Geosci. Model Dev. Discuss., 2, C421–C430, 2009
www.geosci-model-dev-discuss.net/2/C421/2009/


Geoscientific Model Development Discussions

# *Interactive comment on* "Automatic generation of large ensembles for air quality forecasting using the Polyphemus system" *by* D. Garaud and V. Mallet

**D. Garaud and V. Mallet**

damien.garaud@inria.fr

Received and published: 6 November 2009

## 1 General Comments

We thank the reviewer for her/his comments. He/she was right to point out the lack of perspective. He/she asked several very relevant questions (some of which can unfortunately not be answered at the moment).



### 1.1 More Perspective

> " *The paper needs to be much clearer on the purpose and goals of the development, what it achieves and why it is valuable. The first paragraph of the introduction lists previous studies and approaches, but doesn't explore the scientific merits of the ensemble approach, what can be learned from it, or what its strengths and weaknesses are. The paper needs to justify the benefits of the approach adequately, and to be clear about which aspects of the developments described here are new and original and which are merely a simple extension of an already well-developed modeling system. Addressing this major issue in a suitable manner will strengthen the paper greatly and make it much more valuable to the readership of GMD."*

We agree that giving more perspective can help the reader better understand our approach. We therefore added in the introduction:

"In an ideal setting, one should take into account all uncertainty sources based on the best description available. Essentially, this would mean relying on Monte Carlo perturbations for uncertain input data like emissions, on the alternative descriptions available for data like land use cover, on calibrated ensemble weather forecasts, on different formulations for the subgrid parameterizations in the chemistry-transport models, on different numerical schemes in the chemistry-transport models. In this paper, we tend to this ideal setting with a simplified approach: we do not use a meteorological ensemble (the meteorological inputs are treated like other input data), and we rely on an alternative sampling approach to full Monte Carlo simulations. Nevertheless all uncertainty sources can be considered, and they are all taken into account at the numerical-simulation stage: no statistical correction is applied in a postprocessing. The approach described in this paper may be seen as a three-fold extension that of Mallet and Sportisse (2006): new uncertainty sources are included, the uncertainty in input data is specifically taken into account, and the ensemble generation is entirely automatic."

### 1.2 Alternative Model Formulations

> " *Perturbing input data and other variables based on some assessment of the uncertainty involved is very useful in an ensemble framework, but it is not clear that the same method can be adopted for comparing different model parameterizations or formulations. The diversity of results generated by comparing different schemes provides different information on uncertainty from that provided by input data which is continuously distributed; it is possible, for example, that introduction of an 'improved' formulation for some process will generate results outside those generated by the standard schemes currently in place. Although the approach taken here is still valid, the authors need to make this distinction clear and to demonstrate how the results in each case can be interpreted."*

We are not sure what the reviewer is meaning here.

We can interpret the results only when a single change is made in the model formulation. For instance, the vertical diffusion coefficients computed by the Troen & Mahrt parameterization have higher values than that computed by the Louis parameterization, and it will therefore increase the ozone concentrations at ground level. However, it is essentially impossible to provide an interpretation for a hundred simulations in which the model formulation is decided out of 18 options.

The results of the models may surely be *outside those generated by [a reference model with] the standard schemes currently in place*: this is the purpose of the ensemble approach to explore the possible outcomes. The same is true for simulations with perturbation on continuous input data.



We are not *comparing different model parameterizations or formulations*: the different model parameterizations allow us to estimate the uncertainty due to the model formulation, not to compare them. We are glad the reviewer points out the difference between input data that can be continuously perturbed, and the changes in the model formulation. There are indeed two different sources of uncertainty, and our approach is to take into account both sources of uncertainty. We hope the extended introduction (see above) will make the approach clearer.

