**Atmospheric
Chemistry
and Physics
Discussions**

Interactive comment on "Mapping the uncertainty in global CCN using emulation" *by* L. A. Lee et al.

L. A. Lee et al.

l.a.lee@leeds.ac.uk

Dear Dr van Oijen,

Many thanks for your review of our paper and your constructive comments. As a result of your review changes have been made to the paper as detailed below. Please find your comments in bold with a response in normal text. The additions to the paper are in italics.

Incorrect statement: model response to a single parameter can be nonlinear and an OAT sensitivity test can show that. What OAT tests can not do is account for interactions between parameters.
It is true, OAT tests can capture non-linear behaviour in the model. We have made this correction in the paper by saying:

However, OAT tests cannot identify and quantify interactions between parameters and they consider only a small fraction of the total parameter uncertainty space.

Explain the phrase "lognormal modes" which is common in aerosol modelling but not elsewhere and it conflicts with the statistical meaning of the term "mode" of a probability distribution.
A sentence has been added to show that the model uses a "modal" as opposed to a bin (or sectional) representation of the aerosol particle size distribution and that the term "mode" does not refer to statistics.
The modes describe the functional shape of the particle size distribution (particle number versus size). Such a modal model provides a simpler and more computationally efficient representation of the particle distribution than alternative sectional schemes in which the distribution is described as a number versus size histogram (?). The model physical and chemical processes calculate the time dependent evolution of the number concentration and size of particles within seven modes.

How many parameters does the model have besides the 8 selected for sensitivity analysis? And which processes are included in the model that are not controlled by any of those 8 parameters?
The following has been added to the text:
In a follow-up study we identified 37 potentially important uncertain parameters and investigated the uncertainty in CCN and other model outputs to 28 of these parameters after expert elicitation allowing further process representations (such as nucleation and dry deposition) to be included in the analysis; here we use only 8 to illustrate the method more clearly.

The description of the emulation method may easily be misconstrued
We have kept the use of 'the emulator' in the description as each emulator is built the same way but we have added the following to the end of the paragraph to clarify that

the work is carried out with multiple emulators.

In this work, separate emulators are built for each monthly mean CCN value in every grid box on one vertical level of the model, which amounts to 8192 emulators. The sensitivity analysis is carried out in every model grid box taking no account of the spatial correlation.

There should also be a much more in-depth analysis of the limitations of Gaussian Process modelling as applied to aerosol modelling, including the importance of prior assumptions about smoothness of the original model's CCN predictions.

The following has been added to the text.

The prior distributions here are typically used as uninformative priors for the Gaussian process, so in effect all posterior information comes from the training data. It is possible once the model runs are available to build the emulator with different prior assumptions. Here, the emulator was built with different covariance functions and different input distributions with little difference in the results.

The validation will reveal any issues with the prior distribution, in particular the covariance function and the smoothness of the model response. If the emulator is not valid then the prior assumptions can be changed or more training data obtained and the emulator rebuilt. The emulator will not be validated if there are regions of sharp change in the model response to changes in parameters (i.e., the model is not smooth). The validation will show where the discontinuities in the model response lie, providing valuable information on the model behaviour. In such cases other methods of representing the model response for sensitivity analysis will have to be considered. This was not the case in our study.

Evidence of"the" emulator's quality is clearly not sufficient A new subsection has been added to explain how we validated our emulators in this work.

