelation function of the standardized residuals $\frac{v_t}{\sigma_v}$, for the four locations (the value of ACF at lag 0 is 1).

6. Discussion and conclusions

The main goal of this study was to estimate the seasonal factor and long term fluctuation in daily average temperature, as accurately as possible. This was done by developing a Bayesian hierarchical model for the average daily temperature, which takes into account the seasonality, long term fluctuation, temporal correlation and heteroscedasticity. The model is based on normal distribution assumption, resulting in a simple, flexible and computationally efficient model that converges fast.

The seasonal behavior in the data is modeled by giving each calendar day its own parameter and using Markov random field prior distribution to generate dependency structure between the parameters. The same type of prior distribution is used to generate dependency structure for the $\vec{\alpha}$ parameter vector for the B-splines fluctuation. By using this type of prior structure the number of effective parameters in the model is reduced significantly. This results in correlated processes, as seen in Figures 5.2 and 5.3. All other parameters in the model have non-informative prior distributions. It is unlikely that giving these parameters other, more descriptive, prior distributions would result in a better fit.

The analysis of the residuals of the model in Section 5.4 shows that the standardized residuals are close to being white noise and normally distributed, as they were expected to be (although they are not completely white noise, see Figures 5.10 and 5.11). The autocorrelation function of the standardized residuals, in Figure 5.12, also indicates that the residuals are not white noise since the ACF has an apparent structure (nonetheless the values in the ACF are small and mostly within the limits). It could be argued that using a different method to estimate the standard deviation of the PAR model residuals would result in the residuals fitting the normal distribution better and allowing ϕ_1 , ϕ_2 and ϕ_3 in the periodic autoregressive model to vary between calendar days might result in more convincing autocorrelation function for the standardized residuals. This will not be pursued here but left as a future research.

The results in Chapter 5 indicate that the model describes the data well. The seasonal component of the model fits the sample mean of the calendar day reasonably well, see Figure 5.5, and Figure 5.6 shows that the long term fluctuation from the model is close to the long term fluctuation estimated straight from the data using kernel smoothing. The values of the constant of the model (β_0), shown in Table 5.1,

6. Discussion and conclusions

are close to the mean values of the data presented in Table 3.1.

The point estimates of the linear trend γ_0 are between 0.2 and 0.29°C per decade depending on location. This is somewhat consistent with Hrafnkelsson et al. (2012), that estimate a 0.24°C increase per decade for the period 1961 to 2009, and with Jóhannesson et al. (2004) that estimate that the temperature in Iceland will increase by a sinusoidal variation with a maximum of 0.3°C per decade during winter and a minimum of 0.15°C per decade during summer.

7. Future studies

While working on this thesis some ideas about future studies/projects came up. The main ideas are:

- Further development of the model by including a spatial connection between the weather stations.
- Analysis of daily maximum and minimum temperatures based on the Bayesian approach.
- Full Bayesian approach.
- Allow the seasonal effect to change over time.
- Allow the parameters of the periodic autoregressive model to vary between calendar days.
- Link the mean function to smooth outputs from meteorological large scale models which give predictions decades ahead.
- Estimate long term fluctuations in other temperature series, e.g. data created by Crochet and Jóhannesson (2010).

Bibliography

- Akaike, H. (1974, dec). A new look at the statistical model identification. *Automatic Control, IEEE Transactions on* 19(6), 716 – 723.
- Banerjee, S., B. P. Carlin, and A. E. Gelfand (2004). *Hierarchical Modeling and Analysis for Spatial Data*. Chapman & Hall.
- Cox, P., R. Betts, C. Jones, S. Spall, and I. Totterdell (2000). Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. *Nature* 408(6809), 184–187.
- Craigmile, P. F., P. Guttorp, and D. B. Percival (2004). Trend assessment in a long memory dependence model using the discrete wavelet transform. *Environmetrics* 15(4), 313–335.
- Crochet, P. and T. Jóhannesson (2010). A data set of gridded daily temperature in Iceland , 1949 to 2010. pp. 1–17.
- Das, P. (2009). Nonlinear analysis of daily temperature data.
- Gelman, A. (2004). *Bayesian data analysis*. CRC press.
- Gelman, A. and D. B. Rubin (1992). Inference from iterative simulation using multiple sequences. *Statistical Science* 7(4), pp. 457–472.
- Hrafnkelsson, B., J. S. Morris, and V. Baladandayuthapani (2012, December). Spatial modeling of annual minimum and maximum temperatures in Iceland. *Meteorology and Atmospheric Physics* 116(1-2), 43–61.
- Jóhannesson, T., G. Aðalgeirsdóttir, H. Björnsson, C. E. Bøggild, H. Elvehøy, S. Guðmundsson, R. Hock, P. Holmlund, P. Jansson, F. Pálsson, O. Sigurðsson, and . Þorsteinsson (2004). Climate change impacts on glaciers in the nordic countries. *Report by CWE Glaciers group*.
- Jónasson, K. (2005). Spá um meðalhita í Reykjavík á 21. öld (e. prediction for mean temperature in reykjavik on the 21st century, in icelandic). *Raust - Tímarit um raunvísindi og stærðfræði*.
- Jones, R. H. and W. M. Brelsford (1967). Time series with periodic structure. *Biometrika* 54(3-4), 403–408.

BIBLIOGRAPHY

- Király, A. and I. M. János (2002, Apr). Stochastic modeling of daily temperature fluctuations. *Phys. Rev. E* 65, 051102.
- Lindgren, F., H. Rue, and J. Lindström (2011). An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73(4), 423–498.
- Rue, H. and L. Held (2005). *Gaussian Markov random fields: theory and applications*, Volume 104. Chapman & Hall.
- Rue, H. and S. Martino (2007). Approximate bayesian inference for hierarchical gaussian markov random field models. *Journal of Statistical Planning and Inference* 137(10), 3177 – 3192.
- Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. Van Der Linde (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64(4), 583–639.
- Storch, H. v. and F. W. Zwiers (2002). *Statistical analysis in climate research*. Cambridge University Press.
- Tol, R. S. (1996). Autoregressive conditional heteroscedasticity in daily temperature measurements. *Environmetrics* 7(1), 67–75.
- Troutman, B. M. (1979). Some results in periodic autoregression. *Biometrika* 66(2), 219–228.
- Wasserman, L. (2006). *All of nonparametric statistics*. Springer.
- Wikle, C. K. and L. M. Berliner (1998). Hierarchical Bayesian space-time models. *Environmental and Ecological Statistics* 154.

A. Formulations

A.1. Posterior Formulation

The Bayesian hierarchical Gaussian Markov random field model is

$$\begin{aligned}
y|\beta, \beta_0, \gamma_0, \alpha, \sigma^2 &\sim N(X\beta + \beta_0\mathbf{1} + T\gamma_0 + Z\alpha, \sigma^2\Sigma) \\
\beta|\sigma_\beta^2 &\sim N(0, \sigma_\beta^2 P_1^{-1}) \\
\beta_0|\mu_{\beta_0}, \sigma_{\beta_0}^2 &\sim N(\mu_{\beta_0}, \sigma_{\beta_0}^2) \\
\alpha|\sigma_\alpha^2 &\sim N(0, \sigma_\alpha^2 P_2^{-1}) \\
\gamma_0|\mu_\gamma, \sigma_\gamma^2 &\sim N(\mu_\gamma, \sigma_\gamma^2) \\
\sigma^2|V_\sigma, S_\sigma^2 &\sim \text{Inv-}\chi^2(V_\sigma, S_\sigma^2) \\
\sigma_\beta^2|V_{\sigma\beta}, S_{\sigma\beta}^2 &\sim \text{Inv-}\chi^2(V_{\sigma\beta}, S_{\sigma\beta}^2) \\
\sigma_\alpha^2|V_{\sigma\alpha}, S_{\sigma\alpha}^2 &\sim \text{Inv-}\chi^2(V_{\sigma\alpha}, S_{\sigma\alpha}^2)
\end{aligned}$$

where $\Sigma = \mathbf{L}^{-T} \mathbf{W} \mathbf{L}^{-1}$. The posterior distribution of $(\beta, \beta_0, \alpha, \gamma_0, \sigma^2, \sigma_\beta^2, \sigma_\alpha^2)$ is

$$\begin{aligned}
p(\beta, \beta_0, \alpha, \gamma_0, \sigma^2, \sigma_\beta^2, \sigma_\alpha^2|y) &\propto p(y|\beta, \beta_0, \alpha, \gamma_0, \sigma^2) p(\beta, \beta_0, \alpha, \gamma_0, \sigma^2, \sigma_\beta^2, \sigma_\alpha^2) \\
&\propto p(y|\beta, \beta_0, \alpha, \gamma_0, \sigma^2) p(\beta|\sigma_\beta^2) p(\beta_0) p(\alpha|\sigma_\alpha^2) p(\gamma_0) \\
&\quad p(\sigma^2) p(\sigma_\beta^2) p(\sigma_\alpha^2)
\end{aligned}$$

