Atmos. Chem. Phys. Discuss., 7, S6983–S6993, 2007
www.atmos-chem-phys-discuss.net/7/S6983/2007/


Atmospheric Chemistry and Physics Discussions



# *Interactive comment on* "Seeking for the rational basis of the median model: the optimal combination of multi-model ensemble results" *by* A. Riccio et al.

**A. Riccio et al.**

Received and published: 18 November 2007

First, we would to thank you for your comments. In the following you will find the answers to the problems raised by your comments. We first paste your comment (*in italics*) and then our reply.

## 1 Specific comments

*1. The presentation of the estimation procedure (section 4) is obscure. It requires random draws from probability distributions (p. 5717, ll. 15-17) while the basic compu-*

*tations (5-7) of a posteriori probability distributions do not include any random component. Clearly, a form of Monte-Carlo method is used there, but no explanation is given of why such a method is considered as appropriate for the problem at hand, nor of what is exactly achieves. These points must be clarified.*

*3. Uncertainties on the various estimated parameters are given, in Table 1. It is not clear how these uncertainties have been obtained. Are they an output of the Monte-Carlo simulations?*

Probably you refers to section 5 (the estimation procedure). Section 4 introduces the problem of uncorrelatedness as a special case of independence.

We apologize if section 5 sounds 'obscure'; we omit some material, which is quite obvious for a reader expert in Bayesian data analysis, but we agree with your comment: it may be difficult to read for those who are not acquainted with Markov chains in a Bayesian context. We will provide a 'plain' version of section 5 in the revised paper.

Concerning the specific question you addressed, Markov chain Monte Carlo (McMC) is a standard tool to draw samples from the specified posterior pdf. We refer the interested reader to Gelman et al., *Bayesian Data Analysis*, Chapman and Hall/CRC, for an excellent reading on McMC, and to the article written by Geman and Geman, '*Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images*', published on Transactions on Pattern Analysis and Machine Intelligence, **6**(6), pages 721-741, for a deeper insight in Gibbs samplers. We will insert these references also in the revised paper.

Essentially, the basic procedure of Monte Carlo simulation is to draw a large set of samples $\left\{\theta_k^{(l)}\right\}_{l=1}^{L}$, from the target distribution (the posterior pdf in this work). One can



ACPD
7, S6983–S6993, 2007





EGU

then approximate the expectation of any function $f(\theta)$ by the sample mean as follows:

$$\mathrm{E}(f) = \int p(\theta|\cdot) f(\theta) d\theta \approx \frac{1}{L} \sum_{l=1}^{L} f(\theta^{(l)}) \,, \tag{1}$$

$L$ is the number of samples from the target distribution. By the same way, one can obtain an estimate of the variance of this function

$$\mathrm{Var}(f) = \int p(\theta|\cdot)(f(\theta) - \mathrm{E}(f))^2 d\theta \approx \frac{1}{L} \sum_{l=1}^{L} \left( f(\theta^{(l)}) - \mathrm{E}(f) \right)^2 \,, \tag{2}$$

The expected values, along with their standard deviations, are reported in Table 1, and are a straightforward outcome of the McMC procedure.

The Gibbs sampler alternates two major phases: obtaining draws for parameters from the posterior pdf of each model, and obtaining draws for the weights given the model parameters.

In the first phase, we drew a sequence of $L$ samples, $\left\{ (b_k^{(l)}, \sigma_k^{(l)}) \right\}_{l=1}^{L}$, for each model $k$, exploiting the Gibbs sampler. The Gibbs sampler was implemented as follows:

for $k = 1 : K$
  Initialize $b_k^{(1)}$ and $\sigma_k^{(1)}$
  for $l = 2 : L$
    draw $b_k^{(l)}$ from $p(b_k|\sigma_k^{(l-1)}, y_{\cdot k}, z_\cdot)$
    draw $\sigma_k^{(l)}$ from $p(\sigma_k|b_k^{(l)}, y_{\cdot k}, z_\cdot)$
  end
end

