**Atmospheric
Chemistry
and Physics**

# Aerosol model selection and uncertainty modelling by adaptive MCMC technique

**M. Laine and J. Tamminen**

Finnish Meteorological Institute, Helsinki, Finland

**Abstract.** We present a new technique for model selection problem in atmospheric remote sensing. The technique is based on Monte Carlo sampling and it allows model selection, calculation of model posterior probabilities and model averaging in Bayesian way.

The algorithm developed here is called Adaptive Automatic Reversible Jump Markov chain Monte Carlo method (AARJ). It uses Markov chain Monte Carlo (MCMC) technique and its extension called Reversible Jump MCMC. Both of these techniques have been used extensively in statistical parameter estimation problems in wide area of applications since late 1990's. The novel feature in our algorithm is the fact that it is fully automatic and easy to use.

We show how the AARJ algorithm can be implemented and used for model selection and averaging, and to directly incorporate the model uncertainty. We demonstrate the technique by applying it to the statistical inversion problem of gas profile retrieval of GOMOS instrument on board the EN-VISAT satellite. Four simple models are used simultaneously to describe the dependence of the aerosol cross-sections on wavelength. During the AARJ estimation all the models are used and we obtain a probability distribution characterizing how probable each model is. By using model averaging, the uncertainty related to selecting the aerosol model can be taken into account in assessing the uncertainty of the estimates.

## 1 Introduction

Advances in computer resources and algorithms have made the use of increasingly complicated models possible. In geophysical sciences the estimation of unknowns in large mod-els is commonly handled using linearizations and approximations that can effect the uncertainty estimates of the retrievals. Bayesian inference provides a unified and natural framework to consider uncertainty in the estimated values as well as the model uncertainty. In many cases, classical approximative estimation methods can be seen as special cases of some more general Bayesian analyses, see for example Kaipio and Somersalo (2004).

In Bayesian inference, the uncertainty of the estimated value is a primary target of the investigation. Whenever computationally possible, the result of the analysis is the full multi-dimensional posterior probability density of the unknowns. The approach allows the study of many kinds of uncertainties, including uncertainty in the model itself. Prior information from different sources can be directly incorporated and the correlation structure of the unknowns can be fully explored. Practical tools for applying Bayesian inference to modelling problems are provided by Markov chain Monte Carlo (MCMC) methods. MCMC is a common title for algorithms that simulate values from a probability distribution known only up to a normalizing constant. A typical case of such a task is to find the posterior distribution of the unknown parameters of a geophysical model. For application examples and more details on applying Bayesian MCMC methods in geophysical research see, for example, Tamminen and Kyrölä (2001); Tamminen (2004); Haario et al. (2004).

In this article the Bayesian model selection and averaging is applied to the GOMOS (ESA 2007) aerosol model selection problem. GOMOS (Global Ozone Monitoring by Occultation of Stars) is an instrument on board the Envisat satellite that uses stellar occultation to measure the atmosphere (http://envisat.esa.int/instruments/gomos/). The aerosol cross-section model in the GOMOS retrieval algorithm is an approximation of the underlying aerosol extinction process. Indeed, several alternate formulations are possible, depending on the types of aerosol at a given location. Consequently, it is advisable to allow for different types of

models and to use the data to decide which model to use. By adaptive MCMC methods this can be done as a part of general estimation procedure in a statistically correct manner.

This article introduces an adaptive MCMC method, called AARJ, for model selection problems. AARJ is an easy to use and efficient version of the Reversible Jump MCMC algorithm. We demonstrate the technique in the aerosol model selection of the GOMOS remote sensing instrument, but we emphasize that the method is general and applicable to general model selection problems. The structure of this article is the following. In Sect. 2 the basics of Bayesian model selection are reviewed. In Sect. 3 the MCMC method for simulating from a posterior distribution of model parameters is explained and an adaptive automatic reversible jump MCMC algorithm (AARJ) is introduced. The algorithm can be used for model determination problems where a number of different models are fitted and compared. The application example of GOMOS aerosol model selection is explained and the results of computer experiments are given in Sect. 4 and 5.

## 2 Bayesian model selection

Choosing the right model is a complicated matter that can not be solved by purely statistical considerations. Statistical methods can, however, tell if the chosen models and modelling assumptions are highly improbable for the situation and calculate relative merits of different modelling approaches. Here we present a method that is able to tell which of the possible solutions offer the best fit given the set of models to consider, the data observed, and the prior information that is available.

In many cases the ground truth is unknown. We could have several speculative alternative models describing the physical behavior of the system, e.g. depending on some unknown state of nature at the location under consideration. In such cases we can use several models and see if the fits they provide differ significantly. If no single model stands out, then this uncertainty can be taken into account in the results by averaging the predictions over the models according to their posterior weights.

We briefly introduce the main concepts of model determination in the Bayesian framework and discuss various probability distributions of the unknowns concerned. Let $x$ stand for a vector of unknown variables of primary interest and $\eta^{(k)}$ for extra unknown model parameters in the $k$:th model. We assume that $x$ is common to all the models. We want to use the observed data, $y$, to estimate the unknowns $x$ and $\eta^{(k)}$ and also make an inference about the unknown model $k$. In our case, the model index $k$ is a label for a finite set of pre-selected models. In the GOMOS example presented in Sect. 4, the symbol $x$ will stand for the constituent line densities and $\eta^{(k)}$ contains the aerosol cross-section parameters for four cross-section models $k=1, \ldots, 4$.

To apply Bayesian inference we need to assign prior probabilities jointly for all the unknowns, $p(x, \eta^{(k)}, k)$. It can be written as a product of conditional probabilities

$$p(x, \eta^{(k)}, k) = p(x|\eta^{(k)}, k) p(\eta^{(k)}|k) p(k). \tag{1}$$

This formulation reveals the hierarchical structure of the unknowns. Priors can be given sequentially by first assigning prior probabilities for different models, $p(k)$, then prior distributions for the model parameters $p(\eta^{(k)}|k)$ in each model and lastly the priors for the unknown variables $p(x|\eta^{(k)})$. In addition, we must formulate the likelihood function, $p(y|x, \eta^{(k)}, k)$, describing the distribution of the observations using the forward model and the statistical distribution of the observational error.

