Stockholms universitet

Bayesian source apportionment of organic carbon in the East Siberian Sea

Kerstin Nilsson

Kandidatuppsats i matematisk statistik
Bachelor Thesis in Mathematical Statistics

Kandidatuppsats 2014:3
Matematisk statistik
Maj 2014

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Stockholm University

Mathematical Statistics
Stockholm University
Bachelor Thesis **2014:3**
http://www.math.su.se

Bayesian source apportionment of organic carbon in the East Siberian Sea

Kerstin Nilsson*

Maj 2014

Abstract

In this thesis work, data from a climatology study exploring the effects of climate warming in northeastern Siberia is analysed. Thawing of the permafrost and subsequent release of greenhouse gases such as carbon dioxide and methane is considered to be one of the most powerful factors that could have a worsening effect on global warming. The study in question seeks to contribute to improving on the current poor understanding of this situation through investigation of the relative proportions of organic carbon from riverine, coastal erosion and marine sources in surface sediments collected from the East Siberian Sea, at four different sites along a 500 km transect of the Kolyma paleoriver. To this end, a so-called linear mixing model can be used. This type of model is used in many scientific applications where the aim is to investigate the relative contributions of multiple sources to a mixture. Statistical problems include how to estimate uncertainty in the proportion estimates. This thesis work illustrates the use of Bayesian methods to analyse this kind of data and different models are compared. The results for the selected model indicate that the marine organic carbon proportion increases with increased distance from the river mouth, while the riverine or coastal erosion or both decrease, and moreover that at the first measurement station, the marine component is lower than both the riverine and the coastal erosion component, while for the last measurement station, the marine proportion is similar to that of coastal erosion while the riverine proportion is lower.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail:kerstin.c.nilsson@gmail.com . Supervisor: Martin Sköld.

Acknowledgements

I would like to express my gratitude to my supervisor Martin Sköld at the Department of Mathematics, Stockholm University, for giving me the opportunity to work with this project and for his time and help. I would also like to thank my co-supervisor August Andersson at the Department of Applied Environmental Science, Stockholm University, for answering questions related to the project.

Contents

1 Introduction

A research question of interest in many areas is to estimate the relative contributions of multiple sources to a mixture. Isotope analysis is commonly used for this purpose and examples include e.g. determination of food sources in an animal's diet, water sources used by plants and sources of carbon dioxide efflux from forest floor respiration. Here, the application is within climatology, where the aim is to estimate the proportions of different sources of organic carbon in samples obtained from the ocean floor of the East Siberian Sea, to explore effects of climate warming in the Arctic. So-called linear mixing models are often used to analyse this type of data, and statistical challenges include how uncertainty in the proportion estimates should be estimated. This thesis work presents an approach to using Bayesian methods in a linear mixing model application.

2 Background

2.1 Effects of climate warming in Siberia

Since the early 20th century it has been observed that the temperature of the lower atmosphere and oceans of the Earth has risen. This process is expected to continue and it is referred to as global warming. The mean surface temperature of the Earth increased by about 0.8 °C during the 20th century (Intergovernmental Panel on Climate Change (IPCC) AR4 SYR 2007). In an assessment of the scientific literature made in 2007, the IPCC reported that scientists were more than 90% certain that most of the global average warming over the past 50 years is due to increasing concentrations of greenhouse gases produced by human activities (IPCC AR4 SYR 2007). In 2013, the IPCC stated that the evidence for human influence has grown since the 2007 assessment (IPCC AR5 WG1 2013).

There are a number of factors that could have an aggravating effect on the global warming, of which thawing of the permafrost in Siberia is considered to be one of the most important. The permafrost has been intact since the end of the ice age 10000 to 11000 years ago, but during the last years there have been indications that the permafrost has already started to thaw (Sazonova et al., 2004, Chudinova et al., 2006, Romanovsky et al., 2007). As much as half of the global soil organic carbon is stored in the top few meters of the Arctic permafrost (Tarnocai et al., 2009), and upon thawing of the permafrost, organic material that have been frozen could become released and degrade. In the degradation process, green house gases such as carbon dioxide and methane could be released and would thus contribute to the warming of the climate. This process could occur both on land and during enhanced export to the Arctic Ocean. To gain more knowledge on the latter process a study was carried out in the extensive coastal shelf environment of

the East Siberian Arctic. In the study, which is described in more detail in section 3.1, the proportions of organic carbon from different sources along a 500 km transect of the Kolyma River in the East Siberian Sea is investigated.

2.2 Linear mixing models

To estimate the proportions of two sources, a linear mixing model with one isotope could be used. One approach to mixing models is to utilise the law of conservation of mass to model data, and a reasonable model in many applications is to assume the following mass-balance equations to hold:

$$X_{Mixture} = P_A X_A + P_B X_B$$
$$P_A + P_B = 1$$

where X_A and X_B represents the content of the isotope in source A and B respectively, and $X_{Mixture}$ is the content in the mixture, P_A and P_B are the proportions of source A and B in the mixture.

For estimation of source proportions from three sources, two different isotopes are required, the linear mixing model can then be formulated according to

$$X^1_{Mixture} = P_A X^1_A + P_B X^1_B + P_C X^1_C$$

$$X^2_{Mixture} = P_A X^2_A + P_B X^2_B + P_C X^2_C$$

$$P_A + P_B + P_C = 1$$

where X^1 and X^2 indicate the content of the two different isotopes in the respective type of sample, and A, B and C subscripts refer to the sources as above.

2.3 Bayesian data analysis

2.3.1 The Bayesian view and some concepts

The two main approaches used in statistical inference today could be considered to be the classical (frequentist) and the Bayesian. In the frequentist view, a probability for an event corresponds to its relative frequency in a large number of independent random trials. This is an example of an objective probability. Further, parameters are treated as fixed values, with no probability distributions associated with them.

In the Bayesian approach, probabilities have a wider definition, and may include subjective probabilities. Further, in the Bayesian approach one is also willing to assign probability distributions to parameters. The basis for Bayesian framework is Bayes' theorem, published in 1763. Bayes' theorem

may be expressed as:

$$p(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{f(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{m(\boldsymbol{y})},$$

$$m(\boldsymbol{y}) = \int f(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$$

where

$f(\boldsymbol{y}|\boldsymbol{\theta})$ is the sampling model specification (likelihood), $\boldsymbol{Y}|\boldsymbol{\theta} \sim f(\boldsymbol{y}|\boldsymbol{\theta})$
$\pi(\boldsymbol{\theta})$ is the prior probability specification
$p(\boldsymbol{\theta}|\boldsymbol{y})$ is the posterior distribution
$m(\boldsymbol{y})$ is the marginal density of the data $\boldsymbol{y}$

The prior distribution reflects the information we have about the parameters before the data is collected, and this is then used together with the likelihood to compute the conditional distribution of the parameters based on the data (the posterior distribution). Thus, the knowledge that we have about the parameters before the data is collected is updated based on the data through the application of Bayes' theorem. The support of the posterior will be a subset of that of the prior. The prior could further have parameters, these are then referred to as hyperparameters.

When prior information is available, e.g. as data from previous experiments or as the subjective knowledge of an expert, this could be utilised to construct an informative prior for the parameters. If no reliable prior information is available, or if one wishes to base the inference solely on the data, a so-called noninformative prior could be used. With "noninformative" it is meant in this case to set the probabilities equal to all parameter values. An example is when the parameter space is discrete and finite, $\Theta = \{\theta_1, \ldots, \theta_n\}$. By using the distribution

$$p(\theta_i) = 1/n, i = 1, \ldots, n,$$

one assigns equal probability to all possible parameter values.

If a prior distribution could be chosen such that a posterior distribution in the same family as the prior is rendered, it is computationally advantageous, as there is no requirement for numerical integration. This is referred to as a conjugate prior.

Having specified the prior and applied Bayes' theorem to obtain the posterior distribution, this now contains all the current knowledge we have about our parameter(s). There are a number of ways to present the results, e.g. a graph of the distribution, or providing point estimates such as the mean, median or mode. Interval estimation can also be carried out, this is

usually referred to as a credible set or credible interval, but could also be referred to as a Bayesian confidence interval. The definition is according to: a $100 \times (1-\alpha)\%$ credible set for $\boldsymbol{\theta}$ is a subset C of $\boldsymbol{\Theta}$ such that

$$P(\boldsymbol{\theta} \in C|\boldsymbol{y}) = \int_C p(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta} \geq 1-\alpha$$

where integration is replaced by summation for discrete components of $\boldsymbol{\theta}$. This definition enables statements in terms of probability according to:

"The probability that $\boldsymbol{\theta}$ lies in C given the observed data $\boldsymbol{y}$ (and the prior) is at least $(1-\alpha)$."

The definition of the Bayesian credible set could be compared with an interpretation of the frequentist confidence interval for a parameter θ, e.g.

"If the sampling were repeated a large number of times and a confidence interval at the $(1-\alpha)$ level constructed in each case, $(1-\alpha) \times 100\%$ of them would contain the true value of θ".

This may be conceived as more difficult to understand and to explain, and moreover, for a particular dataset, after that data have been collected, the confidence interval either contains the true value of the parameter or not, while in the Bayesian case, it is an actual probability statement. Practically, the limits are often retrieved by taking the $\alpha/2$ and the $(1-\alpha/2)$ quantiles of the posterior as the $100 \times (1-\alpha)$ credible set for θ.

2.3.2 Model selection

Several measures have been developed to aid in the statistical selection among a collection of different models, e.g. Akaike information criterion (AIC) and the Bayesian information criterion (BIC), sometimes called the Schwarz criterion. Both AIC and BIC require specification of the number of parameters in the model, which is problematic in the case of Bayesian models since inclusion of a prior distribution induces a dependence between parameters, that likely has the effect of reducing the effective dimensionality (Spiegelhalter et al., 2002). As a solution to this problem, Spiegelhalter et al. (2002) suggest a generalization of the AIC called the Deviance Information Criterion, DIC, that is based on the posterior distribution of the deviance statistic

$$D(\boldsymbol{\theta}) = -2\log f(\boldsymbol{y}|\boldsymbol{\theta}) + 2\log h(\boldsymbol{y})$$

where $f(\boldsymbol{y}|\boldsymbol{\theta})$ is the likelihood function for the observed data vector $\boldsymbol{y}$ given the parameter vector $\boldsymbol{\theta}$ and $h(\boldsymbol{y})$ is some standardising function of the data

alone, which thus does not impact in model selection. The posterior expectation of the deviance $\overline{D} = E_{\theta|y}[D(\boldsymbol{\theta})]$ is a measure of the model fit where larger values indicate worse fit. The complexity of the model is summarised by the effective number of parameters p_D,

$$p_D = E_{\theta|y}[D(\boldsymbol{\theta})] - D(E_{\theta|y}[\boldsymbol{\theta}])$$

The deviance information criterion (DIC) is defined as

$$DIC = \overline{D} + p_D = 2\overline{D} - D(\overline{\boldsymbol{\theta}}) \tag{1}$$

The p_D term compensates for that the posterior expectation of the deviance will decrease as the number of parameter increases. Smaller values of DIC indicate a better fit, however DIC values have no meaning in themselves and it is only differences in DIC between models that are meaningful. In Carlin and Louis (2009), it is said that differences should be at least between 3 to 5 to be considered interesting (p. 71), in another paragraph they state that practitioners are often of the opinion that differences smaller than 10 may not be relevant (p. 73).

DIC has been widely used in applied Bayesian work due to its generality and that it is relatively easy to compute, however there are some issues that have been raised, such as that DIC is not invariant to parametrisation, and that p_D can sometimes turn out to be negative in certain circumstances where the posterior is clearly non-normal.

2.3.3 Bayesian computation

The determination of posterior distributions and summaries such as moments or quantiles often involve evaluation of complex and high-dimensional integrals. In the early days, approaches to solve this problem included using asymptotic methods to obtain analytic approximations to the posterior density, methods for numerical integration such as Gaussian quadrature and the expectation-maximisation (EM) algorithm. All these methods have their drawbacks however, e.g. the Gaussian approach only works for models of low dimension and the EM algorithm is aimed more at finding the posterior mode rather than the whole posterior distribution. These problems have been resolved with the development of Monte Carlo integration methods, which provide more complete information and are in comparison easier to program, even for very high-dimensional models. Monte Carlo methods include both noniterative methods as well as iterative methods, the latter being more powerful. Examples of iterative methods are Metropolis-Hastings algorithm and the Gibbs sampler. These methods generate a Markov chain, which output corresponds to a correlated sample from the joint posterior distribution, and are referred to as Markov chain Monte Carlo (MCMC) methods.

