Nonlin. Processes Geophys. Discuss.,
https://doi.org/10.5194/npg-2019-1-RC1, 2019

Nonlinear Processes in Geophysics
Discussions
Open Access
EGU

Interactive comment on "Precision Annealing Monte Carlo Methods for Statistical Data Assimilation: Metropolis-Hastings Procedures" *by* Adrian S. et al.

S.G. Penny (Referee)

steve.penny@noaa.gov

Received and published: 12 March 2019

General Comments:

Overall, this is a well presented introduction of the mathematical problem statement of data assimilation from the point of view of the authors, and the proposed DA solution method is interesting and worth pursuing. For this reason, I recommend publication after addressing a few items described in more detail below. Most importantly, I feel that more experiments can be performed to further illustrate the merits of the solution method, including comparison to other common solution methods and exploration of experiment parameter variations. Secondarily, it would be useful to provide some anal-

Printer-friendly version

Discussion paper

ysis of the cost of the algorithm, particularly indicating how man times the full nonlinear model has to be integrated at each stage.

—

The problem statement seems to lay out what would be considered a single “data assimilation cycle” - meaning that observations in one fixed time window are used to estimate the state and model parameters, and then produce a forecast beyond that time window. In forecasting applications, this process is cycled over multiple time windows, and that cycling has a dynamics of its own. The results section would benefit from testing some degree of this cycling process.

I’d like to see some basic analysis of the cost of the Monte Carlo annealing method. Most importantly, since in many realistic applications (e.g. in weather and climate prediction) the model is the most costly part, how many model integrations are needed at each step of the algorithm?

As a reader, it is difficult to evaluate the quality of the solution approach without one or more traditional methods to compare to as a baseline. Can the authors provide a comparison to analyses produced by a simple implementation of a 4D-Var and an EnKF method? Is the analysis and forecast accuracy similar? Are the computational costs significantly more or less expensive?

It should be straightforward to automate the generation of a more thorough set of experiment cases that would provide more robust validation of the solution method. I’d like to see some more sensitivity analysis based on varying the noise level in the observations, the number/distribution of observations used (for example, I expect far fewer observations may be needed if the locations are random at each analysis time), the size of the ‘ensemble’ used in the monte carlo approach, and the starting point using different points on the L96 attractor, for example.

A basic assessment of the Lyapunov spectrum for the L96 model with its given con-

figuration (dimension and forcing) would help to give some context for the DA system performance. The presentation would benefit from a brief description of these at both the global (the maximum LE was referenced briefly) and local (i.e. FTLEs during the analysis and forecast time window) scales.

Specific Comments:

L18:

What is the difference in definition between data assimilation and statistical data assimilation?

L22-23:

Some may also have interest in the probability distribution or higher moments of the distribution.

L33:

It might be worth clarifying:

". . .information from [measured] data. . ."

L 41:

We also want to estimate the measured states which are inaccurately known due to errors in the measurement devices.

L 45:

It might be worth mentioning that in prediction applications, this process is repeated by redefining the time window to start at t_F.

L 52:

Could the bold formatting vector notation from the text for functions F and f be used consistently in equations (1) and (2).

L 57:

Perhaps since ‘F’ is already used as a function name, another symbol could be used for the largest ‘k’ value index.

L 66:

It should maybe be mentioned that this assumes the model parameters are global, i.e. they are not time-dependent or spatially dependent. Though you could also point out that if you were to cycle the analysis window, without any other changes, it would support a slow time varying global forcing parameter as is.

L 91:

Is there a part of the process that is analogous to the validation set?

Assuming these definitions, for example:

“Training Set: this data set is used to adjust the weights on the neural network.

Validation Set: this data set is used to minimize overfitting by verifying that any increase in accuracy over the training data set also yields an increase in accuracy over a data set that has not been shown to the network before.

Testing Set: this data set is used only for testing the final solution in order to confirm the actual predictive power of the network.”

Equation (3):

This may be an item for the editors to address, but the equation is difficult to read and would benefit from improved formatting.

I assume the large ‘dot’ is multiplication. It seems that the second large ‘dot’ (out of three) does not belong.

I like the use of color in the equation, but it might help to explain what it’s purpose is.

L 101:

I think there's an error in the denominator, please revise.

L 105:

Perhaps it is simply semantics, but I interpret equation (4) as being built from propagating forward in time from time t_0 to t_F. Is there a reason for the interpretation of moving backwards in time?

Equation (6):

I'm assuming the summation with index "k" should be inside the left bracket, since the second term doesn't have any "k" indices. Perhaps also double check the placement of the negative signs once the terms are adjusted.

L 115:

Maybe I misinterpreted the notation, but it seemed based on equation (2) that x(n+1) and f(x(n),p) were equal. If this is supposed to represent a measure of model error, it would help if this was explained in more detail.

Equation (7):

Since the argument to the action A is cap X, and this doesn't appear in the right hand side, it is a bit confusing at first sight. Could the authors make the relationship more clear. Again, it would help if somewhere it is clarified how the terms x(n+1) and f(x(n),p) differ. Which is the known quantity and which is the control variable corresponding to X? I'd like a little more explanation of the quantities in the second term and where they come from.

L 132:

Maybe change "perform" to "solve" or "evaluate" the integral.

L 133:

“[occurs at] the [value of X that produces the] maxima of. . .”

L 138:

“methods to [solve] the integral. . .”

L 148:

change “yielding the smallest value of” to “minimizing”

L 170-172:

Could this be written more simply (or additionally) as an equation?

L 173:

“In practice, [the closer] one can choose X[init] to [correspond to] the maximum of P(X|Y), . . .”

