Response to anonymous reviewer comments

Dear Reviewer,

We first thank you for your positive and constructive comments. Below we attempt to address the your comments through a more detailed discussion. We will also make adjustments to the manuscript where necessary.

Many thanks,

Courtney Quinn, Terence O'Kane, and Vassili Kitsios

General comments

This manuscript is exploring the use of covariant Lyapunov vectors (CLVs) to build the error covariance matrix in ensemble Kalman filtring methods. The set of vectors is selected based on the computation of a local Kaplan-Yorke dimension based on the finite time Lyapunov exponents. This approach is implemented in the context of a multi-scale system mimicking the (coupled) dynamics of a coupled tropical ocean-atmosphere system and the extra-tropical atmosphere. Different strategies of observation are then evaluated. It is found that observation within the atmosphere is essential,and that the variable number of CLVs to be used in building the error covariance matrix is a successful strategy for strongly coupled data assimilation. Very interesting results are also obtained with the observation sampling of a shadowing trajectory, leading to measurement correlations.

This is an interesting manuscript exploring many aspects of the strongly coupled data assimilation and I would in principle recommend publication of this work.

We thank the reviewer for the positive evaluation.

I have however an important concern on the use of the local Kaplan Yorke dimension and the CLVs that should be addressed before publication. It seems to me that the use of both is inconsistent. Let me clarify my point. When computing the FTLEs, one can use either the QR (associated with the backward Lyapunov vectors) decomposition,the Forward Lyapunov vectors obtained with backward integration in time, or the local amplification along the CLVs. Although all these are giving the appropriate asymptotic Lyapunov exponents, they are not providing similar variability of these quantities as illustrated for instance in Vannitsem and Lucarini (2016) you quoted. So if you use the QR decomposition and then select the CLVs on that basis this is probably not optimal.

We have considered the reviewer's concern and are in agreement that the use of the two quantities could produce inconsistencies. To clarify, we use the QR decomposition method to calculate the FTLEs, which would give amplification rates for the backward Lyapunov vectors rather than the CLVs. However, in this case we argue that the way in which we utilise the FTLEs does not create an inconsistency with the CLVs in that we are not assigning any particular FTLE to a CLV. Rather, the backward FTLEs are used only to produce an adequate estimate of the number of CLVs to retain in constructing the covariance matrix. The true local dimension of the system may be different from the quantity computed, but the Kaplan-Yorke dimension calculated from the backward FTLEs still produces a sufficient upper bound. We expand on this point further below.

Figure 1 shows the finite-time Kaplan-Yorke dimension of a segment of the example model run from Sections 2 and 3 in the manuscript (compare to Figure 5 in manuscript). The top panel is the dimension measure, while the bottom is the ceiling of the dimension measure which is used to select the number of CLVs for constructing the covariance matrix. We compare the two different methods for calculating the CLVs that the reviewer mentions: growth rate along CLVs (FTCLEs) and QR decomposition (FTBLEs). We observe that for our particular nonlinear system, the QR decomposition method consistently gives a larger estimate for the Kaplan-Yorke measure than the amplification along CLVs, and therefore implies retaining more CLVs in the analysis. We maintain that a higher number of retained CLVs will not harm the DA, while too few CLVs potentially could. Specifically, where the positive CLVs are closely aligned, the DA requires the inclusion of most of the additional neutral and stable CLVs to avoid ensemble collapse (further discussion of this point on the following page). Thus, Kaplan-Yorke is an effective choice to determine the rank of the background covariance matrix despite the any inaccuracies due to degeneracy in the unstable directions. Additionally, we mention that in order to obtain the growth rate of the CLVs, one needs the accurate forward model. Since we are calculating the CLVs from the ensemble mean trajectory within the DA experiments, we do not have an accurate future trajectory and therefore no forward model (due to the state dependence of the Jacobian). For these reasons we maintain that the FTLEs from the QR method are a sufficient for our analysis at hand.