The joint posterior distribution of the unknowns $x$, $\eta^{(k)}$ and $k$ conditional to the observed data $y$ is given by the Bayes formula and can be written as a product of the likelihood and the priors:

$$p(x, \eta^{(k)}, k|y) = \frac{p(y|x, \eta^{(k)}, k) p(x|\eta^{(k)}, k) p(\eta^{(k)}|k) p(k)}{p(y)} \tag{2}$$

For the actual calculation of the posterior density we must solve the well known problem of computing the unconditional probability of the observations $p(y)$ in the denominator of the Bayes formula. As the observed data $y$ are fixed, the term $p(y)$ can be seen as a normalizing constant that makes the product of likelihood and prior to become a probability density function. This means that we can write

$$p(y) = \int p(y|x, \eta^{(k)}, k) p(x|\eta^{(k)}, k) p(\eta^{(k)}|k) p(k) \, d(x, \eta^{(k)}, k) \tag{3}$$

and the calculation involves averaging over all the unknown variables of the model, making it into an integration problem with dimension equal to the number of unknowns in the model. This integration is, in general, impossible without advanced Monte Carlo simulation techniques, like the MCMC.

Let us next consider the problem of selecting the best model $k$ from a set of competing models. Different models can be judged according to the evidence they give to the observations, i.e. we consider the probabilities:

$$p(y|k) = \int p(y|\theta^{(k)}, k) p(\theta^{(k)}|k) p(k) \, d\theta^{(k)}, \tag{4}$$

where $\theta^{(k)}=(x, \eta^{(k)})$ is used as a shorthand for the vector of all unknowns of the model $k$. The posterior model probabilities can be written using the Bayes formula as

$$p(k|y) = \frac{p(y|k) p(k)}{p(y)}. \tag{5}$$

If the values above are available, then model comparisons can be done using posterior odds:

$$\frac{p(k_1|y)}{p(k_2|y)} = \frac{p(y|k_1)}{p(y|k_2)} \frac{p(k_1)}{p(k_2)}, \tag{6}$$

where the first term on the right, $p(y|k_1)/p(y|k_2)$, is called the Bayes factor, the relative evidence of model $k_1$ wrt. model $k_2$ given by the data $y$ and $p(k_1)/p(k_2)$ is the ratio of prior model probabilities.

The calculation of model probability $p(k|y)$, and that of the evidence $p(y|k)$, poses challenges, especially if the class of models considered is large and if there is no natural hierarchy between the models that could be exploited. Several methods for the calculations have been proposed, either by using approximations that avoid the problems of high dimensional integration, or by using results of the MCMC runs on the individual models. The adaptive RJMCMC method, AARJ, presented below allows for a simple method of calculating the model posterior probabilities from an MCMC simulation done simultaneously over all the selected models.

## 3 Markov chain Monte Carlo – MCMC

In the most general setting we are interested in the whole posterior distribution of all the unknowns. Sometimes we are satisfied with some statistics of the distribution, such as the mean and standard deviation. The calculation of most statistics will lead, in general, to a high dimensional integration problem that has no closed form solutions. Markov chain Monte Carlo (MCMC) methods overcome the problems posed by high dimensional integrals by using high dimensional random walks.

The most important MCMC algorithm is the Metropolis-Hasting (MH) algorithm. It has several useful generalizations and important special cases for different purposes. The MH algorithm for sampling from a posterior distribution $p(\theta|y)$ can be described as follows. Again, we let $\theta$ stand for all the unknowns of our model, including unknown state variables, model parameters and the model index, $\theta=(x, \eta^{(k)}, k)$. Starting from an initial guess $\theta_0$ we generate a chain of possible parameter realizations $\theta_0, \theta_1, \ldots$. In each step $i$ with a current value $\theta_i$ we propose a new value $\theta^*$ using a proposal distribution $q(\theta_i, \cdot)$. As the notation suggests, this proposal can depend on the current value $\theta_i$. The proposal could be, for example, a multi dimensional Gaussian distribution centered at the current value $\theta_i$. The new value is accepted using an acceptance probability $\alpha(\theta_i, \theta^*)$ that depends on the ratio of the posteriors and on the chosen proposal distribution:

$$\alpha(\theta_i, \theta^*) = \min\left(1, \frac{p(\theta^*|y)q(\theta^*, \theta_i)}{p(\theta_i|y)q(\theta_i, \theta^*)}\right)$$
$$= \min\left(1, \frac{p(y|\theta^*)p(\theta^*)q(\theta^*, \theta_i)}{p(y|\theta_i)p(\theta_i)q(\theta_i, \theta^*)}\right) \quad (7)$$

If $\theta^*$ is accepted, we set $\theta_{i+1}=\theta^*$, otherwise the chain stays at the current value, that is $\theta_{i+1}=\theta_i$. If the proposal is symmetric, $q(\theta_i, \theta^*)=q(\theta^*, \theta_i)$, as it is the case with the Gaussian density, the $q$ functions cancel out in Eq. (7). A new value $\theta^*$ is then accepted unconditionally if it is better than

the previous value, i.e., if $p(\theta^*|y)/p(\theta_i|y)>1$. If it is not better in the above sense, then $\theta^*$ is accepted with a probability that is equal to the posterior ratio $p(\theta^*|y)/p(\theta_i|y)$. The MH algorithm can be thought as a random walker travelling uphill towards the peak of the posterior distribution, but frequently taking steps downhill, too.

The basic idea behind the MH algorithm is that instead of computing the values of the posterior $p(\theta|y)$ directly, we only need to compute ratios of the posteriors at two distinct parameter values, $p(\theta_2|y)/p(\theta_1|y)$. This cancels out the normalizing constant $p(y)$ and the parts of the likelihood function $p(y|\theta)$ that do not depend on $\theta$. Using standard Markov chain theory (for example Gamerman, 1997), it can be shown that this algorithm produces a chain of values whose distribution approaches the target posterior distribution $p(\theta|y)$. We might need to allow some burn-in time to let the chain reach the limiting distribution.

After the MCMC run we have a chain of values of the parameter vector at our disposal. The inference about the unknowns are made with statistics calculated from the chain of values. The mean of the chain is a Bayesian point estimate for the unknown, a histogram or a kernel density gives an estimate for the marginal posterior density. If we think of the generated chain as a matrix where the number of rows corresponds to the size of the MCMC sample and the number of columns corresponds to the number of unknowns in the model, then each row is a possible realization of the model and these appear in correct proportions corresponding to the posterior distribution. Plotting one-dimensional or two-dimensional scatter plots of the sampled parameter values from the chain produces representations of the respective marginal posterior densities.

