The manuscript addresses an important topic of parameter estimation in coupled atmosphere-ocean climate models. Performance of Ensemble Kalman Filter approach for joint estimation of model state and parameters is assessed in twin data assimilation experiments with a "conceptual climate model". The conceptual model is a system of five first-order ordinary differential equations describing evolution of three "atmospheric" and two "oceanic" state variables. The results of these experiments allow the authors to conclude that "*enhancing estimation accuracy of atmospheric states is very important for the success of coupled model parameter estimation, especially for the parameters in the air-sea interaction processes*". I believe that the test of the technique within the frames of oversimplified model and synthetic observations is a necessary first step in the development of robust and efficient methods for parameter estimation in the modern climate models. The results of such idealized study and the authors' experience in the design of the parameter estimation experiments could be of interest for data assimilation community and is worth publishing. On the other hand I cannot recommend publishing the manuscript in its present form for several reasons listed below.

Major comments

Comment 1. I found the manuscript to be difficult to read. It is not because of grammar errors as in the first version of the manuscript. Some parts of the revised manuscript lack clarity and logic in presenting the material. I had to read some sentences and paragraphs several times trying to understand what the authors intended to say. As an example I will cite here just one paragraph in section 2.3 (P3L16):

"*The different PE experiments can be distinguished in 3 perspectives: 1) 2 state constraint settings (i.e. SE settings) that assimilate the atmosphere "observations" (x2) only, and the ocean "observations" (w) only; 2) 2 parameter settings – a2 in the atmosphere equation, and c2 in the ocean equation; 3) 2 observational settings – one atmosphere (x2), and one ocean (w). Here the SE uses weakly coupled data assimilation as termed in the literature (e.g. Lu et al., 2015), i.e., x2 observations impact on all x variables, and w (η if applicable) observations impact on w (η) itself (considering the different time scales of w and η, no cross-impact between them), while the PE could use different medium observations. Therefore, eventually we have a few PE cases with full SE – both x and w are constrained by their observations, and particularly 8 PE cases with partial SE – only some medium is constrained by its observations. These PE cases have different SE accuracy. Through thoroughly analysing these PE cases, we are able to detect the influence of the SE accuracy in different medium on coupled model PE".*

In the first sentence of this paragraph it is difficult to see the difference between "perspectives" 1) and 3). The second sentence should probably be more appropriate in the section 2.2 describing the ensemble filter. This sentence implies essential modification (simplification) of the algorithm which probably deserves a more detailed discussion in the section 2.2.
The second sentence contradicts the first one since the first sentence mentions only "2 observational settings – one atmosphere (x2), and one ocean (w)" while the second sentence assumes assimilation of η observations "if applicable".
The third sentence ignores assimilation of η again. The reader has to guess what are the "few PE cases with full SE" and why there are exactly 8 PE cases with partial SE. If assimilation of η observations is disregarded, with "2 observational settings" and "2 parameter settings" the reader will not be able to get 8 PE cases with partial SE.

In my view, the authors should perform a significant work on "polishing" the manuscript to improve clarity of the material presentation.

Comment 2. As with any suboptimal data assimilation method, the results of state and parameter estimation presented in the manuscript depend on particular realization of the method. That is why it is important to provide a clear description of the method what is sufficiently detailed to allow other researches to repeat the calculations. I believe that the section 2.2 falls short of these expectations. It seems that at least the Reviewer 3 of the previous version of the manuscript did not get clear understanding of the method either. The Reviewer 3 wrote in his comments:

"*Eq. (2) does not contain observation error statistics, and I am curious how to interpret this equation intuitively. I understand that this equation gives analysis increments for the ith ensemble member. The analysis increments should balance between the observation error and background error. This equation has only the background error variance in the observation space as the denominator, but does not contain the observation error variance which usually appears in the data assimilation equations as an R matrix".*