$p(b_k|\cdot)$, or $p(\sigma_k|\cdot)$, is the posterior pdf of model $k$, conditioned on taking the remaining



ACPD
7, S6983–S6993, 2007

parameter equal to the value of the previous iteration. $y_{ik}$ stands for the log-transformed prediction of model $k$ for the $i$th spatio-temporal location, and $z_i$ for the $i$th observation. Standard algorithms are available to draw from these distributions (the interested reader can have a look at the book by Gelman et al.) and it is guaranteed that the sequence of values $\left\{(b_k^{(l)}, \sigma_k^{(l)})\right\}_{l=1}^{L}$ generated by the Markov chain are $p(b_k, \sigma_k | \cdot)$ identically distributed.

In the second phase, we sampled the posterior distribution to get a sequence of model weights. If we set $\theta_k^{(l)} \equiv (b_k^{(l)}, \sigma_k^{(l)})$, and define $\zeta_i \equiv (\zeta_{i1}, \ldots, \zeta_{iK})$ as a sample from a multinomial process, then the selection of the 'best' model in explaining the $i$th observation can be viewed as the outcome of a multinomial process, i.e.

$$p(\zeta_i) = \text{Multin}(\zeta_i | \, p_{i1}, \ldots, p_{iK}) = \begin{pmatrix} 1 \\ \zeta_{i1}\zeta_{i2}\cdots\zeta_{iK} \end{pmatrix} p_{i1}^{\zeta_{i1}} \cdots p_{iK}^{\zeta_{iK}} \, . \tag{3}$$

The factors $p_{ik}$s in equation (3), are the posterior pdf values of each model, re-normalized so that their sum over index $k$ is equal to 1, i.e.

$$p_{ik} = p(\theta_k^{(l)} | \, y_{ik}, z_i) \sum_{k=1}^{K} p(\theta_k^{(l)} | \, y_{ik}, z_i) \, , \tag{4}$$

which coincides with the Bayes' factor for the $k$th model in explaining the $i$th observation in equation (5) of the paper.

From the properties of a multinomial random process, each model has a probability to be selected as the 'best' model equal to $p_{ik}$, given by equation (4). A draw for $\zeta_i$ from (3) is a vector with $K-1$ components equal to zero, and one component (that corresponding to the 'best' model) equal to one, i.e. $\sum_{k=1}^{K} \zeta_{ik} = 1$ for any $i$.

The selection process was repeated for each observation and iterated for each $\theta_k^{(l)}$ sample, as implemented in the following algorithm:



ACPD
7, S6983–S6993, 2007





EGU

for $l = 1 : L$
  Set $\theta_k^{(l)} = (b_k^{(l)}, \sigma_k^{(l)})$
  for $i = 1 : N$
    set $p_{ik}$ for any $k$ as in equation (4)
    draw $\zeta_i^{(l)}$ from $p(\zeta_i | \, p_{i1}, \ldots, p_{iK})$
  end
end

Table 1 shows the expected values (with their standard deviations) of the fraction of times each model is selected as the 'best' model, averaged over all McMC iterations and observations, i.e. $\frac{1}{LN} \sum_{l=1}^{L} \sum_{i=1}^{N} \zeta_{ik}^{(l)}$ for any model $k$. Again, the standard deviations are calculated from the Monte Carlo iterations, as shown by equation (2).

We will describe the estimation process more clearly in the revised paper.

*2. Concerning the ETEX-1 simulations, important information is missing. How many verifying observations were used? Which error was assumed on those observations? (if the evaluation method does not require quantitative specification of observational errors, say it clearly).*

Concerning errors in the observations, in both cases if you refer to experimental errors or interpolation errors an estimate would be impossible in the first case (since it is a single realization), and in the second not very representative as any interpolation method would lead to an error.