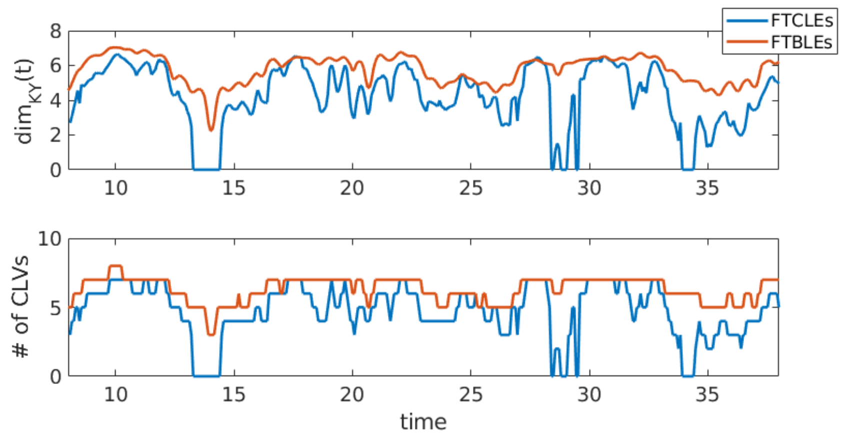

Figure 1: Comparison of finite-time Kaplan-Yorke dimension calculated using the growth rates of CLVs (FTCLEs) and the QR method (FTBLEs).

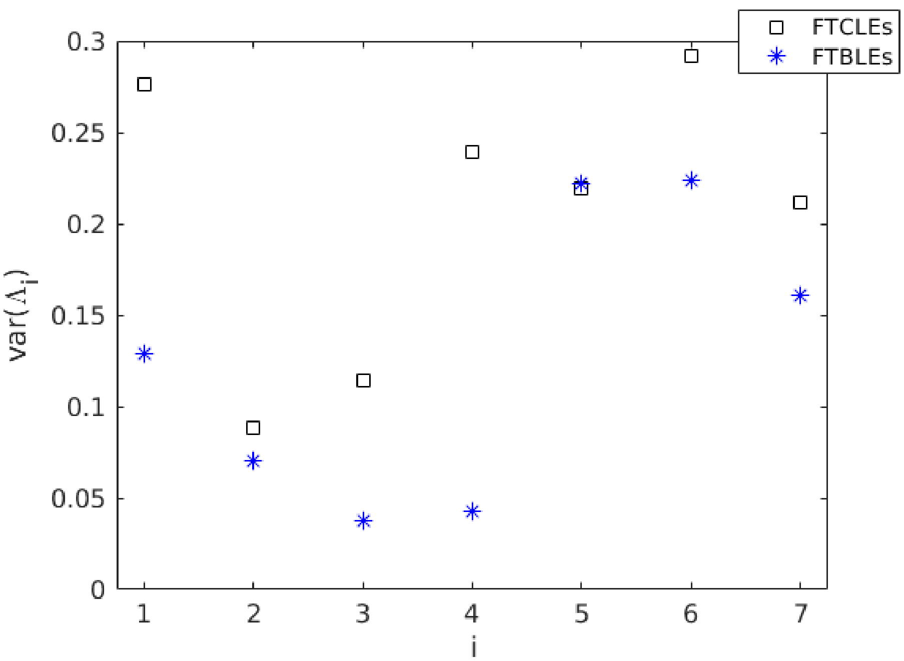

Figure 2: Comparison of variance of the individual FTLEs calculated using the growth rates of CLVs (FTCLEs) and the QR method (FTBLEs).

Alternatively, if you use an estimate of the FTLE using the amplification along the CLVs then the "dimension" of the subspace of instabilities is not the same and one can wonder what is the signification of this quantity. This is related to your comment at line 14 of page 8 indicating that higher dimension is associated with more important alignment of CLVs. Imagine for instance that several CLVs are pointing in almost the same direction with large amplifications, then the dimension would be large but intuitively (as they point all in the same direction) we would expect a low dimension. The specific way you compute the FTLEs and the local dimension should therefore be clarified, and probably I would not call it "dimension". Furthermore, if the KY dimension is computed with the local amplification rates of the backward Lyapunov exponents, a comparison should be made with the use of the backward Lyapunov vectors in building the error covariance matrix. It would have been my first choice in view of the fact this is much less costly than the CLVs.