In my view, the authors' reply to this comment is disappointing and the revisions to the section 2.2 did make the method description even vaguer. The authors' reply to the Reviewer 3 comment states:

"*The observation error variance is calculated before this projection process. The observation are firstly compared to their simulated values, the difference between them are manipulated to produce the observational increments. The production of the observational increments considers the observation error variance and its PDF. The observation error is set as a constant number in our simulation. The standard deviation of "observational" errors are 2 for the atmospheric variables x1,2,3 and 0.2 for the oceanic variable w. New introduction of the EAKF method is added in the section 2.2, P4L1~10".*

The following statements are at least inaccurate in this reply:

"*The observation error variance is calculated before this projection process*" – I believe that the data error variance is set, not computed, since the authors state: "*The standard deviation of "observational" errors are 2 for the atmospheric variables x1,2,3 and 0.2 for the oceanic variable w".*

"…The difference between" observations and their simulated values is *"manipulated to produce the observational increments".* What is this "manipulation"? There is no description of the procedure of this "manipulation" in the revised or original versions of the manuscript.

"*The observation error is set as a constant number in our simulation*" – is it really so? Isn't the observational error a random variable?

The revised version in the section 2.2 is also quite confusing. P4L4 states:

The method "*combines observational probability distribution function (PDF) with model PDF but under an adjustment idea. The algorithm can be sequentially implemented in a two-step procedure (Anderson, 2003): step 1 uses two Gaussian convolution to compute the observational increment at the observational location…*"

First of all, two convolutions of what? Conventionally convolution is an operation on two functions resulting in another function. Do the authors mean the convolution of two Gaussian PDFs? But the observational increment is not a function, so it cannot result from convolution of two PDFs.
Secondly, what is the "observational probability distribution function"? Do the authors mean PDF of data error? That is not the same as the PDF of observations. What is the "model PDF"? Model is not a random variable, it is an operator. Do the authors mean PDF of model forecast error? Do the authors know the PDF of the model forecast error to combine it with observational probability distribution function under an adjustment idea?

Trying to find the answer to the question of the Reviewer 3 I came to the following interpretation of formula (2) in the manuscript:
The authors seem to use well-known stochastic type of ensemble Kalman filter (see e.g. Hamil, 2006) where each ensemble member x_i (model state and parameters) is updated to fit somewhat different realization of observations y_i at the analysis time:
$\Delta x_i = K\,\Delta y_i$; $\Delta y_i = H(x_i) - y_i$, $i=1,\ldots,m$
where $y_i = y + e_i$, and $e_i \sim N(0,R)$ is a realization of Gaussian random variable "observational error" with variance (covariance) R. Realizations e_i are different for each ensemble member. The expression for the Kalman gain matrix K in formula (2) is an ensemble estimate for the case of a single observation of the conventional Kalman gain matrix given by the expression:
$K = PH^T (HPH^T+R)^{-1}$
For example, $std(\Delta y)^2$ in the denominator in formula (2) approximates HPH^T-R:
$std(\Delta y)^2 = <(\Delta y_i)^2> = < (H(x_i) - y_i)\,(H(x_i) - y_i)^T>$,
where < > denotes ensemble averaging. Under the assumptions that the model is unbiased and e and x are uncorrelated, the expression above can be transformed to
$<(Hx'_i - e_i)\,(Hx'_i - e_i)^T> = H<x'_i x'^T_i>H^T + <e_i e_i^T>$
while $<x'_i x'^T_i>$ and $<e_i e_i^T>$ tend to P and R respectively as the dimension of the ensemble increases.

If my guess is wrong and the the analysis for all ensemble members is performed using the same realization of observations, the data errors are effectively neglected by the filter given by equation (2).

