We thank the Referee for the review of the manuscript, and valuable suggestions to improve it. Below you can find the original comments from the Referee (marked with boldface) and our responses to those.
In addition to the replies to the referee, also the following changes were done in the revised manuscript.

- The results were recomputed using the RF algorithm of the freely available scikit-learn library (http://scikit-learn.org). The description of the RF algorithm in Chapter 2 was modified and the Algorithms 2 and 3 were replaced by a scikit-learn code listing. The reason for this change was that it makes the RF chapter shorter and easier for the reader to implement the proposed approach.

- In the new results, also slightly different version of the ARG Fortran codes was used. This change caused minor changes to the running times of the ARG parameterizations.

- Because it turned out non-trivial to evaluate the memory usage of the new RF implementation, it is not included in the revised paper.

- A minor change to the probability distribution of the pressure (from $\mathcal{U}(10^3, 10^5)$) to $\mathcal{U}(10^4, 10^5)$) was done to make it produce more realistic pressure samples.

- Complete result tables of all the different RF models were added as supplementary material of the paper.

This paper introduces a new method for reducing errors in simplified (or reduced) models or parameterizations of physical processes used in complex geophysical models. The method estimates the approximation error in a reduced model in comparison to a more accurate model, then applies this error estimate as a correction to the reduced model. It uses the Random Forest technique to construct a regression of this error on the inputs to the reduced model. The use of the Random Forest regression technique for this type of problem is a new and potentially useful method for geophysical modeling.

General Comments

The Abdul-Razzak and Ghan (2002) aerosol activation parameterization for sectional representations (hereafter ARG-2002) does not predict cloud droplet number concentration (CDNC). It predicts number and activation fractions for each size section. The total number of activated particles gives the number of cloud droplet formed in a rising air parcel, which is equivalent to CDNC only in a non-entraining parcel model with no additional microphysics. In 3-D atmospheric models, this droplet formation, in conjunction with cloud microphysical processes and transport, together determine the CDNC. The activation fractions are important for both droplet formation and in-cloud wet removal of aerosol. The output from ARG-2002 is thus a vector quantity, and the paper does not address vector outputs (and errors) at all. Since many (if not most) atmospheric process models have multiple outputs, this is a serious limitation of the method and the paper.

The reviewer is correct that activated fraction in different size sections should be the output of correction as it is needed also in wet removal calculations. With CDNC-parameterizations activated fraction can be calculated from corrected CDNC with Kohler theory as the hygroscopicity of aerosol in different sections is known, and thus the vector output is not necessarily needed in this case.
Despite the use of a scalar valued function as the test case in this paper, the proposed approach for compensating the approximation errors is not restricted to scalar valued functions. This was clarified in the general description of the approach by changing the notations of functions to bold face indicating vector valued functions and giving definitions of spaces of input and output parameters as real valued vector spaces in Chapter 2.1 of the revised paper. The testing of the proposed approach with other geophysical simulations where the output is vector valued function is left as a topic of future studies.

ARG-2002 is itself a simplified or reduced model, and using a first-principles parcel model of aerosol activation as the accurate model would be more appropriate in this study. In addition to the reasons raised by the first reviewer, I am concerned that the simplicity of ARG-2002 (relative to a first-principles model), combined with the relatively smooth aerosol distributions (3 lognormals) used in the

paper, may produce an oversimplified representation of the aerosol activation problem. As a result, the performance of the approximation error method presented here may be considerably optimistic. Using a first-principles model as the accurate model, they could also demonstrate how the AE method performs for reduction of the internal numerics of the process model, in contrast to reduction of the input parameters.

Results computed with an air parcel model as the accurate model were added. Also in this case, the compensation of the approximation errors significantly improved the results. For example, the mean relative error of the CDNC decreased from 36.6% to 13.1% when using the AE correction for the discrepancy between the air parcel model and the ARG parameterization. The results show that the proposed AE compensation approach is also capable of reducing the errors due to reduction of the internal numerics of the process model.

Title. The title is somewhat inappropriate, as the paper contains no results that directly involve the first indirect effect. (E.g., no plots showing droplet number versus aerosol number.) Correction of approximation errors with Random Forests applied to modeling of aerosol activation would be better.

