Norihiro Miwa[1] and Yohei Sawada[1]

[1] Department of Civil Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

Corresponding author: Yohei Sawada (yohei.sawada@sogo.t.u-tokyo.ac.jp)

Key Points:

- Weakly and strongly coupled data assimilation are compared by applying Local Ensemble Transform Kalman Filter to the joint-Lorenz96 model.
- Strongly coupled data assimilation performs better when cross-domain interaction is large, but it does not increase system's chaoticity.
- Strongly coupled data assimilation is vulnerable to both cross-domain and intra-domain biases.

Abstract

Data assimilation methods in a coupled system, namely coupled data assimilation (CDA), have been attracting researchers' interests to improve Earth system modeling. The CDA methods are classified into two; weakly coupled data assimilation (wCDA), which considers cross-domain interaction only in a model's forecast phase, and strongly coupled data assimilation (sCDA), which additionally uses other domain's information in an analysis phase. Although sCDA can theoretically provide better estimates than wCDA since sCDA fully uses a cross-domain covariance, the effectiveness of sCDA is still in debate. In this paper, we investigated the conditions under which sCDA is effective by applying Local Ensemble Transform Kalman Filter (LETKF) to the joint-Lorenz96 model. By continuously changing the magnitude of the cross-domain interaction of the joint-Lorenz96 model, we found that the superiority of sCDA against wCDA is particularly evident when the cross-domain interaction is large, albeit it does not contribute to increasing the chaoticity of the system. In addition, the performance of sCDA is quite sensitive to the LETKF's hyperparameters (such as localization and inflation parameters) especially when the ensemble size is small, and the insufficient calibration of these parameters deteriorate the sCDA's performance. Furthermore, sCDA is more vulnerable to model bias than wCDA; both cross-domain and intra-domain biases degrade the estimation skills.

Plain Language Summary

Data assimilation (DA) integrates computer simulation and observation to improve the estimate of system states. Since Earth is a coupled system in which different domains such as atmosphere and ocean interact with each other, DA methods in coupled system, namely coupled data assimilation (CDA) have been attracting researchers' interests in Earth system sciences. The CDA methods are classified into two; weakly coupled DA (wCDA) and strongly coupled DA (sCDA). sCDA has a potential as it uses cross-domain interaction information additionally, but previous studies have mixed opinions regarding its effectiveness. Thus, we aimed to reveal under which conditions sCDA is effective. Especially, we focused on the domain interaction intensity, which has been an overlooked perspective in the previous studies. We found that the superiority of sCDA against wCDA depends heavily on the cross-domain interaction. When the cross-domain interaction is large, but it does not increase the chaoticity of the system, sCDA most efficiently worked. Furthermore, sCDA is more vulnerable to model bias than wCDA The findings are important to choose appropriate CDA strategies for general coupled systems in earth system science.

1. Introduction

It has become increasingly important to predict a system of systems (coupled systems, namely systems with two or more domains). Earth system models typically solve such a system of systems and understand the interactions of many different components of Earth such as atmosphere, land, and ocean (Taylor et al., 2012). Such coupled systems are not limited to the earth system models; weather-wildfire coupled system (Bakhshaii & Johnson, 2019), disaster-evacuation coupled system (Liu & Lim, 2018), socio-hydrology (Di Baldassarre et al., 2013; Jia et al., 2021), and socio-ecology (Sun & Hilker, 2021) are other examples. As these coupled systems are strongly non-linear, the development of efficient and practical methods of estimating, predicting, and controlling such systems is the grand challenge for earth system scientists.

Data assimilation (DA) has been recognized as a useful method to estimate the states of these complex coupled systems. It can provide the real-time estimates of state variables by combining model forecast and observation. DA can effectively maximize the potential of the sparsely distributed observations by integrating them into dynamic models, so that it is suitable for earth system models. DA methods are mainly classified into two: ensemble-based methods such as Ensemble Kalman Filter (EnKF; Evensen, 1994) and variational methods such as 4D-Var (Ravier et al., 2000). These two classes of DA methods have been intensively compared (e.g., Kalnay et al., 2007), and many studies proposed combining them (e.g., Zhang et al., 2009). DA methods in coupled systems are called "coupled data assimilation (CDA)". As it is quite challenging to develop adjoint models of the coupled systems, which are required for 4D-Var, ensemble-based methods are preferred in CDA. The CDA methods are classified into two; weakly coupled data assimilation (wCDA) and strongly coupled data assimilation (sCDA). The former considers cross-domain interaction only in the model's forecast phase, while the latter additionally uses other domain's information in the analysis phase.