The Gibbs sampler is used in the JAGS software (see more on software in subsection 2.3.4) and can be described in the following way: Suppose there are k parameters in a model, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$. Assume that samples can be generated from each of the full conditional distributions $\{p(\theta_i|\boldsymbol{\theta}_{j\neq i}, \boldsymbol{y}), i = 1 \ldots, k\}$ either through direct or indirect sampling. In either case, one can show that the collection of full conditional distributions uniquely determine the joint posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{y})$, and hence all marginal posterior distributions $p(\theta_i|\boldsymbol{y})$. For an arbitrary set of starting values $\{\theta_2^{(0)}, \ldots \theta_k^{(0)}\}$ the algorithm works as follows:

Gibbs sampler: For $(t = 1, \ldots, T)$, do:
Step 1: Draw $\theta_1^{(t)}$ from $p(\theta_1|\theta_2^{(t-1)}, \theta_3^{(t-1)}, \ldots \theta_k^{(t-1)}, \boldsymbol{y})$
Step 2: Draw $\theta_2^{(t)}$ from $p(\theta_2|\theta_1^{(t)}, \theta_3^{(t-1)}, \ldots \theta_k^{(t-1)}, \boldsymbol{y})$
.
.
.
Step k: Draw $\theta_k^{(t)}$ from $p(\theta_k|\theta_1^{(t)}, \theta_2^{(t)}, \ldots \theta_{k-1}^{(t)}, \boldsymbol{y})$

It can be shown that the k-tuple $\boldsymbol{\theta}^{(t)}$ obtained at iteration t converges in distribution to a draw from the true joint posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{y})$. This means that for t sufficiently large, (say bigger than t_0), $\{\boldsymbol{\theta}_{(t)}, t = t_0 + 1, \ldots, T\}$ is a correlated sample from the true posterior, from which any posterior quantities of interest may be estimated. The posterior mean may be estimated according to:

$$\hat{E}(\theta_i|\boldsymbol{y}) = \frac{1}{T - t_0} \sum_{t=t_0+1}^{T} \theta_i^{(t)} \tag{2}$$

The time from $t = 0$ to $t = t_0$ is referred to as the burn-in period. If m parallel chains are run, i.e. separate chains from different starting values, the mean may be estimated according to:

$$\hat{E}(\theta_i|\boldsymbol{y}) = \frac{1}{m(T - t_0)} \sum_{j=1}^{m} \sum_{t=t_0+1}^{T} \theta_{i,j}^{(t)}$$

An estimate of the variance for the posterior mean can be calculated (and hence, the standard error). Suppose that for a parameter of interest, λ, we have available a single long chain of MCMC samples $\{\lambda^{(t)}\}_{t=1}^{N}$ which is assumed to come from a stationary distribution of the Markov chain after removal of the burn-in period. According to expression 2 above, an estimate of $E(\lambda|\boldsymbol{y})$ is then given by:

$$\hat{E}(\lambda|\boldsymbol{y}) = \hat{\lambda}_N = \frac{1}{N} \sum_{t=1}^{N} \lambda^{(t)}$$

One approach to estimating the variance would be to use the sample variance, $s^2_\lambda = \frac{1}{N-1}\sum_{t=1}^{N}(\lambda^{(t)} - \hat{\lambda}_N)^2$, divided by N, according to:

$$\widehat{Var}(\hat{\lambda}_N) = s^2_\lambda/N = \frac{1}{N(N-1)}\sum_{t=1}^{N}(\lambda^{(t)} - \hat{\lambda}_N)^2$$

This estimate would however most likely be an underestimate of the variance due to positive autocorrelation in the MCMC chain. A better alternative is to use another estimate that utilises the concept of effective sample size, ESS. ESS is defined as:

$$ESS = N/\kappa(\lambda)$$

where $\kappa(\lambda)$ is the autocorrelation time for λ given by

$$\kappa(\lambda) = 1 + 2\sum_{k=1}^{\infty}\rho_k(\lambda),$$

where $\rho_k(\lambda)$ is the autocorrelation at lag k for the parameter of interest λ. $\kappa(\lambda)$ can be estimated using sample autocorrelations estimated from the MCMC chain, summing up to where the autocorrelations drop to e.g. 0.1 in magnitude. The variance estimate for $\hat{\lambda}_N$ is then

$$\widehat{Var}_{ESS}(\hat{\lambda}_N) = s^2_\lambda/\widehat{ESS}(\lambda) = \frac{\hat{\kappa}(\lambda)}{N(N-1)}\sum_{t=1}^{N}(\lambda^{(t)} - \hat{\lambda}_N)^2$$

To fully explain why the Gibbs sampler works requires advanced Markov chain theory, but in short it can be mentioned that the Markov chain needs to be irreducible (for every set A with positive posterior probability, the probability of the chain ever entering A is positive for every starting point $\theta^{(0)}$) and aperiodic (the chain can move from any state to any other, there can be no absorbing states from which there is no escape). Aperiodicity ensures convergence of the chain to its stationary distribution (the true joint posterior) and irreducibility ensures this stationary distribution is unique.

With the use of MCMC methods, it is necessary to make an assessment of whether the Markov chain has converged. The chain can get stuck in one part of the joint distribution, leading to high autocorrelation ("slow mixing") in the chain. One solution that has been proposed to this problem is to retain only every m^{th} iterate after convergence has been obtained, where m is large enough so that the resulting samples are virtually uncorrelated. This procedure is known as thinning. Assessment of convergence is usually done by visual inspection of the chains in so-called trace plots and investigating the autocorrelation structure. See figure 1 for an example of good convergence and one example of less good convergence. Note however that no diagnostic can prove the convergence of an MCMC chain, since it uses only a finite realisation of the chain.

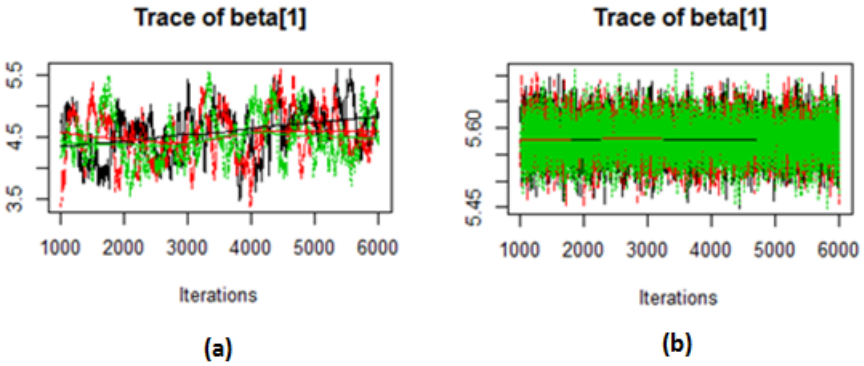

Figure 1: Trace plots: the plot to the right shows an example of a situation where there are problems with convergence (a), the plot to the left show an example where the convergence is good (b). In the plot to the right, the covariates in the linear model were not centred, whereas in the left plot, they were centred and convergence was improved. This example is from modelling the data in Exercise 2.7.12 in Carlin and Louis (2009).

2.3.4 Software

There are a number of different software available for carrying out Bayesian analysis using MCMC. WinBUGS (Lunn et al., 2000) or OpenBUGS, an open source version, is one commonly used software (BUGS stands for Bayesian inference Using Gibbs Sampling), another one is JAGS (Plummer, 2003), which stands for Just Another Gibbs Sampler, that aims for compatibility with WinBugs/OpenBUGS through using a related version of the same modelling language (BUGS), however no graphical user interface for processing of the results is provided with JAGS, and thus another software is required for this purpose. One possibility is to use the R package rjags to connect to JAGS. The main advantage of JAGS compared to WinBUGS/OpenBUGS is its platform independence.

3 Methods

3.1 Description of the current study and data

The data used in this thesis work comes from a study which aims to provide insight in the effects of climate warming in northeastern Siberia through investigation of the sources and degradation of terrestrial organic carbon (OC) in the East Siberian Sea (ESS)(Vonk et al., 2010). Various types of measurements are performed on surface sediment samples (obtained from the ocean floor) and on surface water suspended particle samples. Here, only one part of the measurements is considered, which is the isotope analysis of organic carbon content in surface sediment samples. The objective of this particular analysis is to investigate the proportions of the organic carbon content originating from riverine, coastal erosion and marine sources. A key motivation is that climate warming is expected to increase the input of organic carbon to the ESS from both riverine and coastal erosion sources. The coastal component contains OC from soil-permafrost that is very old, resulting from organic-rich sediments that accumulated during the Pleistocene, a geological epoch which lasted from about 2588000 to 11700 years ago, whereas the riverine source contains more recent organic matter.

In the study, surface sediment samples were collected from the ESS continental shelf between the 3rd to the 5th of September in 2008 as part of a 50 day long international Siberian shelf study conducted on board a research vessel. The samples were taken along a river mouth - midshelf transect following the Kolyma paleoriver canyon, where the water depth is between around 10-50m. This submarine valley was formed during glacial periods when the sea level was lower and since inundation of the shelf, sediments naturally accumulate in the canyon. The samples were collected from the ocean floor at eight stations along a distance of ca 500km, see figure 2 for an overview.

Since there are three sources, two different source markers are required and here ^{13}C and ^{14}C are used. The measurements involve relating the ^{13}C and ^{14}C contents respectively to ^{12}C content, and further relating to carbon content in a reference sample, and measurements are expressed as 'delta': $\delta^{13}C$ and $\Delta^{14}C$. For more detailed information on how these measures are defined, see Libes (1992) for $\delta^{13}C$ and Woods Hole Oceanographic Institution (2013) for $\Delta^{14}C$.

The total δ^{13}C and Δ^{14}C were measured for the surface sediment organic carbon content by using isotope mass spectrometer analysis. For δ^{13}C the analysis was carried out for all stations, but for Δ^{14}C, analyses were only carried out for four of the eight stations. See table 1 for an overview of the sampling locations and the available data.

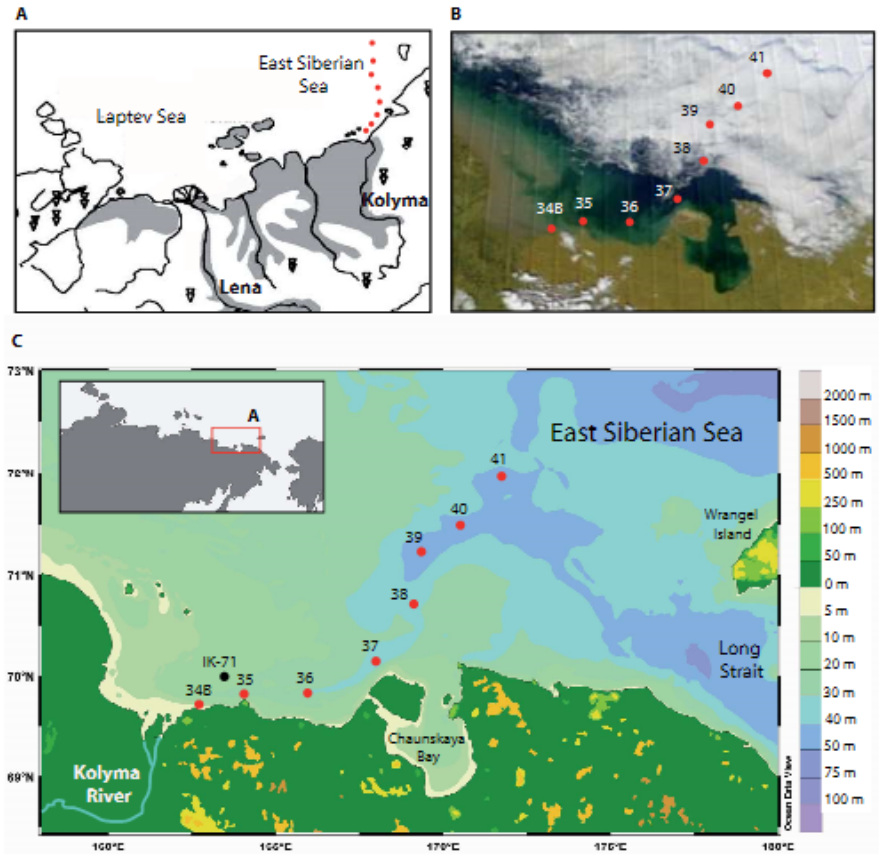

Figure 2: (a) The graph shows the distribution of a certain kind of permafrost-soil that accumulated during the Pleistocene (b) Satellite image taken on 24 August 2000, likely from ongoing erosion. (c) Map of the southern East Siberian Sea showing the locations of the sampling stations (red circles) following the Kolyma paleoriver canyon (figure taken from Vonk et al., 2010)

Table 1: Stations with the distances from the river mouth. Stations with measurements for both δ^{13}C and Δ^{14}C are indicated with an 'x'.