### 3.1 Reversible jump MCMC

To include model selection into the MCMC framework a modification to the basic MH algorithm outlined above is needed. If we want the MCMC chain to explore different models and parametrizations, we must somehow allow the dimension of the unknown to change. This is the motivation behind the Reversible Jump MCMC (RJMCMC) algorithm by Green (1995). In the RJMCMC algorithm the proposal distribution and the acceptance probability are formulated in such a way that the chain can perform reversible jumps between spaces of different dimensions. This means, especially, that the random walk of the MH algorithm can simultaneously explore different models for the same data.

The RJMCMC algorithm can be presented in theoretical framework that extends the standard MH algorithm to a more general state space of the unknowns. We will not present the general theory, but refer to Green (1995). Instead, we show how the method can be succesfully implemented in a situation where we consider several different models for the same data. This approach is also based on the work of Green (2003) and is called automatic RJMCMC.

In automatic RJMCMC a special MCMC sampler is constructed that can jump between different models. For the MCMC chain to move from one model to another, we need a way to transform the model parameters. A simple but general way to do this this is the following. Suppose that for each model $k$, the target posterior distributions can be approximated by a mean vector $\mu_k$ and a covariance matrix $C_k = R_k^T R_k$, where $R_k$ denotes the Cholesky decomposition factor. These approximations are used to transform the unknowns in each model into approximately independent Gaussian variables and they thus provide a common scale to perform the transformations of the parameters between the models. Additionally, as seen below, the covariance matrix $C_k$ can be used to form the proposal distribution of the MH step of the algorithm.

Let again $\theta^{(k)}$ be the vector of all the unknowns in the model $k$ and let the dimension of $\theta^{(k)}$ be $n_k$. Assume that the chain is currently in the model $i$. Using the vector $\mu_i$ and the matrix $R_i$, we can compute a scaled and normalized version of the current chain value as

$$z_i = (\theta^{(i)} - \mu_i) R_i^{-1}. \tag{8}$$

The components of $z_i$ are now approximately independent Gaussian with unit variances. If the model $j$ has the same dimension as the model $i$, we have a simple transformation from the model space $i$ to the model space $j$ as

$$\theta^{(j)} = \mu_j + z_i R_j. \tag{9}$$

If the dimensions of the two models do not match, we either drop some columns of $z_i$ or add new dimensions to it using independent Gaussian random numbers, $u \sim N(0, I)$. The transformations can be written as

$$\theta^{(j)} = \begin{cases} \mu_j + [z_i]_1^{n_j} R_j & \text{if } n_i > n_j \\ \mu_j + z_i R_j & \text{if } n_i = n_j \\ \mu_j + \begin{matrix} z_i \\ u \end{matrix} R_j & \text{if } n_i < n_j. \end{cases} \tag{10}$$

Here $[z]_1^i$ means the first $i$ components of the vector $z$.

The MH acceptance probability for a move from the model $i$ to the model $j$ and from a parameter value $\theta^{(i)}$ to that of $\theta^{(j)}$ is calculated according to the RJMCMC theory. Let $p(i, j)$ be the probability to propose a jump to the model $j$ when the chain is currently at the model $i$, i.e., if the current model is $i$ then the next model is chosen with a draw from a proposal distribution $p(i, \cdot)$. If the model $j$ is selected, then the current parameter vector is transformed to the new model according to Eq. (10).

The acceptance probability for the RJMCMC sampler can be written as

$$\alpha(\theta^{(i)}, \theta^{(j)}) = \max \left( 1, \frac{p(y|\theta^{(j)}, j) p(\theta^{(j)}, j) p(j, i)}{p(y|\theta^{(i)}, i) p(\theta^{(i)}, i) p(i, j)} \frac{|R_j|}{|R_i|} g \right), \tag{11}$$

where $|R|$ is the determinant of the matrix $R$ and the last term $g$ depends on the extra variable $u$ and is given as

$$g = \begin{cases} \phi(u) & \text{if } n_i > n_j, \\ 1 & \text{if } n_i = n_j, \\ \phi(u)^{-1} & \text{if } n_i < n_j, \end{cases} \tag{12}$$

where $\phi$ is the probability density function of independent multi dimensional Gaussian values, $N(0, I)$. Figure 1 illustrates the model moves. Note that when moving from one model to another with equal dimension, the transformation is totally deterministic, no random variables are used to make the move. To introduce more randomness, Green (2003) suggests a random permutation of the components of the normalized $z$ variables at each step. This permutation, if used, does not change the acceptance probability.

For a move inside the same model we use a Gaussian proposal distribution and the standard MH acceptance probability Eq. (7). The approximation of the posterior provided by the matrix $C_i = R_i^T R_i$ is used to make the proposal to have a correlation structure similar to that of the target distribution. If $\xi$ is a random vector of independent Gaussian random variables $\xi \sim N(0, I_{n_i})$, then the proposed value can be written as

$$\theta^{(i)*} = \theta^{(i)} + \xi R_i \sqrt{s}, \tag{13}$$

where $s = 2.4^2/n_i$ is a scaling factor. The acceptance probability Eq. (11) simplifies to that of the standard MH algorithm for symmetric proposal. The scaling constant 2.4 is chosen according to Gelman et al. (1996).

This sampler is easy to implement. Its success depends on how well the Gaussian approximations are able to provide decent proposals for moves from model to model. It is, however, typical in many geophysical applications to have parameter posteriors close to Gaussian. This also is the reason why the classical estimation methods often work quite well. But the use of RJMCMC allows us to incorporate model selection methods together with prior information, such as positivity or smoothness constraints, in a statistically sound manner. Also, we are able to properly deal with nonlinear correlation structures that usually are not found by the classical methods.