The title has been changed to "Correction of approximation errors with Random Forests applied to modelling of cloud droplet formation".

P 2552, L 17-19. This point about a large number of variables should be discussed at more length in the results or conclusions. (Also see later comment listing the actual input variables.)

The discussion in the conclusions about a large number of variables was extended.

P 2554, L 25-28. Some transition and context for the RF method is needed here. E.g., The additive approximation error can be estimated using a number of techniques, such as ... Here mention other methods (e.g., neural network) in addition to RFs, and their

strengths and weaknesses (if possible).

A comment about the possibility to use other methods, such as neural networks, together with references was added to the introduction.

P 2556, L 12-13. Define x here

Definition of x was added.

Section 2.1. Some discussion is needed for when f is a vector quantity.

In general, f is not restricted to scalar valued variables but can also be a vector. This was clarified by changing the notation of f to bold face indicating a vector valued variable and giving definitions of spaces of input and output parameters as real valued vector spaces in Chapter 2.1 of the revised paper.

P 2560, Algorithm 2 definition. Notation. Use a different index variable for trees, such a j, since k is used for other things. (This will just make it easier to follow.) Steps 3 and 8. These need to be explained in considerably more detail (here or in discussion that follows). Step 7. Maybe change to For each of the split variable candidates, construct a random set of Nsplitp split threshold values.

The implementation of the RF regression model was changed to the RF algorithm of the freely available scikit-learn library (http://scikit-learn.org). The description of the RF algorithm in Chapter 2 was modified and the Algorithms 2 and 3 were replaced by a scikit-learn code listing. The reason for this change was that it makes the RF chapter shorter and easier for the reader to implement the proposed approach.

P 2560-2561. I was not able to fully understand the Algorithm 2 description, and I suspect that many GMD readers will have the same problem. The description needs to be lengthened to clarify some of the details in the algorithm steps. Making it 2-3 times longer would not lengthen the paper very much. Algorithms 1 and

3 seem much simpler in comparison, so more detail is appropriate for Algorithm 2.

Algorithms 2 and 3 were replaced by a code listing showing the training and evaluation of the random forest models in scikit-learn library. See the reply to the previous comment.

P 2561, Eqn. 5. It is unclear why Eqn. 5 is an equality, while Eqn. 3 is an approximate equality. If Eqn. 5 is really needed, then more explanation of why Eqn. 3 and 5 differ is needed. Also, it might be clearer to introduce an epsilon-tilde symbol for the approximated error, and reserve epsilon for the exact error as defined by Eqn. 2.

The eqn. 3 of the original manuscript was changed into an equality and the epsilon was replaced by an epsilon-hat symbol to highlight that the computationally fast predictor $\tilde{g}(\tilde{x})$ gives an estimate for the approximation error. In the revised manuscript, the Eqn. 5 of the original manuscript was removed as unnecessary.

P 2562-2563, Section 3. A better description of the activation parameterization is needed. The paper says they are using ARG-2002, but the Kokkola et al. (2008) SALSA model paper describes modifications to ARG-2002. Are they using ARG-2002, which assumes dN/dlogD is uniform within a section, or modifications to it? Do the modifications involve only the calculation of activation fractions for each section (based on the intra-sectional size distributions), or the calculation of maximum supersaturation also? Given that the 4-section model consistently underestimates maximum supersaturation (P 2565, L 15-17), this second type of modification could be important and could reduce the quite large errors.

In this paper, the ARG-2002 is used without any assumptions of for the shape inside the size section. This can produce error in supersaturation calculation like shown in several publications. With very small number of sections it is not straightforward to correct the shape without prior assumption. This will affect also other aerosol processes than activation to form cloud droplets, and we will study this more detail in future for example in the case of aerosol nucleation.

P 2563, L 12. If the ARG-2002 output is treated as CDNC, then some discussion about assumptions involved is needed. (See general comments above.)

Here we refer to CDNC as the number of cloud droplets at the cloud base. We will make it clear in the manuscript that further processes, e.g. entrainment will affect the CDNC within the cloud.