A. Formulations

The conditional distribution of β is

$$\begin{aligned}
p(\beta|\text{rest}) &\propto p(\beta|\sigma_\beta^2) p(y|\beta, \beta_0, \alpha, \gamma_0, \sigma^2) \\
&\propto N(\beta|0, \sigma_\beta^2 P^{-1}) N(X\beta + \beta_0 \vec{1} + T\gamma_0 + Z\alpha, \sigma^2 \Sigma) \\
&\propto \exp \left[-\frac{1}{2} \beta^T \sigma_\beta^{-2} P \beta \right] \cdot \\
&\quad \exp \left[-\frac{1}{2} (y - X\beta - \beta_0 \vec{1} - T\gamma_0 - Z\alpha)^T \sigma^{-2} \Sigma^{-1} (y - X\beta - \beta_0 \vec{1} - T\gamma_0 - Z\alpha) \right] \\
&\propto \exp \left[-\frac{1}{2} (\beta^T \sigma_\beta^{-2} P \beta - 2\beta^T X^T \sigma^{-2} \Sigma^{-1} y + \beta^T X^T \sigma^{-2} \Sigma^{-1} X \beta \right. \\
&\quad \left. + 2\beta^T X^T \sigma^{-2} \Sigma^{-1} \beta_0 \vec{1} + 2\beta^T X^T \sigma^{-2} \Sigma^{-1} T\gamma_0 + 2\beta^T X^T \sigma^{-2} \Sigma^{-1} Z\alpha) \right] \\
&\propto \exp \left[-\frac{1}{2} (\beta^T (\sigma_\beta^{-2} P + X^T \sigma^{-2} \Sigma^{-1} X) \beta - 2\beta^T (X^T \sigma^{-2} \Sigma^{-1} y \right. \\
&\quad \left. - X^T \sigma^{-2} \Sigma^{-1} \beta_0 \vec{1} - X^T \sigma^{-2} \Sigma^{-1} T\gamma_0 - X^T \sigma^{-2} \Sigma^{-1} Z\alpha)) \right]
\end{aligned}$$

This is a normal distribution with covariance matrix and mean

$$\begin{aligned}
\Sigma_{\beta, \text{post}} &= (\sigma_\beta^{-2} P + \sigma^{-2} X^T \Sigma^{-1} X)^{-1} \\
\mu_{\beta, \text{post}} &= \Sigma_{\beta, \text{post}} \cdot \sigma^{-2} X^T \Sigma^{-1} (y - \beta_0 \vec{1} - T\gamma_0 - Z\alpha).
\end{aligned}$$

The conditional distribution of β_0 is

$$\begin{aligned}
p(\beta_0|\text{rest}) &\propto p(\beta_0) p(y|\beta, \beta_0, \alpha, \gamma_0, \sigma^2) \\
&\propto \exp \left[-\frac{1}{2} \sigma_{\beta_0}^{-2} (\beta_0 - \mu_{\beta_0})^2 \right] \cdot \\
&\quad \exp \left[-\frac{1}{2} (y - X\beta - \beta_0 \vec{1} - T\gamma_0 - Z\alpha)^T \sigma^{-2} \Sigma^{-1} (y - X\beta - \beta_0 \vec{1} - T\gamma_0 - Z\alpha) \right] \\
&\propto \exp \left[-\frac{1}{2} (\sigma_{\beta_0}^{-2} \beta_0^2 - 2\beta_0 \sigma_{\beta_0}^{-2} \mu_{\beta_0} - 2\beta_0 \vec{1}^T \sigma^{-2} \Sigma^{-1} y + 2\beta_0 \vec{1}^T \sigma^{-2} \Sigma^{-1} X\beta \right. \\
&\quad \left. + \beta_0 \vec{1}^T \sigma^{-2} \Sigma^{-1} \beta_0 \vec{1} + 2\beta_0 \vec{1}^T \sigma^{-2} \Sigma^{-1} T\gamma_0 + 2\beta_0 \vec{1}^T \sigma^{-2} \Sigma^{-1} Z\alpha) \right] \\
&\propto \exp \left[-\frac{1}{2} (\beta_0^2 (\sigma_{\beta_0}^{-2} + \vec{1}^T \sigma^{-2} \Sigma^{-1} \vec{1}) \right. \\
&\quad \left. - 2\beta_0 (\sigma_{\beta_0}^{-2} \mu_{\beta_0} + \sigma^{-2} (\vec{1}^T \Sigma^{-1} y - \vec{1}^T \Sigma^{-1} X\beta - \vec{1}^T \Sigma^{-1} T\gamma_0 - \vec{1}^T \Sigma^{-1} Z\alpha))) \right]
\end{aligned}$$

This is a normal distribution with covariance matrix and mean

$$\begin{aligned}
\Sigma_{\beta_0, \text{post}} &= (\sigma_{\beta_0}^{-2} + \vec{1}^T \sigma^{-2} \Sigma^{-1} \vec{1})^{-1} \\
\mu_{\beta_0, \text{post}} &= \Sigma_{\beta_0, \text{post}} \cdot (\sigma_{\beta_0}^{-2} \mu_{\beta_0} + \sigma^{-2} \vec{1}^T \Sigma^{-1} (y - X\beta - T\gamma_0 - Z\alpha)).
\end{aligned}$$

The conditional distribution of α is

$$\begin{aligned}
 p(\alpha|\text{rest}) &\propto p(\alpha|\sigma_\alpha^2) p(y|\beta, \beta_0, \alpha, \gamma_0, \sigma^2) \\
 &\propto \exp \left[-\frac{1}{2} \alpha^T \sigma_\alpha^{-2} P_2 \alpha \right] \cdot \\
 &\quad \exp \left[-\frac{1}{2} (y - X\beta - \beta_0 \vec{1} - T\gamma_0 - Z\alpha)^T \sigma^{-2} \Sigma^{-1} (y - X\beta - \beta_0 \vec{1} - T\gamma_0 - Z\alpha) \right] \\
 &\propto \exp \left[-\frac{1}{2} (\alpha^T \sigma_\alpha^{-2} P_2 \alpha - 2\alpha^T Z^T \sigma^{-2} \Sigma^{-1} y + 2\alpha^T Z^T \sigma^{-2} \Sigma^{-1} X\beta \right. \\
 &\quad \left. + 2\alpha^T Z^T \sigma^{-2} \Sigma^{-1} \beta_0 \vec{1} + 2\alpha^T Z^T \sigma^{-2} \Sigma^{-1} T\gamma_0 + \alpha^T Z^T \sigma^{-2} \Sigma^{-1} Z\alpha) \right] \\
 &\propto \exp \left[-\frac{1}{2} (\alpha^T (\sigma_\alpha^{-2} P_2 + Z^T \sigma^{-2} \Sigma^{-1} Z) \alpha \right. \\
 &\quad \left. - 2\alpha^T (Z^T \sigma^{-2} \Sigma^{-1} y - Z^T \sigma^{-2} \Sigma^{-1} X\beta - Z^T \sigma^{-2} \Sigma^{-1} \beta_0 \vec{1} - Z^T \sigma^{-2} \Sigma^{-1} T\gamma_0)) \right]
 \end{aligned}$$

