Authors' response to reviewer 1: Optimization of weather forecasting for cloud cover over the European domain using the meteorological component of the Ensemble for Stochastic Integration of Atmospheric Simulations version 1.0

Yen-Sen Lu[1,4], Garrett Good[2], and Hendrik Elbern[3]

[1]Institute of Energy and Climate Research – Troposphere (IEK-8), Forschungszentrum Jülich GmbH, 52425 Jülich, Germany
[2]Fraunhofer Institute for Energy Economics and Energy System Technology IEE, Königstor 59, 34119 Kassel, Germany
[3]Rhenish Institute for Environmental Research at the University of Cologne, Cologne, Germany
[4]Jülich Supercomputing Centre, Forschungszentrum Jülich, Jülich, Germany

Correspondence: Yen-Sen Lu (ye.lu@fz-juelich.de)

First of all, we authors thank the insight and detailed review by the reviewer 1. Those valuable suggestion will be taken into account for overhauling the manuscript. Please see the replies to the following sections.

1 Reply to the *Major Comments*

The overall manuscript lacks clarity, making it very difficult for the reader to understand the experiment design, analysis, and results. At times, the text is also missing critical justification for decisions made by the authors.

The language has been overhauled by a native speaker for readbility and we have addressed the comments below in the revision. We think the new text better communicates the methodology, in pacticular, the iterative approach to the sensitivity analysis.

It is not very clear why the authors chose to separate the experiments into the three sets described. To help clarify, it would be useful if each set of experiments listed all physics parameterizations used, with the "cluster" name referenced, so they can be compared more easily. Also, the ensemble members that use stochastic physics need to be clearly identified. The authors mention using SPPT as well as SKEBS, but only SKEBS appears to be described in 2.2.

We have clarified in the paper the iterative approach of starting with a limited study of a very large number of physics due to computational resources and narrowing this down to fewer physics studied with more detailed simulations.

We've updated the language and also the table of the listed physics configuration to include the clusters, and thus the reader should be able to understand the connection between the clusters and the optional physics.

The ensemble members that use the stochastic physics in this study only employ the SKEBS since it performed better in a previous study. We update this information with Berndt (2018) reporting that ESIAS-met employing SKEBS can produce more instability effectviely than employing the SPPT scheme. The SPPT scheme exists in the ESIAS-met but we did not used it to conduct the stochastic members in the simulation. We've clarified this in the manuscript.

It is not clear if the authors considered the appropriateness of each physics parameterization for the resolution used in the simulations. Certain physics parameterizations are targeted at specific resolutions, unless they are truly scale-aware. Some parameterizations used in WRF are specifically targeted toward convection-allowing scales (3 km resolution). Therefore, the authors need to thoroughly explain which schemes may not be appropriate for the 20-km resolution they are running, if any, and exclude those from their runs.

We are grateful for raising this important point. From the user's guide (WRF, 2015), Eta microphysics and Morrison are listed as only for the finer resolution ($< 5km$) and cloud-resolving simulations, respectively. We've thus removed these two microphysics from the further results and commented on their exclusion in the model setup section. All the plots in the manuscript and the supplementary material are updated accordingly.

Additional details about the ESIAS and general experiment design would be helpful. For example, why does ESIAS use 1000 members? Does it always run with that many? If it is specifically being run with 1000 members only in this case to sample as many WRF physics parametrizations as possible, please mention that. Which of the 1000 members employ stochastic physics? Is the ESIAS always run at 20 km resolution using the same 180 x 180 grid point domain? Can it be configured to run differently? Is it run operationally?

We have reworked the description in section 2.1. Acknowledging the fact that today's operational ensemble systems at weather centres operate with ensembles sizes < O(100) members, ESIAS is designed to offer the option of ensemble sizes at least one order of magnitude larger, that is O(1000). ESIAS-met does not necessary to run on 1,000 members, but is a flexible system capable of such runs. For instance, in our simulations it was initially run for 672 members in 48-hr simulaitons in Set 1, but also 4 members for a year of simulation in Set 4. The `WRF_TOOLS` can be employed to produce the needed input namelist without working one by one. For the set 1 and set 2 there is no member employing stochastic physics. ESIAS-met can be configured like the normal WRF to employ different nested-domain simulation or other requested time or spatial resolution. ESIAS is further designed to allow for alternative ensemble member definitions in the future, mainly envisioning further flexibility for various SPPT approaches. For our project, we have performed the simulation over the Europe and North Africa to study the sensitivity analysis on the Saharan dust events (unpublished). Our model can be also run operationally to produce

weather forecasting if needed.