### 1.3 Applicability

> " *The paper would be more valuable if it demonstrated an appreciation of the wider applicability of the approach to other air quality models. It is likely that many readers wanting to adopt a similar approach will not be using the Polyphemus system."*

The approach is surely dependent on the modeling framework. The key point is how flexible and featured (in terms of modeling options) the system is. To make this clearer, we added at the end of Section 3:

"The same approach may be applied to another modeling system providing enough options (in the model formulation) are available. This requires that significant diversity is maintained in the system. In particular, when a new formulation (e.g., a more accurate chemistry) is developed, the previous formulation should remain available to the user. The rationale is that, while a formulation may seem better from a deterministic point of view (based on a priori considerations or on performance analysis), the previous formulation still has a significant probability (though lower than that of the new formulation) from a stochastic point of view."

## 2 Specific Comments

### 2.1 Abstract

> " *The abstract describes what was done in the study, but not its purpose, originality or main conclusions. Replacing the procedural aspects with a brief summary of the context, importance and results would strengthen the paper greatly."*

We note this comment and will enhance the abstract in that way.

We agree that we focused too much on the technical aspects in the abstract. There was a lack of perspective. We tried to correct this with the following sentences:

"The objective is to take into account all sources of uncertainty: input data, physical formulation and numerical formulation. The leading idea is to build different chemistry-transport models in the same framework, so that the ensemble generation can be fully controlled. Large ensembles can be generated with a Monte Carlo simulations that address at the same time the uncertainties in the input data and in the model formulation."

We kept the rest of the abstract since the paper primarily addresses the mentioned procedural aspects.

### 2.2 Conclusion

> " *Similarly, the conclusions do not adequately bring out the new and original aspects of this research, and hence its possible value to the geophysical modeling community. In particular, the conclusions need to emphasize what we can learn from the results of this type of ensemble study."*



We completely agree on that. There is again a lack of perspective. Hence we added:

"Contrary to most traditional approaches, which are based on perturbations of input data only, or on small ensembles of models from different teams, our approach takes into account all sources of uncertainties at once: input data, physical formulation and numerical formulation. Each member of the ensemble is a complete chemistry-transport model whose contents are clearly defined within the modeling platform. In this context, the ensemble and the differences between its members can be rigorously analyzed, and also controlled through the probabilities associated with every option. Our approach tries to combine the flexibility of Monte Carlo simulations (large ensembles of simulations with perturbed input data) and the completeness of a multimodel ensemble (models with alternative physical parameterizations, like in ensembles made of a few models from different teams)."

### 2.3 Section 3.1

- p.899, l.14: "*Indication of how the uncertainties were quantified by experts*". We provided further explanations:

  "Estimations of the uncertainties were established by experts and reported in Hanna et al. (1998, 2001), for 18 km and 12 km resolutions, in regions of eastern U.S.A., and for a few days. These estimations should be seen as guidelines to be adapted to the simulation region, to the resolution of the simulation, to the time span, and to other considerations on the quality of the fields. For instance, the uncertainty in the values of a field should decrease when the resolution gets higher. In addition, a few ensembles were generated in order to roughly calibrate the uncertainty parameters, based on comparisons with observations (not reported here)."

  We cannot get further since a complete uncertainty analysis would be a full paper. Here we focus on the generation of the ensemble. The uncertainty parameters need to be adjusted with different methods that are a difficult subject for further research.

- p.900, l.14-15: *Why the selection is based on 3 values rather than a random sampling?* It is due to both technical and historical reasons. We are in the process of integrating random sampling. The selection among 3 values may still be used in case of small ensembles, but without random selection; e.g., with a 3 member ensemble, and depending on the input data, it may be more efficient to select once each of the 3 values than sampling randomly.

- p.901, l.13: *Discussion of probabilities for model formulations.* The explanations are given in the following paragraph, which we extended:

  "Except for the perturbations in the input data, the probabilities are chosen according to the confidence put in each option. There is no direct indicator to determine these probabilities. If two parameterizations are available for a given option, the choice lies between giving a probability one to a parameterization (no uncertainty), and giving 0.5 to both parameterizations (which leads to the largest uncertainty). If one option is supposed to be more accurate (a priori quality of a parameterization, finer grid resolution, . . . ) or if it is usually associated with better model results (comparison with observations), its weight should be higher than that of alternative choices. For example, a time step equal to 600 s is supposed to give more accurate results than 1200 s—the numerical solution converges to the exact solution as the time step tends to 0. Therefore, a higher probability is associated with the time step fixed to 600 s. Another example is the chemical mechanism RACM which is more detailed than RADM 2, and which has shown slightly better results in several studies (Gross and Stockwell, 2003)."