This is a normal distribution with covariance matrix and mean

$$\begin{aligned}
 \Sigma_{\alpha, \text{post}} &= (\sigma_\alpha^{-2} P_2 + \sigma^{-2} Z^T \Sigma^{-1} Z)^{-1} \\
 \mu_{\alpha, \text{post}} &= \Sigma_{\alpha, \text{post}} \cdot \sigma^{-2} Z^T \Sigma^{-1} (y - X\beta - \beta_0 \vec{1} - \gamma_0 T).
 \end{aligned}$$

The conditional distribution of γ_0 is

$$\begin{aligned}
 p(\gamma_0|\text{rest}) &\propto p(\gamma_0|\mu_\gamma, \sigma_\gamma^2) p(y|\beta, \beta_0, \alpha, \gamma_0, \sigma^2) \\
 &\propto \exp \left[-\frac{1}{2} \sigma_\gamma^{-2} (\gamma_0 - \mu_\gamma)^2 \right] \cdot \\
 &\quad \exp \left[-\frac{1}{2} (y - X\beta - \beta_0 \vec{1} - T\gamma_0 - Z\alpha)^T \sigma^{-2} \Sigma^{-1} (y - X\beta - \beta_0 \vec{1} - T\gamma_0 - Z\alpha) \right] \\
 &\propto \exp \left[-\frac{1}{2} (\gamma_0^2 \sigma_\gamma^{-2} - 2\gamma_0 \sigma_\gamma^{-2} \mu_\gamma - 2\gamma_0 T^T \sigma^{-2} \Sigma^{-1} y + 2\gamma_0 T^T \sigma^{-2} \Sigma^{-1} X\beta \right. \\
 &\quad \left. + 2\gamma_0 T^T \sigma^{-2} \Sigma^{-1} \beta_0 \vec{1} + \gamma_0 T^T \sigma^{-2} \Sigma^{-1} T\gamma_0 + 2\gamma_0 T^T \sigma^{-2} \Sigma^{-1} Z\alpha) \right] \\
 &\propto \exp \left[-\frac{1}{2} (\gamma_0^2 (\sigma_\gamma^{-2} + T^T \sigma^{-2} \Sigma^{-1} T) \right. \\
 &\quad \left. - 2\gamma_0 (\sigma_\gamma^{-2} \mu_\gamma + T^T \sigma^{-2} \Sigma^{-1} (y - X\beta - \beta_0 \vec{1} - Z\alpha))) \right]
 \end{aligned}$$

This is a normal distribution with covariance matrix and mean

$$\begin{aligned}
 \Sigma_{\gamma_0, \text{post}} &= (\sigma_\gamma^{-2} + \sigma^{-2} T^T \Sigma^{-1} T)^{-1} \\
 \mu_{\gamma_0, \text{post}} &= \Sigma_{\gamma_0, \text{post}} \cdot \left(\sigma_\gamma^{-2} \mu_\gamma + \sigma^{-2} T^T \Sigma^{-1} (y - X\beta - \beta_0 \vec{1} - Z\alpha) \right).
 \end{aligned}$$

A. Formulations

The conditional distribution of σ^2 is

$$\begin{aligned}
p(\sigma^2|\text{rest}) &\propto p(\sigma^2) p(y|\beta, \beta_0, \alpha, \gamma_0, \sigma^2) \\
&\propto \text{inv-}\chi^2(\sigma^2|V_\sigma, S_\sigma^2) N(X\beta + \beta_0\vec{1} + T\gamma_0 + Z\alpha, \sigma^2\Sigma) \\
&\propto (\sigma^2)^{-(V_\sigma/2+1)} \exp\left[\frac{-V_\sigma S_\sigma^2}{2\sigma^2}\right] |\sigma^2\Sigma|^{-1/2} \\
&\quad \exp\left[-\frac{1}{2}(y - X\beta - \beta_0\vec{1} - T\gamma_0 - Z\alpha)^T \sigma^{-2} \Sigma^{-1} (y - X\beta - \beta_0\vec{1} - T\gamma_0 - Z\alpha)\right] \\
&\propto (\sigma^2)^{-((V_\sigma+n)/2+1)} \cdot \\
&\quad \exp\left[-\frac{1}{2\sigma^2} \left(V_\sigma S_\sigma^2 + (y - X\beta - \beta_0\vec{1} - T\gamma_0 - Z\alpha)^T \Sigma^{-1} (y - X\beta - \beta_0\vec{1} - T\gamma_0 - Z\alpha)\right)\right] \\
&\quad \Sigma^{-1} (y - X\beta - \beta_0\vec{1} - T\gamma_0 - Z\alpha)]
\end{aligned}$$

This is a scaled inverse- χ^2 distribution with degrees of freedom and scale

$$\begin{aligned}
V_{\sigma, \text{post}} &= V_\sigma + n \\
S_{\sigma, \text{post}}^2 &= \frac{V_\sigma S_\sigma^2 + (y - X\beta - \beta_0\vec{1} - T\gamma_0 - Z\alpha)^T \Sigma^{-1} (y - X\beta - \beta_0\vec{1} - T\gamma_0 - Z\alpha)}{V_\sigma + n}.
\end{aligned}$$

The conditional distribution of σ_β^2 is

$$\begin{aligned}
p(\sigma_\beta^2|\text{rest}) &\propto p(\sigma_\beta^2) p(\beta|\sigma_\beta^2) \\
&\propto \text{inv-}\chi^2(\sigma_\beta^2|V_{\sigma\beta}, S_{\sigma\beta}^2) N(\beta|0, \sigma_\beta^2 P^{-1}) \\
&\propto (\sigma_\beta^2)^{-(V_{\sigma\beta}/2+1)} \exp\left[\frac{-V_{\sigma\beta} S_{\sigma\beta}^2}{2\sigma_\beta^2}\right] \exp\left[-\frac{1}{2}\beta^T \sigma_\beta^{-2} P\beta\right] |\sigma_\beta^2 P^{-1}|^{-1/2} \\
&\propto (\sigma_\beta^2)^{-((V_{\sigma\beta}+k)/2+1)} \exp\left[-\frac{1}{2\sigma_\beta^2} (V_{\sigma\beta} S_{\sigma\beta}^2 + \beta^T P\beta)\right]
\end{aligned}$$

This is a scaled inverse- χ^2 distribution with degrees of freedom and scale

$$\begin{aligned}
V_{\sigma\beta, \text{post}} &= V_{\sigma\beta} + k \\
S_{\sigma\beta, \text{post}}^2 &= \frac{V_{\sigma\beta} S_{\sigma\beta}^2 + \beta^T P\beta}{V_{\sigma\beta} + k}
\end{aligned}$$

The conditional distribution of σ_α^2

$$\begin{aligned}
P(\sigma_\alpha^2 | \text{rest}) &\propto P(\sigma_\alpha^2 | V_{\sigma\alpha}, S_{\sigma\alpha}^2) p(\alpha | \sigma_\alpha^2) \\
&\propto (\sigma_\alpha^2)^{-(\frac{V_{\sigma\alpha}}{2} + 1)} \exp \left[\frac{-V_{\sigma\alpha} S_{\sigma\alpha}^2}{2\sigma_\alpha^2} \right] \exp \left[-\frac{1}{2} \alpha^T \sigma_\alpha^{-2} P_2 \alpha \right] |\sigma_\alpha^2 P_2^{-1}|^{-1/2} \\
&\propto (\sigma_\alpha^2)^{-(\frac{V_{\sigma\alpha} + m}{2} + 1)} \exp \left[-\frac{1}{2\sigma_\alpha^2} (V_{\sigma\alpha} S_{\sigma\alpha}^2 + \alpha^T P_2 \alpha) \right]
\end{aligned}$$

This is a scaled inverse- χ^2 distribution with degrees of freedom and scale

$$\begin{aligned}
V_{\sigma\alpha, \text{post}} &= V_{\sigma\alpha} + m \\
S_{\sigma\alpha, \text{post}}^2 &= \frac{V_{\sigma\alpha} S_{\sigma\alpha}^2 + \alpha^T P_2 \alpha}{V_{\sigma\alpha} + m}
\end{aligned}$$