Station	Distance to river mouth (km)	Measurement for both δ^{13}C and Δ^{14}C
YS-34B	46	x
YS-35	100	-
YS-36	174	x
YS-37	258	-
YS-38	334	x
YS-39	392	-
YS-40	443	x
YS-41	513	-

Measurements for the organic carbon content for the sources of interest were collected from analyses reported in the literature or carried out prior to the current study. Table 2 contains a list of the estimated means and standard deviations for the respective measurements and the locations they were sampled from as well as references.

Table 2: Estimated means and standard deviations for the sources, with sampling location and literature references, retrieved from Vonk et al. (2010).

Isotope	Source	Mean (‰)	Std (‰)	Location	Reference
δ^{13}C	Riverine	-29.3	1.7	Kolyma River	McLelland et al., 2008
	Erosion	-25.8	1.7	Kolyma delta	Dutta et al., 2006
	Marine	-21	1	High latitude areas, otherwise not specified	Several references, e.g. Meyers, 1997; Semiletov et al.,2005
Δ^{14}C	Riverine	-296	68	Lena River	Unpublished results
	Erosion	-788	201	Laptev Sea	Sánchez-Garcia et al., 2010
	Marine	-21	25	Northern Pacific Ocean	Key et al.,2004

3.2 Statistical methods

The endmember mixing model can be formulated as

$$X^{13}_{Mixture}[i] = P_R[i]X^{13}_R + P_E[i]X^{13}_E + P_M[i]X^{13}_M$$

$$X^{14}_{Mixture}[i] = P_R[i]X^{14}_R + P_E[i]X^{14}_E + P_M[i]X^{14}_M$$

$$P_R[i] + P_E[i] + P_M[i] = 1$$

where i indicates the station and equals in order the stations YS-34B, YS-36, YS-38 and YS-40; $X^{13}_{Mixture}[i]$ is the total carbon content in the mixture (ocean floor sediment sample) at station i, X^{13}_R is the carbon content originating from riverine sources, X^{13}_E denotes the carbon content from the coastal erosion source and X^{13}_M denotes the carbon content from the marine source, for the δ^{13}C measurements. The variables for Δ^{14}C are denoted analogously. $P_R[i]$, $P_E[i]$ and $P_M[i]$ denote the proportions of the respective source riverine, erosion and marine at station i.

The carbon contents for the sources are assumed to be normally distributed, according to

$$\begin{aligned}
X^{13}_R &\sim N(\mu^{13}_R, (\sigma^{13}_R)^2) \\
X^{13}_E &\sim N(\mu^{13}_E, (\sigma^{13}_E)^2) \\
X^{13}_M &\sim N(\mu^{13}_M, (\sigma^{13}_M)^2) \\
X^{14}_R &\sim N(\mu^{14}_R, (\sigma^{14}_R)^2) \\
X^{14}_E &\sim N(\mu^{14}_E, (\sigma^{14}_E)^2) \\
X^{14}_M &\sim N(\mu^{14}_M, (\sigma^{14}_M)^2)
\end{aligned}$$

Given this assumption, and that the sources can be assumed to be independent, it follows from known results for the Normal distribution (see e.g. theorems 3.25 and 3.47 in Alm and Britton (2008)), that the mixtures will have the following respective distributions

$$X^{13}_{Mixture}[i]|P_R[i], P_E[i], P_M[i] \sim N(\mu^{13}_{Mixture}[i], (\sigma^{13}_{Mixture}[i])^2) \qquad (3)$$

$$X^{14}_{Mixture}[i]|P_R[i], P_E[i], P_M[i] \sim N(\mu^{14}_{Mixture}[i], (\sigma^{14}_{Mixture}[i])^2) \qquad (4)$$

where

$$\mu^{13}_{Mixture}[i] = P_R[i]\mu^{13}_R + P_E[i]\mu^{13}_E + P_M[i]\mu^{13}_M$$

$$\mu^{14}_{Mixture}[i] = P_R[i]\mu^{14}_R + P_E[i]\mu^{14}_E + P_M[i]\mu^{14}_M$$

and

$$(\sigma^{13}_{Mixture}[i])^2 = P_R[i]^2(\sigma^{13}_R)^2 + P_E[i]^2(\sigma^{13}_E)^2 + P_M[i]^2(\sigma^{13}_M)^2$$

$$(\sigma^{14}_{Mixture}[i])^2 = P_R[i]^2(\sigma^{14}_R)^2 + P_E[i]^2(\sigma^{14}_E)^2 + P_M[i]^2(\sigma^{14}_M)^2$$

Here, we use the values in table 2 for the true means and standard deviations for the sources. There could potentially be a correlation between $X^{13}_{Mixture}$ and $X^{14}_{Mixture}$, however we choose not to take that into account here.

We apply three different models to the data, referred to as model 1, 2, and 3. These are described in more detail below. In all the models, the likelihoods are given by expression 3 for $\delta^{13}C$ and 4 for $\Delta^{14}C$ respectively.

Model 1

In this model we set Dirichlet priors for the proportions at all four stations, that are assumed to be independent between stations:

$$(P_R[i], P_E[i], P_M[i]) \sim Dirichlet(\boldsymbol{\alpha}),$$

with $\boldsymbol{\alpha} = (1, 1, 1)$ for station i. When the elements of the parameter vector in the Dirichlet distribution have the same value, it is symmetric in that it does not favour one parameter over another. Further, when all elements are equal to 1, the distribution is uniform over all points in its support. Thus, this prior is noninformative, both with respect to the proportion values within stations, as well as with respect to the proportion relationships between stations. The model specification in rjags is listed in figure 10 in the appendix.

Model 2

Here we also use Dirichlet priors for the proportions with $\boldsymbol{\alpha} = (1, 1, 1)$ for all stations but with an order assumption according to:

$$P_R[1] < P_R[2] < P_R[3] < P_R[4]$$

$$P_E[1] < P_E[2] < P_E[3] < P_E[4]$$

$$P_M[1] > P_M[2] > P_M[3] > P_R[4]$$

That is, it is assumed that the proportions for the riverine and the coastal sources decrease with increased distance from the river mouth, while the marine contribution increases. This model was fitted in rjags by running

model 1, and then selecting the iterations that fulfilled the above order criteria.

Model 3

In this model, we set independent Dirichlet distributions with parameter vector $\boldsymbol{\alpha} = (1, 1, 1)$ for the proportions for the first and last station, and the values for the stations inbetween are estimated using linear interpolation, according to

$$p(s) = (p_0(S - s) + p_1 s)/S$$

where p_0 and p_1 are the proportions at the first and last station respectively, and s is the distance from the first station to the current station and S is the distance between the first and last station. See figure 11 in the appendix for the model specification in rjags.

Model 1 thus makes the least assumptions about the data in that there are no assumptions regarding the relationship of proportions across stations. In model 3, the order of the magnitudes of the proportions for stations 1 and 4 is not specified, but it is assumed that the values for stations 2 and 3 lie along a straight line between the values for stations 1 and 4. In model 2 an order relation is set for all proportion values between stations.

Convergence of the MCMC algorithm was assessed using trace and autocorrelation function (acf) plots. To aid in the selection of the number of sufficient iterations to use for inference, the standard errors of the means of the respective posterior distributions were additionally consulted. The DIC measure was calculated for all models to be used as guidance in statistical model comparison. For model 1 and 3, DIC was calculated using the rjags function dic.samples, and for model 2 it was calculated according to eq. 1.

The analyses were carried out using R version 3.0.2 and rjags version 3-11.

4 Results and Discussion

For an overview of the data, see figure 3 in which the distance to the river mouth is plotted against the values for $\delta^{13}C$ and $\Delta^{14}C$ in the mixtures (ocean floor sediment samples), respectively. Models 1 and 3 were run using three parallel chains, with 150000 iterations and the thinning parameter set to 10. Adaptation was set to 1000 iterations. Starting values for the proportions were set to (1/10, 3/10, 6/10), (4/10, 5/10, 1/10), (1/3, 1/3, 1/3) respectively for the three chains. Of the iterations, 10 % were removed as burn-in and this left 40800 iterations in the respective samples in total. Model 2 was run using one chain for 10^7 iterations, with start values

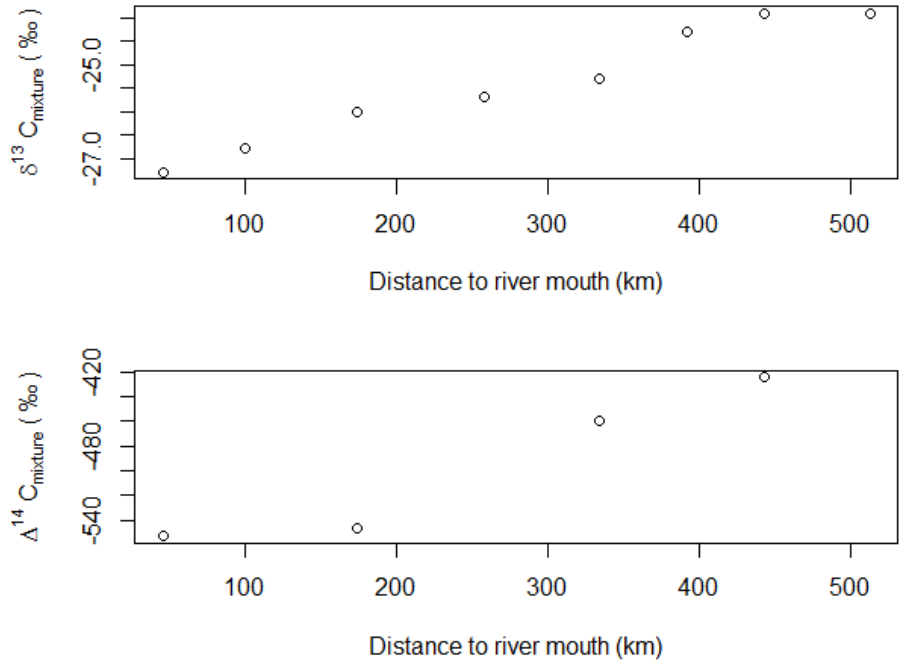

Figure 3: $\delta^{13}C$ and $\Delta^{14}C$ versus distance to river mouth.

(1/3,1/3,1/3), and of these there were 37106 iterations that followed the order criteria as defined in section 3.2.

There appeared to be no issues with non-convergence, and the obtained number of iterations appeared to be sufficient for all models. See the appendix for density, trace and acf plots, and a table with estimated means for the proportion parameters and their associated standard errors, adjusted for autocorrelation.

Graphs of the respective posterior distributions of the proportions from the different models are additionally displayed in figures 4, 5 and 6.

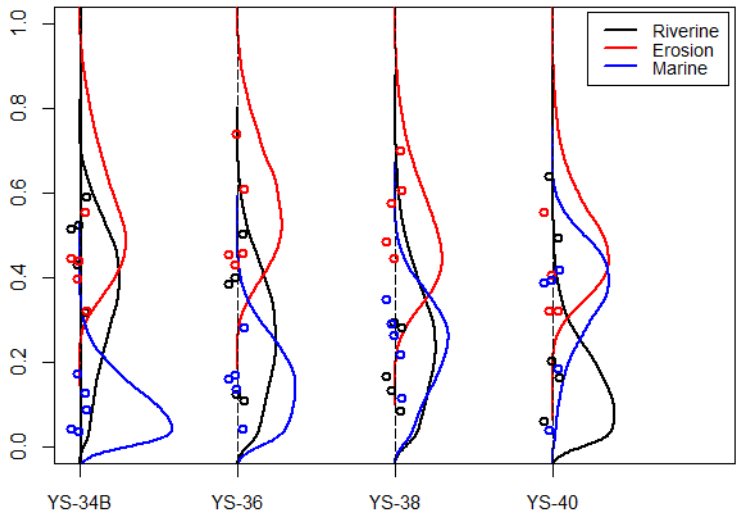

Figure 4: Posterior density distributions for the proportions for model 1. Parameter values are displayed for each source for five randomly selected iterations, indicated by circles. The values have been jittered along the x-axis for visibility.