Although sCDA recently attracts researchers' interests since it is expected to maximize the potential of observation using a cross-domain covariance, the application of sCDA is mainly limited to simplified models or idealized experiments. For example, Sluka et al., (2016) used atmospheric observations to update the state variables of a simplified ocean model in their observation system simulation experiment (OSSE). Suzuki et al. (2017) investigated the covariance between near land-surface temperature and snow temperature using a coupled atmosphere-land model. Sawada et al. (2018) assimilated river flow observations into an atmospheric model using a coupled atmosphere-river model. Note that they performed an OSSE, in which they did not use real observations. These previous studies assimilated observations in one model domain into the other model domain and they did not do the other way around (we called this type of one-way coupling in sCDA quasi-strongly coupled data assimilation in this paper). Thus, sCDA systems proposed previously have not realized the fully strongly coupled data assimilation in the multiple model domains. In addition, Lin & Pu (2020)conducted sCDA in a coupled atmosphere-land model and showed that assimilating 2-m humidity data into soil moisture particularly contributed to improve forecast. Their study is one of the few studies which assimilated real observation data in the sCDA scheme. Although sCDA attracts researchers in various disciplines, the application of sCDA to various types of realistic coupled models is rather limited.

In spite of the sCDA's application to various coupled systems, its effectiveness is still in debate. In previous works, sCDA is not necessarily more effective than wCDA. If cross-domain covariance cannot be adequately estimated (it happens especially when a small ensemble size causes spurious correlation and/or process models have large bias), sCDA degrades the skill to estimate the states of a coupled system. The previous studies have extensively investigated whether sCDA is superior to wCDA but their results are mixed. For example, Ballabrera-Poy et al. (2009) tested sCDA by applying Ensemble Kalman Filter (EnKF) to the multiscale Lorenz96 (Lorenz, 1995; Lorenz & Emanuel, 1998) model in which 8 slower domains have 32 faster sub-domains each. In this coupled system, assimilating observations from the fast domains into the slow domains could not improve the predictability of the whole system, and they attributed it to the spurious covariances (due to the insufficient ensemble size) from the fast variables. Han et al. (2013) coupled atmospheric Lorenz63, oceanic pycnocline, and sea-ice models, and showed the difficulty of assimilating slower domain's observations into the faster domains. On the other hand, Liu et al. (2013) found that in the coupled atmosphere (Lorenz63 model)-ocean (thermocline model) model, assimilating observation of the dominant domain (namely atmosphere) to update oceanic variables was effective. Raboudi et al. (2021) used the one-way coupled multiscale Lorenz96 model to indicate that sCDA is superior to wCDA when the ensemble member is large. Some previous studies evaluated the effectiveness of sCDA with more realistic models than the flavors of the Lorenz63 and Lorenz96 models. Tondeur et al. (2020) used Modular Arbitrary Order Ocean-Atmospheric Model (MAOOAM) (De Cruz et al., 2016), an atmosphere-ocean model with arbitrary spatial resolutions. They suggested that frequent observations of the fast domain can suppress the error evolution, which is necessary for sCDA to improve the estimation of state variables. Thus, it is crucial to identify the conditions under which sCDA is effective than wCDA so that we can understand whether implementing sCDA is worthy of its computational/development cost in each specific coupled system.

Although sCDA has been widely applied in various coupled systems, they seem to lack theoretical backgrounds. Specifically, to the authors' knowledge, there are no studies which focus on cross-domain interaction despite its importance: 1) coupled systems in earth system sciences have diverse cross-domain interactions from duplex to one-way, 2) they may spatiotemporally vary, and 3) they may cause non-linearity or even chaotic nature of the coupled system. Therefore, in this study, we focused on cross-domain interaction t

erms as one of the most important elements to characterize the coupled system. Through an extensive investigation using joint Lorenz96 model, we aimed to reveal the conditions under which sCDA is effective.

2, Method

This study uses the joint Lorenz96 model in which two Lorenz96 models with 40 dimensions are coupled. Although this is a toy model, it captures important dynamics of the atmosphere: advection, dispersion, and attenuation. In addition, it shows chaoticity, which means that small errors of the initial state exponentially grow. Therefore, it is suitable and often used as a testbed of data assimilation methods (e.g. Trevisan et al., 2010). Equation (1a) is the traditional (non-coupled) 40-dimensional Lorenz96 model (Lorenz & Emanuel, 1998):

$$\frac{dX_k}{\mathrm{dt}} = \ -X_{k-1}\left\{X_{k-2} -\ X_{k+1}\right\} - X_k + F_1\ (k\ =\ 1, 2, \ldots, 40)\#(1a)$$