L 176:

“[was] located (Shirman 2018), we [would]. . .”

L 184-186:

it should be acknowledged that this will encounter problems if the model has significant systemic errors relative to the ‘true’ system used to generate the observed data. (The systemic model errors may be more severe than a misspecified model parameter, e.g. a process that was oversimplified in the formation of the model).

I expect there should be some limitations depending on the number of observations that are used. For example, if L=1, then there is still a significant portion of the state estimate that will be unconstrained when Rf=0, and could lead to a quite poor starting condition. If there is not any discussion about this in the coming sections, I’d ask that some be added.

Equation (8):

This definition should probably appear where X first appears in the manuscript.

L 214:

There's an extra space before the period in "Eq. (9) ."

L 217-218:

"However, we substitute for xl(t0 + k[nτ]), whenever it occurs in the equations Eq. (9), . . ."

I assume then that you're relying on a degree of synchronization so that the unobserved variables become dynamically consistent with the observed variables. With respect to this process, I have two questions:

(1) how long does it take for the unobserved variables to reach a dynamic equilibrium with the observed variables so the full states are balanced. Is this performed before the DA analysis window [t_0,t_F], or within this window? If the latter, is sufficient time given in the ensemble generation phase to achieve sufficient 'spin up'?

(2) Assuming there is noise in the observed variables, this could potentially produce states that are not dynamically consistent. What is the sensitivity of this procedure to increasing noise in the observations?

L 227-229:

It would be helpful to explain the fixed iterations of M-H, what a 'burn-in' step is, etc. The ensemble generation stem is important and at the moment it seems a bit ambiguous, which makes analyzing the following steps more difficult as a reader.

L 231:

What does the second argument in N_A(q,0) stand for?

Equation (10):

This looks like an ensemble mean of the N_A accepted paths for each of the N_I initial

paths. From the description in lines 227-230, it's not clear to me how these paths are generated.

L 238:

"All these paths have zero action."

Just to make sure I understand - the paths all have zero action because they all intersect at the observed points, and at this point the Action term only includes the first term that applies a penalty as the state deviates from the observed values. Is this interpretation correct?

L 242-245:

Could you clarify how many integrations of the full nonlinear model are required at each iteration of increasing R_f values? Is it one per each of the N_I ensemble members?

L 249:

missing space: "X[0]for"

L 251:

Is there a guarantee that this procedure finds the true minimum and does not get stuck in one of the local minima?

L 246-264:

In this case, I think this is more explanation than is needed and this iterative processes can probably be consolidated by just defining the first step and the induction step.

Equation (14):

I wonder if there is something that can be learned from computing and examining the error covariance matrix produced by this resulting ensemble. It would be interesting to see how the error covariance matrix changes/converges over each of these iterations from 0 to beta.

L 282-283:

It seems like you will eventually reach a point of 'overfitting' a model with systematic errors (e.g. slightly incorrect fundamental equations, not just parameter errors). Is there a stopping criterion to avoid overfitting? Going back to my earlier question, I wonder if there is an analogous part of the algorithm to the 'validation' phase of the deep learning machine learning process that could help identify overfitting.

L 293:

What is the length of the time window? (i.e. what is t_0 and t_F?)

L 301:

I assume the choices of L=5 to 12 is only applied for the D=20 case. But it mentions above (L290) that there are cases with D=5, does this experiment case also have a similar parameter list?

Equation (16):

I asked before, but I'd like clarification about which 'x' terms in equation (16) are the control variables (i.e. the ones corresponding to X). Based on equation (8), I assume it is the x_I and x_a terms? But I know you are also trying to estimate p. My main point is that it is unclear and needs to be laid out in more detail when first introduced.

It's not really consequential, but the order of (y_I - x_I) changed from the previous action equation (7). It might be better to keep it consistent.

Figure (2):

I wonder if you can limit the ensemble size N_I equal to the number of positive (+neutral) Lyapunov exponents, similarly to the minimum required ensemble size for the EnKF. Have you tried reducing the ensemble size N_I and testing the sensitivity of convergence to this size?

For reference, see Bocquet and Carrassi 2017:

https://www.tandfonline.com/doi/full/10.1080/16000870.2017.1304504

I assume this figure illustrates that the system cannot be observed with as few as 5 observed model grid points per analysis time step. It might be worth clarifying the authors' interpretation in the figure caption.

Could you describe which are the observed variables, as in the Figure (3) caption?

Figure (3):

"This leads the action to become effectively equal to the action itself"

I don't understand this statement.

Figure (4):

It seems like most of the members are not finding the correct value of the forcing parameter. Am I interpreting this correctly?

Figure (5):

Are these results aggregating the cases in Figure (4)? They don't appear to correspond. Or is Figure (4) before convergence?

Figure (6):

dt = 5.0 for the L96 model likely has pretty nonlinear error growth. (1) An estimate of the error growth over this window (e.g. the FTLEs) could be useful to set the context for how you might expect errors to grow during the forecast period. (2) With a systematic model error this might experience even greater sensitivity. I wonder if the authors could attempt a similar experiment with a shorter time window (e.g. with more approximately linear error growth), but cycle the process over multiple time windows like a realistic forecasting application.

Equation (18):

Make the location of the precision terms (Rm, Rf) consistent, and apply consistent use of () versus [] brackets for each term.

L 423:

Put the actual url - a reader of a printed copy of the text will not be able to see the link.

Interactive comment on Nonlin. Processes Geophys. Discuss., https://doi.org/10.5194/npg-2019-1, 2019.

Printer-friendly version

Discussion paper