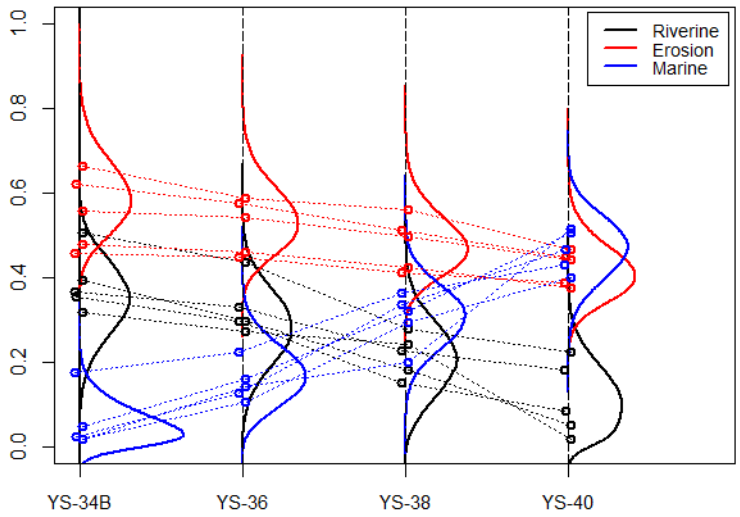

Figure 5: Posterior density distributions for the proportions for model 2. Parameter values for each source are displayed for five randomly selected iterations, indicated by circles. These have been joined with lines between stations. The values have been jittered along the x-axis for visibility.

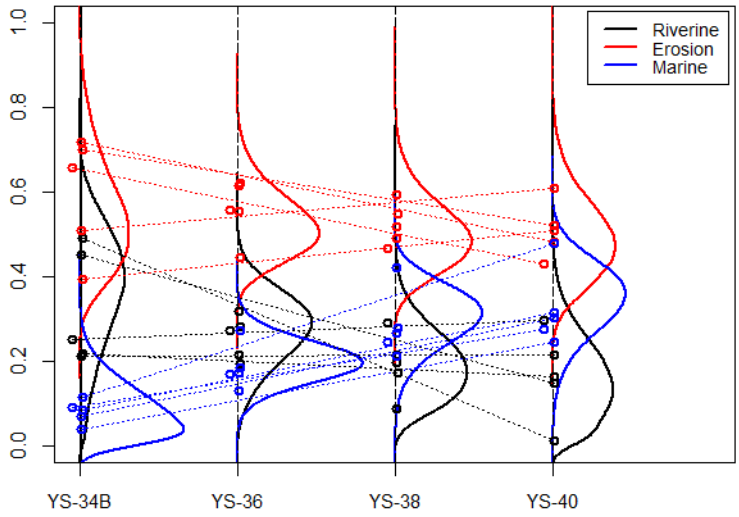

Figure 6: Posterior density distributions for the proportions for model 3. Parameter values are displayed for each source for five randomly selected iterations, indicated by circles, and lines have been drawn between the values of the first and the last station. The values have been jittered along the x-axis for visibility.

In figure 7, the posterior distributions from all three models are overlayed, per station and source. The posterior distributions for model 1 and 3 are quite similar for the first and the last station, while the posterior distributions from model 2 and 3 are quite similar for the two stations inbetween. Looking at the DIC values in table 3, there is a difference of 5.4 between

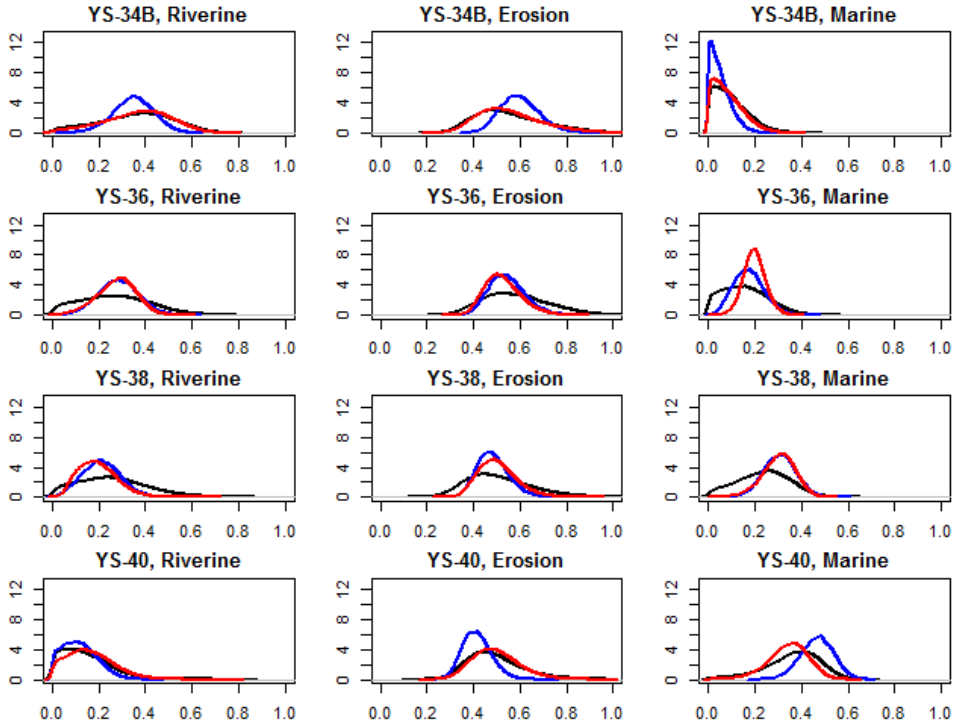

Figure 7: Posterior density distributions for the proportions from the three models. Black - model 1, blue - model 2, red - model 3.

model 1 and 2 and 4.1 between model 1 and 3, thus there is some statistical support in selecting model 2 above model 1 and perhaps also in selecting model 3 over model 1. However, a DIC value of around 5 is not gaged as

Table 3: DIC values.

Model	$\overline{D}$	p_D	DIC
Model 1	58.3	4.8	63.1
Model 2	55.3	2.4	57.7
Model 3	56.3	2.7	59.0

a very important difference according to Carlin and Louis (2009), and in this situation one should perhaps place higher weight on which model is the most plausible from a climatological and geological point of view. If model 2 is believed to be the most plausible, then the choice of model is quite clear.

Based on the credible interval estimates, available in figures 8 (by source) and 9 (by station) as well as in in table 4, one can conclude for model 2 that the marine proportion at station 1 is lower than from those of station 3 and 4 (the estimate is (<0.01,0.16) for station 1 and (0.17,0.44) and (0.32,0.60) for station 3 and 4 respectively). For the riverine and the coastal source, the intervals do not separate, however, since the proportion of the marine source increases with increased distance from the river mouth, then either the riverine or the coastal erosion source, or both, must decrease. Further, the interval estimates at station 1 indicate that the marine proportion is lower than the proportions of riverine (0.18,0.51) and coastal erosion (0.45,0.77). At the last station, the interval estimates are similar for the marine (0.32,0.60) and coastal erosion proportion (0.30,0.55) while the riverine proportion (0.01,0.27) is lower.

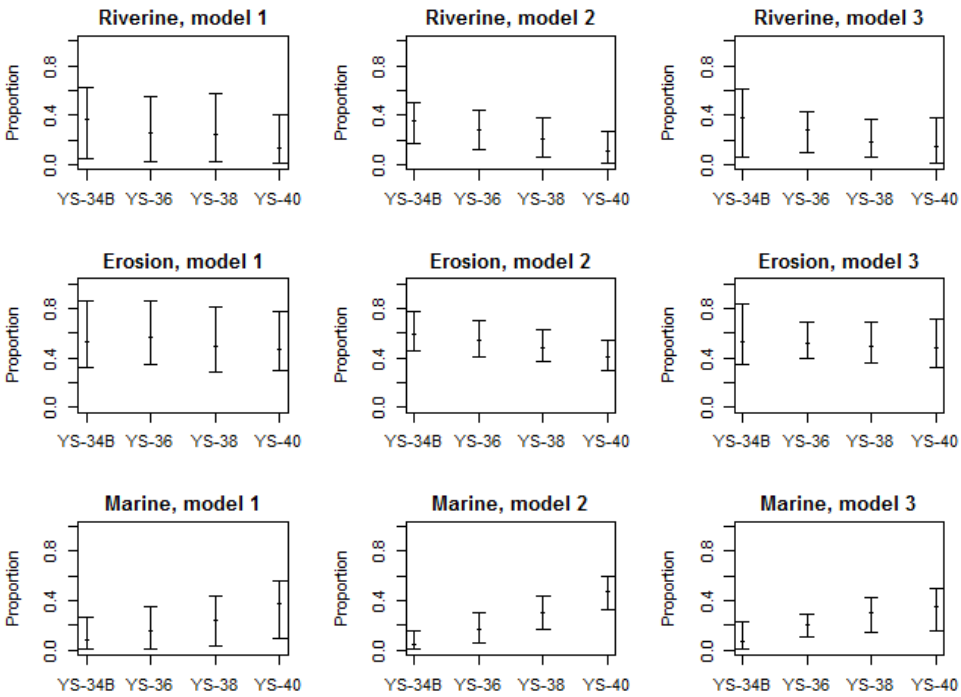

Figure 8: Estimated medians and 95% credible intervals for the proportions at the different stations for the three models, by source.

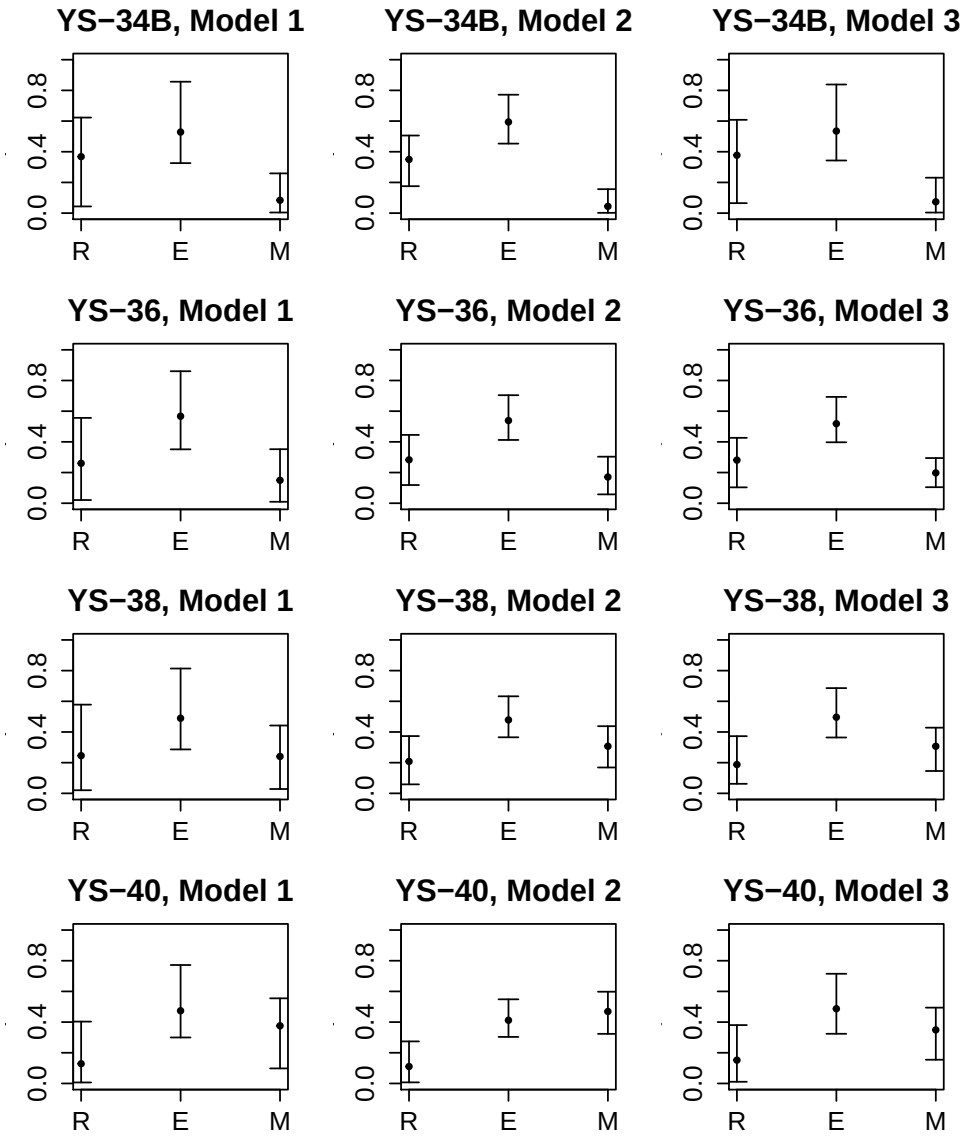

Figure 9: Estimated medians and 95 % credible intervals for the proportions at the different stations for the three models, by station. R - riverine, E - coastal erosion, M - marine.