Section 3.1 is unclear. GEFS data are apparently used, but is it for both ICs and LBCs? What is the frequency at which LBCs are applied? What is the forecast length (48 hours)? What is the frequency of the output (every three hours)?

We hope this is clearer in the reworked text with added details. Yes, the ICs and LBCs of GEFS data are used. The frequency of the LBCs are every 3 hours. The forecast length is 48 hours including the spin-up time. The further used ECMWF ERA5 data for Set 4 (long-term simulation) is also described in the Section 3.1

The work described in this manuscript may represent one of the largest physics sensitivity studies using the WRF to date (certainly this is true for the evaluation of cloud cover?). It may be worthwhile for the authors to highlight this fact in the abstract/conclusions. Grammar should be double-checked throughout the manuscript. Some explanations by the authors are very difficult to understand. There are also numerous typos found throughout the manuscript.

Thank you for highlighting this, we will add it to the abstract. We also work on checking the grammar and typos.

The authors state in the conclusions that they "... offer a recommendation on the choice of physics configurations for studying the European domain and for weather forecasting purposes." The manuscript only focused on the evaluation of cloud cover, which is just one of tens to hundreds of variables that are important for NWP. If the authors wish to provide physics recommendation for general NWP over Europe, many more variables need to be evaluated.

We completely agree that this study can only offer recommendations for the cloud cover and have clarified the focus on energy meteorology. Evaluating further variables and aspects of the NWP performance would be far more complex and was not within the objectives of our funding. We nevertheless hope that the results and technical work presented here can lead to future studies of more meteorological variables and expect more collaboration with experts in the field of observation and analysis. We have modified the wording to limit the scope of the results.

2 Reply to the *Minor Comments*

Lines 41-43: Are the authors saying that most physics combinations will exhibit a bias when compared to surface-based observations? If so, sure, and that's inevitable as model physics will never be completely bias-free. Also, why only surface-based observations? Upper-air observations can be used equally to verify model simulations, with model physics having the potential to impact upper-air variables just as much. Model simulations also include more than just physics, so it's not possible to say that all bias is due to just the model physics. In addition, having some kind of bias doesn't necessarily make a physics

parameterization or suite "unsuitable for deterministic forecasts". All operational models have some kind of physics bias, and work is always ongoing to minimize the error.

We totally agree with this statement and now touch on this in the introduction as: *There is thus potential for optimization, as most physics combinations can be expected to be biased as compared to observations.*

Line 49 - What is "the scientific challenge of proper scoring rules"? Please clarify.

We have reworded this with other comments.

Lines 50-51 – Is the "technical challenge" creating "large supercomputing facilities", finding the resources to run on large supercomputing facilities, the ability of an ensemble to forecast extreme and damaging events, or all of the above? Please clarify.

We clarify as: *Presently, a large supercomputer can produce sufficiently ultra-large ensembles of* $O(1,000)$ *members (at moderate resolution) if challenges in e.g. I/O performance and MPI communication are addressed. The ESIAS framework (Berndt, 2018; Franke et al., 2022) has been developed to accomplish ultra-large ensemble forecasts of up to* $(O)1000$ *members, demonstrated in this study with both multi-physics and stochastic schemes for the probabilistic simulation of the cloud condition.*

Line 52 and elsewhere – A simulation isn't probabilistic by itself, but probabilistic forecasts for a given event can be created from an ensemble forecast. I would replace "probabilistic simulations" with "ensemble-based probabilistic forecasts".

We update the text with: *ensemble-based probabilistic forecasts*

Line 54-55 – It isn't clear whether stochastic physics is used in all 1000 ensemble members described in the ESIAS, or whether there is a subset of additional members that employ stochastic physics. Please clarify.

We have updated the text as mentioned above.

Line 56 – "cope with" or "meet"?

The sentence is rephrased and thus this error is gone.

