Reply to Anonymous Referee #1

We are grateful to the referee for her/his constructive criticism and suggestions to our manuscript. Please find below a detailed point-by-point reply (referee's comment in *italic*).

The model setup, however, exhibits two major problems the authors themselves seem to struggle with in their evaluation at several places. Firstly, the model grid spacing is with 2.8 degree horizontal spacing and 19 vertical layers very coarse nowadays. Based on the publications that are cited, this setup seems to be in use by the EMAC community for more than ten years now. The choice of such a coarse resolution hinders comparability with in-situ measurements.

We are aware of the fact that the T42 (2.8°) resolution chosen for this study is quite coarse, also for the standards of a global model. As mentioned in the abstract, however, we plan to use MADE3 not only for process-oriented studies on aerosol-induced ice formation in the troposphere, but also for a reassessment of the climate effects of anthropogenic aerosol emissions. The latter kind of studies usually requires performing a large number of sensitivity simulations to account for the numerous uncertain parameters characterizing these effects (see for example, Righi et al. 2013). In this case, having a computationally efficient model which is able to capture the main processes also with a coarse resolution is essential, and motivated our choice of the resolution presented here.

We also note that increasing the resolution, for example to T63 (1.8°) would not allow to resolve the scales probed by the in-situ measurements, which are typically of the order of a few kilometers. Nevertheless, in future studies we plan to apply the model at higher resolution, e.g. T63 with 31 vertical levels, to achieve improvements in the representation of atmospheric processes driving the aerosol life cycle.

We have included some additional information on these future plans into the manuscript: "A thorough investigation of such discrepancies would require a large number of sensitivity simulations, including model experiments with different representations of processes and/or different spatial resolutions. Although beyond the scope of the present evaluation, this could be conducted as part of future studies, and also serve for quantification of simulation uncertainties." (last paragraph of P10) and "This should also include simulations with alternative assumptions for the mode widths and with higher spatial resolution." (Sect. 4).

Secondly, the time frame of the model simulation does not align with several of the observations that are used. As a result, temporal collocation is not possible further reducing the comparability.

The referee correctly identified a temporal discrepancy between the applied emission data and some of the observations used for model evaluation. We have applied emission data for year 2000 since a robust inventory is available for that year (Lamarque et al., 2010). These emission data sets were assumed for the years around 2000 (1996-2005). For consistency reasons, we have adopted observational data from this time period in most cases. An exception is the observational data from several field campaigns, particularly some recent aircraft-based field studies which were carried out up to 14 years after 2000. However, the discrepancies between model and observations in these cases are similar to those found for campaigns close to 2000 and there is no systematic trend in the deviations. In addition, these deviations are clearly larger than the changes in emission rates occurring between 2000 and the years of the respective campaigns. Hence internal model discrepancies are probably the main reason for these deviations, rather than trends in the input data. In order to meet this very important referee comment, we have included a corresponding discussion at the end of Section 3.2: "Parts of the discrepancies discussed above could also result from temporal inconsistencies between the simulations and the observational data. We apply emission data for the year 2000 since a robust emission inventory is available for that year (Lamarque et al., 2010). These emissions are assumed valid for the years around 2000 (1996-2005). For consistency reasons, we adopt observational data from this time period in most of the comparisons discussed in this article. An exception is the data from recent aircraft-based field campaigns, which were carried out up to 14 years after 2000. However, deviations between model and observations in the more temporally dislocated cases are similar to those found for campaigns close to 2000. A systematic trend in the deviations does not occur. In addition, the deviations are clearly larger than the changes in emission rates

occurring between 2000 and the years of the respective campaigns. Hence internal model deficiencies, as described above, are probably the main reason for the deviations, rather than trends in the input data."

Despite the shortcomings in the model setup the manuscript provides a thorough evaluation of the MADE3 aerosol scheme. I recommend it for publication after my comments are met. Nevertheless, the authors should consider a change in their setup for future studies.
We agree. See reply to first comment above.

