Geoscientific Model Development Discussions

Geosci. Model Dev. Discuss., referee comment RC2
https://doi.org/10.5194/gmd-2021-221-RC2, 2021

Comment on gmd-2021-221

Anonymous Referee #2

Referee comment on "Training a supermodel with noisy and sparse observations: a case study with CPT and the synch rule on SPEEDO – v.1" by Francine Schevenhoven and Alberto Carrassi, Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2021-221-RC2, 2021

The supermodel approach is an ensemble approach in which models are combined dynamically. The idea is that when the weights that define the combinations are optimized by training on the basis of historical observations, the supermodel could potentially improve upon each of the models in predicting state trajectories. In other words, the supermodel is potentially a model that has better prediction performance than each of the model in the supermodel ensemble. In previous work the authors demonstrated the viability of the supermodel approach in which weights were trained with the CPT (Cross pollination in time) algorithm and with the synch algorithm respectively. Simulations were performed in the context of the SPEEDO model with noiseless data observations at each time step. This paper builds upon that previous work. The extension in this paper is that both the super model dynamics and algorithms are adapted to deal with noisy data that is sparse in time. Furthermore, a modification of the CPT algorithm is proposed to explore the possibility of negative supermodel weights, which were found previously found by the synch algorithm and showed to be beneficial. In this paper, experiments have been performed in the same SPEEDO context as in the earlier work, but now with noisy data that are only available every Delta T time steps. The weights that were found with the adapted CPT and synch rule were compared with the weights found in the previous work. The result is that the adapted versions of the algorithm find similar weights as earlier, however now in the more realistic and difficult setting of the sparse and noisy data.

In my opinion the paper addresses relevant issues and contributes to the field. In particular the proposed methods could make a step forward towards the application of the supermodel approach in real world applications. In my opinion the paper is therefore publishable, however with revisions. Below I state some questions, comments and suggestions for revisions.

Comments:

1. The organization/structure the paper should be improved.

1.a in my opinion, many parts of the 'Results' section rather belong to the 'Training methods' section. To be more specific, in section 3 'Training methods', the authors describe training algorithms, including the synch algorithm and CPT , as well as some novel adaptions in these algorithms. This is fine. However, in section 4, 'Results', the authors continue with describing several adaptations of the algorithm (synch rule adaptations, adaptations to nudging, CPT adaptation, CPT with negative weights). In my opinion, these are all training methods and belong to section 3 'Training methods' and not to section 4 'Results'. To summarize, I would suggest to put all the training methods in the section 'Training methods' and only results (findings from your experiments) in the 'Results' section.

1.b Section 3.1 has the title 'Training in SPEEDO'. It basically contains the SPEEDO imperfect model descriptions, which is important to appreciate the experimental set-up and the results. In my opinion section 3.1 is not about a training, and a renaming of section 3.1, e.g., 'SPEEDO imperfect models' would be appropriate in my opinion. I also would consider to move this subsection 3.1 to section 2 (Supermodels) , since this is more about the models that are to be combined in the SPEEDO supermodel than about Training methods.

There are more of these types of organizational issues. It would be good if the authors would have critical look at the paper and restructure it where necessary.

2. Experimental set-up and interpretation of the outcomes

2.a. Except for table 6, the findings (outcomes of the experiments – I am not talking about proposed training methods) in this paper are only supermodel weights (tables 3,4,5 +figure 3). It seems that the authors are only interested in the resulting weights. Why is that? Is model performance not more interesting? Although it seems nice that the weights found in the experiments are to some extend similar to the weights found in their previous paper, I have no idea about the effect of the difference in weighting on the performance of the super models. It would be good to have a table like table 6, with model performances for all the different combinations of (Delta T/noise/Training algorithm) in order to appreciate the benefit of the supermodel approach in the different cases. Ideally, these results are compared with performances from a weighted MME approach, but maybe that is too much to ask.

2.b. Maybe I overlooked this, but how much training data was used for the synch rule?