## 3.2 Adaptive automatic RJMCMC – AARJ

From a practical point of view the problem with the standard MCMC algorithms is that, in spite of the apparent simplicity of the basic algorithm, it still needs some problem specific tuning. The most important aspect is the choice of the proposal distributions $q$. In the MH algorithm the proposal can,
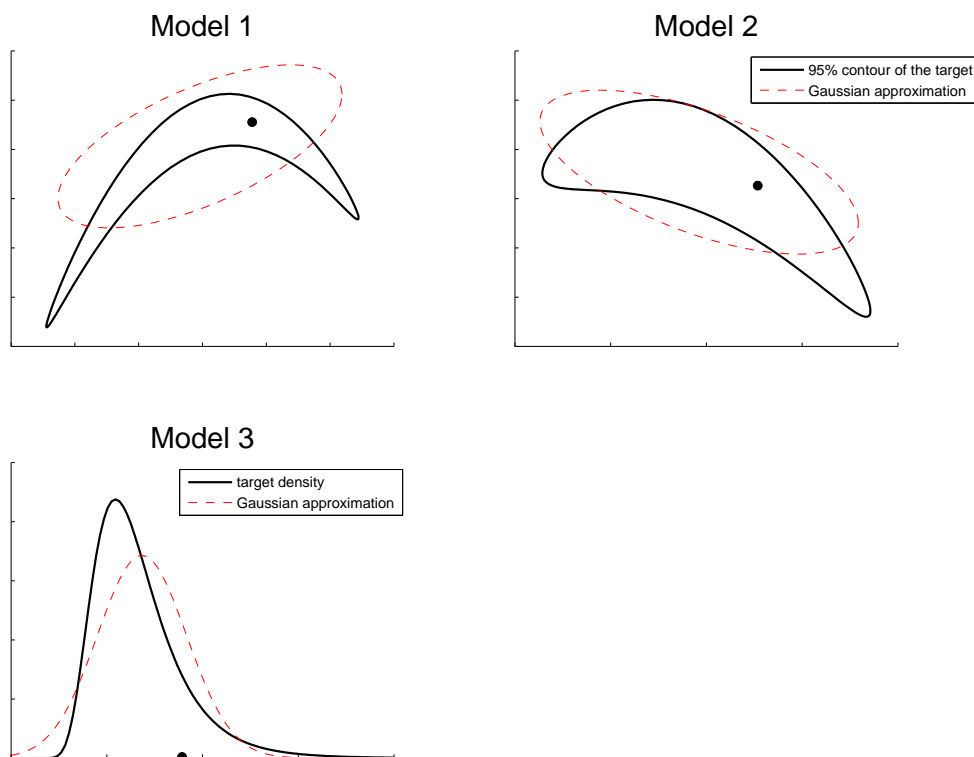
## Model 1

## Model 2

## Model 3

**Fig. 1.** Illustration of the model to model transformations in the automatic RJMCMC algorithm. The contours in 2-dimensional Models 1 and 2 represent 95% probability limits of the distributions. Model 3 is 1-dimensional and is illustrated by its density function. Solid lines give the (unknown in applications) true non-Gaussian density and broken lines the corresponding Gaussian approximations. The dots are values that have the same canonical coordinates given by the covariance matrix of the Gaussian approximation. The arrows shown one possible path from Model 1 to Model 2 and from Model 2 to Model 3. In the AARJ method these approximations are updated as more information on the true target becomes available from the generated MCMC chain.

at least in theory, be quite arbitrary. Choosing a distribution that closely resembles the true posterior distribution can dramatically speed up the convergence of the generated values to the right distribution. The closer the proposal distribution is to the actual posterior distribution $p(\theta|y)$, the better the chain "mixes" and the better a short sequence represents a draw from the posterior. This is especially true in multidimensional cases and when the components of the parameter vector are correlated. A general and computationally efficient choice for the proposal distribution is the multidimensional Gaussian density. As the shape of the Gaussian density is determined by its covariance matrix, the tuning of the algorithm in this case means the selection of the covariance.

In the basic MH algorithm the proposal distribution must not depend on the values generated so far, except for the current value. This is the requirement behind the Markov property of the stochastic process that the MCMC sampler defines. If we allow for adaptation depending on the history, the convergence theorems based on Markov chain theory must be checked. Numerous adaptive strategies for the choice of the proposal distribution have been suggested. In our experiences, the Adaptive Metropolis (AM) and the Delayed Rejection Adaptive Metropolis (DRAM) have proved to per-

form well in several geophysical and environmental modelling applications (Haario et al., 2001, 2006, 2004). These two methods are the building blocks for the new adaptive RJMCMC method presented below, for which we use the acronym AARJ.

In the AM adaptation the Gaussian proposal distribution is tuned using an increasing part of the chain values generated so far. In Haario et al. (2001) this method is shown to be ergodic, so it can be used to accurately sample from the target distribution. A recursive formula for the covariance matrix can be used to ease the computations. The DRAM adaptation (Haario et al., 2006) adds a new component to the AM method that is called Delayed Rejection (DR, Mira, 2001). In the DR method, instead of one proposal distribution we can have several proposals. These can be used in turn, until a new value is accepted. The DR acceptance probability formulation ensures that the generated chain is Markovian and that the so-called reversibility condition holds. This means that all the standard MH distributional convergence statements hold. In the DRAM method the DR algorithm is used together with several different adaptive Gaussian proposals.
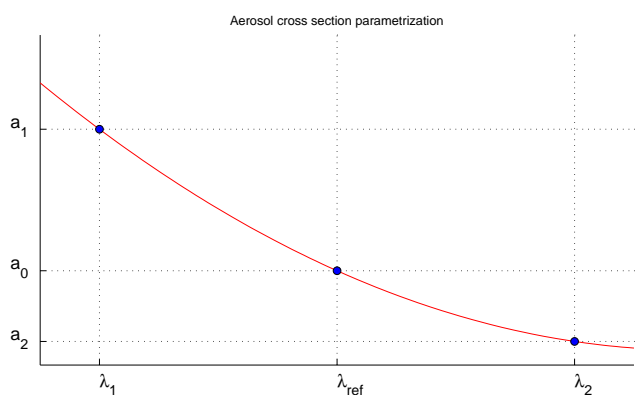
**Fig. 2.** Aerosol model parametrization. Each model is parametrised in such way that the parameters correspond to aerosol extinction at one selected wavelength, 300, 500 and 600 nm for three parameter models and 500 nm for one parameter model. This way we can also require positivity for these values and assure that the resulting estimates provide physically meaningful values.

This helps the algorithm in two ways. First, it enhances the adaptation by providing accepted values that make the adaptation start earlier. Second, it allows the sampler to work better for non Gaussian targets and with non linear correlations between the components. The ergodicity of the DRAM method is proven by Haario et al. (2006).

A new feature presented in this article is the combination of the DRAM and AM adaptations with the automatic RJMCMC. The practical application presented is the aerosol model selection in the GOMOS inversion. Hastie (2005) has also suggested a combination of adaptation and automatic RJMCMC of Green. The adaptation method (so called Adaptive Acceptance Probability, AAP) used in his work is, however, different from the adaptation used here. We regard our AARJ method to be more general and easily applicable to high dimensional nonlinear models typical in geophysical problems.

### 3.3 The AARJ algorithm

Here we present a schema for the algorithm for AARJ, an Adaptive Automatic Reversible Jump MCMC for model selection and model averaging problems with a fixed number of models $M_1, \ldots, M_k$.

