Geosci. Model Dev. Discuss., 6, 2551–2583, 2013
www.geosci-model-dev-discuss.net/6/2551/2013/
doi:10.5194/gmdd-6-2551-2013

Geoscientific Model Development Discussions
Open Access

Correction of approximation errors with Random Forests applied to modelling of aerosol first indirect effect

A. Lipponen[1], **V. Kolehmainen**[1], **S. Romakkaniemi**[1], **and H. Kokkola**[2]

[1]Department of Applied Physics, University of Eastern Finland, P.O. Box 1627, 70211 Kuopio, Finland
[2]Finnish Meteorological Institute, Kuopio Unit, P.O. Box 1627, 70211 Kuopio, Finland

Received: 1 March 2013 – Accepted: 2 April 2013 – Published: 19 April 2013

Correspondence to: A. Lipponen (antti.lipponen@uef.fi)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

In atmospheric models, due to their computational time or resource limitations, physical processes have to be simulated using reduced models. The use of a reduced model, however, induces errors to the simulation results. These errors are referred to as approximation errors. In this paper, we propose a novel approach to correct these approximation errors. We model the approximation error as an additive noise process in the simulation model and employ the Random Forest (RF) regression algorithm for constructing a computationally low cost predictor for the approximation error. In this way, the overall simulation problem is decomposed into two separate and computationally efficient simulation problems: solution of the reduced model and prediction of the approximation error realization. The approach is tested for handling approximation errors due to a reduced coarse sectional representation of aerosol size distribution in a cloud droplet activation calculation. The results show a significant improvement in the accuracy of the simulation compared to the conventional simulation with a reduced model. The proposed approach is rather general and extension of it to different parameterizations or reduced process models that are coupled to geoscientific models is a straightforward task. Another major benefit of this method is that it can be applied to physical processes that are dependent on a large number of variables making them difficult to be parameterized by traditional methods.

1 Introduction

In numerical simulations of complicated physical phenomena, one usually has to balance between the model accuracy and the computation time. Reduction in computation time is typically obtained by using reduced models for some of the functions in the model. The use of reduced models, however, result in errors in model output. The errors are referred to as the approximation errors (AE).

GMDD
6, 2551–2583, 2013

In this paper, we consider the approximation errors caused by coarse discretization of aerosol size distributions in sectional aerosol models. In sectional models, the continuous aerosol particle size distributions are represented with discrete size sections (e.g. Weisenstein et al., 2007; Jacobson, 2001; Rodriguez and Dabdub, 2004; Kokkola et al., 2008). The accuracy of the description of the size distribution increases with increasing number of size sections. The computational demand of the model, however, is heavily increased with the number of the sections. Therefore, a compromise between the model accuracy and the computational time has to be made to construct a feasible model for simulations of atmospheric scale.

The main mechanism by which atmospheric aerosol particles affect the climate is by modifying the concentration of cloud condensation nuclei (CCN) followed by changes in cloud droplet number concentration (the indirect effect of aerosols). While it is well known that the number of CCN in the atmosphere is increasing, the effect of these additional CCN on cloud properties is still the largest single source of uncertainty in the current estimates of the anthropogenic radiative forcing (Forster et al., 2007). Thus, solving the cloud activation of the aerosol particles more accurately, would reduce the uncertainty in the estimated aerosol indirect effect. Current aerosol-climate models include parameterizations for calculating cloud activation of aerosol that use the above mentioned sectional approach (Abdul-Razzak and Ghan, 2002; Fountoukis and Nenes, 2005). Nevertheless, coarse size resolution of the aerosol size distribution that is used as an input for a cloud activation parameterization translate to approximation errors in the calculated aerosol indirect effect.

Recently, an approach for compensating approximation errors in inverse problems was proposed by Kaipio and Somersalo (Kaipio and Somersalo, 2005). The approach is known as the approximation error approach. This far, the approach has mainly been applied to so-called soft field tomography imaging problems that are related to estimation of spatially distributed parameters of partial differential equations from boundary measurements. In such problems, the approach has been successful, for example, in compensation of approximation errors due to coarse finite element

GMDD
6, 2551–2583, 2013

discretization (Arridge et al., 2006; Nissinen et al., 2009), unknown nuisance parameters (Nissinen et al., 2009, 2011; Kolehmainen et al., 2011), and the truncation of the computational domain (Lehikoinen et al., 2007; Kolehmainen et al., 2009). The main idea in the approximation error approach is to model the error between the accurate and approximate computational models as an additive noise process in the measurement model. The realization of the approximation error noise is obviously unknown and cannot be computed without solving the accurate model and knowing the unknown parameters. However, given the prior probability density models of all the unknowns, the inverse problem can be marginalized over the unknown approximation error in an approximate way by utilizing a Gaussian estimate for the joint probability density of the approximation error and the unknown parameters. For a detailed explanation, see Kolehmainen et al. (2011).

In this paper, we propose a novel approach for handling approximation errors in simulation models. The approach is an extension of the approximation error approach. Similarly as in applications of the approximation error approach to inverse problems, the discrepancy between the outputs of accurate and reduced models is modelled as an additive approximation error noise process in the simulation model. However, whereas in the framework of inverse problems the uncertainty related to the approximation errors is taken care of by marginalization, here we propose to construct a computationally low-cost predictor model that computes an estimate for the realization of the approximation error given in the input parameters and solution of the reduced model. This way the solution of the simulation problem is decomposed into a computationally efficient approximation of solving the reduced computation model and estimating the value of the additive approximation error.

One computationally simple and light-weight, and recently widely used function approximation approach is to employ Rfs. The RFs are predictive models introduced in Breiman (2001). An RF model consists of an ensemble of binary tree predictors. Each of these tree predictors is trained based on the training data.

GMDD
6, 2551–2583, 2013

The aim of the RF model construction is to get numerous tree models that slightly differ from each other. This is achieved by introducing randomization in the tree construction. The constructed RF models are further used for the function output prediction. The prediction of the RF model is computed by averaging the predictions of each (almost) unbiased tree models in the ensemble. This averaging should therefore increase the accuracy of the RF model over a single tree prediction accuracy. Recently, the RF models have been applied to classification and regression problems including classification of climate zones (Bechtel and Daneke, 2012), earthquake induced damages (Tesfamariam and Liu, 2010) and remote sensing data (Pal, 2005), disease prediction (Munro et al., 2006; Yao et al., 2013). In some of the cited papers, a comparison between different algorithms were carried out. Despite of its simplicity, the RF was observed to perform at least equally well as the more complicated algorithms in classification and regression problems.

We employ the RF approach for construction of the predictor model for the approximation errors in the simulation model. The training data for the RF algorithm is a set of approximation error realizations between the accurate and reduced models corresponding to a set of random samples of the input parameters that are sampled from the prior probability density models. The computation of the training data involves solution of the computationally demanding accurate model as many times as the number of samples. This step, however, can be done as precomputation and needs to be carried out only once. Given the trained RF model, the accurate model can then be approximated by the sum of the reduced model and the predicted approximation error in the actual simulations.

The proposed approach is evaluated in the case of cloud droplet number concentration (CDNC) estimation from sectional aerosol particle size distribution using the cloud droplet activation parameterization by Abdul-Razzak and Ghan (2002). We consider the approximation errors caused by using a coarse sectional representation of the aerosol particle size distributions. The results show that the proposed approach

GMDD
6, 2551–2583, 2013

gives a significantly improved accuracy over the conventional way of using the reduced model only with the cost of a small increase in the computational burden.