P 2563, L 13-14. Neglecting $D < 50$ nm particles is not a good assumption at high updrafts and low aerosol concentrations. Why make this assumption, as it makes the accurate model even more reduced?

This assumption was done as in some cases the inclusion of small particles increases artificially the saturation maximum and CDNC when ARG-2002 parameterization is used. This is unphysical and should be taken into account. However, the scope of this paper is not to evaluate the goodness of parameterization itself but to present new methodology for correcting approximation errors, and thus the size range was limited.

Section 4.3. The input variables to the reduced models should be listed. I would expect something like aerosol number, mean size, and hygroscopicity for each section, plus T, p, and w. This gives 15 and 24 inputs for the 4 and 7 section models, so how does the Nipcands=25 work with this. Also, given the last sentence of the abstract, there should be some discussion (here or in conclusions) about this being a large number of input variables.

A table listing the input variables was added, see Table 1 in the revised manuscript.

P 2566, L 9-10. Provide some rationale for these values of Ntrees, Nipcands, Nmaxsamples, and Nsplitp.

Based on a preliminary test, limits for the RF parameters that resulted in models with a good accuracy and feasible computation times were selected. Further, it was noted that by using parameters resulting in more

time-consuming models, the accuracy was not significantly improved over the models constructed with the parameters within the selected ranges. Within this range a relatively uniform selection of the parameters was used in the computations. A sentence "These parameter ranges were selected based on a test which showed that selecting values outside these ranges either resulted in poor model accuracy or considerably larger computational burden with no significant improvement on the model accuracy." was added.

P 2566, L15-20. Error metrics. I suggest using RMSE rather than MSE. It provides the same information, but in units that are meaningful to atmospheric scientists. Also state the mean f from the accurate model, to put the RMSE values in perspective. (Or you could normalize the RMSE by the mean of f.) Some bias statistic should also be included.

The MSE values were changed to RMSE values. Also the mean of the CDNC values computed with the accurate model was added to Tables 5 and 6.
The RF regressor is asymptotically unbiased when the number of training data samples and the depth of all the tree models approach to infinity. The bias for the given input parameters and tree model is equivalent to the discrepancy between the true value of the approximation error and the predictor estimate. Hence the RMS errors give mean of the bias over the ensemble.

P 2567-2568. Provide more discussion of how the accuracy and timing depend on the Ntrees, Nipcands, Nmaxsamples, and Nsplitp parameters. This would be of interest to potential users of this method.

The complete result tables of all the different RF models were added as supplementary data for the paper to show the reader the effect of the RF training parameters on the model accuracy.

P 2567, L 26-28. Give an example of these minor variations due to randomness. E.g, for one of the Ntrees=25 cases, state the ranges and/or standard deviations of the error metrics.

An example of the variations was added to the Results chapter: "As an example, in the 7 size sections AE corrected model with the RF training pa-

rameters *n_estimators*=400, *max_features*=2 and *min_samples_split*=15, the RMSE and the MRE varied between values 40.4–40.6 cm^{-3} and 8.91–8.92 %, respectively."

P 2568, L 15-23. The discussion here of the smallest/fastest RF models should be revised. Some things are repeated, such as the .11-.16 and .07-.11 ms times.

The text was reviewed and unnecessary repetition was removed.

Section 5. Can you say more about areas for future work or improvement of the RF technique? One thing that comes to mind is giving different weights to different input variables in the random sampling of Algorithm 2 (more weight where output sensitivity to input is higher).

Some potential future work topics are listed in the conclusions. We thank the referee for suggesting the idea of weighting the input variables. We have added comment about this potential topic of using weighted RFs into the conlusions and plan to study and test the idea in approximation error compensation of geoscientific models.

Minor comments

P 2552, L 2. Define reduced here, or change to ... using reduced (i.e., simplified) models.

Changed as suggested.

P 2554, L 5-6. Use of measurement model here was not clear.

The use of term measurement model was unnecessary and it was removed.