The reviewer is correct in that the finite-time Kaplan-Yorke measure is not necessarily the true local dimension. However, this measure is giving an indication of local dimension combined with alignment. We explain this through an examination of two difference time steps in the benchmark DA experiment (Section 5.2 in the manuscript) where a high Kaplan-Yorke measure is recorded and the FTLE behaviour is similar. The alignment of the CLVs and the FTLEs (computed using the QR decomposition method) are shown in Figure 3. In both cases we see that there are 2 equally strong leading unstable FTLEs, 3 positive but near zero FTLEs, and one weakly stable FTLE. In the first column ($t = 306.24$) we see that the 5 leading CLVs are strongly aligned as well as the two most stable (8 and 9). The dimension based on alignment would then intuitively be around 4, and one would need to select the set of CLVs that are not aligned. In our method since $\dim_{KY} > 7$ we would retain 8 CLVs, therefore accounting for the strong alignment of the leading CLVs and retaining all necessary directions. In the second column ($t = 705.04$) the case is quite different in that the leading CLVs are not strongly aligned, but there is strong alignment of CLVs 4-6 and pairs 3,7 and 8,9. This would give an alignment-based dimension around 5, but again we need to retain up to the 8th CLV to span the local manifold. While one could create a method based on alignment for selecting directions, we point out that the actual criteria for "strong alignment" is arbitrary and one could risk excluding a significant direction.

With regards to using the backwards Lyapunov vectors (BLVs), if one is using the QR method then the backwards Lyapunov vectors computed are orthogonal by construction. We therefore do not see the same behaviour of alignment of the vectors. We ran the benchmark DA experiment (Section 5.2 in the manuscript) using a variable number of BLVs based on the Kaplan-Yorke measure computed using the QR decomposition method (statistics shown in Table 2 of next section). We see that the variable BLVs and variable CLVs perform comparably. While all experiments could be run with the BLVs as suggested by the reviewer, we aim to show the functionality of using the CLVs as they provide additional information about the phase-space dynamics which can be

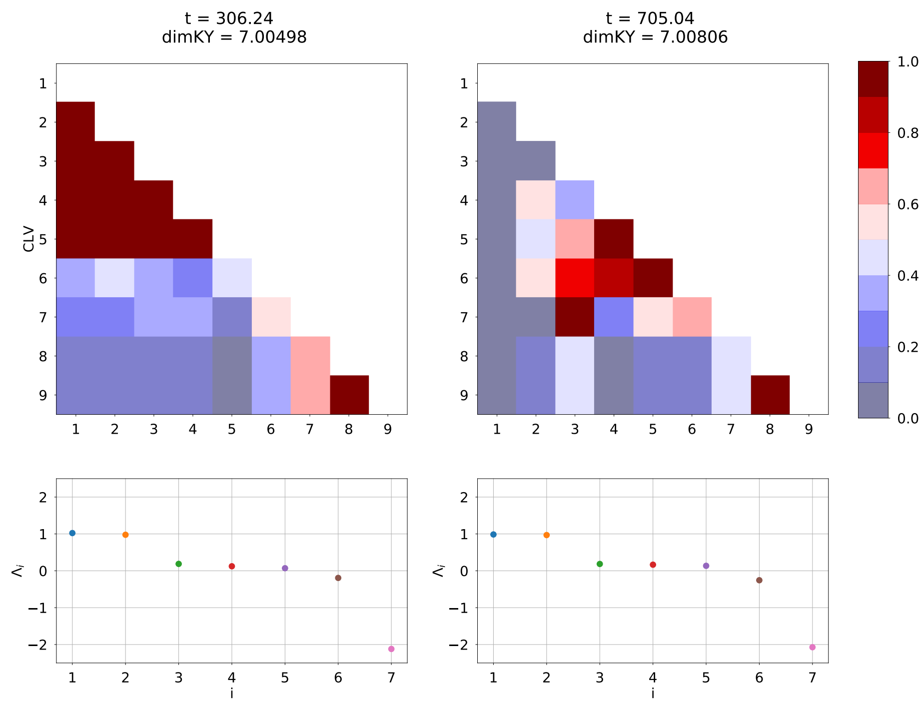

Figure 3: We compare the alignment of the CLVs at two time steps of the benchmark DA experiment (section 5.2 in the manuscript). The two time steps have similar Kaplan-Yorke measure and distribution of leading FTBLEs, shown in the bottom panels. We observe that the behaviour of the alignment can be vastly different for similar FTBLE behaviour, however the method of retaining CLVs based on the Kaplan-Yorke measure gives a reliable way to reflect the true local dimension regardless of alignment.