The rest of the paper is organized as follows. The approximation error approach and the RF models are explained and the approach for prediction of approximation errors using the RF models is proposed in Sect. 2. In Sect. 3, the cloud droplet activation parameterization is briefly reviewed. In Sect. 4, the proposed approach is applied and evaluated in the case of using coarse size resolution aerosol microphysics model together with cloud droplet activation parameterization by Abdul-Razzak and Ghan (2002) referred to as ARG from here on. The conclusions are given in Sect. 5.

2 Correction of approximation errors with Random Forests

2.1 Approximation error model

Let $f(\boldsymbol{x})$ denote the sufficiently accurate but computationally too time consuming computational model. Instead of using the model $f(\boldsymbol{x})$, one wishes to use a computationally low cost reduced model

$$\tilde{f}(\tilde{\boldsymbol{x}}), \quad \tilde{\boldsymbol{x}} = P(\boldsymbol{x}) \tag{1}$$

where P is typically a model reduction mapping from higher dimensional space to a lower dimensional space. However, the approximation errors caused by the model reduction can often render the simulation results unreliable, or even useless.

Using the approximation error model (Kaipio and Somersalo, 2005), we write the simulator as

$$\begin{aligned} f(\boldsymbol{x}) &= \tilde{f}(\tilde{\boldsymbol{x}}) + \left[f(\boldsymbol{x}) - \tilde{f}(\tilde{\boldsymbol{x}}) \right] \\ &= \tilde{f}(\tilde{\boldsymbol{x}}) + \boldsymbol{\epsilon} \end{aligned} \tag{2}$$

where $\boldsymbol{\epsilon}(\boldsymbol{x}) = f(\boldsymbol{x}) - \tilde{f}(\tilde{\boldsymbol{x}})$ represents the approximation error. Notice that model (2) is accurate but the exact realization of the approximation error for a given realization of

GMDD
6, 2551–2583, 2013

input parameters $\boldsymbol{x}$ can only be evaluated by solving the computationally demanding accurate model $f(\boldsymbol{x})$, which we wish to avoid in the first place. In the present work, our objective is to construct a computationally fast predictor model for the realization of the approximation error

$$\boldsymbol{\epsilon} \approx \tilde{g}(\tilde{\boldsymbol{x}}) \tag{3}$$

so that the simulation can be approximated in a computationally efficient form

$$f(\boldsymbol{x}) \approx \tilde{f}(\tilde{\boldsymbol{x}}) + \tilde{g}(\tilde{\boldsymbol{x}}) \tag{4}$$

for a given realization of the reduced parameterization $\tilde{\boldsymbol{x}}$. For this, we model $(\boldsymbol{x}, \boldsymbol{\epsilon})$ as vector valued random variables and utilize the RF model for the construction of the predictor $\tilde{g}(\tilde{\boldsymbol{x}})$.

2.2 Simulation of training data for the Random Forest algorithm

The construction of a predictor model $\tilde{g}(\tilde{\boldsymbol{x}})$ requires a set of feasible realizations of the random variables $\{\tilde{\boldsymbol{x}}_k, \boldsymbol{\epsilon}_k, k = 1, \ldots, N\}$. Firstly, this step involves drawing N random realizations of $\boldsymbol{x}_k$ from the prior probability density model $\pi(\boldsymbol{x})$, or alternatively, one can utilize set of existing data (e.g. measured realizations of $\boldsymbol{x}$) if available. Secondly, one has to compute realizations $\boldsymbol{\epsilon}_k = f(\boldsymbol{x}_k) - \tilde{f}(P(\boldsymbol{x}_k))$ of the approximation error for each of the samples to obtain the training data $\{\tilde{\boldsymbol{x}}_k, \boldsymbol{\epsilon}_k, k = 1, \ldots, N\}$. Obviously, this step involves solving the accurate and computationally demanding model $f(\boldsymbol{x})$ N times. However, this computationally demanding part has to be done only once for the construction of the simulation model (4). This model can then be used to approximate the accurate model $f(\boldsymbol{x})$, for example, within aerosol-climate models where the computational times are a critical issue. The outline of the simulation of the training data is presented in Algorithm 1.

GMDD
6, 2551–2583, 2013

Algorithm 1: Simulation of training data.

Inputs: Accurate and approximative models $f(\boldsymbol{x})$ and $\tilde{f}(\tilde{\boldsymbol{x}})$, respectively, prior probability distribution model $\pi(\boldsymbol{x})$ for the input variable $\boldsymbol{x}$, model reduction mapping P and the number of samples N to be used in the precomputation steps. **Output:** Training data $\{\tilde{\boldsymbol{x}}_k, \boldsymbol{e}_k\}$ for the RF model

1: **for** $i = 1, .., N$ **do**
2: Draw a random sample $\boldsymbol{x}_i$ from the probability distribution $\pi(\boldsymbol{x})$ (or use sample from a set of measured realizations of $\boldsymbol{x}$).
3: Simulate the accurate model, i.e. compute $f(\boldsymbol{x}_i)$.
4: Simulate the approximate model, i.e. compute $\tilde{f}(P(\boldsymbol{x}_i))$.
5: Add a sample $(\tilde{\boldsymbol{x}}_i, \boldsymbol{e}_i)$ where $\tilde{\boldsymbol{x}}_i = P(\boldsymbol{x}_i)$ and $\boldsymbol{e}_i = f(\boldsymbol{x}_i) - \tilde{f}(P(\boldsymbol{x}_i))$ to the training set.
6: **end for**

2.3 Random Forests

Rfs developed by Breiman (2001) are used for classification and regression. The RF algorithm uses training data to construct an RF model used for predicting a class in which the given input belongs (classification) or the output of a function the input would give (regression). An RF model consist of an ensemble of classification or regression trees. Each tree in the RF is grown independently of each other and based on a slightly different training set to avoid overfitting of the model. In particular, each training set is obtained as random subset of the original training set. Further, the reason for constructing an ensemble of tree models, not a single tree model, is to increase the accuracy and reduce the uncertainty of the overall prediction. In this paper, the RF models for regression are considered.

In case of regression, the RF model consists of an ensemble of regression tree models. A regression tree model is a sequence of rules that is used for function output prediction with given inputs. The sequence of rules forms a binary tree structure and it

GMDD
6, 2551–2583, 2013

is evaluated by following the nodes starting from the uppermost node referred to as the root node. Each node rule consists of a pair of input variable index and split threshold value. In the node evaluation, the value of the input variable indicated by the index is compared with the split threshold value. If the input data variable value is less than the threshold value the left branch of the node is followed. In other cases, the right branch is followed. The tree structure is followed until a node that has no child nodes is reached. These nodes are referred to as the leaf nodes. The tree model output prediction is selected as the output value indicated by the leaf node. An illustrative example of a regression tree is shown in Fig. 1.

As stated above, an ensemble of trees is constructed with the training data $\{\tilde{\boldsymbol{x}}_k, \boldsymbol{\epsilon}_k\}$. The samples $\tilde{\boldsymbol{x}}_k$ and $\boldsymbol{\epsilon}_k$ are considered as the inputs and outputs of the function, respectively, which the RF model to be constructed is approximating. We use a slightly modified version of the original RF algorithm. The outline of the modified algorithm is given in Algorithm 2.

