Round 3 Referee Report for
*Data Assimilation using an Ensemble of Models:
A hierarchical approach*

General comments:

The paper is significantly improved, but there are still some issues and ambiguities with Section 2. I can't fully comment on the remainder of the paper without sorting those out first. That said, the paper reads much better, to me at least, from start to finish.

Specific comments:

1. Page 3, lines 10 to 14 and Equation (1): The notation here is (still) ambiguous, and I do not follow how the left and right sides of Equation (1) are equal. You refer to $\mathbf{x}^b$ as the background, which I now understand to be the mean of the prior distribution of $\mathbf{x}$. The prior variance is not specified, so I am not sure how to interpret $p(\mathbf{x}|\mathbf{x}^b)$. You state elsewhere that Gaussian distributions are assumed so the variance would need to be specified. Same issue for $p(\mathbf{y}|\mathbf{y}^o)$ and $p(\mathbf{y}|H(\mathbf{x})$. You say, "$\mathbf{y}$ represents the observations", but then "o represents the observed value". So what is the difference between $\mathbf{y}$ and $\mathbf{y}^o$? My interpretation of Equation (1) as it is written is,

$$\begin{aligned} &\mathbf{x} \sim G(\mathbf{x}^b, \mathbf{V}_{\mathbf{x}}), && \mathbf{y} \sim G(\mathbf{y}^o, \mathbf{V}_{\mathbf{y}1}) && \mathbf{y} \sim G(H(\mathbf{x}), \mathbf{V}_{\mathbf{y}2}), \\ &\mathbf{x} = \mathbf{x}^b + \boldsymbol{\epsilon}, && \mathbf{y} = \mathbf{y}^o + \boldsymbol{\delta}, && \mathbf{y} = H(\mathbf{x}) + \boldsymbol{\eta}, \\ & && && \quad = H(\mathbf{x}^b + \boldsymbol{\epsilon}) + \boldsymbol{\eta}. \end{aligned}$$

   This is as far as I can get without clarification on $\mathbf{y}$ and $\mathbf{y}^o$. In any case, it is not obvious to me why one would take the product of these conditional distributions to obtain $p(\mathbf{x}, \mathbf{y})$. I don't think it would be obvious to a general reader either.

2. Page 4, lines 7 to 9. Here the prior covariance matrices of the distributions $p(\mathbf{x}|\mathbf{x}^b)$ and $p(\mathbf{y}|\mathbf{y}^o)$ are finally defined as $\mathbf{B}$ and $\mathbf{R}$, respectively. On line 9, it says "$\mathbf{y}^o$ is the observed value with uncertainty covariance $\mathbf{R}$", but if $\mathbf{y}^o$ is the mean of the prior distribution $\mathbf{y}$, then $\mathbf{R}$ should be the covariance matrix of the prior distribution of $\mathbf{y}$. There is nothing in your model about $\mathbf{y}^o$ being a random variable (this would imply a hierarchical model) up to this point. Did you mean to say, "$\mathbf{y}$ is the observed value with uncertainty covariance $\mathbf{R}$"?

3. Page 4, line 16, Equation 6: shouldn't this be an equality?

4. Page 4, line 17, This is nit-picking at some level, but you write "Thus, $p(\mathbf{x}, \mathbf{H}_i|\mathbf{y})$ is a sum of Gaussian Distributions..." Technically, it's not the distribution that is a sum of Gaussians, it's the random variable that has a distribution that is a Gaussian mixture.

5. Page 4, line 20, Equation 7: Shouldn't the left-hand side of Equation 7 be $p(\mathbf{H}_{i|\mathbf{y}^o})$? There is no term involving $\mathbf{y}$ on the right. Also, I think you are missing a comma at the end of line 19, and a period at the end of Equation 7.

6. Page 4, line 23: You say "Provided $p(\mathbf{x})$ (the prior distribution for $\mathbf{x}$) and $\mathbf{y}$ are independent...". Another nit-pick: it's not the distributions that are independent, it's the random variables. More importantly, how can $\mathbf{x}$ and $\mathbf{y}$ possibly be independent? Isn't $\mathbf{y} = H(\mathbf{x})$ given the last term on the right-hand side of Equation 1? I don't have the Tarantola book handy, so I am not sure what the Jacobian rule of probabilities is (will attempt to look this up). I think what you are trying to do here is to justify adding the variances and ignoring covariance between $\mathbf{x}$ and $\mathbf{y}$. You could probably just call this an approximation, but then we don't know how good the approximation is. Maybe I am missing something- please clarify.

7. Page 5, line 11: "two two".

8. Page 5, line 18: What do you mean by a "statistically consistent system"?

9. Page 5, line 21, Equation 8: What is K? I do not recall it having been defined previously.