3. Some issues with the formal supermodel definition.

3.a It would be good to have a formal definition of the notion 'synchronized supermodel state'. From the paper I understand that a supermodel is synchronized if its models are in the same state, i.e., $x_1 = x_2$. I assume this in the comments below. I am not sure if this is correct.

3.b I understand that in (1a, 1b, 1c) the supermodel remains synchronized if the states are initialized by $x_1 = x_2 = x_s$. However, in your description, it is not clear to me if this initialization is assumed or not. Although (1a,1b,1c) is not further used in the paper, it would be very helpful if the authors would state this explicitly.

3.c About the redefined supermodel (2a,2b,2c), the authors state in line 98-99 that weighting the states ensures a synchronized supermodel state every Delta T. I understand that the state x_s will be reset to w_1 x_1 + w_2 x_2 after every Delta T time steps, (eqn 2c) but I would expect that the states x_1 and x_2 should be synchronized to x_s as well (or at least nudged to x_s) every Delta T. Otherwise, I would expect that x_1 and x_2 will diverge from each other. In particular if Delta T is large, the effect of the term delta_mod(t,delta T) f (x_s, p_1) on the dynamics of x_1 defined in (2a) will be negligible (the tendency of model 1 is only one time step governed by the superstate x_s and all the other maybe 100s of time steps governed the state x_1.) In other words, the supermodel will basically behave as an unconnected ensemble of models, which I assume will not synchronize at all.

Please indicate whether or not x_1 and x_2 are synchronized (or nudged) to x_s every Delta T steps and if this is not the case, please explain why my reasoning is erroneous.

3.d The authors introduce a Delta T, on the one hand to reduce computational costs of combining models and on the other hand to deal with observational data that is not available at every time step (at least this is what I understand from lines 85-90.) The authors seem to identify the Delta T in both cases. That is to say, the Delta T, which is the frequency of redefining the supermodel state in the supermodel (2a,2b,2c), introduced for computational efficiency, seems to be identified with the observation frequency Delta T as defined in e.g. line 312. Is it obvious that these should be the same? If there would bigger times between observations, would that mean that the dynamics of the free running supermodel should also have a larger Delta T, even if computational resources would allow for a much smaller Delta T, or vice versa? What if training data is available at irregular times? Please explain or comment.

4. Finally, two (minor) comments about statements in the Introduction

4.a) Introduction, line 31-36: The authors remark that weights in a multimodal ensemble (MME) have the issue that weights that are optimal for model behavior in the past may not convert into optimal weights for the future. Then they remark that a dynamical on the fly method is desirable. By the statement 'Along this line, in the supermodel approach …' they seem to suggest that the supermodel approach is such an on-the-fly method that can deal with the issue. I don't see why this would be the case. Both the weights in a weighted MME and weights in a supermodel are necessarily to be trained on historical observations. If a certain type of change in model regime did not occur in the past but only in future, neither MME nor supermodel approach could have taken this into account and in both approaches, the weights might be suboptimal. Do the authors agree? Please comment.

4.b) Line 42-44: The authors make the comment that the if models in the supermodel are well enough synchronized, the supermodel can give improved model trajectories, whereas in MMEs individual variability in trajectories may be cancelled. I think the opposite might be true as well: by considering trajectories in an MME, one can compute a mean trajectory and fluctuations around this to assess the variability and the uncertainty of the prediction. Example: Model 1 predicts a severe storm tomorrow. Model 2 predicts the severe storm the day after tomorrow. It would be very naïve to say that together they predict that there will be a mild storm tomorrow and the day after. To me a more natural interpretation would be: 50% chance that a severe storm arrives tomorrow and 50% chance that it arrives the day after tomorrow. Now back to the supermodel. The authors indeed mentioned that the models in a supermodel must be well enough synchronized. Does this mean that in a supermodel that is not well synchronized, there is risk that models are getting out of phase, counteract and damp each other (in case of the storm example, that indeed the supermodel predicts a mild storm spread over two days?). Please comment.