GMDD
6, 2551–2583, 2013

Algorithm 2: The modified algorithm for growing a Random Forest model.

Input: Training data set $\{\tilde{\boldsymbol{x}}_k, \boldsymbol{\epsilon}_k,\ k = 1,\ldots,N\}$, number of trees in the forest N_{trees}, number of input variable candidates at each split N_{ipcands}, maximum number of training data samples assigned to a leaf node $N_{\text{maxsamples}}$ and the number of split threshold candidates N_{splitp}. **Output:** Random Forest R

1: **for** $k = 1,..,N_{\text{trees}}$ **do**
2: Add root node $R_{k,1}$ to the tree.
3: Assign a random bootstrap sample with replacement (N samples) from the training data to the root node.
4: **while** A non-terminated leaf node exists **do**
5: Select a non-terminated leaf node k.
6: Construct a random set of N_{ipcands} split variable candidates.
7: Construct a random sets of N_{splitp} split threshold values corresponding to each split variable candidate.
8: Find the split variable – split value pair that has the smallest sum of the sample variances of $\boldsymbol{\epsilon}$ in the splitted sets.
9: Split the node according to the best split and assign the training data to the child nodes according to the rule.
10: If a child node has less than $N_{\text{maxsamples}}$ samples assigned to it, terminate it and compute the output of the node as the average of the training samples assigned to it.
11: **end while**
12: **end for**

In our training algorithm, the training data set $\{\tilde{\boldsymbol{x}}_k, \boldsymbol{\epsilon}_k,\ k = 1,\ldots,N\}$, the number of trees in the forest N_{trees}, the number of input variable candidates at each split N_{ipcands}, maximum number of training samples assigned to a leaf node $N_{\text{maxsamples}}$ and the number of split threshold trial points N_{splitp} are given to the training algorithm as inputs. The RF model to be grown consists of N_{trees} regression trees each of which are constructed as follows. First, a bootstrap sample consisting of N training data samples is drawn with replacement from the training data set and assigned to the root node of the tree.

GMDD
6, 2551–2583, 2013

Second, N_{ipcands} unique input variable (elements of vector $\tilde{\boldsymbol{x}}$) candidates are randomly selected. Next, the training data samples assigned to the current node are splitted into two non-empty sets corresponding to N_{splitp} randomly selected split threshold candidates for each input variable candidate. The input variable–split threshold candidate pair that results in the smallest sum of ϵ variances in the non-empty sets is selected for the split. Two child nodes for the current node are created, and the training samples are assigned to them according to the selected input variable and threshold candidates. If a node has less than $N_{\mathrm{maxsamples}}$ training samples assigned to it, it is a leaf node and the sample average of ϵ of the training samples assigned to the node is computed and used as the output value of the node. The splitting of the nodes that have more than $N_{\mathrm{maxsamples}}$ training samples assigned to them is carried out similarly as for the root node until no more nodes in the tree to be splitted are left.

The constructed RF model is used for prediction as follows. All the tree models are evaluated separately by following the tree structures starting from the root nodes. In each node, the value of the input variable indicated by the node rule is compared with the split threshold of the node. If the variable value is less than the threshold, the tree is followed to the left child node. Otherwise the right child node is followed. This procedure is repeated until a leaf node is reached and the output value of the leaf node is taken as the tree output. The overall prediction of the RF model is computed as the average of all the individual tree model outputs, giving us the prediction of the approximation error

$$\boldsymbol{\epsilon} = \tilde{g}(\tilde{\boldsymbol{x}}). \tag{5}$$

The outline of the RF model evaluation is shown in Algorithm 3.

GMDD
6, 2551–2583, 2013

Algorithm 3: The simulation of a Random Forest model.

Input: Model input $\boldsymbol{x}$, and the RF model with N_{trees} trees and nodes with information $\{l, \alpha, \boldsymbol{I}_{\text{L}}, \boldsymbol{I}_{\text{R}}, \hat{\boldsymbol{y}}\}$ where l and α are the input variable index and the related threshold, $\boldsymbol{I}_{\text{L}}$ and $\boldsymbol{I}_{\text{R}}$ indices to left and right child nodes, and $\hat{\boldsymbol{y}}$ the output of the node. **Output:** Random Forest output $\boldsymbol{y}$.

```
for k = 1,..,N_trees do
    Set the root node as the current node to be evaluated
    while The current node is not a leaf node do
        if x(l) < α then
            Set I_L as the current node
        else
            Set I_R as the current node
        end if
    end while
    Set y_k = ŷ of the leaf node reached
end for
Compute y as the average of {y_1, ..., y_N_trees}.
```

3 Cloud droplet activation parameterization

Formation of cloud droplets in the atmosphere is a dynamical process affected by local meteorology and aerosol particles acting as cloud condensation nuclei. In atmospheric models, this process is parameterized. In the most sophisticated parameterizations, CDNC is calculated based on aerosol particle size distribution and chemical composition, pressure, temperature and vertical velocity of air parcel forming the cloud (Abdul-Razzak et al., 1998; Abdul-Razzak and Ghan, 2002; Nenes and Seinfeld, 2003).

The simulations in this study are conducted using the SALSA sectional aerosol model developed for atmospheric models (Kokkola et al., 2008; Bergman et al., 2012).

GMDD
6, 2551–2583, 2013

In SALSA, aerosol size distribution is divided to different subranges based on the particle size (3–50 nm, 50–700 nm, 700–10 000 nm). The size resolution differs between the subranges depending on how sensitive the aerosol processes are to particle sizes of given subrange. When using the default setup of SALSA, it has 10 size sections divided so that there are 3 sections in the first subrange, 4 in the second subrange, and 3 in the third subrange. A more detailed description of the model is given by Kokkola et al. (2008).

SALSA includes all relevant microphysical processes such as condensation of sulfate and organic carbon, nucleation of new particles, hydration, and coagulation. However, in this study we are only interested in the effect of the size resolution on the calculated number of cloud droplets, and the SALSA is used only to create aerosol size distribution and to calculate the CDNC using the ARG parameterization. Also, we are omitting the first subrange as usually the cloud droplet nucleation in the atmosphere is not affected by these particles as they are too small to act as cloud condensation nucleus. For simplicity, in this study we have also assumed that aerosol is composed of only one highly hygroscopic compound (sulphate), one slightly hygroscopic compound (organic carbon) and one non-hygroscopic compound (dust).

4 Models, simulations and results

4.1 Accurate and reduced models

Let $f(\boldsymbol{x}) \in \mathbb{R}$ denote the sufficiently accurate computational ARG cloud droplet activation parameterization that computes the value of the CDNC for the given input $\boldsymbol{x}$. The input parameter vector $\boldsymbol{x}$ contains aerosol particle size and composition distributions, vertical velocity, pressures and temperature information. In the following computations, the number of size sections for the representation of the particle size distributions is 70, see Table 1. With this discretization, the average simulation time of the accurate model is about 1 ms.

GMDD
6, 2551–2583, 2013

In the parameter vector $\tilde{\boldsymbol{x}}$ of the reduced model $\tilde{f}(\tilde{\boldsymbol{x}})$, the number of size sections for the aerosol particle size distributions have been significantly reduced. We consider two different levels of model reduction. In the first one, the number of size sections is 7 and in the second one 4, see Table 1. The average computation times are about 0.11 ms and 0.07 ms for the 7 sections and 4 sections parameterizations, respectively. Thus, when reducing from 70 size sections to 7 or 4 sections the average reductions in computation times are about 89 % and 93 %, respectively.