Line 59 – "ESIAS-met" has not been defined yet. It appears to be defined in line 63, but it's not clear what the difference is between ESIAS and ESIAS-met until the next section.

We have modified the manuscript to introduce ESIAS-met in the next session and thus keep using *the meteorological component of ESIAS* in the introduction.

Lines 59-62 – Please double check grammar. Also, are multi-physics simulations combined with stochastic simulations?

The clarified these details in section 2.2

Line 78 – "to better fit the system" – I'm not sure what the authors are trying to say here. Any specific reason why only SPPT and SKEBS are used? Did the authors also look at using SHUM or SPP?

We have rephrased the sentence as: *The namelist generation by the WRF TOOLS of ESIAS-met is the same as for WRF V3.7.1 and the input and output filenames are flexible.*

We only use SPPT and SKEBS because of it's original inclusion in the ESIAS system. The SPP can be used if ESIAS-met is updated to use the WRF4. The SHUM (stochastic boundary-layer humidity) or other stochastic schemes are not considered since the sensitivity analysis of stochastic schemes themselves is not our main focus. However, ESIAS-met is designed to easily adopt both methods, SHUM taken as a variant of SPP(T).

Line 91 – Can the authors briefly describe the "different approach" used in the other study?

We have extended description with: *which is the simulation on the European domain but with a different performed by changing the physics setup one-by-one through the configuration*

Line 93 – "to investigate the optimal physics configuration for the simulation output" – it might be a bit clearer to say that the optimal physics configuration is for the accurate representation of cloud cover.

We have rephrased the text as: *Large ensemble simulations with members of different physics were created in three sets to iteratively investigate the optimal configuration for cloud cover and the PV forecasting application.*

Line 110 – What does "perform" signify here?

We have rephrased the text as: *The sophisticated land surface models CLM (version 4) and Noah LSM are tested along with RUC LSM, which performs similarly to the other two LSM (Jin et al., 2010)..*

Table 4: Why wouldn't "over-predict" in Table 4 also be a "miss"? The difference between "over" and "over-predict" isn't very clear.

We have rephrase the text in the Section as: *Here, we use the convention that "under" and "over" represent the partial matches between fully missed or false clouds.*

Line 131 – Can the authors define what a "rater" is here? If CFC data aren't available over Northern Europe, why wasn't cloud cover verified over the western and southern portion of the simulation domain?

We have rephrased the text as: *The Kappa (κ) score is used to measure agreement between two or more raters, using determination in large data sets like for subjects in psychological research (Fleiss and Cohen, 1973).*

We did not discuss the cloud cover over the western and southern domain to avoid the uncertainty from the sea-atmospheric interaction. Motivated by the study on the renewable energy in Germany, our simulation use the design domain (Germany-centered) to investigate the simulation results against the observation.

Lines 169-171 – The explanation of how/if the CFC data are upscaled for verification is unclear.

We have added more details and rephrased the text as: *In order to compare the satellite data to the lower resolution model results, we simply average the CFC pixels within each grid cell*[1]*. The target value for any model grid point is then averaged over 12 to 36 observation points, depending on the location.*

Figures 6 and 7, 9 and 10 – Why where the specific dates chosen for these figures?

We have explained in Session 3.1 as: *These are more or less random days in different months without rare conditions, as we target the general forecasting performance for PV. Due to limited computational resources we are only able to demonstrate Sets 1 & 2 & 3 in ESIAS-met for day-ahead simulations beginning on these six days.*

Lines247-250 – This text is unclear. Please clarify.

[1]Other studies like (Bentley et al., 1977) may use all points within a fixed radius, which may or may not overlap

We have clarified it as: *However, the temporal- and spatial-averaged cloud covers provide less information and less variability over time. To determine the simulation skill on the spatial patterns, we score the simulation result by calculating the Kappa score using the pixels in the simulation domain.*

Lines 329-332 – It is unclear how many ensemble members exist in this study.

We have updated the description as: *We identify the most sensitive physics clusters and use the physics configuration (Set 3) to generate totally 1,152 ensemble members. Each physics configuration simulates with 31 additional members employing SKEBS scheme.*

Line 344 – A deterministic simulation is never going to be unbiased.

Agree, we have rephrased the text as: *The most accurate configuration for a deterministic forecast may differ from that for the ensemble with the most accurate mean or that best captures the uncertainty and diversity of possible outcomes*.

