Response to anonymous reviewer 1: Partridge et al., 2012.

The authors thank anonymous reviewer 1 for insightful comments on the manuscript. The reviewer provided several suggestions for improving the readability and quality of the manuscript. We have followed the suggestions in most cases, and our detailed response is outlined below.

Summary of main changes:

- We have extended the discussion in the paper with respect to the parameter sensitivity by adding an entirely new section in which we present results where the updraft velocity, mass accommodation coefficient, and surface tension are included as calibration parameters.
- From these new results we demonstrate that for the cloud parcel model employed and aerosol environments investigated, the updraft velocity is an extremely important parameter whereas the mass accommodation coefficient and surface tension are not.
- It is also demonstrated that the inclusion of these additional parameters does not affect the results or conclusions presented initially in the paper when we included only four calibration parameters.
- We have extended the discussion to include comments on the effects of the choice of synthetic measurement error.
- Figure 5 in the paper has been altered to present the results in a clearer manner.

Response to major comments

M1: *The sensitivity studies of the work are based on perturbing the parameter values through equations 4-6. As seen from the equation, all parameters are perturbed by sampling from a normal distribution in a uniform manner so that the relative "spread" stays constant among the perturbed parameters. In the author's words, this represents a "synthetic measurement error".*

M1A: *How sensitive are the results to the choice of the way how the parameters are perturbed? For example, how sensitive the results are to the choice of the coefficient in equation 4 (0.10 is the current value)?*

M1B: *Also, would be possible to incorporate physical knowledge on the uncertainties to the perturbation scheme? For example, particle size distribution can be relatively easily determined compared to the soluble mass fraction. I do not propose that authors include a comprehensive qualitative study to address the issue, but it would be good to discuss about the point.*

RM1A: Firstly because this is a synthetic study, any error model selected is not going to be 100% applicable to the real world depending on the type of instrumentation on the aircraft used to measure the droplet size distribution and its resolution, i.e. whether forward scattering spectrometer probe (FSSP), fast forward scattering spectrometer probe (FFSSP) or other. The measurement error variance (σ) that we use in this synthetic study should make physical sense and be related to the measurement error of the data. The particular choice of error function used here was guided by experience with real world measurements.

Secondly the focus of this first study in which an MCMC algorithm is coupled to an adiabatic cloud parcel model is the global sensitivity of aerosol measurements with respect to synthetic modeling. As we note in the introduction of the paper, many synthetic modeling studies have been performed to investigate the relative importance of aerosol size vs. chemistry, however, very few have been global and none have used MCMC.

With these points in mind we feel that examining the error model in detail is beyond the scope of this paper. We acknowledge that it is important to understand how the detailed form of the measurement errors impacts the results of the inverse modelling however; we believe that it is more appropriate to concentrate that work on a study that includes real world measurements. This type of study can inform us about the relationship between (real) measurement errors and derived parameter sensitivities.

However, it would be possible to perform different runs with a different magnitude coefficient in which we use to calculate measurement error, e.g. Coef = 0.01; 0.05; **0.10 (used in this study)**; 0.25 and 0.50 and calculate the standard deviation of the parameters of the posterior sample. You will see that the standard deviation of each parameter will go up with increasing value of the error. This was seen when we performed simulations with Coef=0.20 (figures not shown in paper). In other words, the larger the size of the error, the larger the posterior uncertainty, since with a smaller σ the posterior density function probability mass is spread out more over of the parameter space and the parameter uncertainty is increased. However, when the simulations were repeated for Coef=0.20 although the parameter sensitivity did decrease, the relatively sensitivity between the different calibration parameters remained very similar, which is important as this is the focus of our synthetic study.