**The algorithm**

1. Run separate adaptive MCMC chains using the DRAM method for all the proposed models. Collect the mean vectors $\mu^{(i)}$ and the Cholesky factors $R^{(i)}$ of the covariance matrices of the chains, $i=1, \ldots, k$.

2. Run automatic RJMCMC using the target approximations $\mu^{(i)}$ and $R^{(i)}$:

    (a) Assume that the current model is $i$. Select a new model $j$ from distribution $p(i, j)$.

    (b) If $j \neq i$, transform the model parameters from $i$ to $j$ according to Eq. (10) and calculate the acceptance probability according to equation Eq. (11).

    (c) If the current model is kept, $i=j$, propose a new value according to the standard random walk MH with Gaussian proposal distribution as in Eq. (13). The acceptance probability is again as in Eq. (11), however, all but the first terms in the nominator and denominator cancel out.

3. After given (random or fixed) intervals, update the approximations $\mu^{(i)}$ and $R^{(i)}$ for each model by the AM method using those parts of the chain that belong to the particular model.

### 3.4 Computational considerations

The AARJ method is easy to implement. For example, a computer program running the basic MH MCMC simulation can readily be extended to do both DRAM and AARJ. The GOMOS application example below has been coded in Matlab programming environment, using a MCMC toolbox for Matlab (Laine, 2008).

The AARJ algorithm in itself does not add much to the computational burden. The extra calculations involve only some matrix vector products. The forward model calculations needed to calculate the model likelihoods are those that take the most CPU cycles. As several models are tried, some of which possibly do not fit the data, the overall chain will have larger rejection rate than the individual chains would have, and it will, thus, need to be run longer. That is one reason why the adaptivity is needed to make the sampler as efficient as possible.

The algorithm can be applied efficiently even when the competing models are more complex that the ones used in the following example. More complex model, of course, increases the computational burden. If the computations become too heavy for routine use of the AARJ method, it can be used to validate more simple model selection criteria, such as the Deviance Information Criteria (DIC, Spiegelhalter et al., 2002). For the GOMOS problem, for example, we found that the AARJ results agree reasonably well with the DIC which can be calculated directly from separate individual MCMC runs.

## 4 Application: GOMOS aerosol model

To demonstrate the use of MCMC in model selection, we apply the AARJ method to aerosol modelling in the GOMOS retrieval. The forward model is the standard GOMOS model for the spectral transmission according to the Beers law. It is described for example by Bertaux et al. (2000). The cross section that is used for aerosol line density is, however, only an approximation of the underlying aerosol extinction process that actually depends on many unknown factors. The

cross-section is typically modelled by using a function that behaves like $1/\lambda$, where $\lambda$ is the wavelength. See Vanhellemont et al. (2006) for a comparison of different aerosol extinction models for the GOMOS inversion studied using simulated transmission data.

Here we consider four different aerosol cross section models: the standard (operational) $1/\lambda$ model (model 1), a second degree polynomial in $\lambda$ (model 2), $1/\lambda^2$ dependence (model 3), and a second degree polynomial in $1/\lambda$ (model 4). The aerosol models are parametrized using the aerosol extinction at 500 nm (models 1 and 2) or at 300, 500 and 600 nm (models 3 and 4), see Fig. 2. A positivity prior constrains these values. We concentrate on inverting the integrated line densities from the transmission spectra. This is called the spectral inversion step in the GOMOS literature. The so called vertical inversion of transforming the line densities to the actual constituent densities is a linear operation that is done after the line densities for all the heights that have been inverted and is not considered here.

Let $N$ be the vector of integrated line densities of the constituents to be retrieved ($O_3$, $NO_2$, $NO_3$, air, aerosols) and matrix $\alpha$ the corresponding cross sections. The cross section of aerosol depends on the model parameters $\eta^{(k)}$. The model for the observed transmission $T$ is written as

$$T(\lambda, z) = \frac{I(\lambda, z)}{I_0(\lambda)} = \exp\left(-\alpha(\eta^{(k)})N\right) + \epsilon(\lambda),$$
$$\text{with } \epsilon(\lambda) \sim N(0, \sigma_k^2 w_\lambda^2). \tag{14}$$

Here $I_0(\lambda)$ is the spectral intensity measured at a reference height above the atmosphere and $I(\lambda, z)$ is the intensity measured at the tangent height $z$. As the chosen aerosol model will affect the size of the residuals, the error variance is assumed to be of form $\sigma_k^2 w_\lambda^2$, with known weights $w_\lambda$ for each wavelength $\lambda$ and model dependent unknown scalars $\sigma_k^2$, which are also estimated by the MCMC.

The likelihood function assumes the form

$$p(T|N, \eta^{(k)}, \sigma_k^2) \propto \exp\left(-\frac{1}{2\sigma_k^2} SS(N, \eta^{(k)})\right), \tag{15}$$

where $SS(N, \eta^{(k)})$ is the weighted sum of squares,

$$SS(N, \eta^{(k)}) = \sum_\lambda \left(\frac{T(\lambda) - \exp\left(\alpha(\eta^{(k)})N\right)}{w_\lambda}\right)^2. \tag{16}$$

As for priors, only positivity constraint for the line densities is used. For the unknown error variance factors, $\sigma_k^2$, a weakly informative inverse Gamma prior is used (Gamerman, 1997). All the four models are taken, a priori, to be equally likely. A prior for the neutral air would probably help the identification of the aerosol model as the aerosol and neutral air cross-sections resemble each other and thus produce correlated estimates. Note that in the operational GOMOS processing air density is fixed to values provided by European Centre for Medium-Range Weather Forecasts (ECMWF).
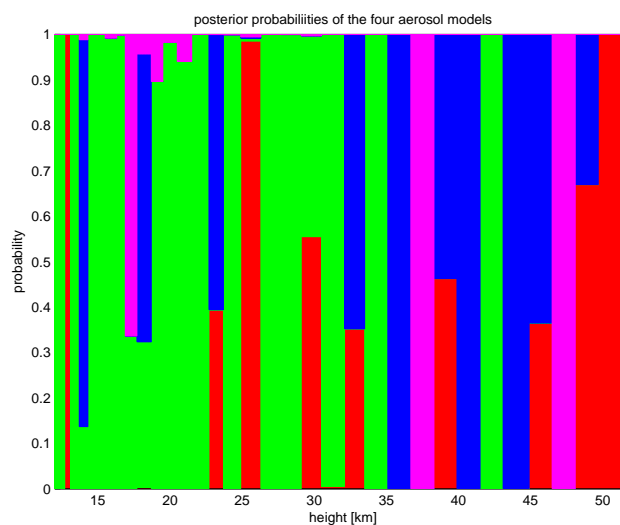


**Fig. 3.** An AARJ run is performed for each height in one GOMOS occultation. The posterior model probabilities are calculated for the four models at each height. The colours show how the different cross sections models are preferred depending on the altitude. The colouring is the same as in Figs. 4 and 5, Model 1: red, model 2 green, model 3 blue, model 4 magenta.