analysed *a posteriori*.

We have made the following additions to the manuscript to address all the above concerns and clarify our methods:

- We have added a paragraph and figure to Section 3 of the manuscript to discuss the differences in calculating the FTLEs.
- We have added Figure 3 along with a discussion of alignment, dimension, and relevance to ensemble DA in Section 4.2.
- We have added the statistics for using BLVs with variable Kaplan-Yorke measure in the benchmark DA case of Section 5.2.

Specific comments:

1. Figure 3, you mentioned 2 neutral Lyapunov exponents. I am wondering why you have two such exponents. Is there any specific symmetries allowing for that? Isn't it a numerical artifact?

   For clarification, the model has a neutral and a near-neutral Lyapunov exponent. The near-neutral exponent comes from the weak coupling to the extratropical subsystem ($c_e = 0.08$). If one were to set $c_e = 0$, the extratropical subsystem would be completely uncoupled from the tropical and ocean subsystems, and it would retain its symmetry $[x_e, y_e, z_e] = [-x_e, -y_e, z_e]$. This implies an additional neutral Lyapunov exponent. However, we retain that the near-neutral exponent is important to consider within the neutral subspace in this study as the timescales of the neutral and near-neutral exponents are indistinguishable and changes in both can significantly affect the local Kaplan-Yorke measure. To improve and assess numerical precision, we have computed the Lyapunov exponents over a longer window (500,000 time steps) and show the results for the original parameter values as well as different combinations of coupling strengths set to zero in Table 1.

	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	λ_9
as in manuscript	0.9043	0.3052	0.0007	-0.0032	-0.4829	-0.8008	-1.8149	-12.2359	-14.5726
$c_e = 0$	0.9083	0.3029	0.0001	-0.0006	-0.4814	-0.7962	-1.8172	-12.2415	-14.5744
$c = 0$	0.9042	0.3491	0.0597	-0.0002	-0.0151	-0.3186	-1.6222	-13.4793	-14.5777
$c_z = 0$	0.9081	-0.0004	-0.0723	-0.0728	-0.1283	-0.1289	-1.1599	-13.4702	-14.5753
$c = c_z = 0$	0.9069	0.8886	0.0902	0.0001	-0.0004	-0.0741	-1.4569	-14.4801	-14.5743
$c_e = c = 0$	0.9083	0.3581	0.0692	-0.0006	-0.0007	-0.3227	-1.6360	-13.5012	-14.5744
$c_e = c = c_z = 0$	0.9083	0.9083	0.0902	0.0001	-0.0006	-0.0006	-1.4569	-14.5744	-14.5744

Table 1: Asymptotic Lyapunov exponents of coupled Lorenz model for different coupling coefficients set to 0. Lyapunov exponents computed over 5000 time units using a QR decomposition method, time step of 0.01, and orthogonalization step of 0.25.

2. Table 2. The RMSE for the extratropics are very close to each other whatever the experiments. Are the differences significant?