Table 4: Estimated medians and 95% credible intervals for the proportions at the different stations for the three models.

Model	Station	Source	Median	Lower 95%	Upper 95%
Model 1	YS-34B	Riverine	0.37	0.04	0.62
		Erosion	0.53	0.33	0.86
		Marine	0.08	<0.01	0.26
	YS-36	Riverine	0.26	0.02	0.56
		Erosion	0.57	0.35	0.86
		Marine	0.15	0.01	0.35
	YS-38	Riverine	0.25	0.02	0.58
		Erosion	0.49	0.29	0.81
		Marine	0.24	0.03	0.44
	YS-40	Riverine	0.13	0.01	0.40
		Erosion	0.47	0.30	0.77
		Marine	0.38	0.10	0.55
Model 2	YS-34B	Riverine	0.35	0.18	0.51
		Erosion	0.59	0.45	0.77
		Marine	0.04	<0.01	0.16
	YS-36	Riverine	0.28	0.12	0.45
		Erosion	0.54	0.41	0.70
		Marine	0.17	0.06	0.30
	YS-38	Riverine	0.21	0.06	0.37
		Erosion	0.48	0.37	0.63
		Marine	0.31	0.17	0.44
	YS-40	Riverine	0.11	0.01	0.27
		Erosion	0.41	0.30	0.55
		Marine	0.47	0.32	0.60
Model 3	YS-34B	Riverine	0.38	0.06	0.61
		Erosion	0.53	0.34	0.84
		Marine	0.07	<0.01	0.23
	YS-36	Riverine	0.28	0.10	0.43
		Erosion	0.52	0.40	0.69
		Marine	0.20	0.10	0.29
	YS-38	Riverine	0.19	0.06	0.37
		Erosion	0.50	0.36	0.69
		Marine	0.31	0.15	0.43
	YS-40	Riverine	0.15	0.01	0.38
		Erosion	0.49	0.32	0.72
		Marine	0.35	0.15	0.49

5 References

Alm, S. E. and Britton, T. (2008). Stokastik. Liber AB, 1st edition.

Carlin, B. and Louis, T. (2009). Bayesian Methods for Data Analysis. Chapman and Hall/CRC Press, 3rd edition.

Chudinova, S.M., Frauenfeld, O.W., Barry, R.G., Zhang, T. J., and Sorokovikov, V.A. (2006): Relationship between air and soil temperature trends and periodicities in the permafrost region of Russia. J. Geophys, Res-Earth, 111, 1-15.

McClelland, J. W., Holmes, R. M., Peterson, B. J., Amon, R., Brabets, T., Cooper, L., Gibson, J., Gordeev, V. V., Guay, C., Milburn, D., Staples, R., Raymond, P. A., Shiklomanov, I., Striegl, R., Zhulidov., A., Gurtovaya, T., and Zimov, S. (2008): Development of a pan-Arctic database for river chemistry. Eos Trans. AGU, 89(24), 217-218, doi:10.1029/2008EO240001.

Dutta, K., Schuur, E. A. G., Neff, J. C., and Zimov, S. A. (2006): Potential carbon release from permafrost soils of Northeastern Siberia. Glob. Change Biol., 12, 2336-2351.

IPCC AR4 SYR (2007): Core Writing Team; Pachauri, R.K; and Reisinger, A., ed., Climate Change 2007: Synthesis Report, Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, IPCC, ISBN 92-9169-122-4.

IPCC AR4 SYR (2007), Treatment of uncertainty:
http://www.ipcc.ch/publications_and_data/ar4/syr/en/mainssyr-introduction.html
(retrieved 2014-05-05)

IPCC AR5 WG1 (2013): Stocker, T.F., et al., ed., Climate Change 2013: The Physical Science Basis. Working Group 1 (WG1) Contribution to the Intergovernmental Panel on Climate Change (IPCC) 5th Assessment Report (AR5), Cambridge University Press

Key, R. M., Kozyr, A., Sabine, C. L., Lee, K., Wanninkhof, R., Bullister, J. L., Feely, R. A., Millero, F. J. Mordy, C., and Peng, T.-H. (2004): A global ocean carbon climatology: Results from Global Data Analysis Project (GLODAP). Global Biogeochem. Cy., 18, GB4031, doi:10.1029/2004GB002247.

Libes, S. (1992): Introduction to Marine Biogeochemistry. Wiley, 1st edition.

Lunn, D.J., Thomas, A., Best, N., and Spiegelhalter, D. (2000): WinBUGS - a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing, 10:325-337

Meyers, P. A. (1997): Organic geochemical proxies of paleoceanographic, paleolimnologic, and paleoclimatic processes. Org. Geochem., 27, 213-250.

Plummer, M. (2003): JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling, Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003), March 20-22, Vienna, Austria.

Romanovsky, V. E., Sazonova, T. S. Balobaev, V. T., Shender, N. I., and Sergueev, D. O. (2007): Past and recent changes in air and permafrost temperatures in eastern Siberia. Global Planet. Change, 56, 399-413.

Sánchez-García, L., Alling, V., Pugach, S., Vonk, J., van Dongen, B., Humborg, C., Dudarev, O., Semiletov, I., and Gustafsson, Ö . (2010a): Inventories and behavior of particulate organic carbon in the Laptev and East Siberian Seas. Global Biogeochem. Cy., in review.

Sazonova, T. S., Romanovsky, V. E., Walsh, J. E., and Sergueev, D.O. (2004): Permafrost dynamics in the 20th and 21st centuries along the East Siberian transect. J. Geophys. Res.-Atmos., 109, D01108, doi:10.1029/2003JD003680.

Semiletov, I., Dudarev, O., Luchin, V., Charkin, A., Shin, K.-H., and Tanaka, N. (2005): The East Siberian Sea as a transition zone between Pacific-derived waters and Arctic shelf waters. Geophys. Res. Lett., 32, L10614, doi:10.1029/2005GL022490.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002): Bayesian measures of model complexity and fit. J. R. Statist. Soc. B, 64, 583-639.

Tarnocai, C., Canadell, J. G., Schuur, E. A. G., Kuhry, P., Mazhitova, G., and Zimov, S. (2009): Soil organic carbon pools in the northern circumpolar permafrost region. Global Biogeochem. Cy., 23, GB2023, doi:10.1029/2008GB003327.

Vonk, J.E., Sánchez-Garcia, L., Semiletov, I., Dudarev, O., Eglinton, T., Andersson, A., Gustafsson, Ö. (2010): Molecular and radiocarbon constraints on sources and degradation of terrestrial organic carbon along the Kolyma paleoriver transect, East Siberian Sea. Biogeosciences, 7, 3155-3166.

Woods Hole Oceanographic Institution
http://www.whoi.edu/nosams/page.do?pid=40146, last updated July 2013 (retrieved 2014-04-06).

6 Appendix

```
model {

#Prior for the proportions
#n - number of stations=4
#1st column in p matrix corresponds to riverine source
#2nd column in p matrix corresponds to erosion source
#3rd column in p matrix corresponds to marine source

for (i in 1:n) {
p[i,1:3] ~ ddirch(alpha[])
}

var_c13 <- sd_c13^2
var_c14 <- sd_c14^2

for (i in 1:n) {
var_c13_mix[i] <- p[i,1]^2*var_c13[1]+p[i,2]^2*var_c13[2]+p[i,3]^2*var_c13[3]
var_c14_mix[i] <- p[i,1]^2*var_c13[1]+p[i,2]^2*var_c14[2]+p[i,3]^2*var_c14[3]

tau_c13[i] <- 1/var_c13_mix[i]
tau_c14[i] <- 1/var_c14_mix[i]

mu_c13_mix[i] <- p[i,1]*mu_c13[1] + p[i,2]*mu_c13[2] + p[i,3]*mu_c13[3]
mu_c14_mix[i] <- p[i,1]*mu_c14[1] + p[i,2]*mu_c14[2] + p[i,3]*mu_c14[3]

x[i,1] ~ dnorm(mu_c13_mix[i],tau_c13[i])
x[i,2] ~ dnorm(mu_c14_mix[i],tau_c14[i])
}
}
```

Figure 10: Model 1 specification in rjags.

```
model {

#Prior for the proportions for the first and last station
p0 ~ ddirch(alpha[])
p1 ~ ddirch(alpha[])

#Second and third station
p2 <- (p0*(S-s[2])+p1*s[2])/S
p3 <- (p0*(S-s[3])+p1*s[3])/S

p[1,1:3] <- p0
p[2,1:3] <- p2
p[3,1:3] <- p3
p[4,1:3] <- p1

var_c13 <- sd_c13^2
var_c14 <- sd_c14^2

for (i in 1:n) {

var_c13_mix[i] <- p[i,1]^2*var_c13[1]+p[i,2]^2*var_c13[2]+p[i,3]^2*var_c13[3]
var_c14_mix[i] <- p[i,1]^2*var_c13[1]+p[i,2]^2*var_c14[2]+p[i,3]^2*var_c14[3]

tau_c13[i] <- 1/var_c13_mix[i]
tau_c14[i] <- 1/var_c14_mix[i]

mu_c13_mix[i] <- p[i,1]*mu_c13[1] + p[i,2]*mu_c13[2] + p[i,3]*mu_c13[3]
mu_c14_mix[i] <- p[i,1]*mu_c14[1] + p[i,2]*mu_c14[2] + p[i,3]*mu_c14[3]

x[i,1] ~ dnorm(mu_c13_mix[i],tau_c13[i])
x[i,2] ~ dnorm(mu_c14_mix[i],tau_c14[i])
}
}
```

Figure 11: Model 3 specification in rjags.

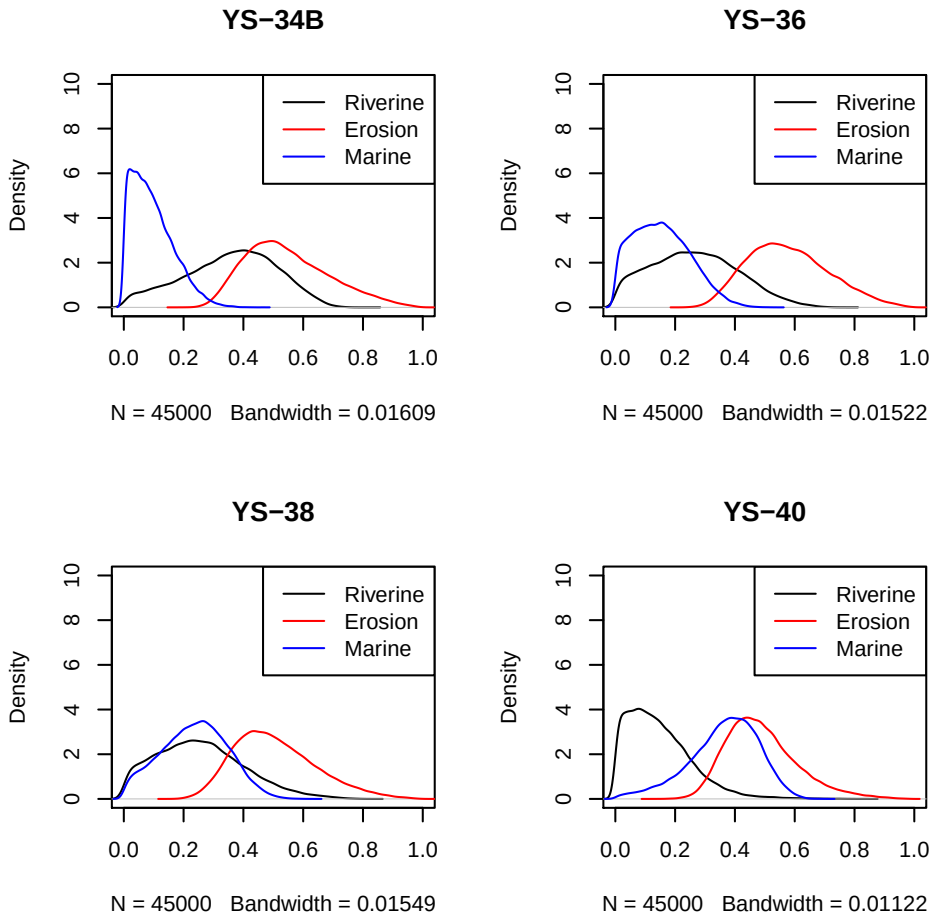

Figure 12: Plots of posterior densities model 1.