*4. The impact of some of the a priori hypotheses made on the errors is not discussed.*



ACPD
7, S6983–S6993, 2007





EGU

*The assumed independence of models is discussed (I understand that what matters here are not the models* per se*, but the errors affecting those models). But, as said, the lognormality of errors and the space-time independence of observational errors are not discussed. A full discussion may of course be very difficult, but still the question must be explicitly stated and commented that the results obtained may significantly depend on those a priori hypotheses.*

You raised two important questions.

**1.0.0.1 1. log-normality** Here, we are facing the problem of assuming a model for "errors". Unfortunately, it is not a simple matter, because there is no physical guidance on how to model these errors. The assumption of log-normality is based on a statistical consideration, i.e. models' deviations look log-normally distributed (look at figure 2 of the paper).

For a deeper insight, you can have a look at the table at the end of this document, where we compare the weights obtained using the log-normality assumption, and those obtained using gamma functions. Results are indeed sensitive to the assumption of the chosen function. However, as you can see, some important aspects emerge: 1) there is no extremely outperforming model, nor an extremely bad model. 2) Apart from a few models, the majority of models perform equally well; this is the main message of our work, i.e. ***the choice of the 'median model' is statistically supported by data evidence***; this statistical evidence is not sensitive to the choice of the assumed model for errors. 3) Weights are calculated using ETEX-1 as a test-bed; a question emerges on to what extent this basecase is representative of atmospheric dispersion conditions; however, there is no reason for considering ETEX-1 as a 'special' one. Weights will depend on the basecase (with variations even larger than those shown in the table at the end of this document), but we can guess that models will continue to be 'well-balanced' for other basecases, too. Results from the ENSEMBLE project



ACPD
7, S6983–S6993, 2007





EGU

(Galmarini et al., 2004a) suggest that this is indeed the case.

**1.0.0.2 2. space-time independence** The space-time independence is a simplification. This is a customary choice in a statistical setting, since it allows to use different observations as replicates of the same process. However, this simplification is not completely supported by data. As already stated in the paper (see table 2 in the paper), there are large correlations between models; in the best case, we can speak of 'block-independent' models, i.e. we can categorize models into several blocks, where each model is correlated with the models belonging to the same block, and uncorrelated with all others. A more complex approach is possible, and we are currently working on this aspect, and (hopefully) it will be the subject of a future work

We will explicitly state in the revised paper that log-normality and space-time independence are simplifications.

*5. The authors claim that their paper provides an a posteriori justification for the 'median model' approach introduced and discussed in previous papers. It would be useful to explain, even if succinctly, what the median model approach exactly is. The reference (p. 5721, l. 15), to the APL50 index used in Galmarini et al. (2004a) is unsufficient.*

The APL50 is already defined (p. 5721, ll. 10-14). In a nutshell the median model is obtained calculating the median value of any of the forecasted variables (instantaneous and time integrated concentrations, depositions etc) in every point in space and time where an ensemble of model have predicted the presence of a cloud. The median model will therefore be the composite of the median values of the results of all models.



ACPD
7, S6983–S6993, 2007





EGU

## 2 Technical corrections

*6. The paper is inconsistent as to what the reader is expected to know. While some developments describe material that belongs to elementary probability theory (Section 4, Independence and correlation, for instance) and can be expected to be known by the reader of a scientific paper, other parts are technically much more difficult. For instance, the first equation after eqs (3) (that is before Section 4) very succinctly states, without real explanation that the information entropy (the expression is not used) decreases in the estimation (3. Reference is also made, without explanation, to the Kullback-Leibler information divergence theorem. These parts may be difficult to understand, even by readers who are reasonably familiar with probability theory.*
*Another example is 'Gibbs sampling', which is mentioned (p. 5717, l. 12) without explanation nor reference. I am not sure many readers will know what Gibbs sampling is.*