In Lorenz (1995), he assumed multiscale coupled system shown below:

$$\frac{dX_k}{\mathrm{dt}} = X_{k-1}\left\{X_{k+1} -\ X_{k-2}\right\} - X_k - \frac{\mathrm{hc}}{b}\sum_j Y_{j,k}\#(2a)$$

$$\frac{dY_{k,j}}{\mathrm{dt}} = cbY_{j+1,k}\left\{Y_{j-1,k} -\ Y_{j+2,k}\right\} - \mathrm{cY}_{k,j} + \frac{\mathrm{hc}}{b}X_k\#(2b)$$

In this study, we set k to be 40, b = c, added forcing term (F=8.0) and set j (the number of faster domains connected to one slow domain) to 1. This would lead:

$$\frac{dX_k}{\mathrm{dt}} = \ -X_{k-1}\left\{X_{k-2} -\ X_{k+1}\right\} - X_k + F_1 + \alpha Y_k\ (k\ =\ 1\ldots\ldots, 40)\#(3a)$$

$$\frac{dY_k}{\mathrm{dt}} = \ -c^2 Y_{k-1}\left\{Y_{k-2} -\ Y_{k+1}\right\} - \mathrm{cY}_k + F_2 + \beta X_k\ (k\ \ldots, 2, \ldots, 40)\#(3b)$$

By eliminating the multiscale interactions, we mainly assumed that spatial scales are identical between two domains. During the time integration, timestep was set to 0.005 (as 0.2 corresponds to one day, it means roughly 40 timesteps a day), and the 4$^{\mathrm{th}}$-order Runge-Kutta method was applied for time integration.

We initially set c to 1 (which means there is no time-scale difference between the domains) and investigated the impact of cross-domain interaction on the DA skill. As the previous studies implies that the time-scale differences are the cause of poorly estimated inter-domain covariances (Han et al. 2013), we later adopted different c (namely, 0.5, 0.3 and 0.2) so that we could analyze the impact of time-scale difference.

In this experiment, we adopted the observation system simulation experiment (OSSE). In OSSE, we adopt simulated values as the true state, and try to estimate the simulated values by DA. It is initially used to check whether the new observation systems are useful, but it is often used as a method to verify the skills of newly developed DA methods (Penny, 2014; Raboudi et al., 2018; Yoshida & Kalnay, 2018).

The coupled system was initialized with random initial conditions following 40-dimensional gaussian distribution N(0,5) and spun up for 36000 days (0.2/0.005*36000=1.44 million timesteps). Then, DA was conducted every 6 hours (10 timesteps) for 180 days. To generate observations, we assumed that 13 out of 40 dimensions are observable (namely dimension 1, 4, 7, 10, 13, 16, 19, 23, 26, 29, 32, 35 and 38). Gaussian errors were added to the synthetic truth; when c = 1, the standard deviation of a gaussian error is set to 1.0 so that we can assume a 20% sampling error (note that the standard deviation of the state variable is approximately five). When c is not equal to 1, we assumed the 20% sampling error in the same manners.

Local Ensemble Transform Kalman Filter (LETKF; Hunt et al., 2007; Miyoshi & Yamane, 2007) was adopted as a DA method. In LETKF, we can easily choose a certain subset of the observations to be assimilated into the specific model domains, and thus it has been applied in the previous sCDA studies (Yoshida & Kalnay, 2018). By utilizing this characteristic, we can implement both wCDA and sCDA by changing the observations to be assimilated.

In ensemble-based filtering, the dimensions of the analysis space in k-dimensional system is less than $k - 1$, and more than the number of non-negative Lyapunov Exponents (LE) of the system (Ng et al., 2011). After calculating the number of positive LEs in the coupled system (results shown in the next section), this study adopted 20 ensemble members as the insufficient ensemble size and 80 ensemble members as the sufficient ensemble size. As the previous studies show that the score does not improve further when the ensemble member exceeds a certain threshold (Sakov & Oke, 2008; Yoshida & Kalnay, 2018), we did not conduct experiments with ensemble sizes larger than 80.

