Data-driven Reconstruction of Partially Observed Dynamical Systems – review report

16th January 2023

In this manuscript, the authors derive a method to reconstruct the dynamics of a system from partial observations, in which data assimilation and machine learning steps are alternate. The data assimilation steps are used to estimate the state from observations using the surrogate model, while the machine learning steps are used to estimate the surrogate model from the data assimilation analysis. This method is the same as the one derived by Brajard et al. 2020, with the exception that, on top of this method, the authors propose a new, innovative state augmentation process. The entire method is illustrated using numerical experiments with the 3-variable Lorenz 1963 system.

I am overall positive about this manuscript. The text reads very well and is easy to follow. To my knowledge, the state augmentation process is new and deserves to be published. However, I have some concerns, in particular about the methodology and about the experiments, that needs to be fixed before I can recommend publication.

1 General comments

1.1 How the methodology differs from that of Brajard et al. (2021)

As far as I understand, the method derived in this manuscript proposes to alternate data assimilation steps (with the ensemble Kalman smoother) and machine learning steps (with a linear regression) on a given dataset of observations until convergence. This is exactly what has been originally proposed by Brajard et al. (2020) and later formalised by Bocquet et al. (2020). Pushing further the comparison, I see only three significant differences with the original method:

- in the present method the machine learning step is restricted to linear regression, while in the original method, nonlinear regression tools (such as neural networks) are used;
- in the present method observations are assumed to be perfect (even though they are sparse), while in the original method, sparse and noisy observations are used;

- the state augmentation process added on top the data assimilation / machine learning iterations.

I do not see the first two points as a major limitations, in fact I am rather confident that the present method should also work with neural networks replacing the linear regression and with noisy observations. By contrast, the third point is in my opinion the real added value of the present work, and this should be emphasised.

Additional questions about the methodology

1. Is there a fundamental reason to use only linear regression and perfect observations? If not, I would suggest to get rid of these assumptions in the methodological section.
2. How does the state augmentation scale with the system dimension?
3. Can the additional state components be added all at once? Did you try that in the numerical experiments?
4. In the experiments, 30 iterations seem sufficient to reach convergence. Do you have an idea how this number would scale with the system dimension?
5. The text is ambiguous about the data assimilation method used: 'and thus uses the classic Kalman filter and smoother equations' (L 71-72), 'by the Kalman filter' (77-78) 'a Kalman smoother is applied' (L 87) 'Kalman filter and smoother' (L 161). Kalman filter or smoother, you have to choose (I assume it is Kalman smoother).

1.2 About the numerical experiments

The description of the experiments is incomplete, in such a way that the experiments cannot be reproduced without further assumptions. For example, what numerical method is used to integrate in time the model equations to compute the truth?

Furthermore, I have a serious concern about the 'model distance' introduced by equation (6). Without further details, I assume that it is computed using the same trajectory as the training step. Using the same data for training and testing should be avoided by all means. Moreover, in this context where observation are perfect, I am not sure to see the point of this metric: observations are required to initialise the model (for the hidden components), but if we have observations, we do not need the forecasting system any more since observations are perfect... Therefore, I think that the metric used to evaluate the accuracy of the model should be reconsidered.

Additional questions about the experiments

1. '10 loops of the Lorenz-63 system' (L 104-105) Do you mean 10 model time units or 10 revolutions on the model attractor? In any case, I would not say that this is a small period of time, compared to the doubling time which is 0.78 MTU.

2. From what I understand (L 104-106), you have access to the true x_2 and x_3 (no observation noise) every $dt = 0.001$ (which is probably the integration time step for the truth). This seems to be very strong requirements. Can you discuss this?

3. What is the choice of the data assimilation window length for the ensemble Kalman smoother? Without further details, I assume that it covers the entire experiment, i.e. 10^4 observation steps. This is really huge. Can you discuss this?

2 Technical comments and suggestions

L 17-18 'using Bayesian framework' → 'using a Bayesian framework'?

