Comment on the paper*:*
"Towards Strongly-coupled Ensemble Data Assimilation with Additional Improvements from Machine Learning"

General comments:
This paper compiles and comprehensively shows developments towards strongly coupled data assimilation. The studies included in this paper are organized according to the coupled model complexity, exploring the benefit of strongly coupled data assimilation over other approaches. They also include some studies in which one of the caveats of strongly coupled data assimilation is addressed by exploring low-order and intermediate-complexity coupled models. The last one is explored with the aid of neural networks. The paper is well- written and clear in general. However, it only focuses on the authors' research group developments and understanding of the field.

Thank you very much for your comments. All the line numbers in our response refers to the line numbers in the manuscript with tracked changes.

Although the title and the abstract are pertinent, the Authors should clearly state that the paper reviews their previous work. In its present form, the title indicates a novel contribution.

R: Thank you for this suggestion. We have revised the original title "Towards Strongly-coupled Ensemble Data Assimilation with Additional Improvements from Machine Learning"

to

"Review Article: Towards Strongly-coupled Ensemble Data Assimilation with Additional Improvements from Machine Learning"

At the same time, in being a review, the manuscript should also mention other key contributions to the field of coupled data assimilation.

R: In the Introduction Section (L39-L92), we have mentioned numerous important contributions in both variational and ensemble coupled data assimilation. We also mentioned two other excellent workshop or review papers by Penny et al. [2017] and Zhang et al. [2020].

We made the following revisions to narrow down the scope of our paper:

- We emphases that we focus on 1) model state estimations and 2) investigating the impacts of coupled atmosphere-ocean data assimilation on the coupled analysis and short-range weather forecasts.
- We refer to the review paper by Zhang et al. [2020] to the readers who are interested in the parameter estimation using coupled model and other coupled DA applications.

We add the following discussion (in Bold) to L93:

"…with increasing complexities. **We focus on model state estimations and impact of atmosphere-ocean CDA on coupled analysis and short-range weather forecast. Besides model state estimation, Zhang et al. [2020] recently reviewed parameter estimations and other important applications of CDA.**"

Reference:
Zhang, S., Liu, Z., Zhang, X., Wu, X., Han, G., Zhao, Y., Yu, X., Liu, C., Liu, Y., Wu, S., Lu, F., Li, M., and Deng, X.: Coupled data assimilation and parameter estimation in coupled ocean–atmosphere models: a review, Climate Dynamics, 54, 5127–5144, https://doi.org/10.1007/s00382-020-05275-6, 2020.

Besides the above point, the paper is acceptable, with some technical corrections listed below. The corrections are listed, in order of importance:

- The title and abstract should state clearly that this paper is a review (see general comment).

R: We have revised the title as "**Review Article**: Towards Strongly-coupled Ensemble Data Assimilation with Additional Improvements from Machine Learning"

- In some of the experiments listed the details of how the DA is performed the variables assimilated are missing.
  1. Section 2: The SC ETKF and SC 4D-Var do not specify which variables are assimilated, and the components towards they are assimilated

R: The analyzed variables are the full 9-element state vector of the Pena and Kalany [2004] model, and the synthetic observations are created for all these 9 elements.

We make the following addition (in bold) to L131 as:
"Singleton [2011] obtained the nature run by integrating the model using the 4th-order Runge-Kutta method with a time step $\Delta t = 0.01$. **The analyzed variables in the data assimilation experiments are the full 9-element state vector.** Observations are generated every 8 time-steps by adding to the true **9-variable** model states the uncorrelated Gaussian errors with zero mean and a standard deviation of $\sqrt{2}$."

  2. Section 4: experiments do not explain the variables assimilated.

R: The analyzed variables are the 36 nondimensionalized coefficients of spectral modes for the atmosphere (N_a=20) and ocean (N_o=16). Synthetic observations are created from the full true state vectors plus 10% natural variability. We add the following sentence to L238 as

"It also includes Ekman dynamics at the atmosphere–ocean interface and the simplified radiation parameterizations. **The analyzed variables are the 36 nondimensionalized coefficients of spectral modes for the atmosphere (N_a=20) and ocean (N_o=16).**"

We also added the following sentence to L242 as
"Figure 4 (a)-(b) compares the atmosphere and ocean analyses by 3D-Var under three CDA strategies. **Each experiment assimilates the synthetic observations of the full state vector.**"

3. Section 5: Missing which atmospheric variables are assimilated to atmosphere and ocean in both WCDA and SCDA.

R: The atmospheric observations in Section 5 include: surface pressure, vertical profile of temperature, humidity, zonal and meridional winds. We revise L312 as:

"… Synthetic atmosphere observations **(i.e., surface pressure, vertical profile of temperature, humidity, zonal and meridional winds)** are assimilated into the atmosphere in both experiments."