If my guess regarding the meaning of the formula (2) is correct, the authors should describe the procedure of generation of the ensemble of observations in section 2.2. This is an essential part of the method (the method cannot use unperturbed observations). Unfortunately, the authors discuss the observations only in section 2.4 (P5L9):

"*The output of last 103 TUs is then used as the "truth" to produce "observations" by superimposing a white noise on the "observed" variables. The standard deviation of "observational" errors are 2 for the atmospheric variables x1,2,3 and 0.2 for the oceanic variable w.*"

Note that this statement does not specify that the statistics of observational noise is Gaussian. The term "white noise" probably assumes no time correlations in observational errors which is always taken for granted in applications of the Kalman filter. Also it is not clear if the authors "produce" only one realization of observations by superimposing a white noise and "truth" or they produce an ensemble of perturbed observations.

In my view, the authors should revise the description of the data assimilation method considerably. The present description of the Ensemble Kalman filter method misses some important details which are critical for understanding the data assimilation results.

Comment 3. A number of inconsistencies can be found also in the description of the data assimilation results. For example, the authors mention several times that only the values of x2, w, and "*(η if applicable)"* are perturbed with noise and used as observations. The formula (2) assumes assimilation of scalar (single) data at the analysis time with separate data assimilation for "atmospheric" and "oceanic" sub-models. Conversely, in the statement:

P5L9 "*The standard deviation of "observational" errors are 2 for the atmospheric variables x1,2,3 and 0.2 for the oceanic variable w."*

the authors specify standard deviations for observational errors in x1 and x3 as if these variables were also used as observations (and do not provide standard deviation for η). Caption to Figure 1 also declares assimilation of x1,2,3 data: "*a) both the atmosphere (x1,2,3) and ocean (w) from their observations (x1, 2, 3 and w)*".

Another inconsistency in the description of the results is related to the authors' discussion of the time scales of variability of "*fast-varying variables of the atmosphere*" (P3L13) and "*the low-frequency variables of the ocean"* (P3L13).
The authors state that "*the time scale of the w is nearly 10 times of the time scale of the x2*" (P3L13), and that the "*conceptual model mimics very fundamental natures of interactions of three typical time scales in the real world: synoptic (chaotic) atmosphere, seasonal interannual upper tropical oceans and decadal/multi-decadal deep ocean*". This time scale separation was used by the authors as an argument for utilization of a simplified "*weakly coupled data assimilation"* approach (P5L20):
"x2 observations impact on all x variables, and w (η if applicable) observations impact on w (η) itself (considering the different time scales of w and η, no cross-impact between them)".
The time scale separation is addressed several times in the interpretation of the results and is used as an explanation of the performance of different parameter estimation experiments, for example: "*The energy of x2 is in the high frequency band and the energy of w is in the low frequency band. x2 varies fast and represents the most uncertain mode, transferrable to low frequency w through the "air-sea" interaction"*(P8L19).

In the first reading of the manuscript, the description of the experiments and analysis of the results convinced me that the atmospheric variable x2 has a well determined spike of energy at periods of few TU, the variable w reacts to coupling with chaotic x2, but is mainly driven by the periodical forcing at the period of 10TU, while η has even longer scales of variability. The understanding that this impression is completely wrong came to me only when I read Appendix 2 there the authors describe an application of the **high-pass filter** to time series of η observations and states to remove "*chaotic signal*" and improve parameter estimation. The authors write:

P12L30 *"… the periodic signal produced by the cosine function has a period of 10TUs (1000 time steps) (defined by S_{pd} in Eq. (1), also see Fig. 10) and the* ***chaotic signal is much slower*** *than the periodic signal".*

Since "*chaotic signal*" comes only from atmospheric variables, I was puzzled how could it be "slow". I had to read the results of the experiments with one-way coupled model presented in the section 3.2 several times to find this out. The turning point was Figure 10 showing time evolution of η in the experiments where the amplitude of the periodical forcing was increased 100 to 1000 times compared to the fully coupled model experiments. This figure clearly shows that influence of stochastic atmosphere on oceanic variables is not confined to the periods of few TUs, but dominates the whole range of resolved periods in time series of η. Chaotic atmospheric signal clearly larger than periodic signal in Figure 10 for amplitude of periodical forcing Ss=100, while in the fully coupled model Ss =1. Also, the authors mention that:

