Review Data Driven Regional Weather Forecasting

This paper uses data driven techniques to forecast partially observed systems. The proposed framework relies on phase space reconstruction and defines an approximate discrete vector field on the reconstructed phase space to forecast the regional observations. The model was tested on a Shallow Water Equation setting with various configurations of the observations. Overall, the paper is well written and suitable for the audience of NPG. However, several aspects need some investigation to further strengthen the paper.

Major comments :

1. Positioning of the paper and bibliography: I liked your text in the 1.1 section, especially the shift in the machine learning community to account for physical constraints in the models. However, your state-of-the-art section is missing two crucial aspects. You did not discuss the related works on machine learning for regional forecasting systems (for example: [1, 2, 3, 4]). These references do not have explicit titles like yours, but are very similar in spirit. Furthermore, you also need to discuss the state of the art in data driven modeling for partially observed systems (for example: old traditional methods [5], new ones [6, 7, 8, 9, 10]). These works do use phase space reconstruction techniques, as you do for predicting partially observed systems.

2. Notation: overall I found the paper very smooth and pleasent to read, except when you introduced the models equations. I think there is a way to simplify the notations to make the paper even smoother. For example, you use for the "plane coordinates" the variable $\mathbf{r} = \{x, y\}$ which is equal to $\{r_1, r_2\}$ which is equal to $\{r_{10} + i\Delta x, r_{20} + j\Delta x\}$ these are lots of symbols to keep track of. I think you can only keep $\mathbf{r} = \{r_{10} + i\Delta x, r_{20} + j\Delta x\}$ and explain what is $\Delta x, \Delta x$ and i, j. Also, you use both Δt and h for the time step. You can also explain that you use capital letters for the observations and regular letters for the states. In equations (3), (4) ...etc. when you write that
$$\frac{d\mathbf{S}(i,j,t)}{dt} = \mathbf{F}_{i,j}(\mathbf{S}(i,j,t), \theta) + [\mathcal{F}(i,j,t), 0]$$
I understand that the state $\mathbf{S}(i,j,t)$ at the grid point i, j depends only on $\mathbf{S}(i,j,t)$ and the forcing. In order to avoid confusion, make the above equation depend on a defined global state, for example $\mathbf{S}(t) = [\mathbf{S}(0,0,t), \ldots, \mathbf{S}(n_x, n_y, t)]^T \in \mathbb{R}^{3(n_x \times n_y)}$. Also, make sure the forcing notation is consistent with the definition of the global state vector $\mathbf{S}(t)$.

3. Regarding the optimization of the embedding parameters: after reading lines 347 to 352, I don't really understand how you parameterized your embedding parameters. In the appendix, it says that you use FNN to get the dimension, but it's not clear how you calibrate the delay. Please further explain your methodology here.

4. On the delay embedding parameterization: You state in lines 379 to 382 that in the Takens delay embedding theorem, there is no underlying assumption on the time delay. I have an example that contradicts this. Imagine that your limit-set is periodic with period T and that you use a time delay of T. In this situation, your reconstruction will not be diffeomorphic to the underlying limit set. Actually, this is written the other way around in the Fractal Delay Embedding Prevalence Theorem (Theorem 2.5 in [11]).

5. Benchmark: This is extremely important. You should provide a benchmark of your results to help readers better understand the performance of your work. At least compare to persistence, but I would recommend testing against other state-of-the-art machine learning models.

6. Presentation of the experiments: For the experimental section, I recommend including only the sensors configuration figure and a single forecasting figure per experiment. The remaining figures should be placed in the appendix.

7. On the dimension of the embedding: You state in lines 384 to 388 that the result of Sauer et al. indicates that if $D_E > 2D_A + 1$ the unprojection would work. I think that this result is from Whitney Embedding Prevalence Theorem (theorem 2.2 in [11]). Also, when you assume that D_A is near 300 (lets say $D_A = 300$) then $D_E \approx 10, 20$ is very far from satisfying the condition $D_E > 2D_A + 1$. Please rectify this sentence.