   We preface this response with pointing out that the statistics of Table 2 have slightly changed on fixing a numerical error in the code (see following section for revised statistics). In order to discuss significance of differences in the average RMSE we have to compute error bounds on the means. To do this, we perform a bootstrapping of the time series of extratropical RMSE for each of the five experiments listed in Table 2 of the next section. We resample the data (with replacement) 10,000 times

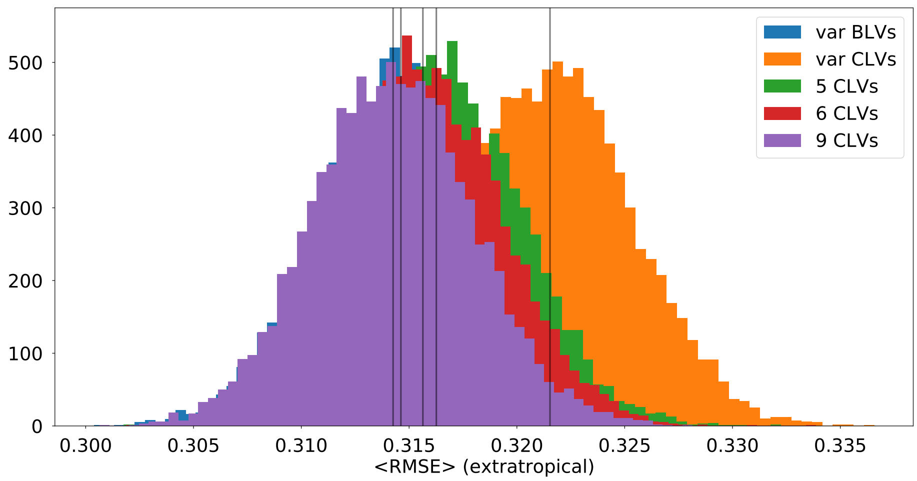

Figure 4: Histograms for the means of the resampled time-dependent extratropical RMSEs for each respective experiment. The black lines show the means of the histograms, which are approximately the values of extratropical average RMSE in Table 2 of the manuscript.

and then compute the average RMSE for each resample. This produces a distribution of average RMSE for each experiment, shown in Figure 4. Each distribution has standard deviation $\sigma \approx 0.004$, and we see that the mean of the distributions shift by less than σ between experiments with the exception of the variable CLVs experiment. The mean of the variable CLVs experiment still lies within 2σ of all the experiment distributions. In this sense we can conclude that the values listed in Table 2 are not significantly different. Additionally, we note that in the sense of quantifying the DA performance, all five methods constrain the extratropical system to a similar degree.

3. Also in Table 2. An average dimension is computed. This average is based on the QR decomposition or some estimate with the CLVs? This is related to my main point. Please clarify how it is computed.

   The average dimension comes from the FTLEs which are calculated using the QR decomposition. We have clarified this in the manuscript.

4. FTLEs are computed for 4 time units. What is happening to your analysis when this window is changed? And why choosing this specific window?

   The window of 4 time units was chosen in order to capture dimension changes related to the ENSO-like excursions in the ocean subsystem (occurring anywhere between approximately 6.5 and 27 time units). The window should therefore be long enough to filter out the intrinsic oscillatory behaviour of the ocean subsystem but short enough to capture the transition to the excursion state. The period of intrinsic variability is approximately 3 time units, therefore any window between 3 and 6.5 time units should perform comparably. Decreasing and increasing the window leads to higher and lower temporal variability of the FTLEs, respectively.

5. At page 12, in the two first paragraphs of Section 5, you present how the experiment is done. As far as I understood, the CLVs are computed during a limited period of time during the assimilation period. Am I right? At first reading it was not very clear to me and it would be nice to improve the presentation of that part. In particular, a sketch of the whole process in a figure would be really useful.

   We have clarified the paragraph where we explain the computation of the CLVs within the DA framework.

6. In the partially observed CDA, you also compute a local dimension. I am wondering how the CLVs are computed there since the trajectory of the model is probably very far from reality. Moreover, it was not clear to me whether you are using the CLVs of the reality or of the model integration. Would you please clarify how you do this? It can maybe be incorporated in the general description of the experiments mentioned at point 5 above.

