Round 2 Referee Report for *Data Assimilation using an Ensemble of Models: A hierarchical approach*

General comments:

I read the new version of the paper from the top, and have a new set of specific comments and questions that are given below. I am still stuck on Section 2. However, we have made substantial progress because the increased clarity of the new version makes it now possible to identify the ambiguities that are hanging me up.

First, though, I have an emphatic comment: I *do* think the paper should be self-contained. I am referring here to the author's response to my item 11 in the first round of review, in which he states, "I have relied heavily on the notation and explanations in Rayner et. al. (2016) thus making the current paper less self-contained. Should I move away from that and define the notation locally?" My opinion is a definite "yes". Without those definitions and explanations, it's quite difficult to follow what's going on. Unless you want this paper to appeal only to those who are quite familiar with previous work, I feel it's <u>essential</u> that this paper explain all the notational conventions that are used, and provide adequate background.

Specific comments:

1. Page 2, lines 24 to 28: What does "This required running optimized fluxes through the forward model used to generate the Jacobians" have to do with challenging equal weighting? What is TM3?

2. Page 3, Equation (1): Here is a case where you have referenced Rayner et. al. (2016). I find the material here impossible to understand without going back to that 2016 article, and even so, it's not clear where this formula is coming from. In Rayner et. al. (2016), a similar version of current paper's Equation (1) appears as Equation (2), but it doesn't look right to me: is it missing and integral? For the left-hand side to be $p(x)$, you would have to integrate out all the other variables. You appear to be headed that way in Equation (1) of the current paper by integrating out $\mathbf{y}^t$, as if $H(\cdot)$ is a deterministic function of $\mathbf{x}$. However, this is never stated, and is contrary to both the notation and treatment of H later.

3. It's difficult to tell what the derivations in the first part of Section 2 are trying to show. It looks to me that you want to end up with Equation (3), which is an expression for

$P(\mathbf{x}, \mathbf{H}_i|\mathbf{y})$, but this is only an intermediate step towards getting $p(\mathbf{x}|\mathbf{y})$:

$$p(\mathbf{x}|\mathbf{y}) = \frac{\sum_i p(\mathbf{x}, \mathbf{H}_i, \mathbf{y})}{p(\mathbf{y})} = \frac{\sum_i p(\mathbf{x}|\mathbf{H}_i, \mathbf{y})\, p(\mathbf{H}_i, \mathbf{y})}{p(\mathbf{y})},$$

$$= \frac{\sum_i p(\mathbf{x}|\mathbf{H}_i, \mathbf{y})\, p(\mathbf{H}_i|\mathbf{y})\, p(\mathbf{y})}{p(\mathbf{y})} = \sum_i p(\mathbf{x}|\mathbf{H}_i, \mathbf{y})\, p(\mathbf{H}_i|\mathbf{y}). \tag{1}$$

Moreover, you want the expression for $p(\mathbf{x}|\mathbf{y})$ to factor in such a way that it involves estimating $p(\mathbf{H}_i|\mathbf{y})$ because those are the weights for the transport models, and you are interested in those for their own sakes. I think this argument could be made more clearly if you started Section 2 by stating that the ultimate goal is to obtain the moments of $p(\mathbf{x}|\mathbf{y})$, which can be factored in different ways, and the particular factorization above is the most informative because it involves estimating the weights $p(\mathbf{H}_i|\mathbf{y})$.

4. In the footnote on page 3 you explain that $\mathbf{H}_i$ is intended to be an indicator variable that really represents the index into a set of transport models. You also say that $\mathbf{H}_1, \ldots, \mathbf{H}_N$ are the Jacobians of those transport models (line 26). Elsewhere, H_i is not bold (Equation (2)). These conventions should all be described in the main text (no footnote) and the meaning of bold versus non-bold should be clarified. I suspect your use of non-bold H_i and non-bold x in Equation (2) is because you are stating a generic result, and you are not specifically referring to $\mathbf{H}_i$ and $\mathbf{x}$ used the rest of the text in this section. Please explain that.

5. Line 27, page 3: Please define $\mathbf{y}^O$. I get that it is the mean of the random vector that represents the observations, but is it different that $\mathbf{y}^t$? It probably could be, but are you making any assumptions about that? Also, here you treat $\mathbf{x}^b$ as the mean of the Gaussian distribution of the random variable $\mathbf{x}$, but Equation (1) treats it like a random variable ($p(\mathbf{x}|\mathbf{x}^b)$). Of course, it is possible that it could be both if the model was hierarchical and specified a prior distribution on $\mathbf{x}^b$, but if that's the case it should be stated. I suspect that this is really just notation given that you write, $G(\mathbf{x}|\boldsymbol{\mu}, \mathbf{C})$ on line 28 (if $\mathbf{C}$ is bold, then $\boldsymbol{\mu}$ should also be bold). Finally, the expression "uncertainty covariance" is somewhat confusing, at least to me: should it just be "covariance"?