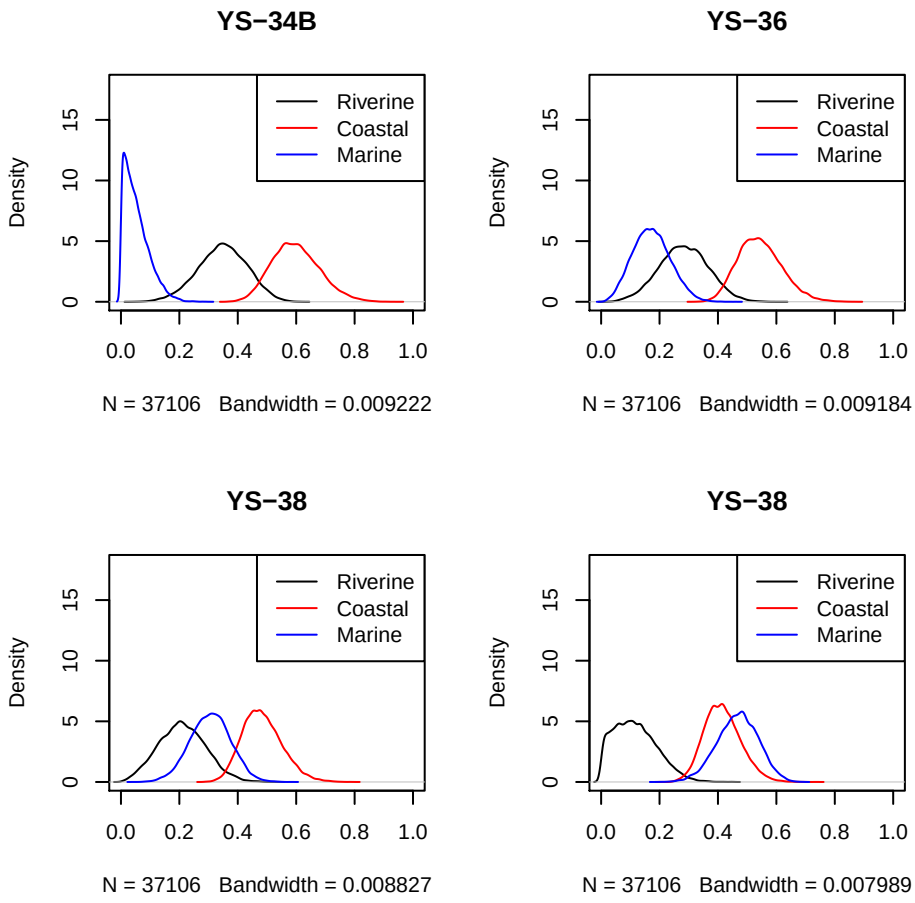

Figure 13: Plots of posterior densities model 2.

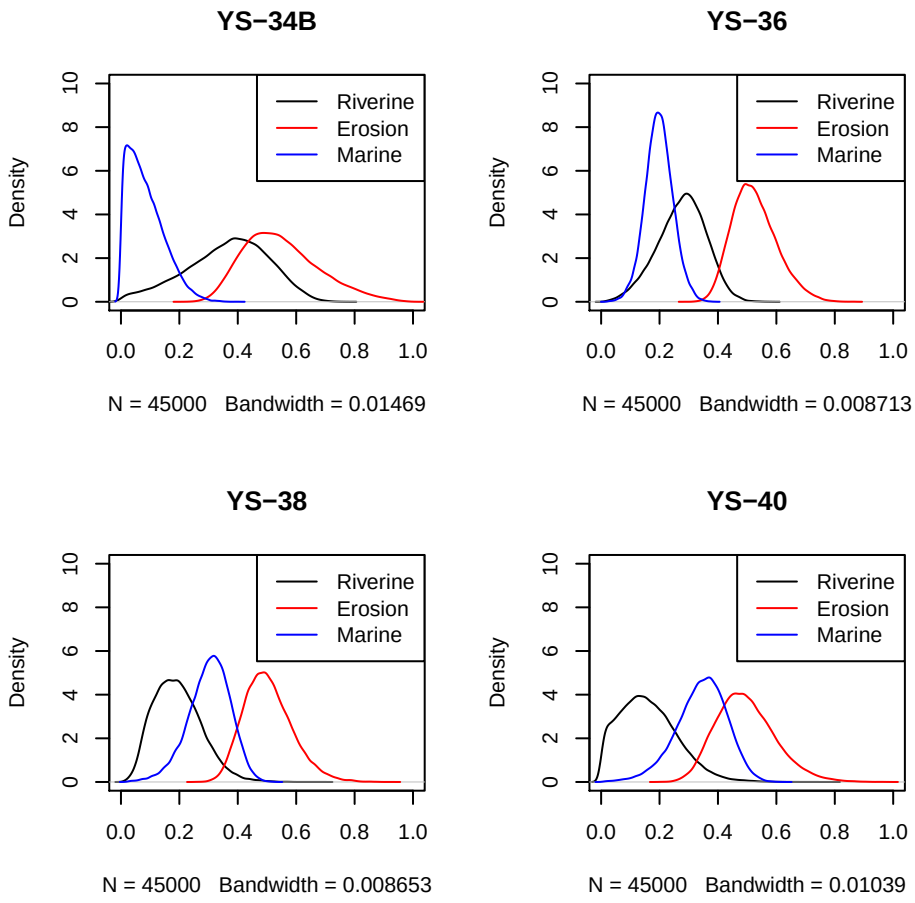

Figure 14: Plots of posterior densities model 3.