A.2. DIC formulation

To simplify the calculations let's introduce a new variable \vec{u} and \vec{v} , as the vectors

$$\vec{u} = (\vec{y} - \mathbf{X}\vec{\beta} - \hat{\beta}_0\vec{1} - \gamma_0\vec{T} - \mathbf{Z}\vec{\alpha})$$

and

$$\vec{v} = (\vec{y} - \mathbf{X}\vec{\bar{\beta}} - \bar{\beta}_0\vec{1} - \bar{\gamma}_0\vec{T} - \mathbf{Z}\vec{\bar{\alpha}})$$

where $\vec{\bar{\beta}}, \bar{\beta}_0, \bar{\gamma}_0$ and $\vec{\bar{\alpha}}$ are posterior mean of $\vec{\beta}, \beta_0, \gamma_0$ and $\vec{\alpha}$.

$$\begin{aligned} D(y, \beta, \beta_0, \gamma_0, \alpha, \sigma^2) &= -2 \log [p(y|\beta, \beta_0, \gamma_0, \alpha, \sigma^2)] \\ &= -2 \log [(2\pi)^{-\frac{mn}{2}} |\sigma^2 \Sigma|^{-\frac{1}{2}} \exp(-\frac{1}{2} \vec{u}^T \Sigma^{-1} \vec{u})] \\ &= \vec{u}^T \Sigma^{-1} \vec{u} + \log(|\Sigma|) + mn \log(2\pi) + mn \log(\sigma^2) \end{aligned}$$

$$D_{\hat{\theta}}(y) = \vec{v}^T \Sigma \vec{v} + \log(|\Sigma|) + mn \log(2\pi) + mn \log(\sigma^2)$$

$$D_{avg}(y) = \frac{1}{L} \sum_{l=1}^L D(y, \theta^l) = \frac{1}{L} \sum_{l=1}^L D(y, \beta^l, \beta_0^l, \gamma_0^l, \alpha^l, (\sigma^2)^l)$$

$$\begin{aligned} p_D &= D_{avg}(y) - D_{\hat{\theta}}(y) \\ &= \frac{1}{L} \sum_{l=1}^L \vec{u}^T \Sigma^{-1} \vec{u} - \vec{v}^T \Sigma \vec{v} \end{aligned}$$

$$\begin{aligned} \text{DIC} &= 2D_{avg}(y) - D_{\hat{\theta}}(y) \\ &= \frac{2}{L} \sum_{l=1}^L (\vec{u}^l)^T \Sigma^{-1} \vec{u}^l + \vec{v}^T \Sigma^{-1} \vec{v} + \log(|\Sigma|) + mn \log(2\pi) + mn \log(\sigma^2) \end{aligned}$$

A.3. The error term ϵ

Assume that $\vec{\epsilon}$ can be described by

$$\vec{\epsilon} \sim N(0, \sigma^2 \mathbf{L}^{-T} \mathbf{W} \mathbf{L}^{-1})$$

where \mathbf{L} is a matrix containing the AR parameters ϕ_1, ϕ_2, ϕ_3 and \mathbf{P} is a matrix containing variance information from (4.9).

The conditional distribution of $\vec{\epsilon}$ is

$$\begin{aligned} p(\vec{\epsilon} | \text{rest}) &\propto p([\epsilon_1, \epsilon_2, \epsilon_3]) \prod_{t=4}^{n \cdot m} p(\epsilon_t | \epsilon_{t-1}, \epsilon_{t-2}, \epsilon_{t-3}) \\ &\propto p([\epsilon_1, \epsilon_2, \epsilon_3]) \prod_{t=4}^{n \cdot m} N(\epsilon_t | \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + \phi_3 \epsilon_{t-3}, \sigma_v^2(t)) \\ &\propto p([\epsilon_1, \epsilon_2, \epsilon_3]) \prod_{t=4}^{n \cdot m} \frac{1}{\sqrt{2\pi\sigma_v^2(t)}} \exp \left\{ -\frac{1}{2\sigma_v^2(t)} (\epsilon_t - \phi_1 \epsilon_{t-1} - \phi_2 \epsilon_{t-2} - \phi_3 \epsilon_{t-3})^2 \right\} \\ &\propto \prod_{v=1}^n \frac{1}{\sigma_v^2} \exp \left\{ -\frac{1}{2} \vec{\epsilon}^T \mathbf{L}^T \mathbf{W}^{-1} \mathbf{L} \vec{\epsilon} \right\} \end{aligned}$$

From viewing the last to lines in the conditional distribution for $\vec{\epsilon}$ the matrices \mathbf{L} and \mathbf{W} can be written as

$$L_{i,j} = \begin{cases} 1 & \text{for } (i, j = i) \\ -\phi_1 & \text{for } (i, j = i + 1) \\ -\phi_2 & \text{for } (i, j = i + 2) \\ -\phi_3 & \text{for } (i, j = i + 3) \\ 0 & \text{Otherwise} \end{cases} \quad (\text{A.1})$$

and

$$\mathbf{W} = \text{diag}(\vec{1} \otimes \vec{\sigma}_v^2).$$

B. Figures

Histograms and MCMC trace plots for some parameters (see discussion in Section 5.1.1).

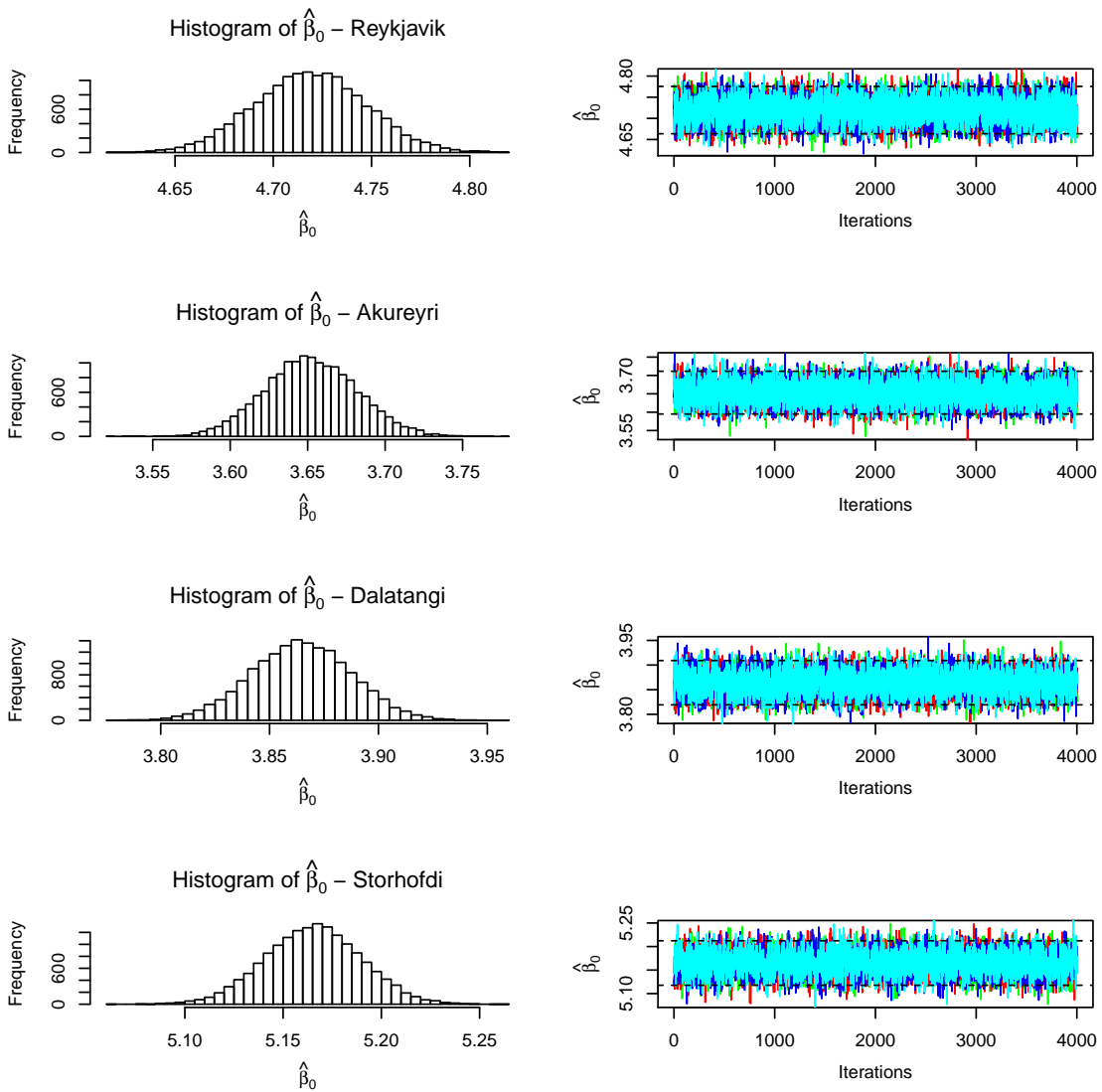


Figure B.1: Histograms and MCMC trace plots after burn-in for the parameter β_0 .

