Comment on remarks of first Referee:

These are very helpful comments, and we appreciate them indeed.

We agree that we should remove any references to “global” minimum. We did try to be a bit cautious, calling it a “consistent” global minimum, but we can certainly see how it might be too small a distinction. Indeed, others have asked us to be very much more careful on this point as those who know finding a global minimum of an NP complete problem will wonder, rightly, what we are actually talking about.

This will definitely be corrected in the rewrite of the paper.

We appreciate the brevity of our discussion of how the paths were found as we changed R_f We will enlarge that section of the paper as well.

We focused on the variational approach to evaluating the path integral only. We have some remarks on the Monte-Carlo methods which you are describing in your referee report, and they are below.

For purposes of explanation we did this:

We work in this paper only within the variational principle, call it Laplace’s method or 4DVar.

At $R_f = 0$, the solution to the variational principle is easy $x_l(n) = y_l(n); l = 1,2,...,L$ for the observed variables.

So we start with small R_{f0}. We chose $R_{f0} = 0.01$ in the case of Lorenz96. Now solve for the saddle points of the action $A_0(X)$ starting with 100 initial choices for the path. All of the initial paths for the interative method for finding saddle points of $A_0(X)$ are solutions at $R_f = 0$; so $x_l(n) = y_l(n); l = 1,2,...,L$ for the measured variables, and selections of the unmeasured variables uniformly distributed over the dynamical rage of the Lorenz96 model. Here they are chosen from a uniform distribution in [-10,10].

This results in 100 new paths that break the degeneracy of the paths at $R_f = 0$.

=======

Then we increase R_f to $R_f = R_{f0}2$, and use the paths from the solutions to the variational principle at $R_f = R_{f0}$ as initial points for the minimization procedure for $A_0(X)$. This gives us 100 new paths.

Then we move R_f to $R_f = R_{f0}2^2$, and use the paths from the previous value of $R_f = R_{f0}2^1$, arriving at another set of 100 paths.

Keep doing this, evaluating the action $A_0(X)$ on each of the paths at each R_f, until we have $R_f = R_{f0}2^\beta; \beta = 0,1,2,...$well into the regime where the values of the action on the saddle point paths have become independent of R_f.

We do this for each number of measured variables L = 1, 2, …

We then plot $A_0(X)$ versus R_f for all the paths we found by using the variational principle at each value R_f. That is what is shown in Figure 1.

So, we never used a Metropolis-Hastings like search in path space, which it seems you were supposed we had done.

This actually raised our awareness that calling our procedure an “annealing” might cause confusion, so we will change the name in the revised manuscript in addition to putting this enlarged explanation into more polished language.

We also realized, in thinking through your comment (3) that we could do the same thing for a Monte Carlo procedure, and we will put this into the manuscript, though we want to work out a set of examples which we will report in a follow-up paper so as not to lose focus on variational principles (4DVar or Laplace) here.

We very much appreciate your having raised this question.

On item (4), we did vary the forcing term over a range $7.5 \le f \le 8.5$ including our specific choice of f, and it had little effect. We will introduce this into the revised manuscript. We have also checked the effect of creating data with f = 8.17 and then presenting that to the model with f fixed at 15. This had a significant effect on the outcome. Further we changed the damping coefficient in the Lorenz96 model from -1 in the Lorenz96 equations to -2, and this too had an effect on the outcome.

These are then signals to the user that one should let the procedure search for the parameter values.

The result of these calculations will be in the manuscript.

Equation (7) for the Lorenz96 model was used to define the vector field f(x(n)) appearing in the model error term of the action. In the twin experiment, we solved

the dynamical equations to get orbits for the Lorenz96 equation, then we added Gaussian noise of variance 0.25 to these solutions to get our data.

We will make this clear in the revised manuscript.

On item (5) we will move this comment into the motivation paragraphs along with the reason. We hoped it might be enough to refer to Quinn's dissertation, but this is a good point to make.

On item (6). We effectively have done the 4DVar part of this suggestion as every step taken as we increase R_f from very small to order 10,000 or so was a variational calculation.

We are considering an EnKF calculation, though our initial reaction is that it takes attention away from this discussion of variational approaches.

Our method is clearly an ensemble variational method, and we will mention that.

More important in our opinion is that our approach does three things which are the core new results of our approach: (1) It tracks through slowly increasing R_f saddle paths where $\frac{\partial A_0(X)}{\partial X} = 0$ which contribute to Laplace's method. (2) It allows the explicit exposure of how increasing the number of measurements at any observation time L changes the values of the action levels produced and their splitting, if it happens. (3) It permits explicit calculation of the corrections, as an expansion (possibly asymptotic) in R_f^{-1}.

We hesitate to distract from these three things.

Our inclination at this point is to make the changes we have discussed in this comment for this paper. Then we plan to use the same method of continuation in R_f (our new name for annealing) but in a Monte-Carlo context. This would then make a second paper focused on ensemble methods.

Actually, as we do each 4DVar calculation at each R_f with 100 initial paths, our procedure in this paper is a kind of ensemble variational procedure. Of course, this is required because the nonlinear optimization problem: minimize $A_0(X)$ has many solutions depending on the initial value of the path.

==========

In addition to responding to your comments in a manner which we hope addresses the issues you raise, we will note that the identification of the

distribution of the model error term as chi squared is found in an earlier paper by Bennett and Chua. We will add this reference in our revised manuscript.

Jinxing Ye, Narig Kadakia, Paul Rozdeba, Henry Abarbanel, and Jack Quinn
23 November 2014