4.2 Construction of the RF predictor model

The size of sample set $\{\boldsymbol{x}_k\}$ was selected as $N = 50\,000$ for the construction of the training data (Algorithm 1). The realizations $\{\boldsymbol{x}_k\}$ of the input parameters were drawn from their prior probablity distribution models, which were selected so that the realizations are plausible representations of their values in the nature. The aerosol particle number distribution $n = n(d)$, where d is the diameter of the particle, was modelled as a sum of three log-normal modes representing the Aitken, accumulation and coarse mode aerosols:

$$n(d) = \sum_{i=1}^{3} n_i(d) \tag{6}$$

where each of the modes was modelled by

$$n_i(d) = \frac{n_{\mathrm{tot},i}}{d\sqrt{2\pi(\log(\sigma_i))^2}} \exp\left\{-(\log(d/\mu_i))^2 / \left(2\sigma_i^2\right)\right\} \tag{7}$$

where the $n_{\mathrm{tot},i}$ is the total number of particles in mode i, and σ_i and μ_i the shape and log-scale parameters of mode i. The parameters of the prior probability distribution models used in the generation of the vertical velocity w, pressure p, temperature T, and the particle number distribution parameter n_i, σ_i, μ_i samples are shown in Table 2

GMDD
6, 2551–2583, 2013

and the respective probability density functions are shown in Table 3. The histograms of the temperature, pressure and vertical velocity samples, and the particle number distribution parameters in the training sample set $\{\boldsymbol{x}_k\}$ are shown in Figs. 2 and 3, respectively. The aerosol particle volume size distributions were constructed with the particle number distributions of the modes and randomly distributed volume fractions of each compound. The volume fractions for the sulphate was drawn from an uniform distribution $\mathcal{U}(0.01, 1)$. Further, the fractions of dust and organic carbon were drawn from uniform distributions such that the sum of the compound fractions was 1.

Figure 4 shows the output values of the accurate parameterization against the output of the approximate parameterization for the set of training samples $\{\boldsymbol{x}_k\}$ (i.e. the points in the figure are $(x_j^*, y_j^*) = \left(\tilde{f}\left(\tilde{\boldsymbol{x}}_j, f(\boldsymbol{x}_j)\right)\right)$). In the top panel, the reduced model uses 7 size sections for the size distributions and in the bottom 4 size sections. The black line shows the identity line $y = x$ corresponding to the case that accurate and reduced models match. The average relative errors in the CDNC values were 20.4 % and 55.7 % for the 7 and 4 size sections parameterizations, respectively. The reason for the lower CDNC with the smaller number of size sections is the lower maximum supersaturation when using the ARG parameterization.

Given the samples $\{\boldsymbol{x}_k\}$, the realizations of the approximation error were simulated as

$$\{\epsilon_k = \log(f(\boldsymbol{x}_k)) - \log(\tilde{f}(P(\boldsymbol{x}_k))), k = 1, \ldots, N\}. \quad (8)$$

Here the logarithmic scale for the CDNC values was selected based on preliminary tests in which this selection slightly improved the accuracy of the RF models. The histograms of the approximation errors ϵ for both the 7 and 4 size sections parameterizations are shown in Fig. 5.

Finally, the sample sets $\{\boldsymbol{x}_k, \log(\tilde{f}(P(\boldsymbol{x}_k)))\}$ and $\{\epsilon_k\}$ were used as the RF training set inputs and outputs, respectively, and the RF models were trained as described in the Sect. 2.3. Also here, the addition of logarithms of the coarse parameterization outputs in the training set slightly improved the RF model accuracy and was therefore used.

GMDD
6, 2551–2583, 2013

Once the RF predictor $\tilde{g}$ was constructed, the output of the accurate simulator $f(\boldsymbol{x})$ was approximated with

$$f(\boldsymbol{x}) \approx \hat{f}(\tilde{\boldsymbol{x}}) = \exp\left(\log(\tilde{f}(\tilde{\boldsymbol{x}})) + \tilde{g}(\tilde{\boldsymbol{x}}, \tilde{f}(\tilde{\boldsymbol{x}}))\right). \quad (9)$$

4.3 Results

To evaluate the proposed approach, multiple RF predictor models for the approximation errors corresponding to both approximate ARG parameterizations, with 7 and 4 size sections, were constructed with different parameters of the Algorithm 2. All possible combinations of parameter sets $\{25, 50, 100, 200\}$, $\{5, 10, 15, 25\}$, $\{5, 15, 25, 100\}$, and $\{25, 75\}$ for N_{trees}, N_{ipcands}, $N_{\text{maxsamples}}$, and N_{splitp}, respectively, were used. To avoid overoptimistic results, the constructed AE models were evaluated with a separate validation set of 25 000 samples of ARG model inputs. The validation set was sampled similarly as the training set but the samples were not included in the training of the RF model.

All predictor models were evaluated using the validation set, and the mean squared error (MSE) ϵ_{MSE} and mean relative error (MRE) ϵ_{MRE} estimates were computed. The error estimates were computed as

$$\epsilon_{\text{MSE}} = \frac{1}{N}\sum_{i=1}^{N}(f(\boldsymbol{x}_i) - \hat{f}(\tilde{\boldsymbol{x}}_i))^2 \quad (10)$$

and

$$\epsilon_{\text{MRE}} = \frac{1}{N}\sum_{i=1}^{N}\frac{|f(\boldsymbol{x}_i) - \hat{f}(\tilde{\boldsymbol{x}}_i)|}{|f(\boldsymbol{x}_i)|}. \quad (11)$$

As the construction of an RF model is random, the tests were repeated 50 times for each AE model to also evaluate the random variations in the results. The average

GMDD
6, 2551–2583, 2013

MSE and MRE, average computation time of the approximation error model, memory required to store the RF model, and the model parameters for both of the approximate parameterizations $\tilde{f}(\tilde{\boldsymbol{x}})$ corresponding to 20 different combinations of RF training parameters (N_{trees}, N_{ipcands}, $N_{\text{maxsamples}}$, N_{splitp}) are given in Table 4 for the parameterization with 7 size sections and 5 for the parameterization with 4 size sections. The bottom row in both Tables gives the respective errors between the accurate parameterization $f(\boldsymbol{x})$ and reduced parameterization $\tilde{f}(\tilde{\boldsymbol{x}})$ without approximation error correction. The CDNC values computed with the accurate parameterization $f(\boldsymbol{x}_j)$ as a function of the AE corrected CDNC values using the predictor $\tilde{g}$ with the lowest MSE error are shown in Fig. 6. Top row shows the case for the reduced model with 7 size sections and bottom row the case with 4 size sections for the particle size distributions.