- **Figure 1**: has a very poor quality and labels do not correspond to the experiments described in the manuscript, also the caption is confusing.

R: We have revised the captions of Figure 1 as

"**Figure 1:** Time-averaged Analysis RMSE for SC 4D-Var (green), SC ETKF-QOL (red), SC ETKF with the "atmos-coupling" (Figure 2 of Yoshida and Kalnay [2018]) as the localization pattern (cyan) and its 4D extension (blue) for the extratropical atmosphere (top left), tropical atmosphere (top right), and ocean (bottom). Adapted from Singleton [2011]."

- In some results there is a confusion between high/low "accuracy" and low/large "error". This, in lines:
  1. ln 255 - 256: "... the SC 40-member ETKF and 4D-Var have similar accuracies for the atmosphere and ocean analyses, lower than SC 3D-Var." Change "lower" for "higher".

  R: corrected.

  2. ln 262: "...ETFK SC 4D-Var and CERA present similar analysis accuracies smaller than SC 3D-Var." This statement is wrong: the accuracy of 3D-Var is lower than for the other experiments.

  R: corrected.

  3. ln 471: "... EnKF and 4D-Var reach similar analysis accuracy smaller than 3D-Var." This statement is wrong: 3D-Var accuracy is smaller than EnKF and 4D-Var.

  R: corrected.

- **Figure 9:** The caption mentions lines that are not included in the figure. This is: the dashed lines for WCDA experiments, for the left panel. Besides, there is no indication on the meaning of the colors. Red=reduction, Blue=increase?

R: We have revised the caption of Figure 9 as:

"**Figure 9:** RMSD reduction of observation minus 6-hour forecast (O-F) for ocean temperature by switching from WC to SC CFSv2-LETKF. The left panel shows the spatially averaged RMSD change (improvements with positive value, and degradation with negative value) that varies with the ocean depth over the Northern Hemisphere (NH, blue) and tropics (TR, green). The right panel shows the spatial distribution of the RMSD by switching from WC to SC (improvements in blue, and degradation in red) at selected ocean depth. Adapted from Sluka et al. [2016] and Sluka [2018]."

- **ln 215:** There is not an explanation on what or which are the "quasi-SCDA" methods.

R: The quasi-SCDA method first appears at L65:

"The European Centre for Medium-Range Weather Forecasts (ECMWF) implemented a *Quasi-SCDA* system **through the "outer loop coupling", where the incremental 4D-Var atmospheric and 3D-Var with the First Guess at the Appropriate Time (3D-FGAT, Lee et al., [2004]; Lawless [2010]) oceanic analyses share the same outer loops so that their updated analyses will be used together to acquire the new model trajectory for the next round** [Laloyaux et al., 2016; 2018]."

Compared to the WCDA, the Quasi-SCDA method such as outer loop coupling still performs separate analysis update using their own domain observations for the atmosphere and ocean. However, since the atmosphere and ocean share the same outer loop, their updated analysis will be used to acquire the new model trajectory for the next outer loop. This can improve the atmosphere and ocean analyses through the dynamical coupling at the data assimilation step.

Based on the definition by Penny et al. [2017], which defines Quasi-Strongly Coupled DA as "observations are assimilated from a subset of components of the coupled system. The observations are permitted to influence other components during the analysis phase, but the coupled system is not necessarily treated as a single integrated system at all stages of the process", the outer loop coupling used by CERA belongs to the quasi-SCDA method.

We have revised L65 as

"The European Centre for Medium-Range Weather Forecasts (ECMWF) implemented the "outer loop coupling", where the incremental 4D-Var atmospheric and 3D-Var with the First Guess at the Appropriate Time (3D-FGAT, Lee et al., [2004]; Lawless [2010]) oceanic analyses share the same outer loops so that their updated analyses will be used together to acquire the new model trajectory for the next outer loop [Laloyaux et al., 2016; 2018]. **Though cross-domain observations are not directly assimilated into separate earth components, separate earth component analysis benefit from a more coherent coupled-state through the dynamical coupling at the data assimilation step. Based on Penny et al. [2017], outer loop coupling belongs to the *Quasi-SCDA* methods**."