The authors should think about removing 'model description' from the title, as the paper clearly focusses on evaluation of the model.
Although the paper mainly focusses on evaluation, it also describes the global version of MADE3 within EMAC. Since no previous description of this setup has been published, we think it is appropriate to have "model description" in the title.

(P1 L16-17, P11 L22-24, P20 L22, P22 L27, P24 L28-29, P28 L3-4): Several sections are ended with an outlook and further plans. In my opinion this devaluates the results presented in the respective sections. Especially because this is done too frequently. These plans can be part of the summary section though.
The statement at P1 L16-17 serves as motivation for the model development presented in the paper and we would prefer to keep it. We have also kept the paragraph at P11 L22-24, since it addresses the first issue raised by the reviewer (see above), and the sentence at L28 L3-4, since it specifically concerns the topic discussed in that paragraph and also addresses a comment raised by both referees.
We agree with the referee about the other sentences and we have removed them from the manuscript. As suggested, we have added a paragraph to the Conclusions section about the plans for future studies: "Future studies with MADE3 should focus on the analysis and reduction of the model discrepancies highlighted in the present evaluation. This could include, for example, the consideration of observational uncertainties, a detailed analysis of the scavenging efficiency and its dependency on the aerosol size distributions and the underlying microphysical processes, as well as simulations with higher spatial resolution and model experiments focusing on new particle formation processes considering different nucleation parameterizations."

(P2 L20): The formation efficiency of ice particles depends not only on size but also strongly on surface area (see Hoose and Möhler, 2012, which is cited in the same line).
Thanks for noting this, it has been added.

(P2 L29): A two-moment aerosol microphysics scheme does not really 'explicitly simulate the size distribution'. The shortcomings of the approach with a fix standard deviation of the distribution are obvious in section 3.3.
We have made this statement more precise by adding "(assuming lognormal modes with fixed widths)".

(P3 L22): ECHAM5 seems to be a rather outdated version of this model. A more recent version might also be used easier at higher resolutions. Please comment.
Indeed, ECHAM6 is the newest version of this model. This version is, however, not available as base model in EMAC. Nevertheless, both ECHAM5 and ECHAM6 can be operated at the same resolutions.

(P3 L27): As I understand output was written in 12h intervals. Depending on what exactly is written this can lead to a bias. For example, in one time zone the output is always written at 0 and 12 local time, in another one at 6 and 18 local time. Could the authors please clarify?
The authors agree with the referee that the current method could lead to discrepancies in the 'local' output times. From our experience this is, however, not too critical in the case of aerosol tracers, since their variability is largely driven by changes in synoptic scale dynamics and diurnal variations mostly

are of secondary importance in the simulations. Nevertheless we will try to increase the output frequency in future simulations.

(P3 L31): Please add more information on the emission inventory, i.e. which one is used and at which resolution is the raw data.

Details of each emissions inventory used for this study are given in Sect. 2.4. Since this is essentially based on a previous model study (Righi et al., 2013), we prefer not to give too many details here, but rather refer to that publication and only highlight the differences in the present manuscript. These mostly concern the assumptions for the size distribution of the emitted particles, which were outlined in detail in Table 2.

(P5 L7): Please add a short description of the "big leaf" approach.

We have extended this sentence as follows: "…DDEP, which uses the so-called big leaf approach assuming that deposition fluxes within the canopy have the same relative responses to the environment as any single leaf, and that the scaling from leaf to canopy is therefore linear (Sellers et al., 1996), …"

(P6 L8): From the text and also from Figure 1 it did not become clear to me how the authors are dealing with SOA in MADE3. Please extend!

We have tried to clarify this by extending this part as follows: "Sulfuric acid (H_2SO_4) and secondary organic aerosol (SOA) precursors are assumed to condense irreversibly on the particles. The amount of condensable H_2SO_4 is calculated online by the model using the corresponding production rate provided by the chemical scheme. However, the amount of condensable SOA is prescribed in terms of an effective emission of SOA from natural terpenes based on Dentener et al. (2006)."

(P6 L15): What does the neglection of POM mean for the diameters that are used to calculate the aging process?