B. Figures

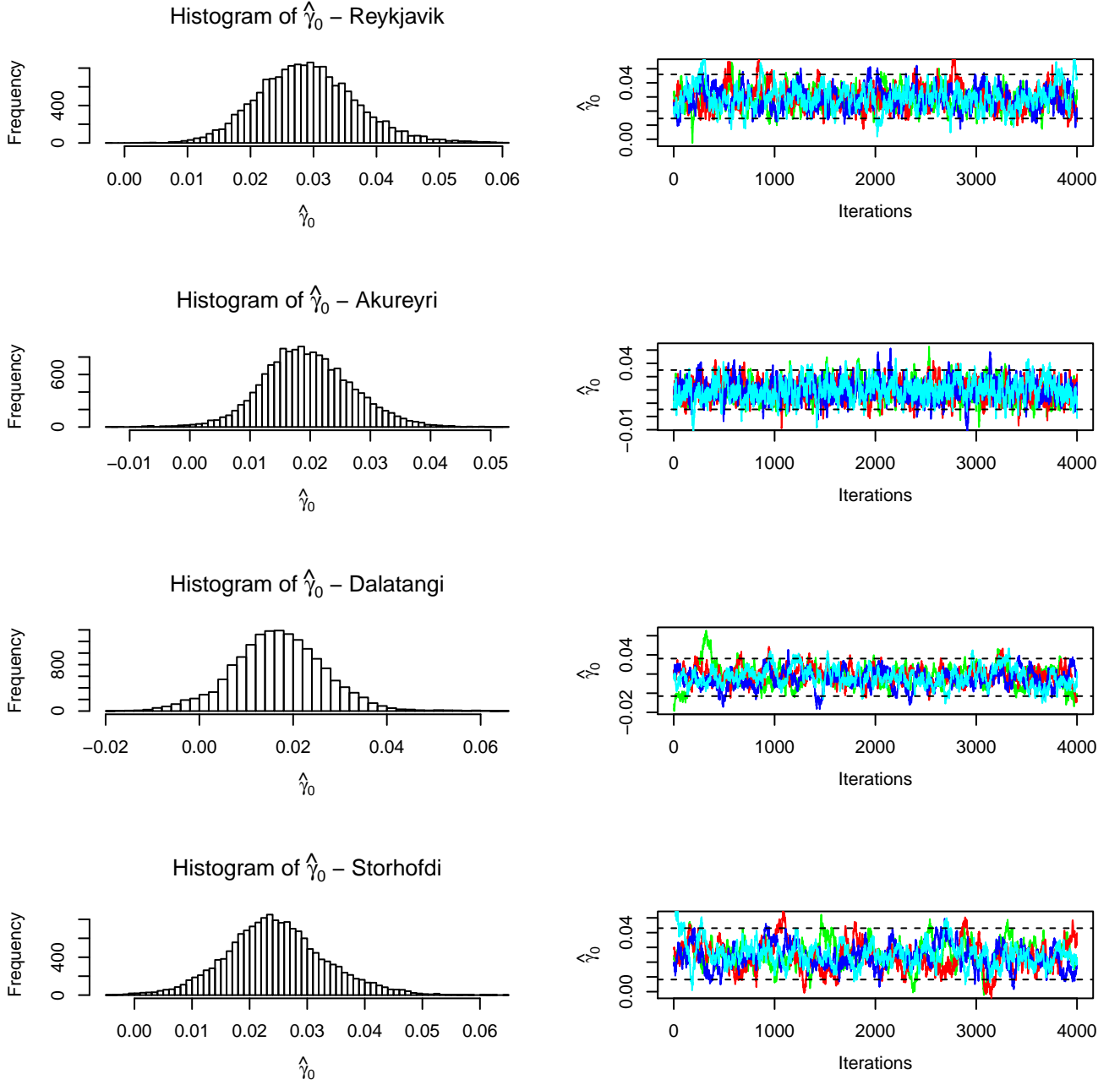


Figure B.2: Histograms and MCMC trace plots after burn-in for the parameter γ_0 .

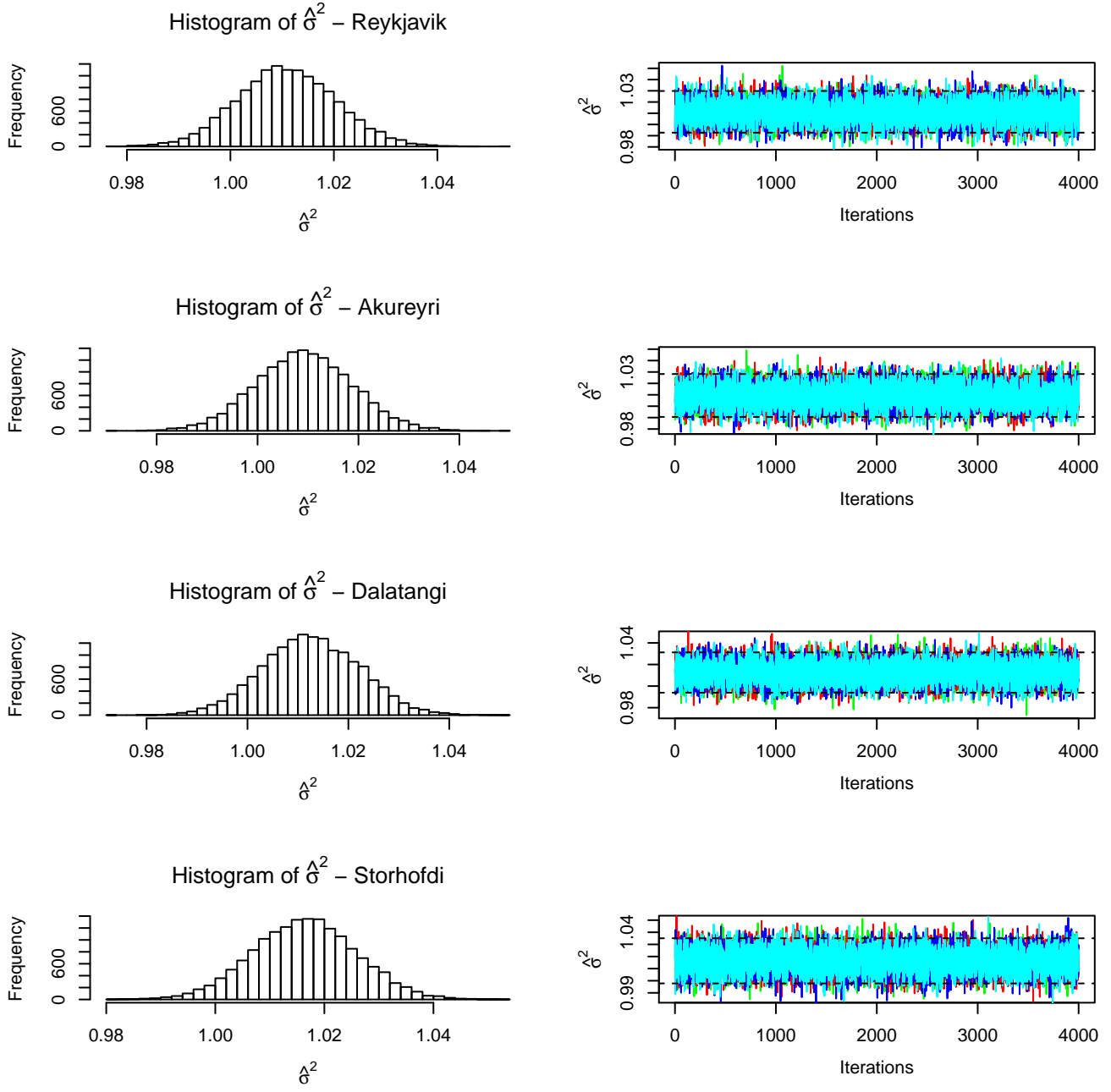


Figure B.3: Histograms and MCMC trace plots after burn-in for the parameter σ^2 .