## 5  Results

For each line of sight (tangent height), and given one fixed aerosol model, the problem of inverting the line densities from the transmittance is a nonlinear problem with 5 unknowns. This is a fairly easy problem, assuming we have appropriate initial guesses and the noise level in the transmission spectra is low. The estimation problem can be solved in a least-squares sense as a nonlinear optimization problem using, e.g., the Levenberg-Marquardt method. This is basically the method used in the operational GOMOS algorithm. In this article we use MCMC to replace the operational inversion and take in account the model uncertainty. The MCMC method can also be extended to a one step solution, where all the heights are solved simultaneously, with regularization (smoothness) priors on the vertical structure of the profiles, see e.g. Haario et al. (2004).

To use the AARJ method for model selection we use the following strategy. First, for each occultation height and for each aerosol model, separate MCMC runs are performed using the DRAM method (Haario et al., 2006) to find the individual posterior distributions. Chains of length 10 000 each were generated and the last 5000 values were used for subsequent analyses. From the MCMC chains of these runs the mean vectors and covariance matrices together with their Cholesky factors are calculated to produce the mean vectors $\mu_i$, and Cholesky factor matrices $R_i$, $i = 1, \ldots, 4$ needed in the RJMCMC stage. Second, an MCMC run is done for a chain of length 50 000 using the AARJ algorithm for further adaptation of the approximations. The resulting chains are
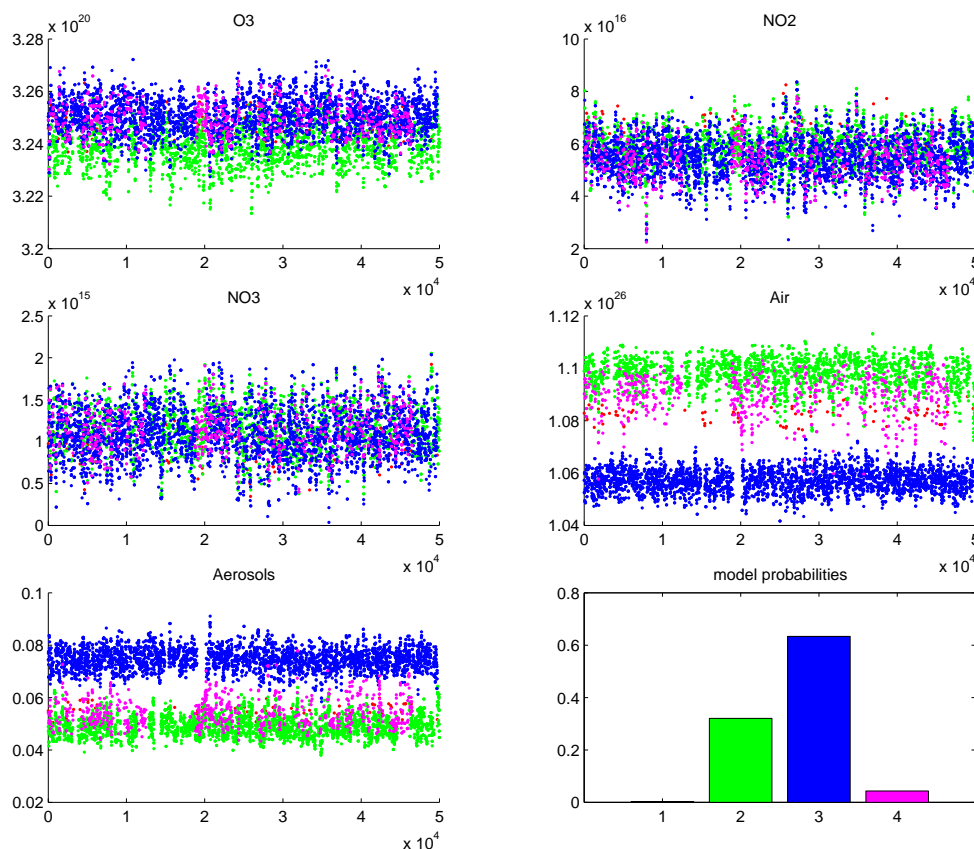
**Fig. 4.** The MCMC chains for the line densities for one selected GOMOS occultation. The horizontal axis runs with the simulation indexes, vertical axis being the simulated and accepted values for the line density for each constituent. The color indicates in which model the algorithm is in each step. Plot on the lower left corner labeled "Aerosol" show the relative aerosol extinction at 500 nm for all models. The last plot shows relative times spent in each model. Of the total 50 000 MCMC simulations of this particular run the models 1, 2, 3 and 4 are visited 129, 16 035, 31 679 and 2157 times, which makes the corresponding marginal model posterior probabilities $p(k_i|y)$, $i=1,\ldots,4$ to be 0.003, 0.321, 0.634, and 0.043.

visually investigated using 1-D plots like those in Fig. 4, in order to judge if the chains have converged. Some automatic convergence criteria could be used as well.

For the model selection, we calculate the relative times the MCMC chain has spent in each model. In Fig. 3 the results for each altitude of one GOMOS occultation are shown. For most of the heights one model stands out as the main candidate, but no single model can be used for all the heights. For altitudes from 14 to 22 km the second order polynomial (Model 2, coloured green) is prevailing. Each of the four models become selected as the most probable one at some of the altitudes. The second order polynomial over $1/\lambda^2$ (Model 4, magenta) seems to be less favoured. The amount of aerosols above 30 km is typically very low and the choice of the aerosol model does not significantly affect the retrievals. Certainly, a more thorough investigation would be needed to determine the relative merits of different aerosol models for the GOMOS inversion algorithm.

As an illustration of the model averaging we select one altitude at about 18 km where all the four models have gained

some posterior probability. Figures 4, 5 and 6 show the MCMC chains, the estimated posterior distributions and the fitted cross-sections for this selected altitude. Model averaging is useful when the best model can not be determined. The model used for estimation is then a mixture of different models each weighted according to its posterior weight. The uncertainty in the model is taken into account in the predictions and in the posterior inference for the constituents. In Fig. 6 the uncertainty in the cross-section of each model is illustrated. The cross-section curve is calculated for each model parameter in the MCMC chain. Then the corresponding posterior distribution for each wavelength is estimated. Together these provide predictive envelopes of the aerosol extinctions. These are drawn as different grey regions in the plots.