   The CLVs are calculated from the ensemble mean trajectory. In the cases where the full system is constrained (atmosphere-only observations), the ensemble mean remains close enough to the truth such that the CLVs can still be calculated to a sufficient accuracy. On the other hand, this is not the case when some of the subsystems are unconstrained (i.e. ENSO and extratropical-only observations). In such cases we cannot use the variable dimension method, and instead we use all 9 computed CLVs. This is valid since the forecast error covariance matrix has rank 9 in this set-up and we are projecting onto a basis of equivalent rank, regardless of the accuracy of the computed CLVs. The local dimension we are calculating comes from the QR decomposition as mentioned in the response to the main comment. While the CLVs (or BLVs) may be quite inaccurate at a given point in time, the average dimension can give an idea of variance in the unobserved subsystems. In cases where an unobserved subsystem is unconstrained, it is common to observe variance collapse in the ensemble mean of that subsystem and we subsequently record a lower average local dimension. We therefore explore methods to maintain the variance in the unobserved subsystems (i.e. increase the average local dimension), with the caveat that the instantaneous dimension will not necessarily be reflective of the true dynamics of the system. We have provided additional clarification in the manuscript about the calculation of the dynamical properties within the DA experiments.

Minor points:

1. Line 7, page 11. Please modify the notation on the brackets. It looks like a vertical rectangle.

   We have changed the text to read "…and the **subscript** $*, n$ denotes taking …".

2. Line 20, page 11. I suppose that λ should be larger than 1.

   We have added a condition to Equation 19 specifying $\lambda > 1$.

3. Line 22, page 11. Please do not use the terminology "model error". It is confusing as model error is used when there is structural uncertainty in the model.

   We have removed the term "model" as we agree that the usage is confusing in this instance.

Minor corrections to DA experiments

After submission we discovered a small error in the orthogonalisation step of the CLV calculation within the DA experiments. In correcting we found that using the CLV without orthogonalisation to the backwards subspaces (Algorithm 2.1 in [1]) was more appropriate due to the lack of an accurate forward model within the DA implementation. This has minor impact on the results of the experiments, however we include the tables with the updated statistics below which we will also update in the new manuscript.

References

[1] Gary Froyland, Thorsten Hüls, Gary P Morriss, and Thomas M Watson. Computing covariant lyapunov vectors, oseledets vectors, and dichotomy projectors: A comparative numerical study. *Physica D: Nonlinear Phenomena*, 247(1):18–39, 2013.

[2] Takuma Yoshida and Eugenia Kalnay. Correlation-cutoff method for covariance localization in strongly coupled data assimilation. *Monthly Weather Review*, 146(9):2881–2889, 2018.

Method	Observations [error variance]	⟨RMSE⟩ extratropical	⟨RMSE⟩ tropical	⟨RMSE⟩ ocean	⟨RMSE⟩ full	⟨$\dim_{KY}$⟩
CLVs - 9 (full rank)	y_e, y_t, Y [1, 1, 25]	0.3142	0.1598	0.4948	0.4027	5.8928
CLVs - 5 (unstable/neutral subspace + 1)	y_e, y_t, Y [1, 1, 25]	0.3123	0.1843	0.5920	0.4550	5.8870
CLVs - 6 (global dimension)	y_e, y_t, Y [1, 1, 25]	0.3156	0.1674	0.5513	0.4310	5.8892
CLVs - variable (local dimension)	y_e, y_t, Y [1, 1, 25]	0.3215	0.1688	0.5346	0.4272	5.8863
BLVs - variable (local dimension)	y_e, y_t, Y [1, 1, 25]	0.3149	0.1658	0.5122	0.4141	5.8895

Table 2: Summary metrics of DA experiments using right-transform matrix (ETKF) and benchmark observations (y_e, y_t, Y). The angle brackets $\langle\cdot\rangle$ denote average over analysis steps. Compare to results in [2]. Parameters: analysis window 0.08, inflation factor 1%, 10 ensemble members.

Method	Observations [error variance]	⟨RMSE⟩ extratropical	⟨RMSE⟩ tropical	⟨RMSE⟩ ocean	⟨RMSE⟩ full	⟨$\dim_{KY}$⟩
CLVs - 9	y_e, z_e, y_t, z_t [1, 1, 1, 1]	0.1734	0.1332	0.5782	0.4515	5.8672
CLVs - 6	y_e, z_e, y_t, z_t [1, 1, 1, 1]	0.1715	0.1386	0.5807	0.4524	5.8718
CLVs - variable	y_e, z_e, y_t, z_t [1, 1, 1, 1]	0.1697	0.1383	0.5659	0.4433	5.8681

Table 3: Summary metrics of DA experiments using right-transform matrix (ETKF) and atmosphere observations (y_e, z_e, y_t, z_t). Parameters: analysis window 0.08, inflation factor 1%, 10 ensemble members.