The results show that by using the AE correction with the RF predictor model, both the MSE and MRE errors are significantly decreased. In the case of the reduced parameterization $\tilde{f}(\tilde{\boldsymbol{x}})$ with 7 size sections, the RF training parameter selections $N_{\text{ipcands}} = 25$, $N_{\text{maxsamples}} = 5$, $N_{\text{splitp}} = 25$ and $N_{\text{trees}} = 200$ resulted in the overall model in which both the MSE and MRE were the smallest. Here, the approximation error correction decreased the MSE error by more than one order of magnitude and the MRE was decreased by more than 10 %. In the case of the reduced parameterization $\tilde{f}(\tilde{\boldsymbol{x}})$ with 4 size sections, the lowest MSE and MRE were obtained with the RF training parameters $N_{\text{ipcands}} = 25$, $N_{\text{maxsamples}} = 15$, $N_{\text{splitp}} = 75$, and $N_{\text{trees}} = 200$. Also here, both the MSE and MRE errors were significantly decreased. Notice that the MSE errors of the 4 size sections parameterization with the approximation error correction are smaller than the MSE errors of the uncorrected 7 size sections parameterization.

The results also show that the RF model training parameters did not significantly affect the accuracy of the AE model. The RF training parameter affecting the accuracy of the model most was the number of trees in the forest. The randomness in the RF model training caused only minor variations in the resulting RF models showing the robustness of the approach.

GMDD
6, 2551–2583, 2013

The average times to simulate the AE models varied between 0.05 and 0.58 ms in the case of 7 size sections parameterization and between 0.04 and 0.49 ms in 4 size sections parameterization on a standard desktop computer. The average time to simulate the reduced model $\tilde{f}(\tilde{\boldsymbol{x}})$ with 7 and 4 size sections were about 0.11 ms and 0.07 ms, respectively. These AE model running times for computing $\hat{f}(\tilde{\boldsymbol{x}})$ resulted in overall average runtimes of 0.16–0.69 ms for the 7 size sections and 0.11–0.56 ms for the 4 size sections AE corrected parameterizations. Thus, the reduction in computation times of the approximation error corrected models $\hat{f}(\tilde{\boldsymbol{x}})$ is in the range 31–89 % compared to the run time of the accurate model $f(\boldsymbol{x})$. Note that the errors using the fastest RF predictor model with the least number of trees is only sligthly larger (less than 1 % in the MRE error) compared to the slowest RF model with the largest number of trees. By using the RF models with the least number of trees, one would still get more than an order of magnitude improvement in the accuracy compared to the reduced model $\tilde{f}(\tilde{\boldsymbol{x}})$ with an increment of computation time from 0.11 ms to 0.16 ms for the 7 size sections model and from 0.07 to 0.11 ms for the 4 sections model. Note that the use of, for example, the RF predictor model with the training parameters $N_{\text{trees}} = 25$, $N_{\text{ipcands}} = 25$, $N_{\text{maxsamples}} = 15$, and $N_{\text{splitp}} = 15$ resulted in the overall model with only slightly larger (about 0.3 %) MRE error and 0.46 ms faster running time compared to the RF model with the smallest MRE error in the case of 7 size sections parameterization. By using the fastest RF models listed in Tables 4 and 5, one would still get more than an order of magnitude improvement in the MSE error compared to the reduced model $\tilde{f}(\tilde{\boldsymbol{x}})$ with an increment of computation time from 0.11 ms to 0.18 ms for the 7 size sections model and 0.07–0.11 ms for the 4 sections model. Notice that the computation time of the error prediction by the RF model is independent of the computation times of f or $\tilde{f}$. Thus, the relative time saving by the proposed approach will increase as the computation time of f increases. The memory requirement for storing the RF models varied between 10 to 140 MB depending on the number of trees. This can be considered as a low amount for modern computers.

GMDD
6, 2551–2583, 2013

5 Conclusions

Due to computational time and resource limitations related to atmospheric models, several physical processes have to be simulated using reduced models. The use of a reduced model, however, induces approximation errors to the simulation results. In this study, we presented a novel approach to correct these approximation errors and applied it in the calculation of cloud droplet number concentration (CDNC). In the studied case, the approximation error (in CDNC) is caused by coarse sectional representation the aerosol particle distribution.

In our approach, the approximation errors caused by model reduction are modelled as an additive approximation error noise process in the simulation model and the RF algorithm is utilized for construction of a predictor for the realization of the approximation error for given model input parameters. This way the accurate simulation model can be approximated in a computationally fast form by evaluating the reduced model and the prediction of the approximation error.

It was found that the RF approach gives significantly smaller errors in the CDNC calculation than using the reduced model alone with a small increment in the computational cost. Also the systematic errors caused by reduced model accuracy can be efficiently eliminated.

Another significant result in this study was that if the number of size sections were further decreased from 7 to 4, the errors in the RF corrected CDNC of the 4 sections model were lower than the errors of the uncorrected 7 sections model. This shows that the RF method could be useful in reducing the number of size distribution parameters, when aerosol models are developed for simulations of decades or centuries. As the method is in no way limited to sectional approach, it could be applied for reducing number of modes in modal models as has been done by, e.g. Liu et al. (2012).

Here the RF method was employed in the calculation of CDNC with variables typical to atmospheric models. The method can be easily and efficiently extended to take account more complex aerosol including for example surface active Sorjamaa

GMDD
6, 2551–2583, 2013

et al. (2004) or semivolatile aerosol compounds Romakkaniemi et al. (2005) by simply adding new variables to the training data. The method is highly efficient especially in the case of physical processes, which have been found to be difficult to parameterize with traditional methods due to high dependence of the processes on several parameters. Further, the proposed approach is rather general and extension of it to different physical simulation models is a straightforward task.

Acknowledgements. The financial support by the Academy of Finland (project 119270 and Centre of Excellence programs 1118615 and 250215) and by the strategic funding of the University of Eastern Finland are gratefully acknowledged.

References

Abdul-Razzak, H. and Ghan, S. J.: A parameterization of aerosol activation 3. Sectional representation, J. Geophys. Res., 107, AAC 1-1–AAC 1-6, doi:10.1029/2001JD000483, 2002. 2553, 2555, 2556, 2562

Abdul-Razzak, H., Ghan, S. J., and Rivera-Carpio, C.: A parameterization of aerosol activation 1. single aerosol type, J. Geophys. Res., 103, 6123–6131, doi:10.1029/97JD03735, 1998. 2562

Arridge, S., Kaipio, J., Kolehmainen, V., Schweiger, M., Somersalo, E., Tarvainen, T., and Vauhkonen, M.: Approximation errors and model reduction with an application in optical diffusion tomography, Inverse Probl., 22, 175–195, doi:10.1088/0266-5611/22/1/010, 2006. 2554

Bechtel, B. and Daneke, C.: Classification of local climate zones based on multiple earth observation data, IEEE J. Sel. Top. Appl., 5, 1191–1202, doi:10.1109/JSTARS.2012.2189873, 2012. 2555

Bergman, T., Kerminen, V.-M., Korhonen, H., Lehtinen, K. J., Makkonen, R., Arola, A., Mielonen, T., Romakkaniemi, S., Kulmala, M., and Kokkola, H.: Evaluation of the sectional aerosol microphysics module SALSA implementation in ECHAM5-HAM aerosol-climate model, Geosci. Model Dev., 5, 845–868, doi:10.5194/gmd-5-845-2012, 2012. 2562

Breiman, L.: Random Forests, Mach. Learn., 45, 5–32, doi:10.1023/A:1010933404324, 2001. 2554, 2558

GMDD

6, 2551–2583, 2013

Forster, P., Ramaswamy, V., Artaxo, P., Berntsen, T., Betts, R., Fahey, D., Haywood, J., Lean, J., Lowe, D., Myhre, G., Nganga, J., Prinn, R., Raga, G., Schulz, M., and Van Dorland, R.: Changes in atmospheric constituents and in radiative forcing, in: Climate Change 2007: The Physical Science Basis, Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K., Tignor, M., and Miller, H., Cambridge University Press, Cambridge, UK and New York, NY, USA, 2007. 2553