Lines 349-350 – Note that Jankov et al. (2019) don't necessarily advocate for multi-physics over stochastic-based ensembles. The authors describe the practical and theoretical deficiencies of multi-physics ensembles as well.

Firstly we appologize for mistaking the reference – We refer it to Jankov et al. (2017). Second we agree they did not advocate the multi-physics ensembles since they describe the disadvantage of multi-physics ensembles in the introduction session.

Line 352 – Spread produced by a multi-physics-based ensemble is mostly due to physics biases, not physics uncertainty.

Agree, and we have rephrased it as: *In our simulations, combining the ensembles into one multi-physics ensemble would enhance the spread, but this would be somewhat artificially due to the different biases of the model physics*.

Lines 354-355 – I wouldn't call this random. It's a specific result of the different physics parameterizations. It's also possible that probability matched mean could be calculated for the cloud field instead of just the standard mean to alleviate some of these problems.

Agree, we have modified it as *biases of the model physics*.

Line 356 – Jankov et al. (2019) used an eight-member ensemble, not four.

We have fixed it to *yielding eight members in total*. However, in Jankov et al. (2017) it's four ensemble members for one physics configuration.

Abstract – the following two sentences aren't clear: "We then selectively conduct stochastic simulations to assess the best choice for ensemble forecasts. The results indicate a high variability in terms of physics and parameterization."

We have rephrased the abstract and changed it to *on six test cases*.

Line 23 – "negative wind energy prices" – This topic needs to be briefly explained

We explained it by rephrasing it as: *e.g. in negative prices during high wind events*.

Line 23 - to study to the -> "to study the"

The sentence is rephrased and thus this error is not existed

Line 26-29 - The introduction to deterministic models should be followed by a reference to the WRF model as being deterministic. Something like "Various global and regional deterministic weather models"

We have changed to use *Various global and regional deterministic weather models*.

Line 31 – "optimal meteorological models," – I would say "optimal model configuration" instead

We have changed the text to *optimal model configuration*.

Lines 39-41 – Double check for typos and correct comma placement.

The typos and comma are double-checked.

Table 4 – Typo in description -> "Indaddition"

We have fixed the typo.

Line 206 – Typo - "most" -> "the most"

We have fixed the typo.

References

User's Guide for the Advanced Research WRF (ARW) Modeling System Version 3.7, https://www2.mmm.ucar.edu/wrf/users/docs/user_guide_V3/user_guide_V3.7/users_guide_chap5.htm, 2015.

Bentley, J. L., Stanat, D. F., and Williams, E.: The complexity of finding fixed-radius near neighbors, Information Processing Letters, 6, 209–212, https://doi.org/https://doi.org/10.1016/0020-0190(77)90070-9, https://www.sciencedirect.com/science/article/pii/0020019077900709, 1977.

Berndt, J.: On the predictability of exceptional error events in wind power forecasting–an ultra large ensemble approach–, 2018.

Fleiss, J. L. and Cohen, J.: The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability, Educational and psychological measurement, 33, 613–619, 1973.

Franke, P., Lange, A. C., and Elbern, H.: Particle-filter-based volcanic ash emission inversion applied to a hypothetical sub-Plinian Eyjafjallajökull eruption using the Ensemble for Stochastic Integration of Atmospheric Simulations (ESIAS-chem) version 1.0, Geoscientific Model Development, 15, 1037–1060, https://doi.org/10.5194/gmd-15-1037-2022, https://gmd.copernicus.org/articles/15/1037/2022/, 2022.

Jankov, I., Berner, J., Beck, J., Jiang, H., Olson, J. B., Grell, G., Smirnova, T. G., Benjamin, S. G., and Brown, J. M.: A Performance Comparison between Multiphysics and Stochastic Approaches within a North American RAP Ensemble, Monthly Weather Review, 145, 1161–1179, https://doi.org/10.1175/MWR-D-16-0160.1, https://journals.ametsoc.org/view/journals/mwre/145/4/mwr-d-16-0160.1.xml, 2017.

Jin, J., Miller, N. L., and Schlegel, N.: Sensitivity study of four land surface schemes in the WRF model, Advances in Meteorology, 2010, 2010.