Geosci. Model Dev. Discuss., 2, C257–C260, 2009
www.geosci-model-dev-discuss.net/2/C257/2009/

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

Anonymous Referee #1

Received and published: 14 September 2009

This paper describes the use of an automatic approach in generating ensembles of runs within an air quality forecasting model. It describes the different options available within the model for a number of variables and uses a probability approach to weight the choice of variables during ensemble generation. It then describes an example and makes a preliminary analysis of the results obtained. While the scientific value of the approach adopted here is evident in part, I believe that there are major weaknesses in the description that need to be addressed before the paper is ready for publication.

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 pre-

vious 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.

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.

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.

Specific Comments

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.

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

GMDD

2, C257–C260, 2009

this type of ensemble study.

p.899, l.14: "established by experts": it would be helpful to give some indication of how the uncertainties were quantified, and whether they apply to the model systems/locations under consideration in this study and/or how they might be expected to differ.

p.900, l.14-15: indicate why the selection is based on 3 values rather than a random sampling (for computational ease in the given example?)

p.901, l.13: The discussion of probabilities is important here for anyone attempting to reproduce this study, and the discussion should be expanded so that it can be applied in other situations. How are the probabilities for the different model formulations arrived at?

p.903, l.5: details of routine names not needed here.

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?

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?

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).

Table 1: Add a comment on the probabilities in the table caption.

Table 6 should be removed, as it has too much detail, none of which is referred to in

GMDD

2, C257–C260, 2009

the text. It would be more valuable to highlight the statistics of each metric (mean, variance, extrema) instead.

Tables 9 and 10 do not aid in interpretation of the results of the paper and therefore are not needed.

Figure 9: Label graphs with relevant species. In the caption, one "NO2" should be replaced by "SO2".

Figure 10: This figure is not useful, as it simply demonstrates spatial and temporal variability in performance which can be described in a single sentence in the text.

Typos, etc.

p.886, l.17: COV -> VOC

p.898, l.2: units -> s

p.899, l.2: 'be' after 'could'

p.904, l.9: 'Go' -> 'Gb'

Use of ellipsis (...) throughout the text should be avoided.

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