Fountoukis, C. and Nenes, A.: Continued development of a cloud formation parameterization for global climate models, J. Geophys. Res., 110, D11212, doi:10.1029/2004JD005591, 2005. 2553

Jacobson, M. Z.: GATOR-GCMM: a global through urban scale air pollution and weather forecast model. 1. Model design and treatment of subgrid soil, vegetation, roads, rooftops, water, sea ice, and snow, J. Geophys. Res., 106, 5385–5402, doi:10.1029/2000JD900560, 2001. 2553

Kaipio, J. and Somersalo, E.: Statistical and Computational Inverse Problems, Springer, New York, 2005. 2553, 2556

Kokkola, H., Korhonen, H., Lehtinen, K. E. J., Makkonen, R., Asmi, A., Järvenoja, S., Anttila, T., Partanen, A.-I., Kulmala, M., Järvinen, H., Laaksonen, A., and Kerminen, V.-M.: SALSA – a Sectional Aerosol module for Large Scale Applications, Atmos. Chem. Phys., 8, 2469–2483, doi:10.5194/acp-8-2469-2008, 2008. 2553, 2562, 2563

Kolehmainen, V., Schweiger, M., Nissilä, I., Tarvainen, T., Arridge, S., and Kaipio, J.: Approximation errors and model reduction in three-dimensional diffuse optical tomography, J. Opt. Soc. Am. A, 10, 2257–2267, doi:10.1364/JOSAA.26.002257, 2009. 2554

Kolehmainen, V., Tarvainen, T., Arridge, S., and Kaipio, J.: Marginalization of uninteresting distributed parameters in inverse problems – application to diffuse optical tomography, Int. J. Uncertain. Quantif., 1, 1–17, doi:10.1615/Int.J.UncertaintyQuantification.v1.i1.10, 2011. 2554

Lehikoinen, A., Finsterle, S., Voutilainen, A., Heikkinen, L., Vauhkonen, M., and Kaipio, J.: Approximation errors and truncation of computational domains with application to geophysical tomography, Inverse Probl. Imag., 1, 371–389, doi:10.3934/ipi.2007.1.371, 2007. 2554

Liu, X., Easter, R. C., Ghan, S. J., Zaveri, R., Rasch, P., Shi, X., Lamarque, J.-F., Gettelman, A., Morrison, H., Vitt, F., Conley, A., Park, S., Neale, R., Hannay, C., Ekman, A. M. L., Hess, P., Mahowald, N., Collins, W., Iacono, M. J., Bretherton, C. S., Flanner, M. G., and Mitchell, D.: Toward a minimal representation of aerosols in climate models: description and

GMDD
6, 2551–2583, 2013

evaluation in the Community Atmosphere Model CAM5, Geosci. Model Dev., 5, 709–739, doi:10.5194/gmd-5-709-2012, 2012. 2569
Munro, N. P., Cairns, D. A., Clarke, P., Rogers, M., Stanley, A. J., Barrett, J. H., Harnden, P., Thompson, D., Eardley, I., Banks, R. E., and Knowles, M. A.: Urinary biomarker profiling in transitional cell carcinoma, Int. J. Cancer, 119, 2642–2650, 2006. 2555
Nenes, A. and Seinfeld, J.: Parameterization of cloud dropletformation in global climate models, J. Geophys. Res., 108, 4415, doi:10.1029/2002JD002911, 2003. 2562
Nissinen, A., Heikkinen, L., Kolehmainen, V., and Kaipio, J.: Compensation of errors due to discretization, domain truncation and unknown contact impedances in electrical impedance tomography, Meas. Sci. Technol., 20, 105504, doi:10.1088/0957-0233/20/10/105504, 2009. 2554
Nissinen, A., Kolehmainen, V., and Kaipio, J.: Compensation of modelling errors due to unknown domain boundary in electrical impedance tomography, IEEE T. Med. Imaging, 30, 231–242, 2011. 2554
Pal, M.: Random Forest classifier for remote sensing classification, Int. J. Remote Sens., 26, 217–222, doi:10.1080/01431160412331269698, 2005. 2555
Rodriguez, M. and Dabdub, D. J.: IMAGES-SCAPE2: A modeling study of size and chemically resolved aerosol thermodynamics in a global chemical transport model, J. Geophys. Res., 109, D02203, doi:10.1029/2003JD003639, 2004. 2553
Romakkaniemi, S., Kokkola, H., and Laaksonen, A.: Parameterization of the nitric acid effect on CCN activation, Atmos. Chem. Phys., 5, 879–885, doi:10.5194/acp-5-879-2005, 2005. 2570
Sorjamaa, R., Svenningsson, B., Raatikainen, T., Henning, S., Bilde, M., and Laaksonen, A.: The role of surfactants in Köhler theory reconsidered, Atmos. Chem. Phys., 4, 2107–2117, doi:10.5194/acp-4-2107-2004, 2004. 2569
Tesfamariam, S. and Liu, Z.: Earthquake induced damage classification for reinforced concrete buildings, Struct. Saf., 32, 154–164, doi:10.1016/j.strusafe.2009.10.002, 2010. 2555
Weisenstein, D. K., Penner, J. E., Herzog, M., and Liu, X.: Global 2-D intercomparison of sectional and modal aerosol modules, Atmos. Chem. Phys., 7, 2339–2355, doi:10.5194/acp-7-2339-2007, 2007. 2553
Yao, D., Yang, J., and Zhan, X.: A novel method for disease prediction: hybrid of Random Forest and multivariate adaptive regression splines, J. Computers, 8, 170–177, doi:10.4304/jcp.8.1.170-177, 2013. 2555

GMDD
6, 2551–2583, 2013

Table 1. Size section configurations of the cloud droplet activation parameterizations used in simulations.

Total number of size sections in the model	Size sections in the diameter range 50–700 nm	Size sections in the diameter range 0.7–10 µm
70	40	30
7	4	3
4	2	2

GMDD
6, 2551–2583, 2013

Table 2. The prior probability distribution models used for the cloud droplet activation parameterization inputs. The $\mathcal{U}$, $\mathcal{N}$ and Γ denote the uniform, Gaussian and gamma distributions, respectively. The details of the probability distribution functions are shown in Table 3.

Variable	Distribution	Unit
w	$\Gamma(1.25, 0.75)$	$\mathrm{m\,s^{-1}}$
p	$\mathcal{U}(1000, 100\,000)$	P
T	$\mathcal{U}(240, 300)$	K
$n_{\mathrm{tot},1}$	$\Gamma(2, 800)$	$\mathrm{cm^{-3}}$
μ_1	$\mathcal{U}(50, 80)$	nm
σ_1	$\mathcal{N}(1.5, 0.125)$	
$n_{\mathrm{tot},2}$	$\Gamma(3, 200)$	$\mathrm{cm^{-3}}$
μ_2	$\mathcal{U}(100, 200)$	nm
σ_2	$\mathcal{N}(1.5, 0.125)$	
$n_{\mathrm{tot},3}$	$\Gamma(1.25, 0.75)$	$\mathrm{cm^{-3}}$
μ_3	$\mathcal{U}(500, 1500)$	nm
σ_3	$\mathcal{N}(1.5, 0.125)$	

GMDD
6, 2551–2583, 2013

Table 3. The notations used for the probability distributions and their probability density functions. $\Gamma(k)$ denotes the Gamma function.