## 2.4 Section 3.2

- p.903, l.5: *Routine names not needed.* The routine names has been removed.



## 2.5 Section 4

- p.903, l.26: *What sample size would be statistically appropriate for a well-characterized ensemble given the number of variables perturbed in the present study?* This is of course an important question to which we have no answer yet. A convergence study will be undertaken. The convergence will depend on the variance of the target concentrations and on the number of members (but not on the number of variables). In the paper, we give insights in the structure of the ensemble, but we do not give accurate estimations of the uncertainties (of the target concentrations). We are currently working on the calibration of the ensemble in order to provide such estimates. This work should eventually be published on its own.

## 2.6 Section 4.2

- p.906, l.18: *exactly one model, or at least one model? Is the result described here significant, other than providing a way of characterizing the performance of the reference models?* There always is at least one model in the ensemble which is better than the six reference models. We changed the text accordingly. We believe that this result is interesting: it was not obvious to us that the random procedure would generate models with good performance, not to mention better models (over?)tuned by the developers, on the basis of comparisons with observations.

- p.907, l.1: *while it is fine to identify a "best model", there needs to be some statement about the significance of this. Does it mean anything, and if so, what can we learn from it? Or does it indicate that the observational comparison is not sufficiently broad and the metric chosen for comparison (RMSE) is not appropriate? The paper needs to be clear about this given that different models show* *similar performance (as stated on line 23).* We believe a model evaluation cannot be solely based on the RMSE, but this is still a good indicator that do not allow error cancelations. It is usually a value of interest for the modelers to rank their models. We added:

"It is interesting to note that (1) the random sampling generates several models with good performance (compared to the observations, with the RMSE), (2) the random sampling generates a model with lower square errors (over a long time period) than the models tuned by the modelers."

It is perfectly reasonable to have models with similar performance. In ensembles built with models from different teams, this is observed too. It is noteworthy that these models with similar performance may be built with settings that significantly differ.

### 2.7 Tables and Figures

- Table 1: we added:

  "The numbers enclosed in brackets correspond to the occurrence probability of an option."

- About Table 6, we added mean, variance and extrema to the caption.

- About Table 9 and Table 10: we removed Table 9 and put the SNAP description in the caption of Table 10.

- Figure 9: "NO2" has been replaced with "SO2" in the caption.

- Figure 10: We believe these figures are very relevant. We agree it could be described in the text, but we doubt the reader can imagine how fragmented the maps are, how much they change from one day to another. We have noticed in talks, for example, that these figures were well appreciated.



### 2.8 Typos, etc.

We made all the corrections suggested by the reviewer.

We thank the reviewer for her or his useful corrections.

## References

Gross, A. and Stockwell, W. R.: Comparison of the EMEP, RADM2 and RACM mechanisms, Journal of Atmospheric Chemistry, 44, 151–170, 2003.

Hanna, S. R., Chang, J. C., and Fernau, M. E.: Monte Carlo estimates of uncertainties in predictions by a photochemical grid model (UAM-IV) due to uncertainties in input variables, Atmospheric Environment, 32, 3,619–3,628, 1998.

Hanna, S. R., Lu, Z., Frey, H. C., Wheeler, N., Vukovich, J., Arunachalam, S., Fernau, M., and Hansen, D. A.: Uncertainties in predicted ozone concentrations due to input uncertainties for the UAM-V photochemical grid model applied to the July 1995 OTAG domain, Atmospheric Environment, 35, 891–903, 2001.

Mallet, V. and Sportisse, B.: Uncertainty in a chemistry-transport model due to physical parameterizations and numerical approximations: An ensemble approach applied to ozone modeling, Journal of Geophysical Research, 111, doi:10.1029/2005JD006149, 2006.

Interactive comment on Geosci. Model Dev. Discuss., 2, 889, 2009.