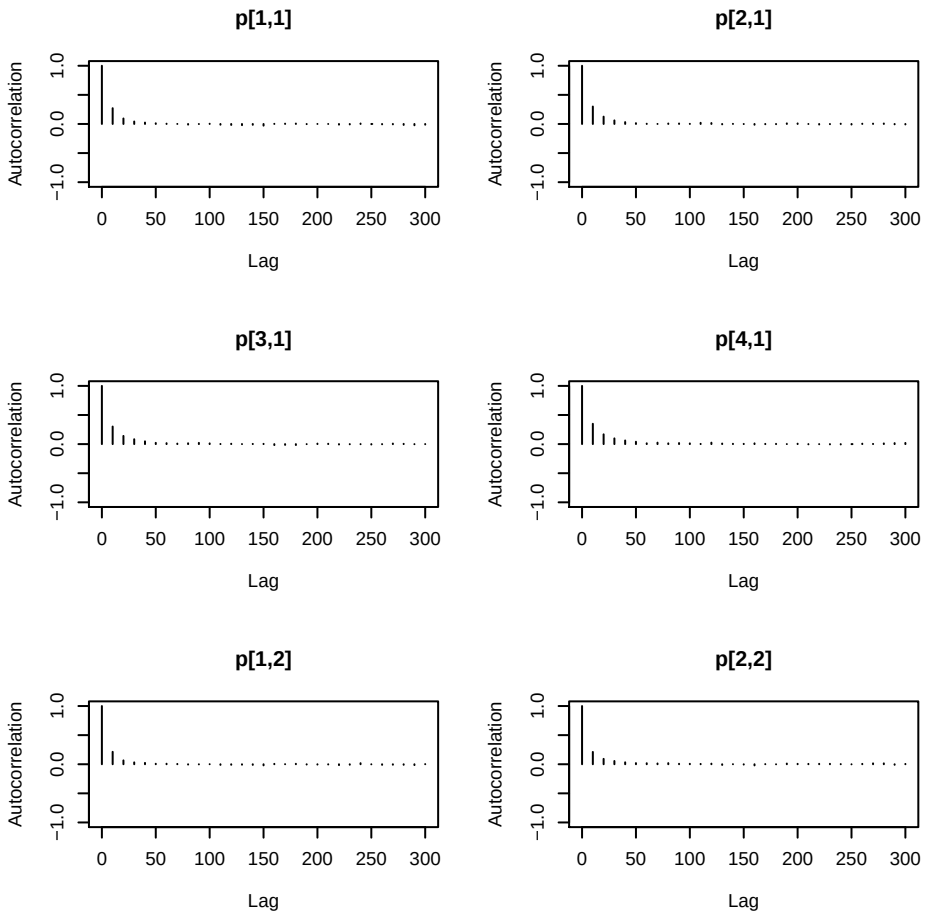

Figure 15: Acf plot model 1, chain 1

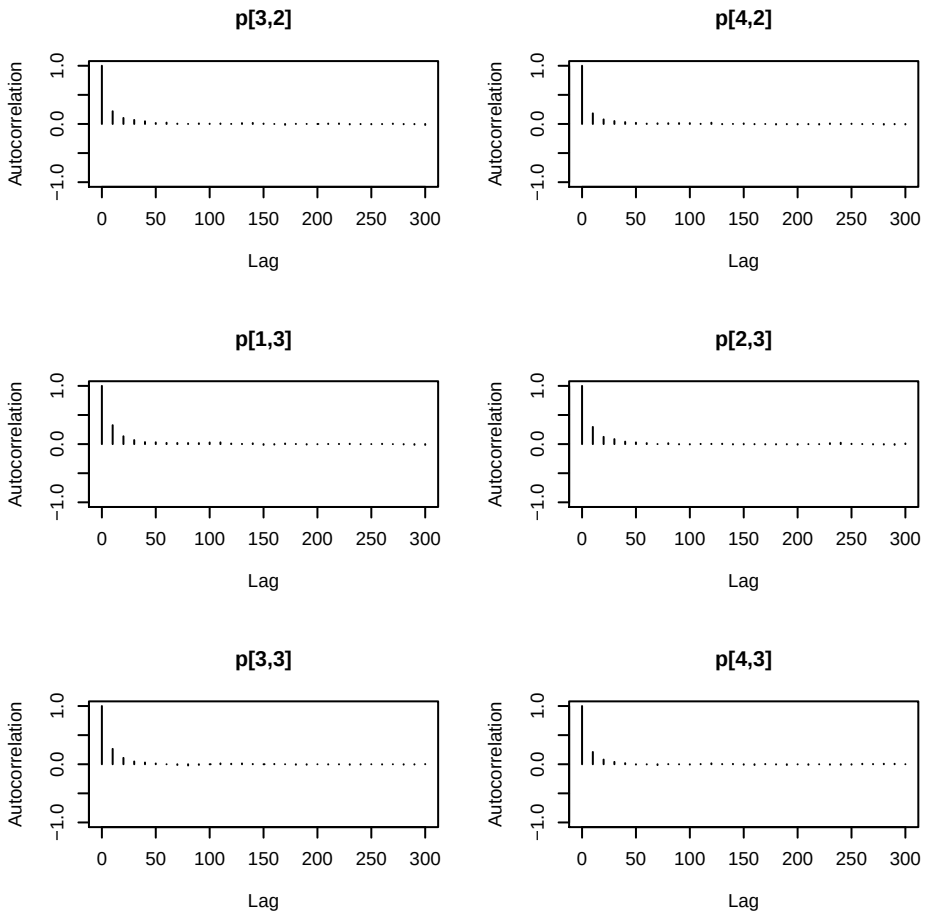

Figure 16: Acf plot model 1, chain 1

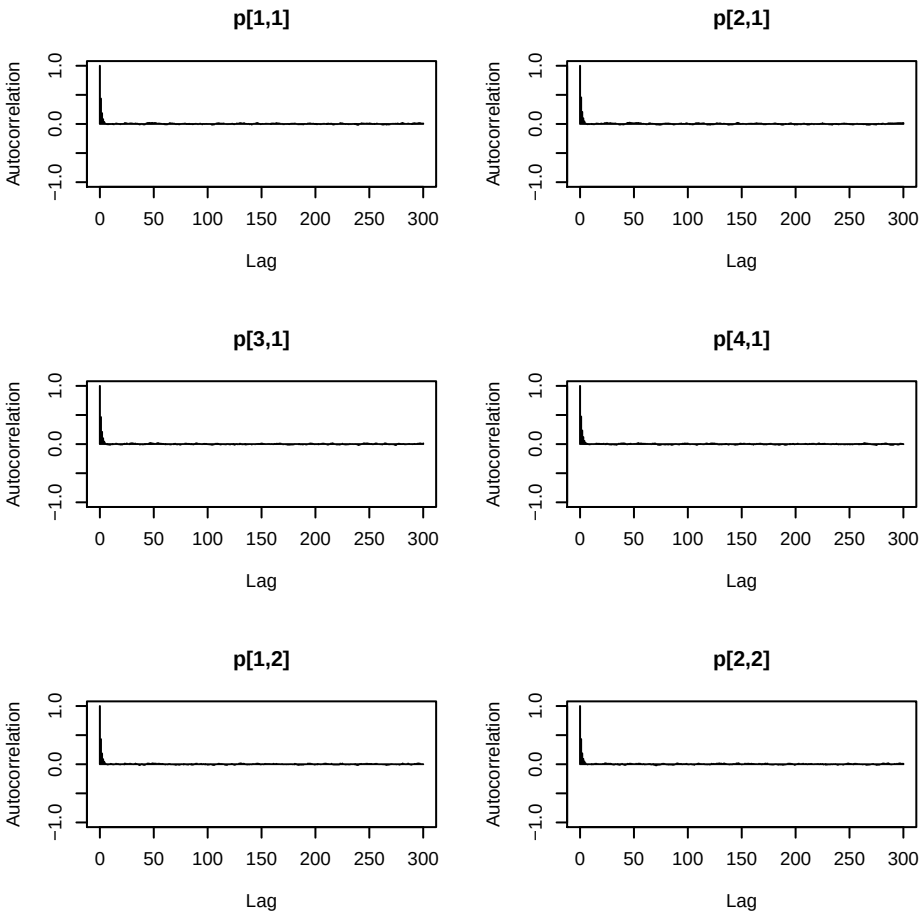

Figure 17: Acf plot model 2

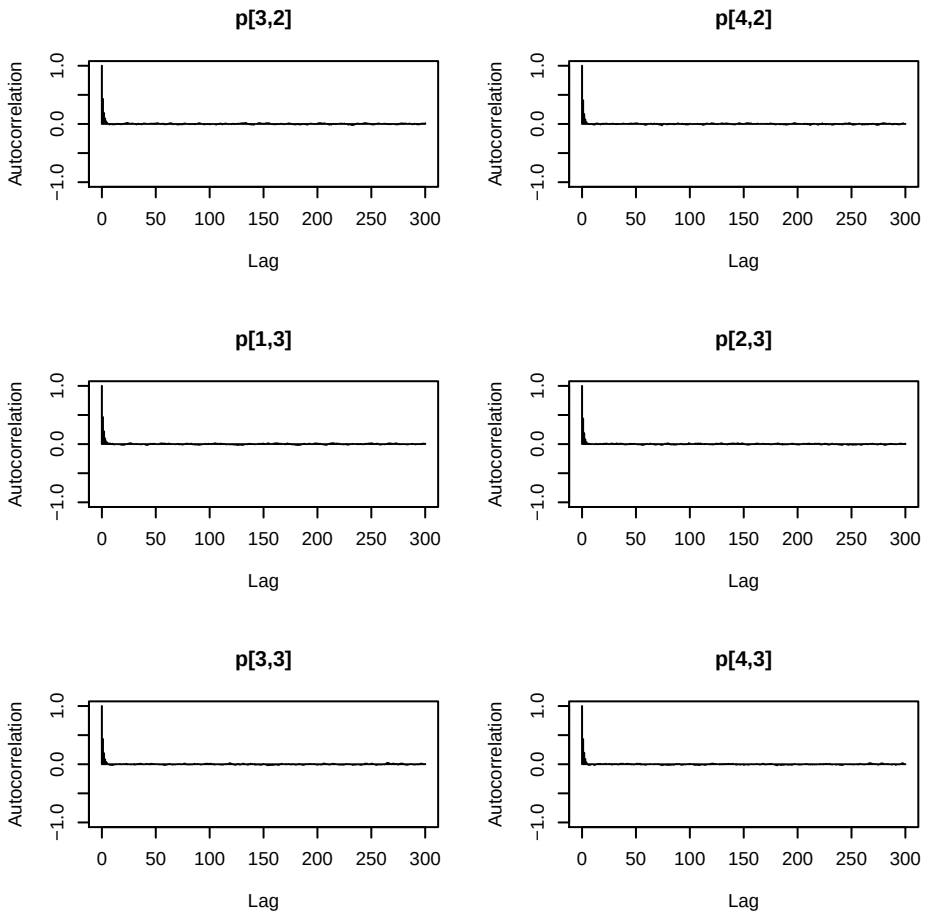

Figure 18: Acf plot model 2

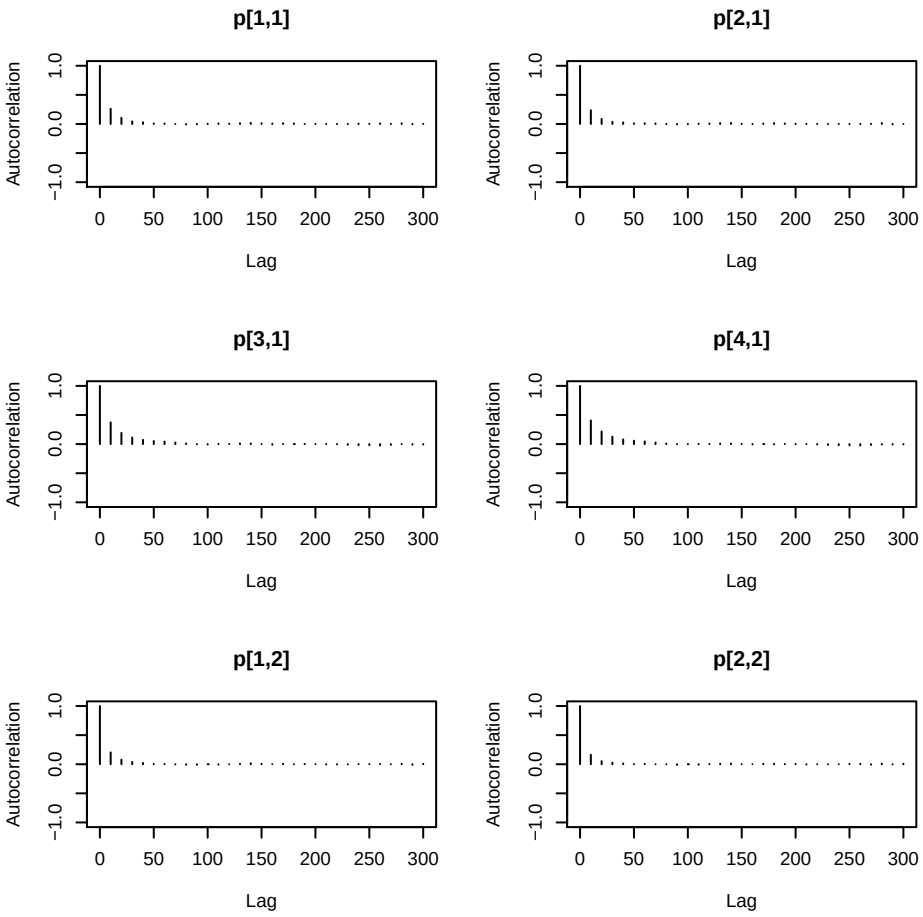

Figure 19: Acf plot model 3, chain 1

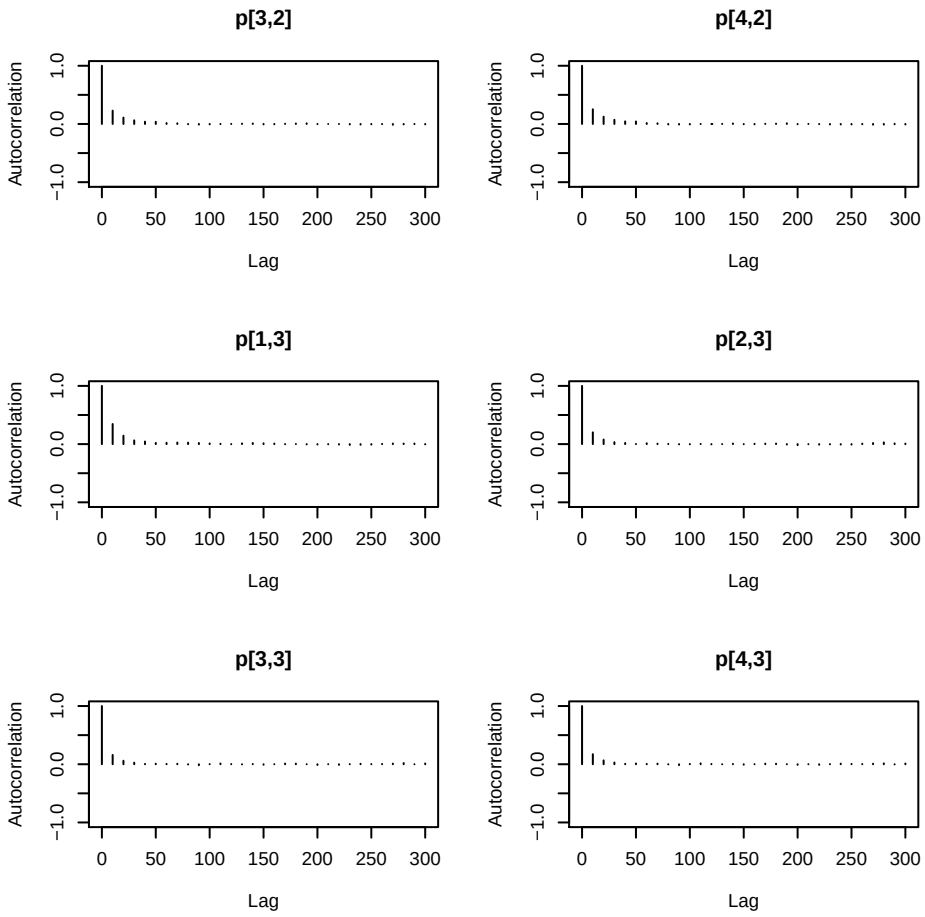

Figure 20: Acf plot model 3, chain 1

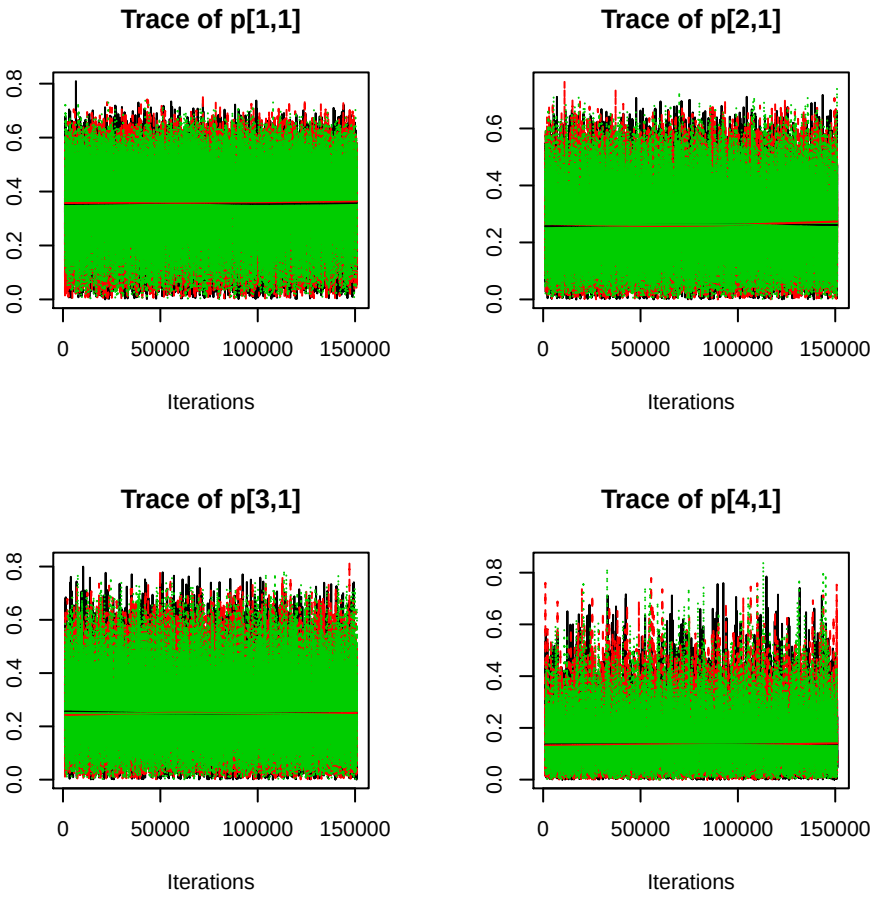

Figure 21: Trace plot from model 1 (riverine component)