*7. There are inconsistencies of notations. For instance, the notation zetaik is used in section 5 (top of p. 5718) for denoting a quantity that is said to take the values 0 and 1 only (incidentally, it is not said that this quantity, which is called an 'unobserved latent variable', and is linked to 'observations drawn from models', is itself obtained by a random draw). Then, the same notation $\zeta_{ik}$ is used a few lines further down (eq. 21) to denote a quantity which can take all integer values between 0 and some undefined upper limit $N$. From what I understand, that new $\zeta_{ik}$ must be the sum of the previously defined $\zeta_{ik}$ over $N$ independent draws. I must say all that is totally confusing for the reader.*

We will take your comments into consideration in the revised paper. However, as already said, we expect that the reader is acquainted with Bayesian data analysis and Markov chain Monte Carlo procedures. See also our answer to your comments n. 1 and 3.



ACPD
7, S6983–S6993, 2007





EGU

*8. P. 5706, l. 17. The reference to the 'least squares method' ('least variance' would here be more appropriate here anyway), following considerations on bayesian estimation, is ill-placed, It would be preferable to speak of gaussianity first, and to say that bayesian estimation amounts to least variance estimation in the linear and gaussian case.*

We will take your suggestion into consideration in the revised paper.

*9. It could be useful to state explicitly that eq. (18) is the same as eq. (6). And there should no index k on theta in the right-hand-side of eq. (18).*

Not exactly! $\theta_k$ represents the expected value of $\theta$ conditioned on model $k$ alone, and the notation is consistent with that used throughout the paper.

*10. Contrary to what is implied in the paper (last line of p. 5710), gaussianity is not necessary for eqs (8) to be valid. Actually, eqs (8) are only a very slightly modified reformulation of eqs (7). Either one of those two sets of equations is sufficient.*

We agree with you, and we will drop equations (8) in the revised paper.

*11. Figures 3 and 4 should preferably be merged into one figure.*

Figure 4 is the continuation of figure 3; we will merge the two figures in the revised paper.



ACPD
7, S6983–S6993, 2007





EGU

*12. P. 5703, l. 3. To be frank, the sentence 'ensemble estimates average out nonpredictable components' seems more like wishful thinking than like the statement of an objectively established fact.*

Apparently, there is a certain degree of subjectivity in this sentence, but the removal of non-predictable components is one of the goals of the 'breeding method'. Quoting from Toth and Kalnay, *"Ensemble Forecasting at NCEP and the Breeding Method"*, published on Monthly Weather Review, 1997, page 3317, "*... We also pointed out that ensemble averaging removes the unpredictable components of the flow while leaving the predictable part virtually intact.*"

---

Interactive comment on Atmos. Chem. Phys. Discuss., 7, 5701, 2007.

Table 1: Model weights obtained assuming gamma distributed errors (second column) or log-normally distributed errors (third column).

| # | gamma | log-normal |
|---|---|---|
| m01 | 0.0347 | 0.0377 |
| m02 | 0.1453 | 0.0652 |
| m03 | 0.0172 | 0.0366 |
| m04 | 0.0109 | 0.0100 |
| m05 | 0.0198 | 0.0397 |
| m06 | 0.0050 | 0.0417 |
| m07 | 0.0779 | 0.0379 |
| m08 | 0.0272 | 0.0152 |
| m09 | 0.0278 | 0.0356 |
| m10 | 0.0853 | 0.0409 |
| m11 | 0.0240 | 0.0357 |
| m12 | 0.0321 | 0.0508 |
| m13 | 0.0254 | 0.0427 |
| m14 | 0.0174 | 0.0356 |
| m15 | 0.0063 | 0.0393 |
| m16 | 0.0086 | 0.0424 |
| m17 | 0.0729 | 0.0288 |
| m18 | 0.0224 | 0.0409 |
| m19 | 0.1317 | 0.0541 |
| m20 | 0.0646 | 0.0716 |
| m21 | 0.0076 | 0.0404 |
| m22 | 0.0278 | 0.0466 |
| m23 | 0.0519 | 0.0353 |
| m24 | 0.0178 | 0.0401 |
| m26 | 0.0383 | 0.0351 |



ACPD
7, S6983–S6993, 2007