- **ln 222 - 224:** They talk about an experiment that uses 6-hour forcing in Figure 4 (a)-(b), but these subfigures do not present such experiment.

R: Thanks for the comments. For the 3D-Var UCDA experiment, the most frequent forcing update is 1-day (experiment UC_1day), not 6-hour forcing experiment. we choose the slowest forcing update as 1 day to mimic the common approach adopted by operational centers where their uncoupled atmosphere model uses the daily SST products as the surface forcing.

We have revised L245 as:

"However, the analysis error with a **1-day** forcing update is still one order of magnitude greater …"

- **Figure 7:** The labels of the subfigures are wrong. Panels (c)-(d) do not exist.

R: We have revised the labels of the subfigures in Figure 7.

- **ln 338:** It is mentioned the result for WCDA experiment, but the figure does not show it, or it is not clear.

R: We have revised the caption of Figure 9. Now the improvement/degradation by switching from WC to SC CFSv2-LETKF are shown in Figure 9.

Technical corrections:

- Make sure all the captions have the same style for the citations.

R: We have revised the citation styles in the captions of Figures 1, 8, 9, 10, 12.

- **ln 63**: Citation for the 3D-FGAT method is needed.

R: 3D-Var with the First-Guess at the Appropriate Time (3D-FGAT, Lee et al., 2004; Lawless, 2010)
We revise L66 "…, where the incremental 4D-Var atmospheric and 3D-FGAT oceanic analyses …"

as

"…, where the incremental 4D-Var atmospheric and 3D-Var with the First Guess at the Appropriate Time (3D-FGAT, Lee et al., [2004]; Lawless [2010]) oceanic analyses…"
Reference:
Lee, M.-S. D. Barker, W. Huang, and Y.-H. Kuo: First guess at appropriate time (FGAT) with WRF 3DVAR. *WRF/MM5 Users Workshop*, pp. 22-25. 2004.

Lawless A. S.: A note on the analysis error associated with 3D-FGAT. *Q. J. R. Meteorol. Soc.* **136**: 1094–1098, doi: 10.1002/qj.619. 2010.

- **ln 104:** "... Kalnay [2004], of which equations are written as." Replace the period with a colon.

R: Corrected.

- **ln 123:** "…standard deviation of √2. Besides, Assimilation experiments…" Change capital to low-case in word 'Assimilation'.

R: Corrected.

- **ln 124:** "... experiments with the Ensemble Transform Kalman Filter (ETKF) in this section uses 9 members." Mismatch between subject and verb. Change "uses" to "use".

R: Corrected.

- **ln 177:** "..., and constant surface fluxes [$f_{X,i}$, $f_{Y,i}$, $f_{Z,i}$] that forces ..." Change "forces" to "force",

R: Corrected.

- **ln 178:** "... x_0^b represents the initial background states, ..." Change "states" for "state".

R: Corrected.

- **ln 183:** NMC is an acronym not defined.

R: NMC method refers to "National Meteorological Center" method (Parrish and Derber, 1992).

We have revised L184: "Where $\mathbf{B}_{x,0}$ is the background error covariance of the initial ocean states estimated by the NMC methods"

to

"Where $\mathbf{B}_{x,0}$ is the background error covariance of the initial ocean states estimated by the National Meteorological Center (NMC) method [Parrish and Derber, 1992]."

Reference:

Parrish, D. F, and Derber J.C.: The National Meteorological Center's spectral statistical interpolation analysis system. *Mon. Weather Rev.,* 120:1747-1763

- **ln 225:** "... update is still one order greater than... " Change to "... update is still one order of magnitude greater than... "

R: Corrected.

- **ln 287:** Acronym OSSE is not defined.

R: We revise L310 "A 6-year perfect model OSSE is then conducted …"

To

"A 6-year perfect model **Observation System Simulation Experiment (OSSE)** is then conducted…"

- **ln 313:** missing indent at beginning of paragraph.

R: Thanks for the comment. Since this sentence is the 1st sentence under Section 6, we do not indent. To make the style consistent through the whole context, we also remove the indentation before the 1st sentence under Section 5.