Figure 5 reveals the effect of the aerosol model on other retrievals. The plots show the marginal posterior distributions of the constituent line densities separately for each model and the posterior distribution of the averaged model. For the retrieval of ozone the difference between posterior mean of
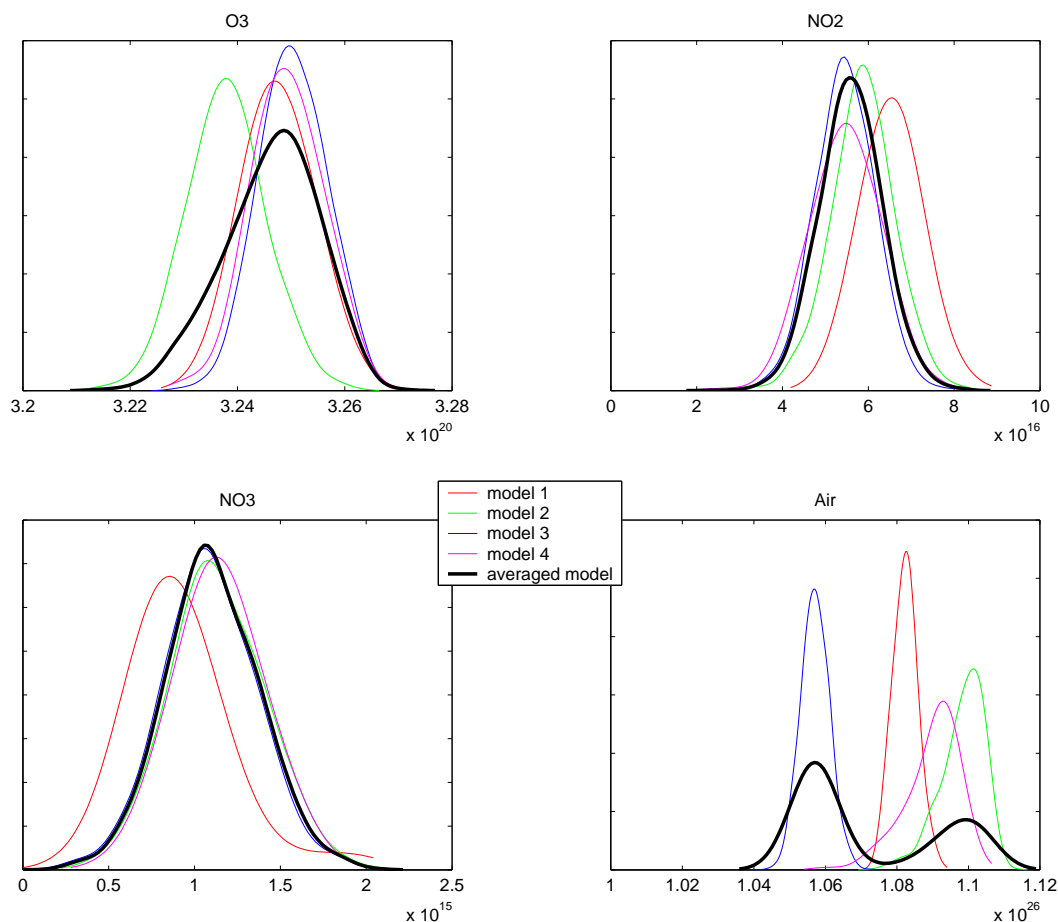
**Fig. 5.** Marginal posterior density estimates of the constituent line densities calculated from the MCMC chains of Fig. 4. The thicker line is the uncertainty coming from the averaged model that takes into account the model uncertainty. The posterior probabilities of the models are the relative times the chain has spent on each model. This depend on given prior weights for each model and on how well each different model fit the data compared to other models. In the present example, all themodels are taken a priori to be equally likely , so $p(k)=1/4$ for $k=1,\dots,4$. The $x$ axis value is the integrated number density $[1/cm^3]$.

Model 2 and of the other models is about twice the estimated posterior standard deviation of the estimated value. The most notable effect is seen on the estimated neutral air density (the lower right plot in Fig. 5). The averaged uncertainty of neutral air over all the models is a distribution with two distinct modes. This is mainly due to the similarity of the cross-section of air and that of the aerosols models. An accurate prior for neutral air, if available, would help this unidentifiability.

The study of aerosols in the GOMOS inversion is further complicated by the fact that, in addition to aerosols, parts of the unmodelled variations in the GOMOS spectra are due to the scintillation effects caused by turbulence. These effects are actively studied at Finnish Meteorological Institute, and the methods presented in this article will give useful methodological tools for these studies, too.