P 2552, L 13; P 2555, L 26; Section 3 heading; and other places. Cloud droplet activation is incorrect terminology, although frequently used. Aerosol particles activate, and this leads to formation (or nucleation) of cloud droplets. Use aerosol activation or

cloud droplet nucleation or cloud droplet formation.

Cloud droplet activation was changed to cloud droplet formation.

P 2555, L 5. Change models to model.

Corrected as suggested.

P 2556, L 24-25. Suggest deleting Note that model (2) is accurate but. Model (2) is never used (why would it be), so this point can only add confusion.

The purpose of the sentence is to clarify that it is not feasible to compute the exact value of the approximation error as it would require evaluation of the accurate model. The sentence was left unchanged.

P 2557, Eqn. 3. Swapping left and right hand sides would seem more appropriate. Epsilon is already defined, and you are constructing g-tilde to approximate it.

Corrected as suggested.

P 2558, L 12. Maybe change are considered to are used to approximate g(x).

Corrected to are used to approximate $\tilde{g}(\tilde{x})$

P 2562 L 4-5. Change to In many atmospheric models, this process is parameterized. Some LES and cloud resolving models treat this process explicitly.

Corrected as suggested.

P 2563, L 2-4. State the resolution within each subrange (e.g., constant delta-logD).

The resolutions within each subrange were stated by adding a sentence: "In this study, the size sections within subranges have a constant volume ratio

between the adjacent sections."

P 2563, L 15. Change nucleus to nuclei

Corrected as suggested.

P 2563, L 20. The mathematical notation here (that f(x) is a real number) is unnecessary. Use simple text.

$f(x)$ could also be vector valued, see earlier comments. The mathematical notation was added to emphasize that in this case $f(x)$ is real valued.

P 2563, L 26. Add more digits to the accurate model timing value (1 ms).

More digits were added, the average time was 0.92 ms.

P 2565, L 4-8. Clarify the description of volume fractions. Is it that for a given mode within a given sample, they are randomly selected but fixed within a mode, and the volume fractions differ between the three modes? Also, what hygroscopicities were used for organic carbon and dust.

"The volume fractions for the sulphate was drawn from an uniform distribution $\mathcal{U}(0.01, 1)$" was changed to "The volume fractions for the sulphate were drawn from an uniform distribution $\mathcal{U}(0.01, 1)$ separately for each mode."
Dust was assumed to be insoluble, the hygroscopicity of organic carbon in cloud activation was calculated assuming that it is fully soluble with dissociation constant of 1 in Kohler equation.

P 2565, L 11-13. Suggest you drop the (x, y) notation used to describe Fig. 4, as it is unnecessary, and these variables are used for other things. So on L 13, ...the identity line f(x) = f(x)-tilde corresponding...

Corrected as suggested.

P 2565, Eqn. 8 and 9. The notation involving log(f) could be

improved. You are just transforming f and f-tilde to f-prime and f-tilde-prime, calculating epsilon-prime from the transformed fs, then applying the RF technique with these transformed quantities. Very simple, but not clear from the equations.

The Equations 8 and 9 were reviewed and the notation was simplified by defining f-prime and f-tilde-prime that are the logarithms of f and f-tilde.

P 2565, L 26. The notation here (two groups of brackets) is inconsistent with the notation used for inputs to Algorithm 2 on page 2560.

The algorithm 2 was replaced with a code listing and this inconsistency was removed.

P 2568. It would be helpful if somewhere in this paragraph, you restate the timing for the accurate model, as it was given several pages earlier (so perhaps forgotten by readers).

The timing for the accurate model was restated to help the reader with the comparison of running times.

P 2568, L 15-23. Change increment of computation time to increase in computation time. The increments (due to the AE calculations) are .07 and .04 for the 7 and 4 section models.

Corrected as suggested.

Figures. Axes labels should be larger size in many figures.

The figures were re-plotted and axes labels should now be larger.

Fig. 2, 3, 5. Add labels to vertical axes.

The labels were added.

Fig. 4, 6. I would expect to see horizontal axes used for accurate model results (the truth), and vertical axes for reduced model re-

sults (the approximations).

The horizontal and vertical axes were swiched.