6. Lines 29–30 on page 3 and Equation (3): I don't understand why this is here, but perhaps that is because my understanding of what you are trying to do relies on expressions I wrote above for item 3 (my Equation (1)). The final expression for $p(\mathbf{x}|\mathbf{y})$ there is already in terms of $p(\mathbf{x}|\mathbf{H}_i, \mathbf{y})$. You then write, "Thus our posterior for the ensemble is a mixture of Gaussians...", which I agree with. We both have $p(\mathbf{H}_i|\mathbf{y})$ (I note that you have now switched to using capital "P" for probability instead of "p" used earlier– it's a minor thing, but it would be better to be consistent), and the remaining term I call $p(\mathbf{x}|\mathbf{H}_i, \mathbf{y})$ and you call $G(\mathbf{x}|\mathbf{x}_i^a, \mathbf{A}_i)$. It might be helpful to clarify this correspondence in the text since it ties back to the ultimate objective of expressing the posterior $p(\mathbf{x}|\mathbf{y})$ in a special way that admits the mixture of Gaussians representation.

7. Lines 2–3, page 4: When you say, "As usual with a joint PDF we obtain the marginal probability for a variable by integrating over all others", to what are you referring? Are you justifying Equation (4)?

8. Equation (4): There are a few things about this that need to be addressed or explained. First, you stated earlier in the footnote on page 3 that $\mathbf{H}_i$ is a stand-in for an index random variable that distinguishes between transport models, but you use $\mathbf{H}_i$ anyway to remind the reader to what this index refers. If that remains true, then $\mathbf{H}_i$ is a discrete variable here, not a continuous one. If that's the case, then $P(\mathbf{H}_i)$ is not Gaussian, and I don't think the right-hand-side of the equation makes sense. In Michalak et. al. (2005), the target of inference is $\boldsymbol{\theta}$ which is a vector of continuously-valued variance parameters, so it makes sense there. I think what you are trying to do with this expression is to obtain the set weights associated with models represented by $\mathbf{H}_i$ as in Raftery et. al., (2005) which you cite. Alternatively, maybe you have changed the notation implicitly to treat $\mathbf{H}_i$ (or more properly $\text{vec}(\mathbf{H}_i)$) as a Gaussian random vector. If so, please explain.

9. Lines 7–10, page 4: Several issues here. First, the words of the first sentence in Section 2.1 provide an example of where $\mathbf{x}^b$ is now discussed as if it were a random variable rather a parameter (in contrast to its use earlier in the paper). Is $\mathbf{x}^b$ a parameter of the prior distribution of $\mathbf{x}$ or is it a random draw from that distribution? Only in the latter case does the notion of independence from $\mathbf{y}$ make sense. Second, Equation (4) as stated is not the probability of simulating the observations ($\mathbf{y}$); it is the probability of $\mathbf{H}_i$ given the observations. Should it be $p(\mathbf{y}|\mathbf{H}_i, \mathbf{x})$? Third, I question assertion made in Michalak et. al. (2005), Section 6.4, Equation (4) that Equation (2) of that paper can be written,

$$p(\mathbf{x}) \propto G(\mathbf{x} - \mathbf{x}^b, \mathbf{B}) \times G(\mathbf{H}(\mathbf{x}) - \mathbf{y}, \mathbf{R}).$$

Equation (2) in Michalak et. al. (2005) is $p(x) \propto p(x|x^b) \times p(y^t|y) \times p(y^t|H(x))$. It appears to me that $p(x|x^b)$ (or $p(\mathbf{x}|\mathbf{x}^b)$ using the notation of the paper under review) is missing from the expression above. Finally, also $G(\mathbf{x} - \mathbf{x}^b, \mathbf{B})$ is ambiguous at best and nonsense at worst: do they mean $G(\mathbf{x} - \mathbf{x}^b|\mathbf{x}^b, \mathbf{B})$ and $G(\mathbf{H}(\mathbf{x}) - \mathbf{y}|\mathbf{H}(\mathbf{x}), \mathbf{R})$?

10. Lines 17–19, page 4: $P(\mathbf{H}_i)$ appears several times in this passage. Do you mean $P(\mathbf{H}_i|\mathbf{y})$?

11. Lines 22–28, page 4: What's the point of this second-to-last paragraph of Section 2.1? Is it simply to draw a line between the more familiar concept of χ^2 in the literature and the work here? You do use it in the next paragraph (and in Section 2.2), so perhaps these should all be combined into one paragraph? That would make it clear why χ^2 is being defined. Also, I don't understand the calculation given in lines 25–27.

12. Section 2.2: The statement that neither AIC nor BIC "take account of different prior uncertainties among parameters or different sensitivities of the observations to these parameters" is mysterious to me. That is certainly true, but that's not their purpose. Since I am confused about what $\mathbf{H}_i$ means here notationally, and that makes it hard to understand what you are driving at.