Notation	Probability density function $\pi(x)$
$x \sim \mathcal{U}(a,b)$	$\begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0, & \text{otherwise} \end{cases}$
$x \sim \mathcal{N}(\bar{x},\sigma^2)$	$\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\bar{x})^2}{2\sigma^2}\right)$
$x \sim \Gamma(k,\theta)$	$\frac{1}{\Gamma(k)\theta^k} x^{k-1} \exp\left(-\frac{x}{\theta}\right)$

GMDD
6, 2551–2583, 2013

Table 4. Training parameters and results of the AE correction in the case of 7 size sections parameterization: number of trees in the RF model N_{trees}, the RF training parameters $N_{ipcands}$, $N_{maxsamples}$, N_{splitp}, the minimum, mean and maximum value for the mean squared errors (MSE) and mean relative error (MRE), average time used for evaluating the RF model t, and the memory required to store the RF model M.

N_{trees}	$N_{ipcands}$	$N_{maxsamples}$	N_{splitp}	mean(ϵ_{MSE}) (cm^{-6})	mean(ϵ_{MRE}) (%)	t (ms)	M (Mb)
200	25	5	25	1.18×10^3	8.1	0.52	139.5
200	25	15	25	1.20×10^3	8.1	0.41	56.4
200	25	15	75	1.21×10^3	8.1	0.41	56.5
100	25	15	75	1.22×10^3	8.2	0.19	28.2
50	25	5	75	1.23×10^3	8.2	0.13	34.9
100	25	25	25	1.24×10^3	8.3	0.16	18.0
100	25	25	75	1.26×10^3	8.3	0.17	18.0
25	25	15	25	1.27×10^3	8.4	0.06	7.1
25	25	15	75	1.28×10^3	8.4	0.06	7.1
25	25	25	75	1.32×10^3	8.4	0.06	4.5
200	15	5	75	1.39×10^3	8.9	0.54	140.9
200	15	15	75	1.41×10^3	8.9	0.43	57.2
100	15	15	25	1.42×10^3	9.0	0.20	28.5
50	15	5	75	1.45×10^3	9.0	0.13	35.2
50	15	15	25	1.45×10^3	9.1	0.10	14.3
100	15	25	75	1.46×10^3	9.1	0.17	18.3
25	15	5	25	1.47×10^3	9.2	0.08	17.6
200	25	100	75	1.48×10^3	8.8	0.19	9.7
50	15	25	75	1.49×10^3	9.1	0.09	9.1
50	25	100	75	1.50×10^3	8.9	0.06	2.4
7 size sections parameterization without AE correction				1.76×10^4	20.4	0.11	

GMDD
6, 2551–2583, 2013

Table 5. Training parameters and results of the AE correction in the case of 4 size sections parameterization: number of trees in the RF model N_{trees}, the RF training parameters $N_{ipcands}$, $N_{maxsamples}$, N_{splitp}, the minimum, mean and maximum value for the mean squared errors (MSE) and mean relative error (MRE), average time used for evaluating the RF model t, and the memory required to store the RF model M.

N_{trees}	$N_{ipcands}$	$N_{maxsamples}$	N_{splitp}	mean(ϵ_{MSE}) (cm^{-6})	mean(ϵ_{MRE}) (%)	t (ms)	M (Mb)
200	25	5	25	4.25×10^3	24.0	0.45	139.4
200	25	5	75	4.29×10^3	24.2	0.45	139.5
200	25	15	75	4.32×10^3	24.2	0.36	55.6
100	25	15	75	4.35×10^3	24.3	0.17	27.8
50	25	15	25	4.37×10^3	24.2	0.08	13.9
50	25	5	75	4.39×10^3	24.4	0.11	34.9
100	25	25	75	4.41×10^3	24.4	0.14	17.4
25	25	15	25	4.47×10^3	24.5	0.05	6.9
200	15	5	75	4.51×10^3	24.5	0.46	140.5
25	25	5	75	4.52×10^3	24.7	0.06	17.4
100	15	5	75	4.55×10^3	24.6	0.22	70.2
100	15	5	25	4.56×10^3	24.5	0.22	70.2
100	15	15	75	4.61×10^3	24.6	0.17	27.8
50	15	15	25	4.66×10^3	24.7	0.08	13.9
50	15	15	75	4.67×10^3	24.7	0.08	13.9
100	15	25	75	4.70×10^3	24.7	0.14	17.4
50	15	25	75	4.76×10^3	24.8	0.07	8.7
25	15	15	75	4.81×10^3	25.0	0.05	7.0
25	15	25	25	4.87×10^3	25.1	0.04	4.3
200	25	100	75	4.98×10^3	25.1	0.15	8.4
4 size sections parameterization without AE correction				1.05×10^5	55.7	0.07	

GMDD
6, 2551–2583, 2013

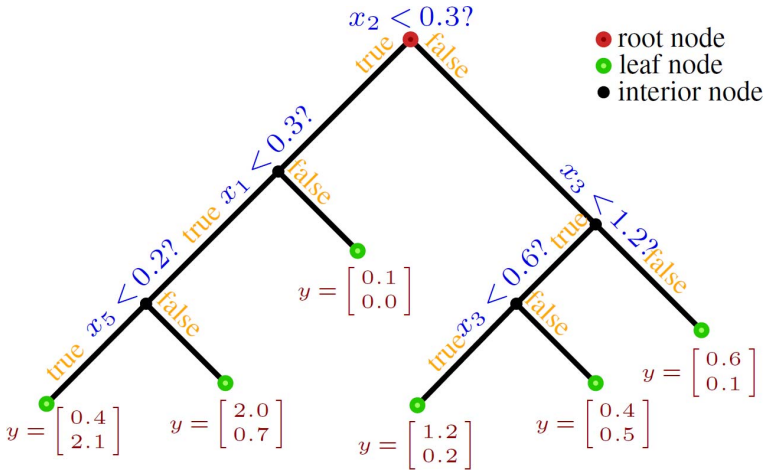

Fig. 1. An illustrative example of a regression tree.

GMDD
6, 2551–2583, 2013

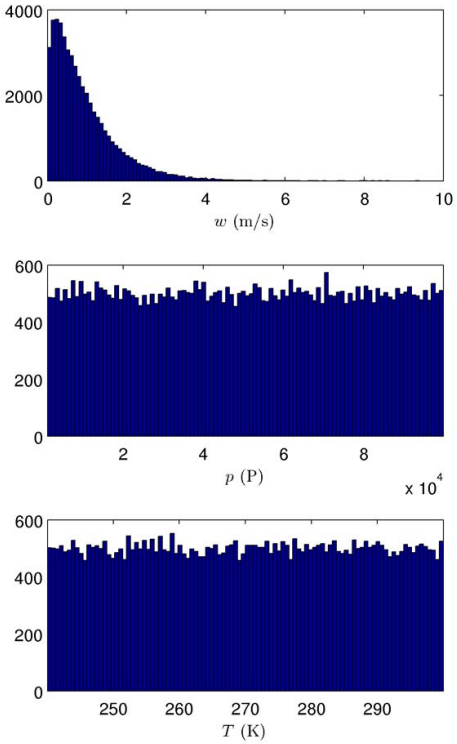

Fig. 2. Histograms of vertical velocity w, pressure p and temperature T in the sample set used for constructing the approximation error samples.