RM1B: Yes you can use the deviation from some a-priori known parameter values in the likelihood function. You have a combined objective function that then not only contains the distance of the model to the calibration data but also the distance of the selected parameter values to their prior values. To do this in a statistically accurate manner, we need detailed information about the error of each parameter. This is required for a proper weighting of both entities in the objective function. However, a problem arises if the number of calibration data points, n, is much larger than the number of parameters, p or $n >> p$. In this case, the MCMC approach will still tend to fit the observations, because the size of the residuals is much larger than the size of the parameter deviations. This is somewhat resolved by using more restrictive prior distributions that mimic the information available about the parameters prior to the calibration data. The main thrust of this paper however relies in testing the models ability to properly fit the data, and invert the parameters. We assume that the parameters are not known a-priori, a typical situation in many modelling studies, where parameters often represent non-measurable entities. Even if they represent physical properties, we typically cannot measure them directly at the scale of interest, and calibration is required because of problems with temporal and spatial aggregation.

The thrust of our work relies on moving towards a new framework by which the identification of structural errors in the adiabatic cloud parcel model is possible. However, in this part of the paper series we outline our methodology using synthetically generated CDNC distributions and utilise the information stored in the posterior sample to calculate the global parameter sensitivity. It is important that this step is taken for synthetic observations, as many synthetic studies have been performed that only consider local sensitivity. In a forthcoming paper we will investigate aerosol-CDNC distribution closure by utilising measurements from the Marine Stratus/Stratocumulus Experiment (MASE II) campaign within the framework we have developed. This will allow us to assess model structural errors. These become apparent when the parameters are allowed to take on any value within their physically realistic ranges. The deviation from the inversely estimated parameters and their "measured" values is a useful diagnostic to assess the error in the model itself. For a perfect model, model input and calibration data, the inversely estimated parameters will match with those independently observed in the field. If we assume that the forcing data of the model is observed accurately, then the distance of the inversely estimated values to their measured values is thus a direct estimate of the error in the cloud-aerosol model. Investigating these differences provides the necessary inspiration to improve our models, and provide a better description of cloud-aerosol interactions. This should reduce predictive uncertainty in global-climate models.

M2: *As can be seen from tables 1 and 2, several important parameters are fixed in the sensitivity studies. Would it be possible to repeat the sensitivity study while perturbing e.g. mass accommodation coefficient and/or updraft velocity. Although the authors touch this topic briefly in the text, I believe that a more extended discussion would strengthen the manuscript. In particular, are the conclusions of the study sensitive to the choice of the perturbed parameters?*

Several parameters were indeed fixed in the main analysis of the paper, most notably the updraft velocity which was shown in P11 to be very important.

The updraft is known to be important; however the main interest was to investigate relative importance of the aerosol size versus chemistry in relation to "synthetic studies". As we perform this investigation for a range of different base updraft velocities, the main sensitivity analysis was made simpler and clearer by keeping this parameter fixed.

However, we believe that the reviewer makes a very good point and have therefore added a section in the paper in which the updraft, mass accommodation coefficient and surface tension are also investigated. This section is titled: Inclusion of additional calibration parameters.

The main results are not found to be significantly affected by the choice of number of perturbed parameters unless the aerosol Aitken lognormal parameters are also included. The correlation between these additional lognormal aerosol parameters and the original four calibration parameters results in a reduction in the relative sensitivity of the accumulation mode lognormal parameters.

Response to minor comments

1.

The reference has been updated.

2.

The concept of the marginal distribution has been extended in Section 3.5.

3.

The associated text in section 3.4 has been modified accordingly to clarify this point, specifically regarding the difference in the prior ranges presented in Tables 1+2 compared to P11.

The true values defined in Tables 1 and 2 are the values used for the calibration parameters for each environment when defining the calibration data before it is perturbed with a 10% heteroscedastic error.

The prior limits defined in Tables 1 and 2 apply to the range over which DREAM is allowed to search for during on route to the posterior distribution. The algorithm can only try calibration parameter values that lie within these prior limits.

The prior limits are also used in the ensuing sensitivity analysis since the relative sensitivity for each calibration parameter that is perturbed is calculated by normalizing the posterior range by the prior range.

4.

The caption has been updated accordingly.

5.

The caption has been updated accordingly.