B. Figures

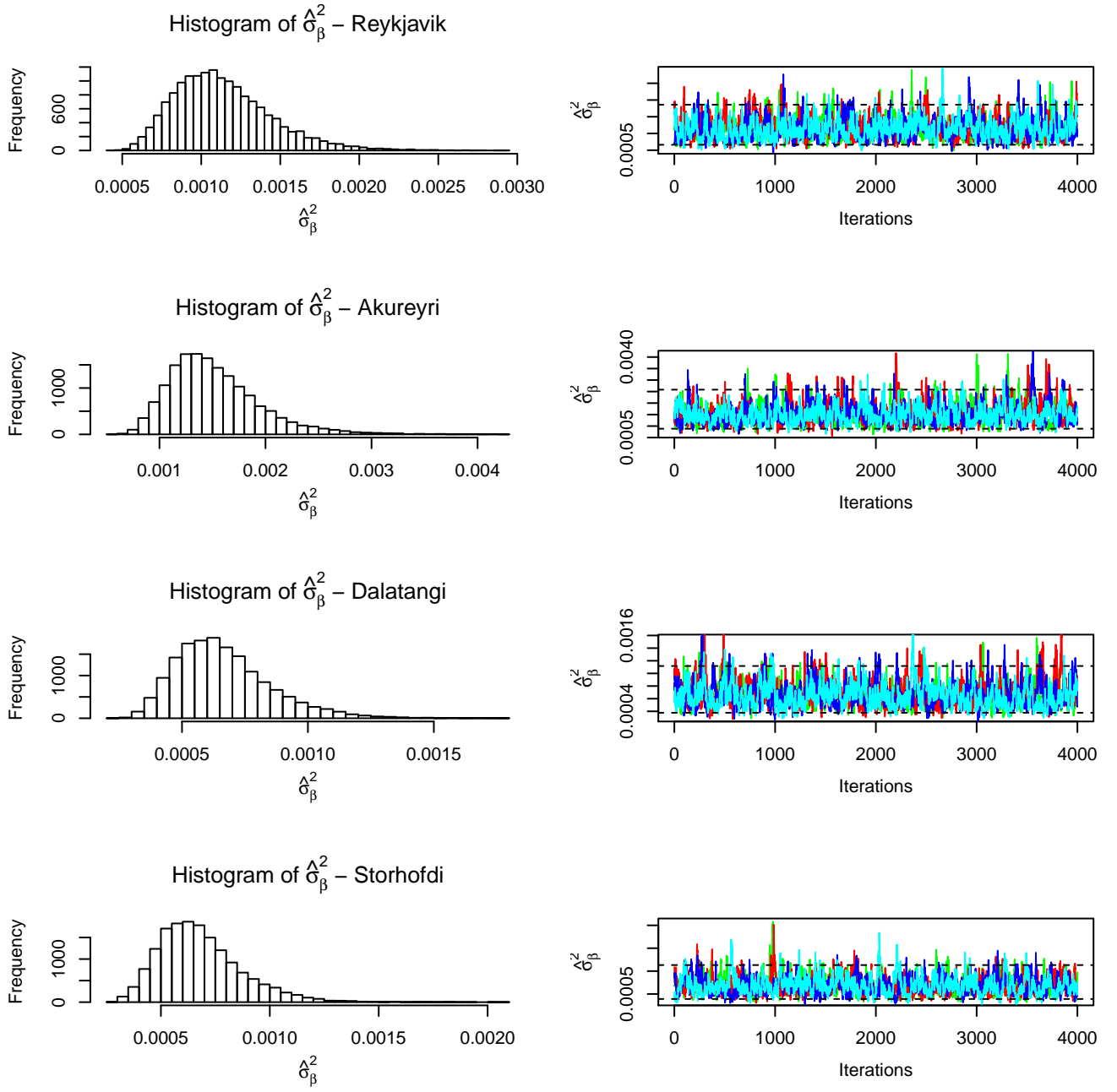


Figure B.4: Histograms and MCMC trace plots after burn-in for the parameter σ_β^2 .

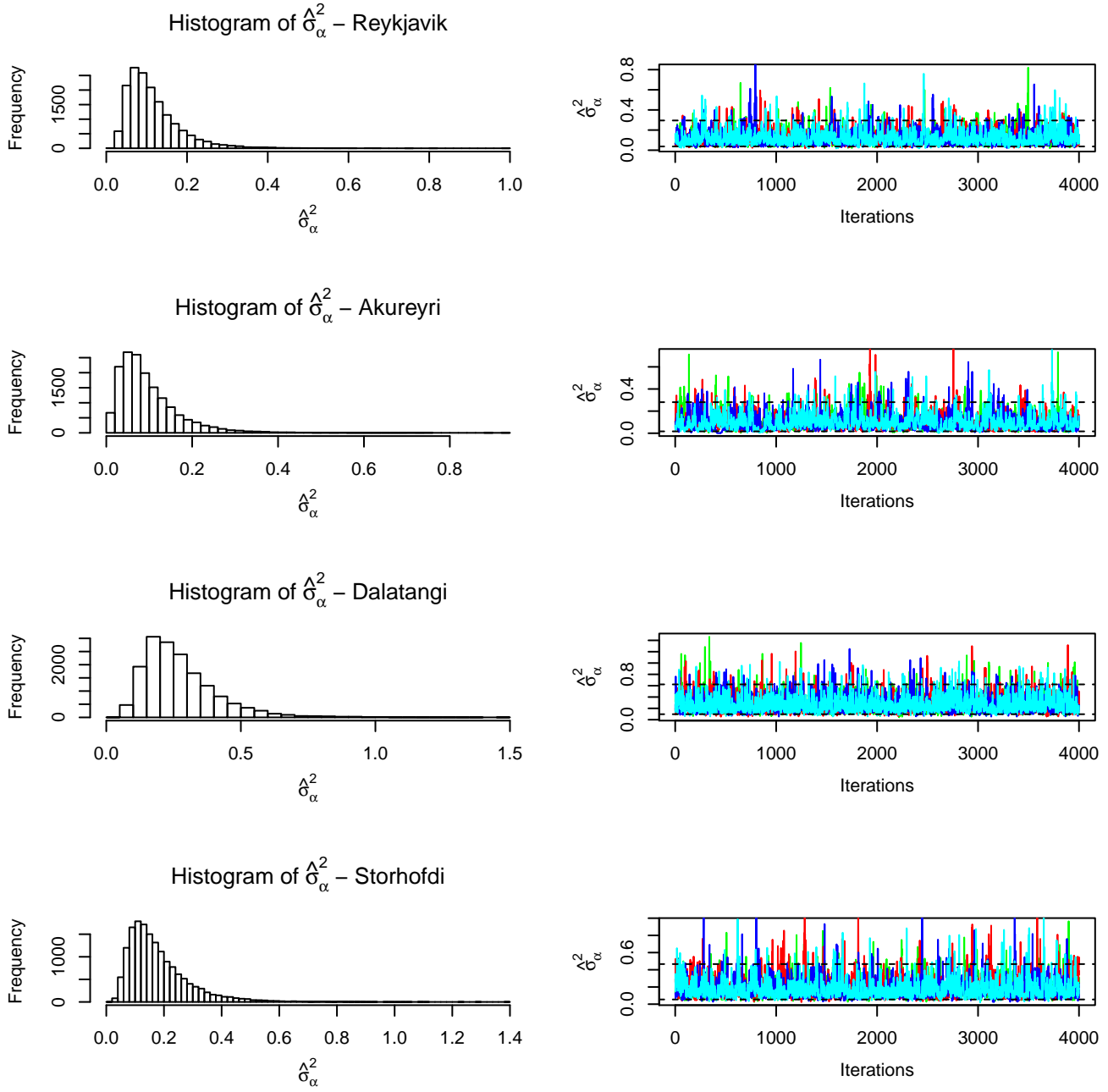


Figure B.5: Histograms and MCMC trace plots after burn-in for the parameter σ_α^2 .

