Comment on “Improving weather and climate predictions by training of supermodels” by Schevenhoven et al.

Reply to Reviewer 2:
We would like to thank Reviewer 2 for his/her helpful comments. We revised our manuscript accordingly and provide point-wise replies to the Reviewer below.

1. *Eq. 6c, synch role: it seems that the derivative of the model (f) with respect to a certain parameter (q) is needed for sync role. I may have overlooked it, but how is this derivative obtained. It somehow looked like one may have to use an adjoint (or at least linear) version of the model.*

In the paper the synch rule is applied to optimize the weights of the weighted supermodel. In our case **f** in Eq. 6c represents the supermodel tendency and **q** the weights. The derivative of **f** with respect to a certain weight is simply the tendency of the imperfect model belonging to that weight which is readily available. This because the supermodel tendency is a weighted superposition of the time derivatives of the imperfect models, as shown in Eq. 2c for an example of two imperfect models: $\dot{\boldsymbol{x}}_s = \boldsymbol{W}_1\dot{\boldsymbol{x}}_1 + \boldsymbol{W}_2\dot{\boldsymbol{x}}_2$, where $\dot{\boldsymbol{x}}_s$ denotes the supermodel tendency, $\boldsymbol{W}_{1,2}$ the weights and $\dot{\boldsymbol{x}}_{1,2}$ the imperfect model tendencies. Hence taking the derivative of the supermodel tendency with respect to the weight of an imperfect model results in the tendency of that imperfect model. We clarified this in the paper in section 4.2.

2. *Coupling to the same ocean/land: I’m wondering how much the use of the same (perfect) ocean/land component for all models affect the training/behavior of the super- model, as, in my view, it may constrain the variability on long time scales to be similar in all models. In addition, the coupling to land/ocean may act similar (though more complex) to the nudging term (K(as-a0)). Perhaps, the authors have tried to build a supermodel from a set of totally independent SPEEDOs, and can comment on this.*

Indeed, coupling to the same ocean/land constrains the variability near surface on longer time-scales in the different atmosphere models since the coupling acts as a nudging to the true ocean/land temperature through the surface heat fluxes during training. The reviewer asks how this affects the training of the supermodel and the behavior after training. In CPT training it helps to keep the CPT trajectory in the neighbourhood of the truth. A longer CPT trajectory allows for a more robust estimate of the weights. In case the imperfect models include imperfect representations of the land and ocean components as well, a stronger drift from the truth can be expected during training since model errors in the evolution of the ocean and land state are introduced in addition to the model errors in the atmospheric evolution. We have not built a supermodel from a set of SPEEDOs with imperfections in the ocean and land components included. It remains to be seen whether in this more general case CPT training results in a weighted supermodel with improved prediction quality compared to the constituent imperfect models. Application of the synch rule in this more general case includes nudging of the ocean and land state to the truth so during training the supermodel cannot drift from the truth on longer timescales. But it also remains to be seen how much the long-term evolution of the supermodel can be improved on the basis of the minimization of the short-term synchronization errors in the presence of imperfections in the slow components of the climate systems. In this paper we have demonstrated that in the absence of imperfections in the slow components of the climate system, the long-term behavior of the supermodel improves while training only short-term prediction errors through the CPT or synch rule approach. A natural next step is to include imperfections in the slow components of the climate system as well.