As mentioned in the manuscript, we have modified the aging criterion following the results of laboratory measurements which identified a critical inorganic coating mass for the particles to become hygroscopic. To lean on these measurements, POM needs to be neglected. The consequences for the considered coating mass depend on the respective amount of POM which can show strong spatio-temporal variations. However, neglecting POM leads to an aging criterion consistent to the underlying measurements.

(P6 L33): Are the ice nucleation scavenging ratio assumptions necessary because feedback is excluded? It could also be calculated explicitly by ice nucleation parameterizations.

Correct. This version of the model does not include an ice nucleation parameterization; therefore scavenging ratios need to be assumed. However, even the inclusion of an ice nucleation parameterization would not automatically lead to full consistency since, in the current model version, ice nucleation is part of the cloud module and scavenging is calculated by a separate submodel (SCAV) This submodel currently receives only basic parameters from the cloud module, such as cloud and ice water content, cloud cover, and precipitation formation rates, which are usually calculated by all different cloud modules available in EMAC. Specific parameters like the aerosol activation rates are not yet transferred since this would require extensive code modifications in the scavenging scheme. This is planned to be subject to future model improvement activities.

(P7): Please provide some details on how the time integration of the aerosol dynamics equation is implemented.

We have included the following explanation in the revised version of the manuscript (Section 2.2): "The aerosol dynamics equation is solved by applying a combination of analytical approximations and process-specific numerical solvers. For the details of this approach, we refer to Kaiser et al. (2014) and references therein."

(P11 L1): Aerosol optical properties and assumptions with respect to the mixing state are also important.

Good point, we have added “mixing state” to this sentence.

(P12 Table 3 description, L4): Why are there arithmetic averages in the inequation? Shouldn’t the criterion apply to each single data point?

Yes, correct. We have fixed it, thanks for spotting.

(P14 L4): With a high bias and a wrong gradient the distribution is not reproduced. Please weaken the statement.

We have rephrased this sentence, also in view of a similar remark by Referee #2: “The model mostly reproduces the spatial pattern in this region, but it does not capture the west-east gradient seen in the observations, and is biased high.”

(P14 L17): Do the authors have an idea why the simulation provides such a high bias for Spain and the western Mediterranean? Are misrepresentations of the land use in the emission data leading to too high precursor gas emissions?

Aerosol ammonium is mostly controlled by NH_3 emissions, with agriculture being by far the most important source. The CMIP5 dataset used here shows large NH_3 agriculture emission over Spain in the year 2000, but it is hard to judge whether these values are realistic. A misrepresentation of the land use or an overestimate of the agricultural activity in this region could explain this bias.

(P16 L19): With the assumption of a homogeneous mixture between different particle sizes within one mode, you can calculate the modelled concentration below 2.6 microns. Did you try this?

This is a very good point. We have done a similar kind of analysis for the number concentrations as shown in Figure 6, considering only particles larger than detection size limits. We will adapt our post-processing workflow for including the suggested analysis for NO_3 in future studies.

(P17 L5): Please extent the findings of other studies.

We have extended the last paragraph of section 3.1 to include more detailed information: “The high-bias of near-surface mass concentrations of secondary species found here is not typically seen in studies using other global aerosol models. Although (relative) discrepancies are often of similar magnitude as those obtained here, the deviations are typically more variable in their directions for different species (e.g., Bauer et al., 2008; Mann et al., 2010; Pozzer et al., 2012; Lee et al., 2015). For instance, EMAC (MADE3) simulates larger average sulfate concentrations than observed by all considered station networks. The corresponding biases amounts to 13%, 38%, 34%, and 92% compared to EANET, EMEP, CASTNET, and IMPROVE, respectively. In contrast, Lee et al. (2015) found a similar high bias compared to IMPROVE (95%) but a low bias compared to observations from European sites (-13%). Other studies even show a general low bias. For example, the average sulfate concentrations obtained by Pozzer et al. (2012) show low biases of -45%, -16%, and -28% compared to EANET, EMEP, and CASTNET, respectively. Hence, EMAC (MADE3) shows a tendency towards enhanced sulfate concentrations. Nevertheless, the ability to simulate several tens of percent of monthly mean values within a factor of two of the observations indicates a quality of EMAC (MADE3) that is similar to that found in other model studies which performed this kind of analysis (Pozzer et al., 2012; Kirkevåg et al., 2013). It should also be mentioned that, in contrast to many other global aerosol models, EMAC (MADE3) performs quite well in case of black and organic carbon. However, we note that the primary goal of the present study was not to improve on previous aerosol climatologies, but rather, to show that our new model, with its additional capabilities in terms of particle mixing state representation and coarse mode particle interactions, also produces reasonable climatologies and hence is ready for investigating new topics that could not be addressed with the former versions of the model.”