P10L14: "*Comparing Fig. 9a to Fig. 6b, it can be seen that the chaotic signal in the one-way coupling model is much smaller than in the original two-way coupling model…"*

My suppositions are that:
a) for the experiments with fully coupled model (amplitude of the periodical force Ss=1) the ocean variables are driven almost entirely by the stochastic atmosphere while the periodical forcing in the equation for w seems to be negligible.
b) the influence of stochastic atmosphere on oceanic variables is not confined to the periods of few TUs, but dominates the whole range of resolved periods in the ocean.
c) time scale separation does not exist and cannot be used as a validation of a simplified "*weakly coupled data assimilation"* approach.
In the light of these suppositions, the authors' conclusions that:

"*enhancing estimation accuracy of atmospheric states is very important for the success of coupled model parameter estimation, especially for the parameters in the air-sea interaction processes*"
or
P8L9: *"This seems different from our previous intuition that in-situ ocean data are always considered as the first important piece of information for determining the oceanic coefficients. Our results here strongly suggest that in the future real coupled model PE experiments, for determining the best coefficient values, no matter the atmospheric or oceanic, sufficient and accurate atmospheric measurements are crucially important".*

looks rather as the consequence of the particular settings of the "conceptual model" parameters and utilization of "*weakly coupled data assimilation"* approach, than as a general feature of parameter estimation that can be extrapolated to realistic climate models.

Finally, note that the amplitude of the periodical force is stated to be 1 in the section 2.1:
P3L18: "(c_1, c_3, c_4, S_m, S_s, S_{pd}, Γ, c_5, c_6) are
(10^{-1}, 10^{-2}, 10^{-2}, 10, 1, 10, 10^2, 1, 10^{-3})"
while in section 3.2 the authors state that Ss=10:
P10L14: "*Comparing Fig. 9a to Fig. 6b, it can be seen that the chaotic signal in the one-way coupling model is much smaller than in the original two-way coupling model (with an identical S_s value of 10)*".

In my view, the authors should revise the description of the model and the sections presenting the results of data assimilation experiments considerably (and possibly redo the experiments to change the relative role of the periodical forcing). The present description may mislead the reader in the same way as it happened to me in my first reading of the manuscript.

Minor remarks

1. Some terms used in the manuscript seem to be misleading and/or not well defined.

These terms are:

a) "*signal-to-noise ratio of error covariance between the model state and parameter*". The term signal-to-noise ratio is commonly introduced for time series of state variables, but not for covariances. I agree that it can be used for time series of covariances but that may mislead the reader. Why the authors do not use R shown in Figures 7 and 8 to analyze "relationship between the states and the parameters" quantitatively (I believe R in Figure 7 and 8 is R2= 1 - ||Hx-y||2/||y-mean(y)||2 which is conventionally called the coefficient of determination of linear regression or Least Squares fit). It is possible to call R the time-mean "correlation coefficient". I do not see the reason to multiply R with Sf/Sp and call it the "signal-to-noise ratio", especially because it is not shown in the manuscript that rs2n is a better indicator of parameter estimation performance. Also, in more realistic applications, it might be impossible to assess the ratio Sf/Sp since the full set of observations is not available.