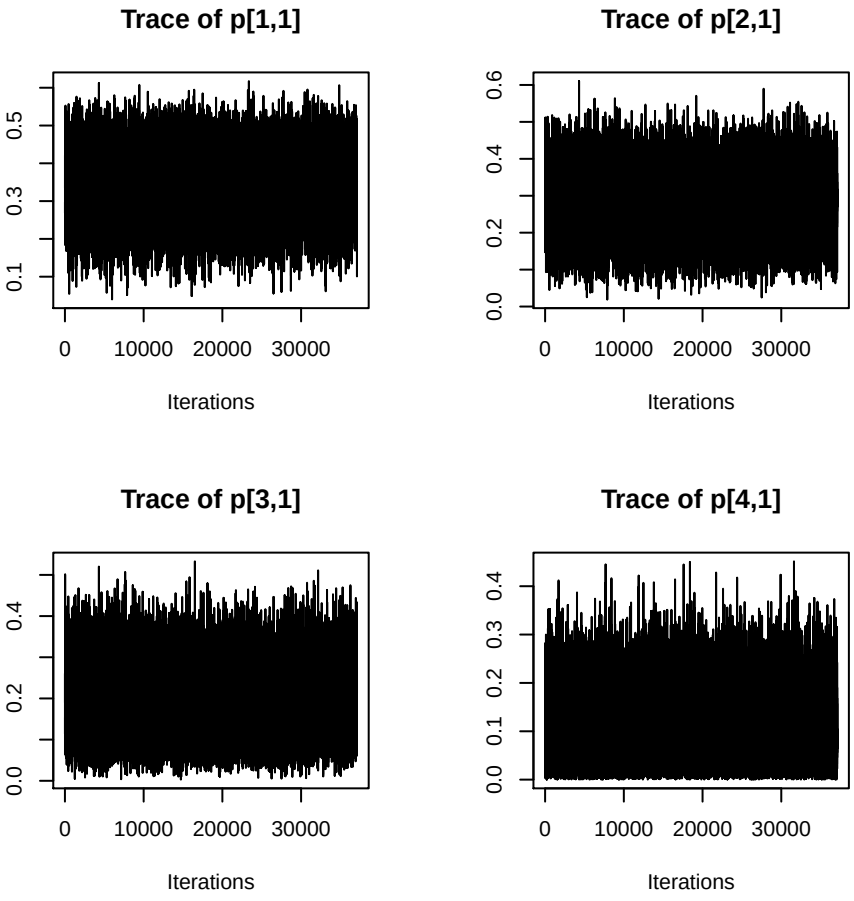

Figure 22: Trace plot model 2 (riverine component)

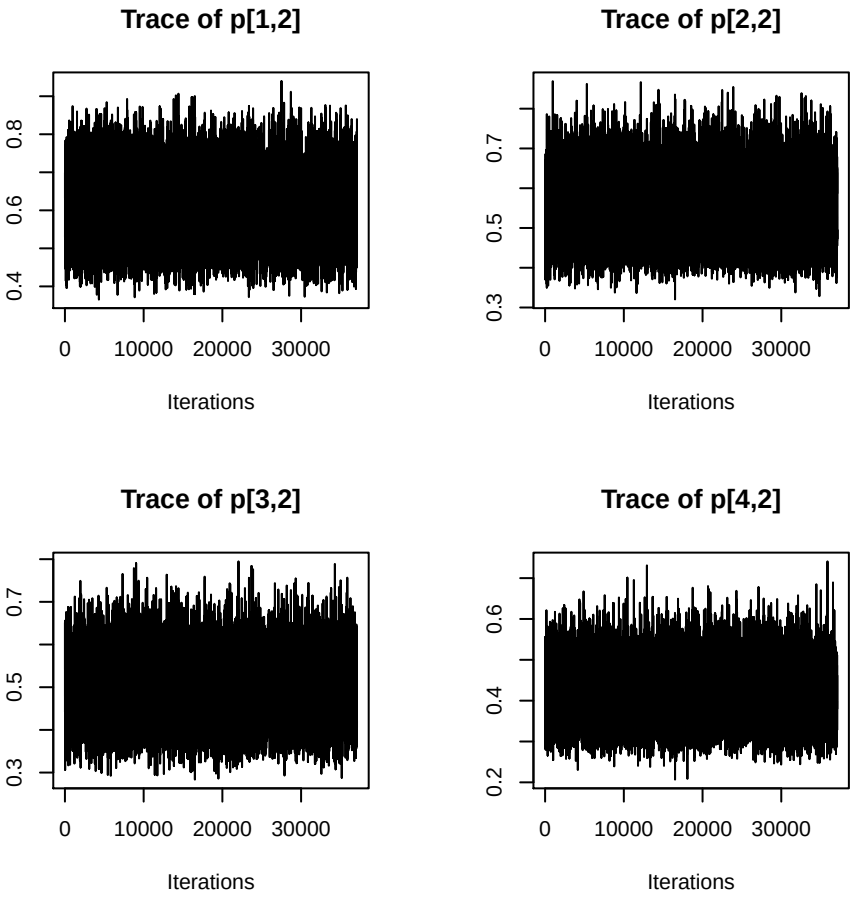

Figure 23: Trace plot model 2 (coastal erosion component)

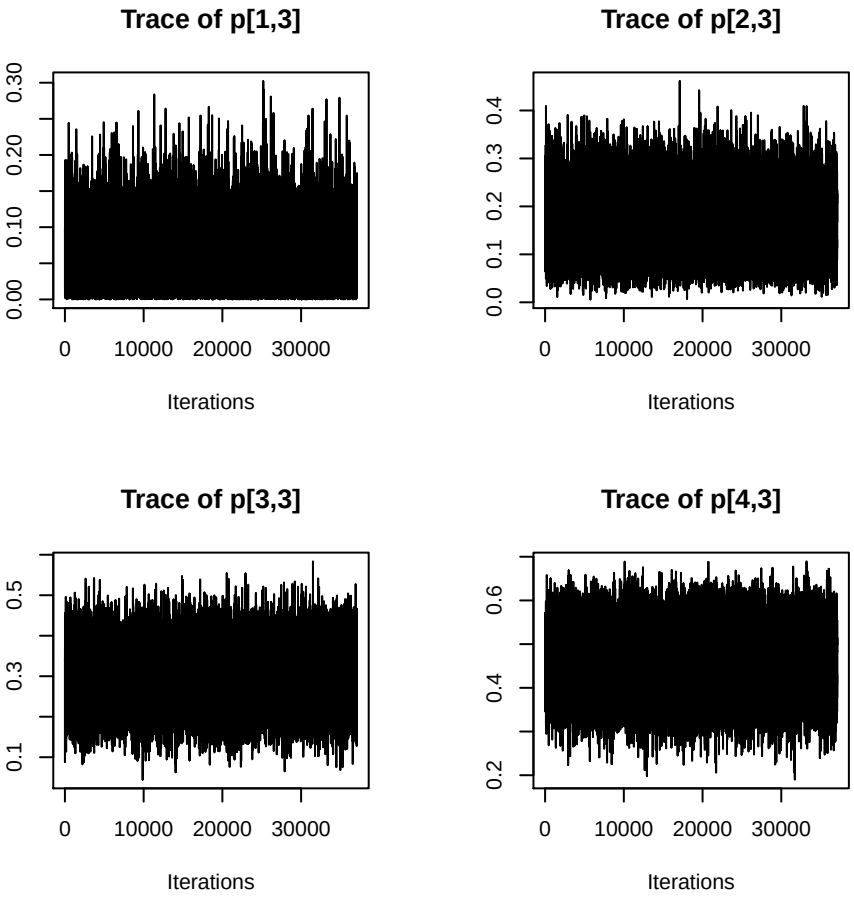

Figure 24: Trace plot model 2 (marine component)

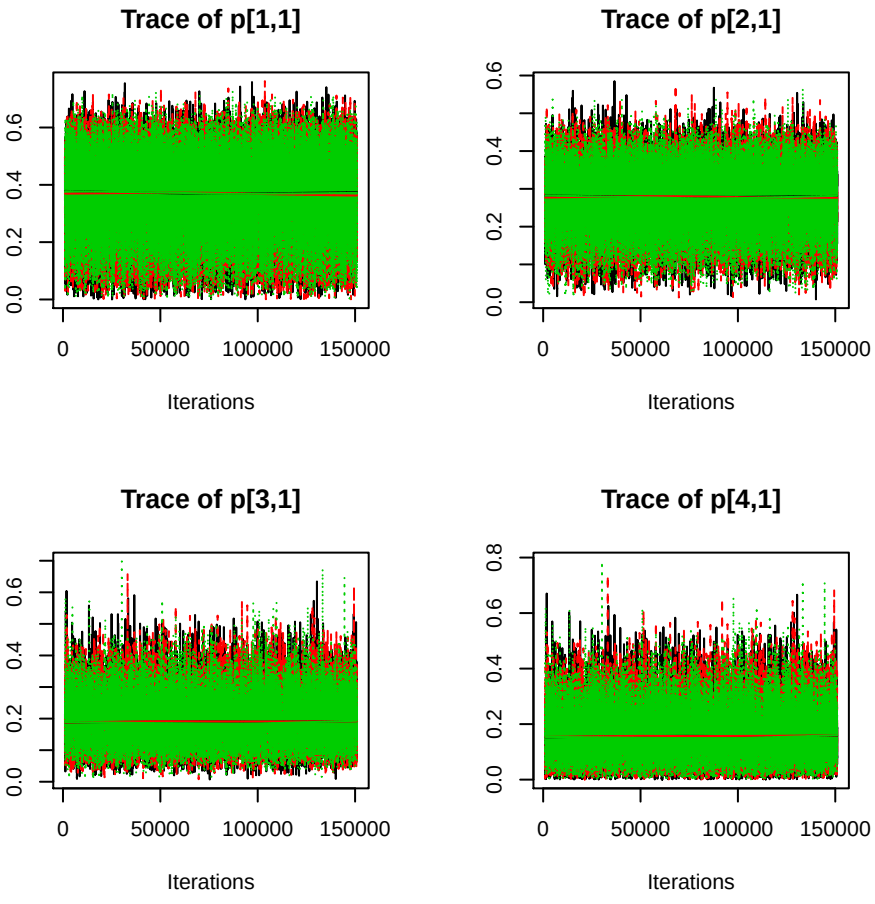

Figure 25: Trace plot from model 3 (riverine component)

Table 5: Estimated means and standard errors (adjusted for autocorrelation).

Model	Station	Source	Mean	SE
Model 1	YS-34B	Riverine	0.36	0.0010
		Erosion	0.55	0.0009
		Marine	0.10	0.0005
	YS-36	Riverine	0.27	0.0010
		Erosion	0.58	0.0008
		Marine	0.16	0.0007
	YS-38	Riverine	0.26	0.0010
		Erosion	0.51	0.0009
		Marine	0.24	0.0007
	YS-40	Riverine	0.15	0.0009
		Erosion	0.49	0.0008
		Marine	0.36	0.0007
Model 2	YS-34B	Riverine	0.35	0.0009
		Erosion	0.60	0.0009
		Marine	0.05	0.0005
	YS-36	Riverine	0.28	0.0010
		Erosion	0.54	0.0008
		Marine	0.17	0.0007
	YS-38	Riverine	0.21	0.0009
		Erosion	0.48	0.0008
		Marine	0.31	0.0008
	YS-40	Riverine	0.12	0.0008
		Erosion	0.42	0.0007
		Marine	0.47	0.0008
Model 3	YS-34B	Riverine	0.36	0.0010
		Erosion	0.55	0.0008
		Marine	0.09	0.0005
	YS-36	Riverine	0.28	0.0005
		Erosion	0.53	0.0005
		Marine	0.20	0.0003
	YS-38	Riverine	0.19	0.0007
		Erosion	0.50	0.0006
		Marine	0.30	0.0004
	YS-40	Riverine	0.16	0.0008
		Erosion	0.49	0.0007
		Marine	0.34	0.0005