8. On the scalability of the model: In your experiments, your initial state is of dimension 300. However, the embedding you used is of dimension $27 * 10$ to $27 * 20$. On higher dimensional systems as you mensioned, the dimension is much larger that 300 ($\mathbf{O}(10^6)$ for example). Do you expect your model to be in even larger dimensions than $\mathbf{O}(10^6)$. If it is the case, what is the cost of the model ? If not, why would you expect to get a smaller dimension.

Minor comments,

1. Do you mean embedding when you write unprojection ? If so, use embedding.

2. in lines 100 to 101, It's good to explain what one loses when bypassing equations.

3. in lines 96 to 97, I don't understand why you avoid uncertainties in the data. If your model is optimized from data and there is some uncertainty in the data, you would have uncertainty in the models.

4. I found the sentence from line 253 to 257 difficult to read.

5. Equation 6 has two commas.

I hope that the authors will understand my comments in a constructive way, and that I value their work and the time they invested in the preparation of the manuscript. It might be that I have misunderstood something, in this case, if something wasn't clear for me as a reviewer, it is possible that it wouldn't be clear also for the readers.

References

[1] James Knighton, Geoff Pleiss, Elizabeth Carter, Steven Lyon, M Todd Walter, and Scott Steinschneider. Potential predictability of regional precipitation and discharge extremes using synoptic-scale climate information via machine learning: An evaluation for the eastern continental united states. *Journal of Hydrometeorology*, 20(5):883–900, 2019.

[2] Veronica Nieves, Cristina Radin, and Gustau Camps-Valls. Predicting regional coastal sea level changes with machine learning. *Scientific Reports*, 11(1):1–6, 2021.

[3] Changjiang Xiao, Nengcheng Chen, Chuli Hu, Ke Wang, Zewei Xu, Yaping Cai, Lei Xu, Zeqiang Chen, and Jianya Gong. A spatiotemporal deep learning model for sea surface temperature field prediction using time-series satellite data. *Environmental Modelling & Software*, 120:104502, 2019.

[4] Serkan Kartal. Assessment of the spatiotemporal prediction capabilities of machine learning algorithms on sea surface temperature data: A comprehensive study. *Engineering Applications of Artificial Intelligence*, 118:105675, 2023.

[5] Henry D. I. Abarbanel and Upmanu Lall. Nonlinear dynamics of the great salt lake: system identification and prediction. *Climate Dynamics*, 12(4):287–297, Mar 1996.

[6] Ahmad Kazem, Ebrahim Sharifi, Farookh Khadeer Hussain, Morteza Saberi, and Omar Khadeer Hussain. Support vector regression with chaos-based firefly algorithm for stock market price forecasting. *Applied Soft Computing*, 13(2):947 – 958, 2013.

[7] Jordan Frank, Shie Mannor, and Doina Precup. Activity and gait recognition with time-delay embeddings. In *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence*, AAAI'10, pages 1581–1586. AAAI Press, 2010.

[8] S Ouala, D Nguyen, L Drumetz, B Chapron, A Pascual, F Collard, L Gaultier, and R Fablet. Learning latent dynamics for partially observed chaotic systems. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 30(10):103121, 2020.

[9] Georg A Gottwald and Sebastian Reich. Combining machine learning and data assimilation to forecast dynamical systems from noisy partial observations¡? a3b2 show [editpick]?¿. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 31(10):101103, 2021.

[10] Charles D Young and Michael D Graham. Deep learning delay coordinate dynamics for chaotic attractors from partial observable data. *arXiv preprint arXiv:2211.11061*, 2022.

[11] Tim Sauer, James A. Yorke, and Martin Casdagli. Embedology. *Journal of Statistical Physics*, 65(3):579–616, Nov 1991.