GMDD
6, 2551–2583, 2013




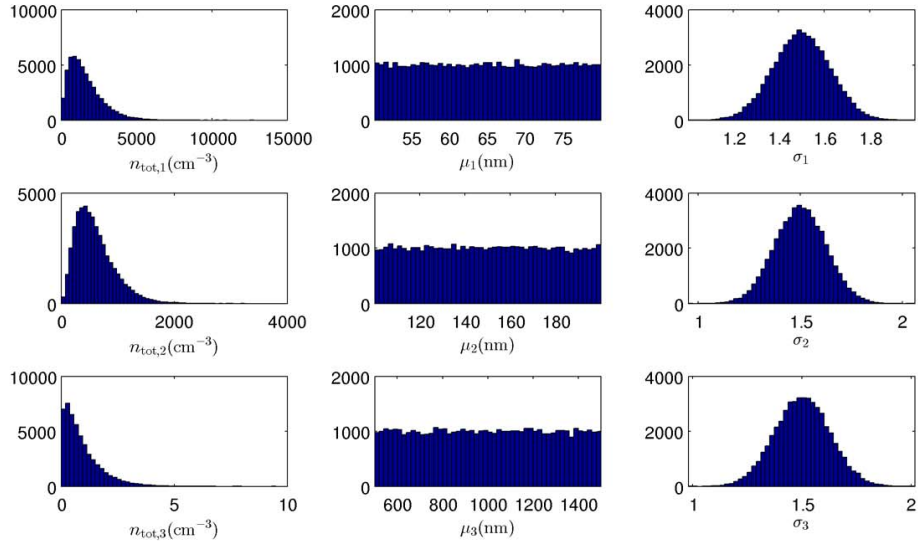

Fig. 3. Histograms for number concentrations of particles n_i, scale parameters μ_i and shape parameters σ_i for the log-normal modes $i = 1,2,3$.

GMDD
6, 2551–2583, 2013

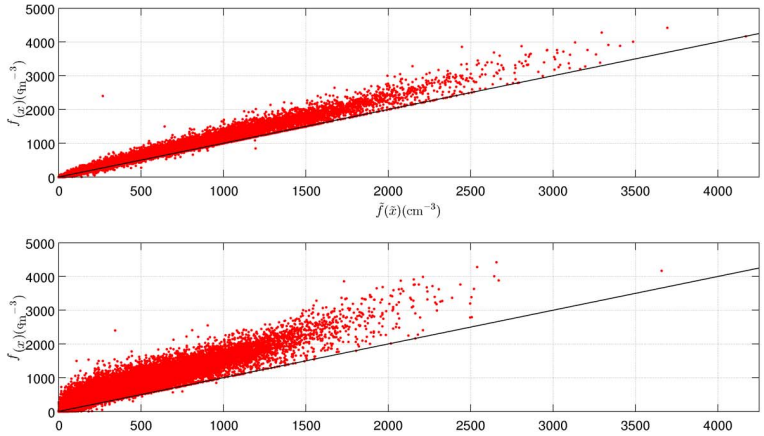

Fig. 4. Computed cloud droplet number concentrations (CDNC) computed with the accurate model $f(\boldsymbol{x})$ as functions of CDNCs given by the approximate model $\tilde{f}(\tilde{\boldsymbol{x}})$. Top: approximate parameterization with 7 size sections for the aerosol particle size distributions. Bottom: approximate parameterization with 4 size sections for the aerosol particle size distributions. Black solid lines represent the identity lines.

GMDD
6, 2551–2583, 2013

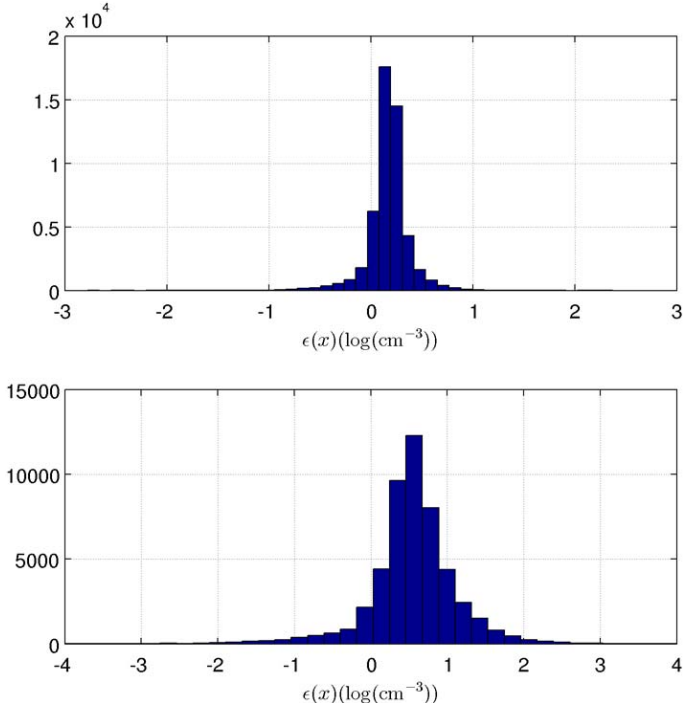

Fig. 5. Histograms of the approximation errors $\epsilon(\boldsymbol{x})$. Top: approximate parameterization with 7 size sections for the aerosol particle size distributions. Bottom: approximate parameterization with 4 size sections for the aerosol particle size distributions.

GMDD
6, 2551–2583, 2013

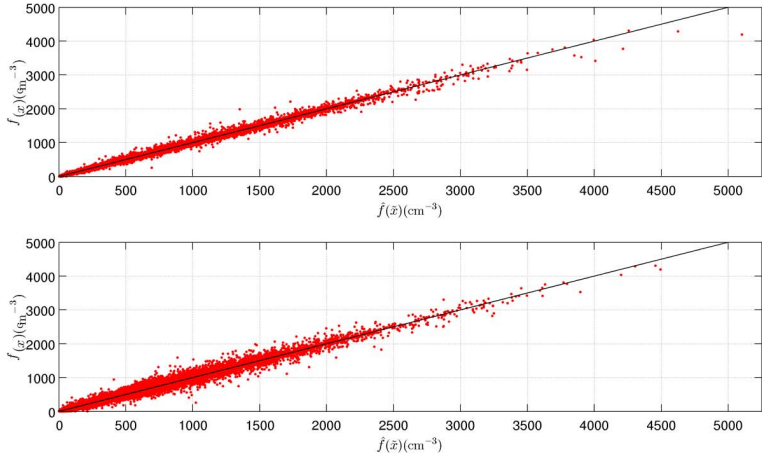

Fig. 6. Computed cloud droplet number concentrations (CDNC) computed with the accurate parameterization $f(\boldsymbol{x}_j)$ as functions of CDNCs given by the approximation error corrected parameterization $\hat{f}(\tilde{\boldsymbol{x}}_j)$. Top: reduced parameterization $\tilde{f}(\tilde{\boldsymbol{x}})$ with 7 size sections for the representation of the aerosol particle size distributions. Bottom: reduced parameterization $\tilde{f}(\tilde{\boldsymbol{x}})$ with 4 size sections for the representation of the aerosol particle size distributions. Black solid lines represent the identity lines.

GMDD
6, 2551–2583, 2013