(P22 L17): Although this is not a contradiction, it still means that the size distribution is not captured well.

We have added a statement to point this out: " Nevertheless, this could be an indication of a misrepresentation of the size distribution of such particles."

(P22 L24): A possible reason could also be that upper tropospheric temperatures are not captured well by the model. As a result, nucleation rates are not represented well. As I understand, the nudging takes place only at the surface. Did you check for biases in the model climate? The rather low model top at roughly 30 km may also play a role.
The nudging is applied using constant coefficients from the surface to the upper model boundary, therefore temperature should be correctly constrained through the whole model domain. As mentioned above, simulations with higher vertical resolution (and a higher model top) are planned for future studies, as well as further developments considering alternative parametrization for the nucleation process, as mentioned at the end of Section 2.2.

(P24 L13): Did you check if the assumptions made in the SCAV module (with regard to the release of particles) are causing the unimodal size distributions?
Since this phenomenon also occurs in our box model tests which do not take into account aerosol cloud interactions, we think that it is likely related to the chosen lognormal approach, as already discussed in the manuscript. Since it also occurs at near-source locations with a comparatively fresh pollution level, we do not think that cloud or precipitation processing plays a primary role here.

(Section3.3): A high resolution implementation of MADE3 could provide valuable insight into the performance of the aerosol scheme with respect to the size distributions shown. If the authors see any chance to realize this with the available emission data, the manuscript could be substantially improved.
As mentioned above, we agree that a higher resolution study would provide more insights. This is indeed planned as a follow-up study, including a more precise evaluation against aircraft data.

(P27 L22): Are there many particles? I would not expect too many particles in this region at this size range. An underestimation of small mineral dust particles could also explain this behaviour.
We rewrote the corresponding discussion as follows (last paragraph of section 3.4): "We have also analyzed the model biases in the individual years, but the interannual variability (not shown) was found to be small, hence meteorology alone cannot explain the discrepancies. Model misrepresentations, for instance, of the mineral dust particle size distribution, the local sulfate concentration, or the competition between nucleation and condensation of gaseous H_2SO_4 could also play a role. On the other hand, the SEM analysis in particular of the smallest size fraction might have a bias towards an underestimation of sulfate particles due to their instability under the electron beam. Since the number concentration of particles in this size fraction is comparatively high, a thorough analysis, including comparisons of the measured and simulated size distributions and also measurement uncertainties, should be the subject of a separate study."

(P29 Figure10): Please provide also relative differences plots.
Thanks for this suggestion. We have added two panels to Fig. 10 with the relative difference plots.

(P30 L30): Please add sentences highlighting differences of MADE3 to its predecessors.
We have added the following sentence: "With respect to its predecessors, MADE3 now explicitly simulates the partitioning between the gas and the aerosol phase in the coarse mode, as well as the interactions between the coarse and fine modes, and includes a fully revised coupling to the scavenging submodel accounting for the wet deposition processes."

(P3 L6): Please add turbulent to diffusion.
(P8 L22): leads
(P22 L2): There are some words missing in this sentence.
(P26 L33): be derived

(P27 L14): Please remove the word preliminary. If the results are preliminary you should not publish them.
(P32 L15): should
All done. Thanks for spotting these typos.