B. Figures

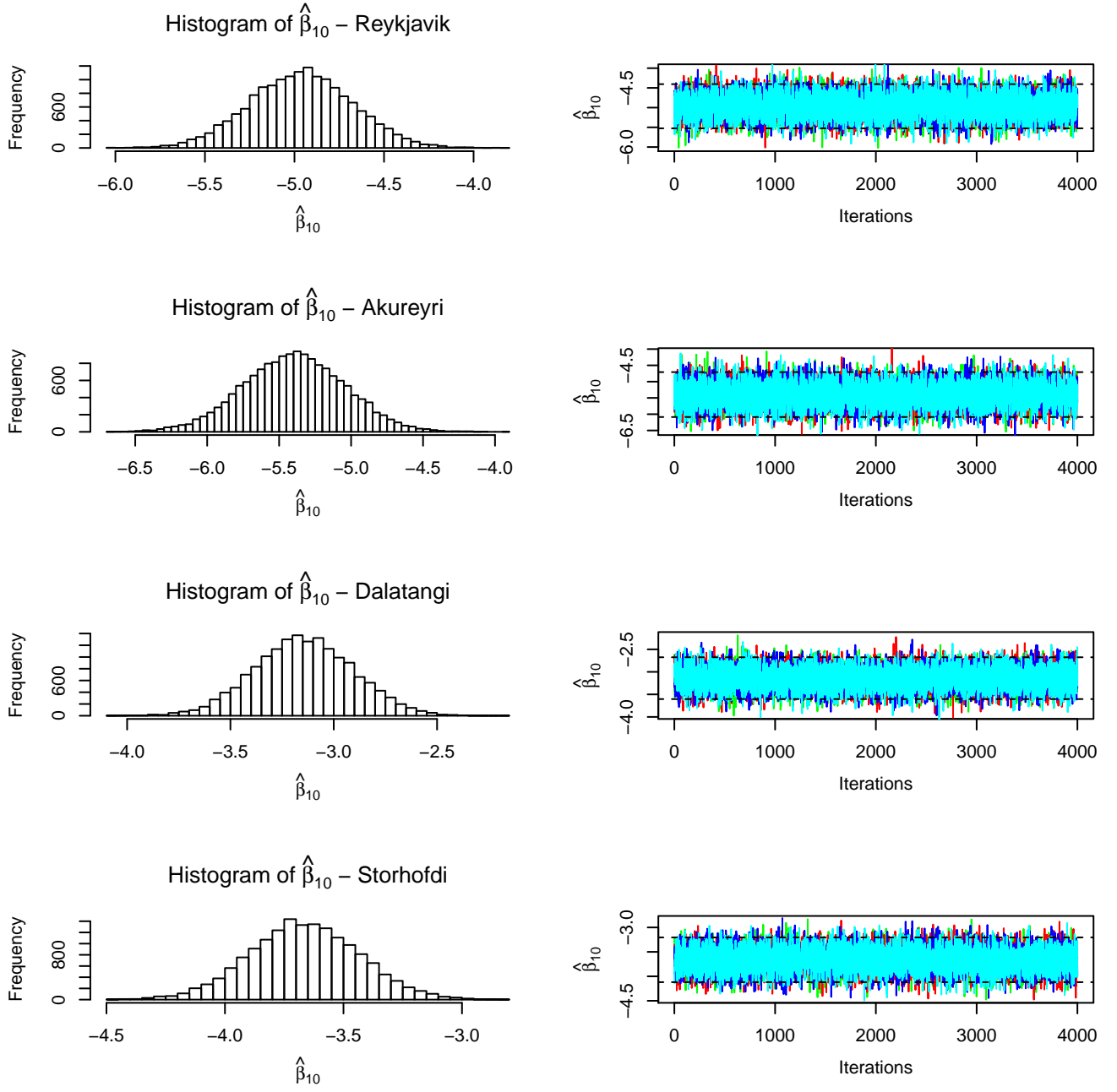


Figure B.6: Histograms and MCMC trace plots after burn-in for the parameter β_{10} .

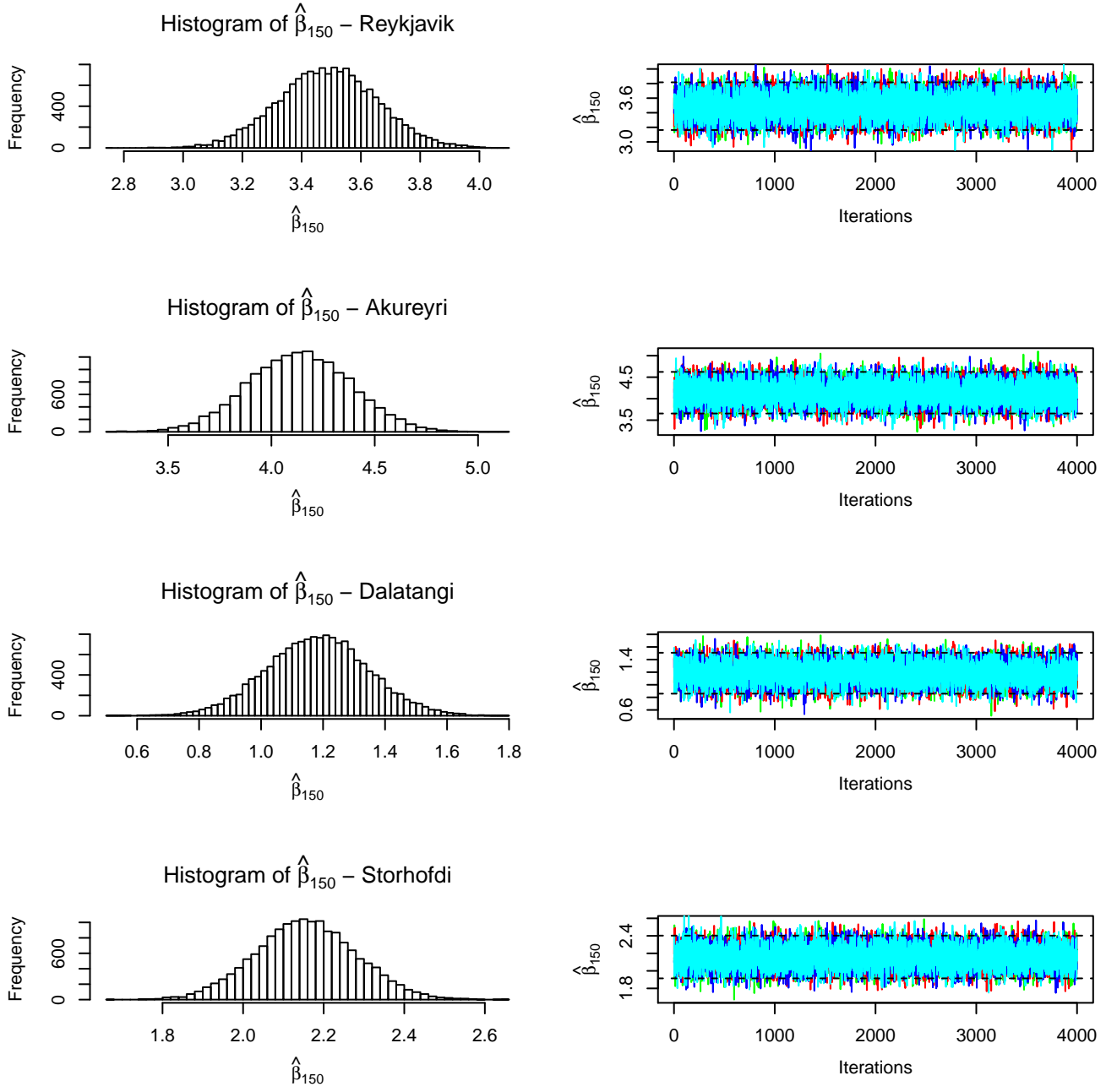


Figure B.7: Histograms and MCMC trace plots after burn-in for the parameter β_{150} .