Separation of the signal and noise assumes that where is a model for noise or signal. Here this model is a linear relationship between state variable anomalies and parameter anomalies. I do not see why state variable anomalies and parameter anomalies should exhibit linear relationship at a particular time step. The authors' explanation of this linear relationship is a bit confusing:

P8L26: "*PE completely relies on the covariance between the parameter and model states for projecting the observational information of states onto the parameter. While the PE projection is carried out by a linear regression equation based on the state-parameter covariance (EnKF/EAKF, for instance), only a linear or quasilinear* *relationship between parameters and states in ensemble is recognized. A hypothesis for all the failed cases without direct atmospheric SE could be that, under such a circumstance, the chaotic disturbances in the atmosphere (Lorenz equations in this case) continuously interacting with the parameter make difficulties for the system to build up a quasi-linear relationship between the state variable and the parameter."*

I agree completely with the first sentence. The second sentence is misleading because the results of linear regression shown in Figures 7 and 8 do not correspond to formula (2). But I agree that "*only a linear or quasilinear relationship between parameters and states in ensemble is recognized"*. I do not like the last sentence since no "interaction" between parameters and state is possible. Model states do not act on parameters. Instead of the last sentence I would suggest the following explanation:

PE completely relies on the covariance between the parameter and model states for projecting the observational information of states onto the parameter. The ensemble gets the information on parameter-state covariance entirely from dynamics (not from data): ensemble members obtained in the model forecast with different values of parameters become different with time. The ensemble of states accumulates this covariance information gradually, over time periods which are significantly longer than the data assimilation window of 0.05TUs. Partial assimilation of ocean data effectively restarts the model at every analysis step since oceanic variables are driven by x2 which is not controlled by observations (see Comment 3). As a result, almost no "dynamical" information is accumulated in ensemble covariance. Alternatively, in full data assimilation or in assimilation of x2 data, state corrections at the analysis step are more gentle and some dynamical information on state-parameter covariance has a chance to accumulate with time. And it is not that important if the relationship between parameters and states at a given time step is linear, non-linear or in general there is no relationship at all. It is important that ensemble estimates properly dynamically induced state-parameter covariance.

b) *"state-parameter covariance uncertainties"*- Use of this term assumes that covariance is a random variable that has a "true", mean, or the most probable value. Then "uncertainties" are assessed as some measure of the higher moments of this random variable. This is not discussed in the manuscript.

c) *"chaotic-to-periodic ratio"*- The term itself is fine, but no expression to compute this ratio is given. No numerical values of this ratio for different experiments are presented in the manuscript. It is only mentioned how the authors change this ratio:

P10L11: "*Then we define a chaotic-to-periodic ratio (CPR) in the signals of w (η) by manipulating the coefficient S_s. Eight experiments are performed here, four for w-to-c_2 PE and four for η-to-c_6 PE. Each experiment has a different S_s value of 100, 250, 500 and 1000 and thus a reducing CPR in w and η*".

It is worth to note that by changing the amplitude of Ss the authors also change RMS variability of w and η but do not change variance of corresponding observational error. The observations may become much more "accurate" with increase of Ss by the factor of 1000.

Due to vague definition of *signal-to-noise ratio* and *chaotic-to-periodic ratio* the authors mix these two terms. In the Appendix 2 *signal-to-noise ratio* is used instead of *chaotic-to-periodic ratio*.

Also please correct the typo: P12L25 "signal-to-ratio ratio".

3. P3 Eq(1): one of the coefficients Om, Gamma, or Od is redundant.

4. P4L26: "*Thus a practical ensemble size of 20 (applicable for a CGCM) is chosen as a basic experiment setting*". No comparison with CGCM is possible. In the presented research the ensemble size exceeds the dimensions of both the state vector and data. This is not the case for CGCM.

5. P4L30: “as 0.05 TU (i.e. 5 time steps)”. Notation “TU” is not defined on page 4. It is introduced only on page 5.

6. P5L31: “*The ensemble initial values of c_2 are set as N(0.8, 0.5)*”. What if a particular realization of c2 is negative? The authors state that c2 is “analogous to the drag coefficient c_d” which must be positive.

7. P7L20: “*We also performed the experiments under different update interval settings. Test results show that for the issue we are addressing, the conclusion is not sensitive to the update interval if it is within a reasonable range*.” The range of the tested update intervals should be given here.