## 6 Conclusions

The adaptive automatic RJMCMC method, AARJ, is a novel combination of previous adaptive MCMC methodologies that have been found to work reliably in various statistical inverse problems applications. AARJ provides an easy-to-use adaptive reversible jump MCMC method for Bayesian model selection. It can be used as a tool for automatic model determination and for making simultaneous inference about the model and the model parameters. If one model clearly stands out, we can select it as the "true" model. If the data do not give any definite indication of the right model, and no accurate prior for the model is available, the uncertainty in the modelling can be taken into account in the model predictions by using a weighted mixture of the models. The method itself is a general one and not limited to geophysical applications. It can be used to solve the model selection problem for a set of models having different parameters of different
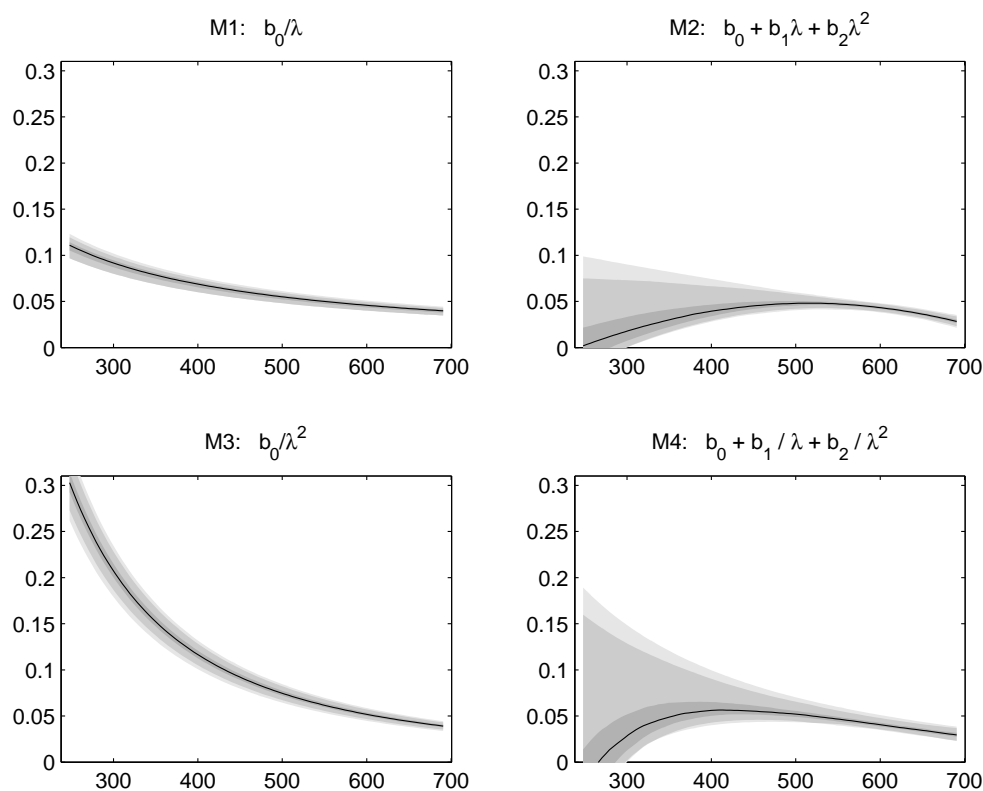
**Fig. 6.** Estimated aerosol extinctions for the selected altitude of the example given in the text. Solid line is the fitted median cross section. Grey areas correspond to 50%, 95% and 95% posterior limits of the extinctions. The model are the following. Model 1: linear for $1/\lambda$, Model 2: a second degree polynomial on $\lambda$, Model 3: linear for $1/\lambda^2$, Model 4: a second degree polynomial on $1/\lambda$.

dimensions. The new algorithm will make it possible to use Bayesian methods in more realistic modelling settings than before, thus further widening the scope of statistical inversion methodology.

The GOMOS aerosol model selection problem can be successfully studied with the AARJ method. For the GOMOS inversion problem it is natural to consider a set of competing aerosol cross section models, as the most suitable model will depend on the unknown type of aerosols present in the corresponding location. In the present example the number of aerosol cross-section models is four, but the method could be used to study a larger number of models. The current operational GOMOS algorithm uses a fixed aerosol model. It would be advisable to further study the effect of the chosen aerosol model of the retrieval of various gas constituents. Different aerosol models could be used depending on the location.

This model selection technique can be used in different applications. The inversion algorithm of the OMI ozone instrument onboard EOS-Aura satellite, for example, has five main aerosol models, each having several sub models (Veihelmann et al., 2007). In the OMI inversion the aerosol model is chosen from a few (2–3) pre-selected models according to the minimum $\chi^2$ criteria. Both the GOMOS and OMI inversions could benefit from the model averaging approach that takes into account the uncertainty in the model selection.

## References

Bertaux, J. L., Kyrölä, E., and Wehr, T.: Stellar Occultation Technique for Atmospheric Ozone Monitoring: GOMOS on Envisat, Earth Observation Quarterly, 67, 17–20, 2000.

ESA 2007: GOMOS Product Handbook Issue 3.0, European Space Agency, http://envisat.esa.int/dataproducts/, 2007.

Gamerman, D.: Markov Chain Monte Carlo – Stochastic simulation for Bayesian inference, Chapman & Hall, 245 pp., 1997.

Gelman, A., Roberts, G. O., and Gilks, W. R.: Efficient Metropolis Jumping Rules, Bayesian Statistics, 5, 599–607, 1996.

Green, P. J.: Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, Biometrika, 82, 711–732, 1995.

Green, P. J.: Trans-dimensional Markov chain Monte Carlo, in: Highly Structured Stochastic Systems, edited by: Green, P. J.,

Hjort, N. L., and Richardson, S., no. 27, Oxford Statistical Science Series, pp. 179–198, Oxford University Press, 2003.

Haario, H., Saksman, E., and Tamminen, J.: An adaptive Metropolis algorithm, Bernoulli, 7, 223–242, 2001.

Haario, H., Laine, M., Lehtinen, M., Saksman, E., and Tamminen, J.: MCMC methods for high dimensional inversion in remote sensing, J. Roy. Stat. Soc. B., 66, 591–607, doi:10.1111/j.1467-9868.2004.02053.x, 2004.

Haario, H., Laine, M., Mira, A., and Saksman, E.: DRAM: Efficient adaptive MCMC, Stat. Comput., 16, 339–354, doi:10.1007/s11222-006-9438-0, 2006.

Hastie, D.: Towards Automatic Reversible Jump Markov Chain Monte Carlo, Ph.D. thesis, University of Bristol Department of Mathematics, 2005.

Kaipio, J. P. and Somersalo, E.: Computational and Statistical Methods for Inverse Problems, Springer, 353 pp., 2004.

Laine, M.: MCMC toolbox for Matlab website, http://www.helsinki.fi/~mjlaine/mcmc/, 2008.

Mira, A.: On Metropolis-Hastings algorithms with delayed rejection, Metron, LIX, 231–241, 2001.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A.: Bayesian measures of model complexity and fit, J. Roy. Stat. Soc. B., 64, 583–639, 2002.

Tamminen, J.: Adaptive Markov chain Monte Carlo algorithms with geophysical applications, Finnish Meteorological Institute Contributions 47, Finnish Meteorological Institute, Helsinki, http://urn.fi/URN:ISBN:952-10-2016-4, 2004.

Tamminen, J. and Kyrölä, E.: Bayesian solution for nonlinear and non-Gaussian inverse problem by Markov chain Monte Carlo method, J. Geophys. Res., 106, 14377–14390, 2001.

Vanhellemont, F., Fussen, D., Dodion, J., Bingen, C., and Mateshvili, N.: Choosing a suitable analytical model for aerosol extinction spectra in the retrieval of UV/visible satellite occultation measurements, J. Geophys. Res., 111, 1–8, doi:10.1029/2005JD0069412006.

Veihelmann, B., Levelt, P. F., Stammes, P., and Veefkind, J. P.: Simulation study of the aerosol information content in OMI spectral reflectance measurements, Atmos. Chem. Phys., 7, 3115–3127, 2007, http://www.atmos-chem-phys.net/7/3115/2007/.