B. Figures

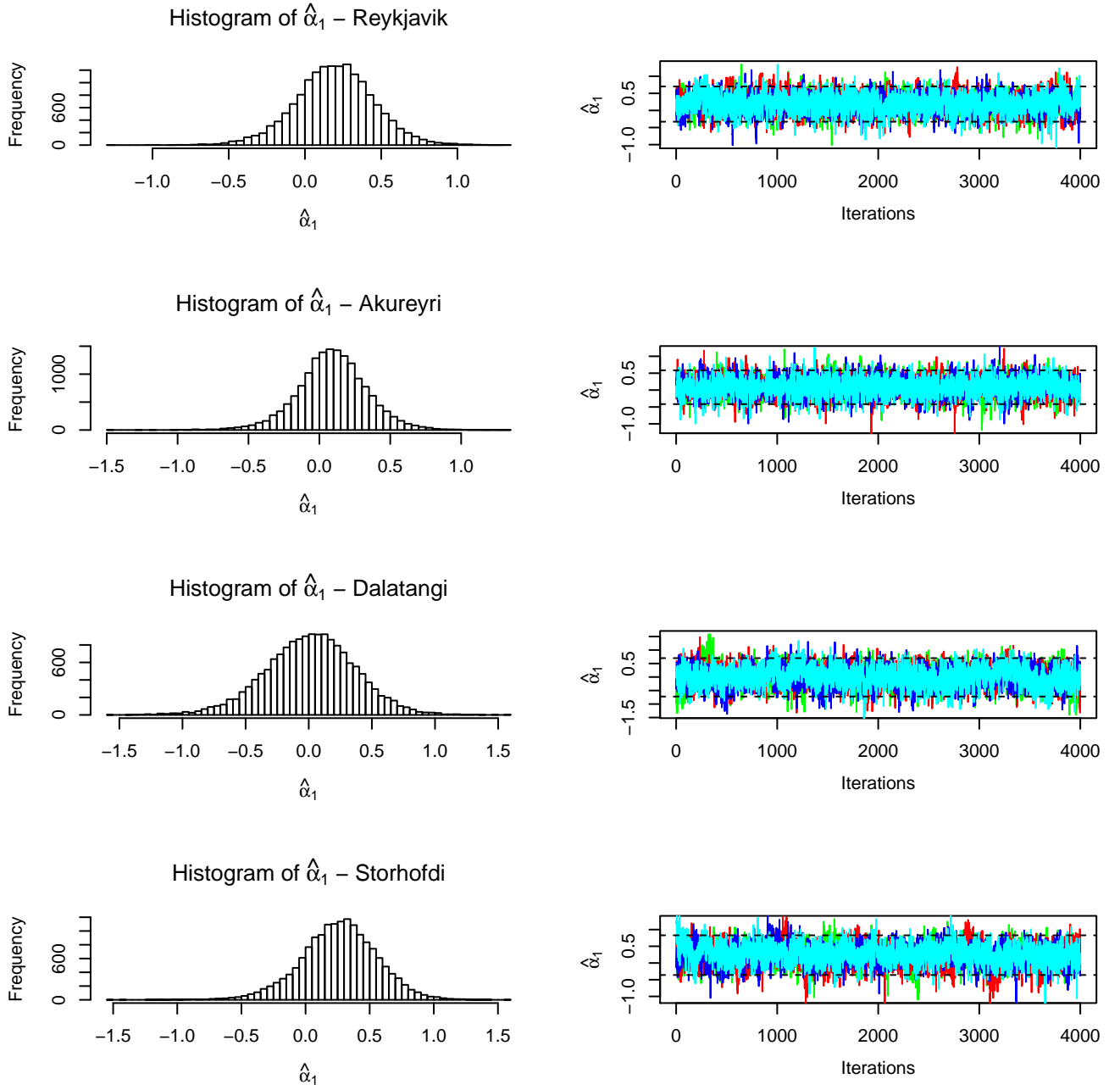


Figure B.8: Histograms and MCMC trace plots after burn-in for the parameter α_1 .

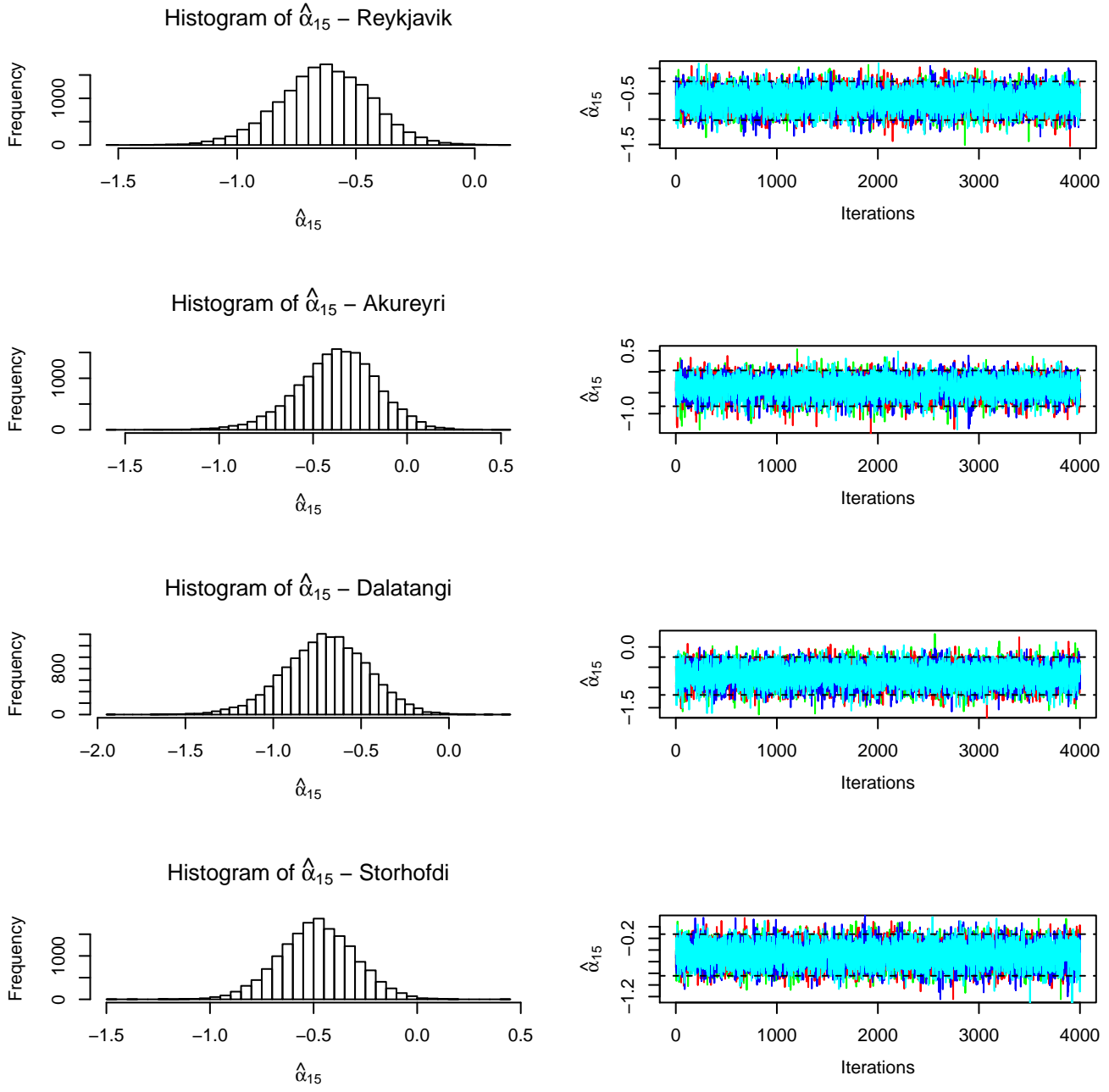


Figure B.9: Histograms and MCMC trace plots after burn-in for the parameter α_{15} .