L 21 'All the approaches cited above are assuming that the full state of the system is observed' This is not true: at least Tandeo et al. (2015), Lguensat et al. (2017), Bocquet et al. (2019), Brajard et al. (2020), Fablet et al. (2021) use sparse observation operators in their methods. I would replace 'All the approaches cited above' by 'Many approaches'.

L 23-24 'To deal with those strong constraints' I would replace here 'constraints' by 'assumptions' in order to avoid a potential confusion with strong-constraint methods in variational data assimilation.

L 24-26 'An option is to [...] whereas an other option is to [...]' I would suggest to also mention here the combination of data assimilation and machine learning, because (i) this is what is used in some of the previously cited papers (the ones that can handle sparse and noisy observations), and (ii) this is what is used in the present manuscript!

L 29 'with a dynamical model (model- or data-driven)' I would replace here 'model-driven' by 'based on physical knowledge' or something like this (to avoid a *model-driven model*).

L 31 'estimation of the parameters' Which parameters?

L 40-41 'from data assimilation, machine learning, and theory of dynamical systems' → 'from data assimilation, machine learning, and dynamical systems'?

L 42 'from partial observations $\mathbf{y}$' In data assimilation, observations are usually noisy in addition to being partial.

L 42-46 In this paragraph, why didn't you mention the crucial role of the background error statistics?

L 48 'to mathematically approximate the dynamic of the system' → 'to mathematically approximate the system dynamics'.

L 76 In equation (2), I would suggest to explicit the definition of $\mathcal{L}$, i.e. use something like that:

$$\mathcal{L} \triangleq p(\mathbf{y}_1, \ldots \mathbf{y}_T | \mathbf{x}_1^f, \ldots \mathbf{y}_T^f) \propto \prod_{t=1}^{T} \cdots \tag{1}$$

Furthermore, T is undefined in this equation.

L 78-79 'The innovation likelihood given in Eq. (2) is interesting because it corresponds to the squared distance between the observations and the forecast normalized by their uncertainties, represented by the covariance $\mathbf{\Sigma}_t$.' In data assimilation, this quantity is simply called 'the likelihood'.

L 89-90 'This random sampling is used to exploit the correlations between the components of the state vector' I do not understand why this is necessary. Could you elaborate?

L 110 'After 30 iterations of the algorithm presented in section 2, the hidden component z_1 is stabilized.' Can you please explain the exact meaning of 'stabilized' in this context?

L 114 'this augmented state procedure is repeated' $\rightarrow$ 'this state augmentation process is repeated'.

L 117 'z_3 is very flat' I would replace 'very' by 'rather' in this statement.

L 125 'Finally, the inclusion of z_3 reduces the likelihood (purple lines).' Do you have an explanation for this phenomenon?

L 131 In equation (6), I would explicit the dependence on time, i.e. replace $\mathrm{dist}(\mathbf{M})$ by $\mathrm{dist}(\mathbf{M})(t)$.

L 137-138 'Are they correlated with the unobserved component x_1 or with the observed one x_2 and x_3?' $\rightarrow$ 'Are they correlated to the unobserved component x_1 or to the observed ones x_2 and x_3?'?

L 139 'It has been found that...' How did you come up with this? As it is presented, it looks like something pulled out of a hat.

L 47-48 'This is illustrated in Fig. (3), with 50 independent realizations of the proposed algorithm.' Strictly speaking, this is not the case since a and b are not represented in this figure.

L 152-153 'Then, when considering z_1 and z_2 (red lines), the 50 independent realizations reach the same likelihood after 30 iterations.' What about a and b? Are they similar over the 50 realisations?

L 153-154 'it will then focuses' → 'it will then focus'.

L 175-176 'the dynamical evolution of the system is retrieved with our methodology' This is not clearly shown in the experiments.

References

Bocquet, Marc, Julien Brajard, Alberto Carrassi and Laurent Bertino (2020). 'Bayesian inference of chaotic dynamics by merging data assimilation, machine learning and expectation-maximization'. In: *Foundations of Data Science* 2.1, pp. 55–80. DOI: 10.3934/fods.2020004.