Method	Observations [error variance]	⟨RMSE⟩ extratropical	⟨RMSE⟩ tropical	⟨RMSE⟩ ocean	⟨RMSE⟩ full	⟨$\dim_{KY}$⟩
CLVs - 9	y_t, z_t, Y, Z [1, 1, 25, 25]	7.0467	0.1571	0.3675	4.7182	4.5449
CLVs - 9	y_t, z_t, Y, Z [0.01, 0.01, 25, 25]	5.5203	0.0480	0.1764	3.6860	5.4052
CLVs - 9	y_t, z_t, Y, Z [1, 1, 0.25, 0.25]	7.1308	0.1386	0.0873	4.7566	4.4959
CLVs - 9	y_t, z_t, Y, Z [0.01, 0.01, 0.25, 0.25]	5.1697	0.0431	0.0822	3.42478	5.4899

Table 4: Summary metrics of DA experiments using right-transform matrix (ETKF) and ENSO observations (y_t, z_t, Y, Z). Parameters: analysis window 0.08, inflation factor 1%, 10 ensemble members.

Method	Observations [error variance]	⟨RMSE⟩ extratropical	⟨RMSE⟩ tropical	⟨RMSE⟩ ocean	⟨RMSE⟩ full	$\langle \dim_{KY} \rangle$
CLVs - 9 3% inflation	$\tilde{y}_e, \tilde{y}_t, \tilde{Y}$ $[1, 1, 25]$	0.5200	0.1635	0.4255	0.4552	5.9576
CLVs - variable 3% inflation	$\tilde{y}_e, \tilde{y}_t, \tilde{Y}$ $[1, 1, 25]$	0.4848	0.1568	0.4095	0.4263	5.9371
CLVs - 9 3% inflation	$\tilde{y}_e, \tilde{z}_e, \tilde{y}_t, \tilde{z}_t$ $[1, 1, 1, 1]$	0.4258	0.1342	0.4300	0.4687	5.9237
CLVs - variable 3% inflation	$\tilde{y}_e, \tilde{z}_e, \tilde{y}_t, \tilde{z}_t$ $[1, 1, 1, 1]$	0.4337	0.1373	0.4027	0.4649	5.9191
CLVs - 9 4% inflation	$\tilde{y}_t, \tilde{z}_t, \tilde{Y}, \tilde{Z}$ $[1, 1, 25, 25]$	6.8613	0.1431	0.3025	4.5961	4.5309
CLVs - variable 4% inflation	$\tilde{y}_t, \tilde{z}_t, \tilde{Y}, \tilde{Z}$ $[1, 1, 25, 25]$	6.8722	0.1463	0.3370	4.6054	4.5541

Table 5: Summary metrics of DA experiments using right-transform matrix (ETKF) and shadowed trajectory as observations. We set the observation error covariances to the standard values as in [2]. Parameters: analysis window 0.08, 11 ensemble members, inflation as noted in table.

Method	Observations [error variance]	⟨RMSE⟩ extratropical	⟨RMSE⟩ tropical	⟨RMSE⟩ ocean	⟨RMSE⟩ full	$\langle \dim_{KY} \rangle$
CLVs - 9	x_e, y_e, z_e $[1, 1, 1]$	0.0640	8.4752	36.3662	21.7108	4.1332
CLVs - 9 adaptive gain	x_e, y_e, z_e $[1, 1, 1]$	0.0032	0.7241	3.5757	2.1504	5.9991

Table 6: Summary metrics of DA experiments using left-transform matrix (ESRF) and the full extratropical subsystem as observations (x_e, y_e, z_e). We use perfect observations (no random error added to the control run) with the observation error covariances set to the standard values as in [2]. Parameters: analysis window 0.02, inflation factor 1%, 10 ensemble members.