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

Interactive
Comment

As in Lee11 emulation can only be used for sensitivity analysis in every grid box if it is validated according to some criteria. In Lee11 the validation was simple since there were only two emulators to validate (one for each grid box). Here we have 8192 emulators to validate. The validation procedure is the same as in Lee11 whereby extra runs of GLOMAP are compared to the emulator predictions but a summary of the results is required. The first summary result does not involve these extra GLOMAP runs but quantifies the uncertainty around the emulator predicted mean due to the emulation as which is done analytically in Lee11 using GEM-SA (?). The uncertainty around the emulator predicted mean output in every grid box has to be small compared to the uncertainty around the emulator mean output due to the uncertain parameters (in Lee11 these values are defined as $V^*(E)$ and $E^*(V)$). In the DiceKriging package the emulator uncertainty has to be simulated. The posterior mean function is used to carry out the sensitivity analysis but there are infinitely many possible functions within the emulator uncertainty which we can simulate and compare to obtain an estimate of the emulator uncertainty. For the 14 grid boxes shown in Figure ?? the simulated uncertainty around the emulator mean due to the emulation ($V^*(E)$) and due to the uncertain parameters ($E^*(V)$) is shown in Table ?? (to show that these values are a result of simulation rather than analytically calculated they are defined by $V^{*\hat{}}(E)$ and $E^{*\hat{}}(V)$ here). The values in Table ?? are calculated by simulating 1000 possible functions from the posterior Gaussian process conditioned on the training data. Table ?? shows that for the 14 grid boxes in Figure ?? the uncertainty around the emulator estimated July CCN is small compared to the uncertainty due to the uncertain parameters and so the signal to noise ratio of the function is large enough to get a meaningful sensitivity analysis. For comparison the same results were obtained by GEM-SA showing consistency in the results. As in Lee11 the emulator prediction and its uncertainty can be plotted versus the GLOMAP prediction and investigated. Figure ??a shows an example of the grid box validation. It should be noted here that 3 of the original 24 validation runs in Lee11 have been removed as the input settings were incorrect in the GLOMAP runs. The validation plot shown in Figure ??a is summarised for every grid box in Figure ??b by

counting how many of the emulator uncertainty intervals contain the GLOMAP prediction. Figure ??b shows some regions where the emulator uncertainty does not contain the GLOMAP prediction at least 90% of the time, further investigation of these regions shows that the emulator prediction and the GLOMAP prediction are in fact very close but the emulator is too confident and so the interval too small. An example of regions where the emulator is too confident but is nonetheless doing a good job of predicting the GLOMAP value is shown in Figure ??c. In this case it is clear that the validation points close to the training data do not match the GLOMAP run and contain small bias indicating that σ^2 is underestimated here. In such cases the emulator prior assumptions can be changed to increase the emulator uncertainty but in this case this would not improve the sensitivity analysis which depends only on the mean values (marked by dots in Figures ??a and c). Diagnostic plots are also created by DiceKriging but as with the in-built diagnostics in GEM-SA the validation is not out-of-sample, despite this the diagnostics available in DiceKriging and GEM-SA (GEM-SA is only checked for the 14 grid boxes in Table ??) were checked and show no reason to declare the emulators used here invalid. Given all the results together in Table ??, Figure ?? and the software diagnostic checks we declare our emulators valid for purpose and carry out sensitivity analysis for every grid box using its associated emulator.

"clear and testable assumptions"
This sentence has been changed and together the additional text in the statistical methods should help to make the point clearer.
It is based on well established statistical theory with clear prior assumptions as detailed in Section ??. The choice of priors used in GP emulation can be tested with no further model runs to test for statistical robustness.

Technical corrections Techinical changes have been made throughout the text.

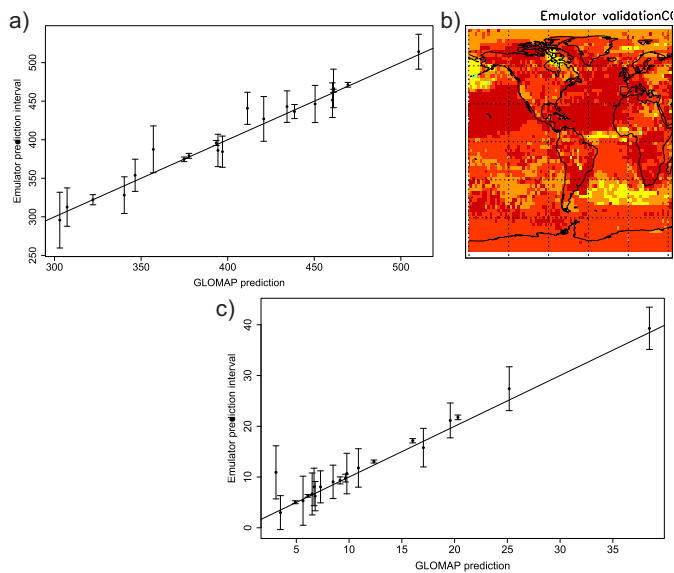Interactive comment on Atmos. Chem. Phys. Discuss., 12, 14089, 2012.

Fig. 1. New validation plots