In LETKF, we have hyperparameters such as localization and covariance inflation, both of which are introduced to accurately estimate background covariance matrix. In this study, the observation localization (Miyoshi et al., 2007) was applied, so that the observation covariance matrix R^{-1} was multiplied by the localization function $L(r)$, taking the physical distance between the analyzed dimension and observed dimension (eq. (3)) into account. By introducing this method, we can set $L(r)$ to zero for faraway observations, and thus we can prevent those observations from being assimilated.

$$L(r) = \begin{cases} \exp\left(\frac{1}{2}\left(\frac{r}{\sigma}\right)^2\right) & if\ r < 2\sqrt{10/3}\sigma \\ 0 & if\ r \geqq 2\sqrt{10/3}\sigma \end{cases} \#(3)$$

While some previous studies proposed dynamically estimating the localization scales (Wang & Bishop, 2003; Yoshida & Kalnay, 2018), in this study we applied the hyperparameter sweep (in other words, we evaluated many possible combinations of the hyperparameters). This is because this study's scope includes the comprehensive analysis of the LETKF's stability to the hyperparameters in two settings of CDA (i.e., sCDA and wCDA). We used four localization parameters σ (3, 5, 7, 9) for each domain and four inflation parameters (1.02, 1.06, 1.10, 1.14). We applied multiplicative inflation (Anderson & Anderson, 1999).

Four types of the CDA methods are implemented. We defined the quasi-sCDA to distinguish it from the fully coupled sCDA. In quasi-sCDA, we strongly coupled only one domain and the other domain was weakly coupled. For example, in quasi-sCDA (domain A only, namely qsCDA_A), we used observations from both domain A and domain B to update domain A's forecast but used observations from domain B only to update domain B's observation (see the lower left and upper right squares of Figure 1-c). Therefore, observations in domain A were not used to update domain B. Note that the previous studies of sCDA which used realistic models mainly adopted quasi-sCDA.

Table 1 summarizes our experiments. They are classified into two: perfect model experiments and imperfect model experiments. In the perfect model experiments, we assumed that the dynamics had no errors; uncertainties were attributed to state variables only. In the other experiments, we assumed imperfect models. In the experiment 2, we assumed the bias in cross-domain interaction terms and intra-domain forcing terms. In experiment 3, we assumed that we did not know the magnitude of cross-domain interaction term and tested if CDAs could estimate these cross-domain interaction parameters. In the experiment 3, both state variables and interaction parameters were jointly estimated online. In other words, we applied state augmentation; we concatenated the parameter vector to the state vector and estimated both of them simultaneously. As for the domain interaction term which needs to be estimated, we adopted the similar method to Relaxation to Prior Spread method (Whitaker & Hamill, 2012). We set $\alpha = 1$ (see equation 3 in their study) to maintain the spread of the domain interaction term. The initial conditions of parameter ensembles were derived from gaussian distribution whose mean is 0.2 larger than the truth and whose standard deviation is 0.01. Note that the spreads of the state variables are inflated with multiplicative inflation (same method as the other experiments).

The observations were assimilated for 180 days with 5 different initial conditions.

As its interval was set to be six hours, DA was conducted as a total of 720 times. The first quarter of the analysis result was discarded as the spin-up period of DA, and the rest (540 times) were compared with the synthetic truth.

The adopted score metric is the domain average of tRMSE ($\overline{\mathrm{tRMSE}}$), which is defined below:

$$\overline{\mathrm{tRMSE}} = \frac{1}{40}\sum_k \sqrt{\frac{1}{540}\sum_t \left(x_{k,t}^{\mathrm{anl}} - x_{k,t}^{\mathrm{truth}}\right)^2}\#(4)$$

where $x_{k,t}^{\mathrm{anl}}$ stands for the analysis value of dimension k in time t, and $x_{k,t}^{\mathrm{truth}}$ stands for the true value of dimension k in time t. DA was conducted and $\overline{\mathrm{tRMSE}}$ was calculated from 5 different initial conditions, and the score was defined as the trimmed mean of five $\overline{\mathrm{tRMSE}}$s (largest and smallest ones are discarded). If some calculations failed due to filter divergence, the trimmed mean was calculated in the same manner. When only two or less experiments were converged, we considered that we could not get a result.

To summarize, this research introduced the four patterns of localization radius in domain A, the four patterns of localization radius in domain B, the four patterns of covariance inflation, the five initial conditions, and the two ensemble numbers (20 and 80). Totally we had 640 implementations of CDA for the single β value. The calculation was conducted with Oak-forest PACS (the supercomputer in the University of Tokyo); MPI parallelization is introduced in the hyperparameter sweep for efficient calculation.

Table 1. Model settings in the experiments

Exp. ID	Experiment	Coupled-DA method used	Model settings
-1	Perfect model	wCDA, sCDA, qsCDA_A, qsCDA_B	Symmetric coupling with same temporal scales: = , -1.4 1.4, c=1,0.5,0.2
-2			One-way coupling with same temporal scales: =0, -1.4 1.4, c=1
-3		wCDA, sCDA	Inverse coupling with same temporal scales: = - , -1.4 1.4, c=1

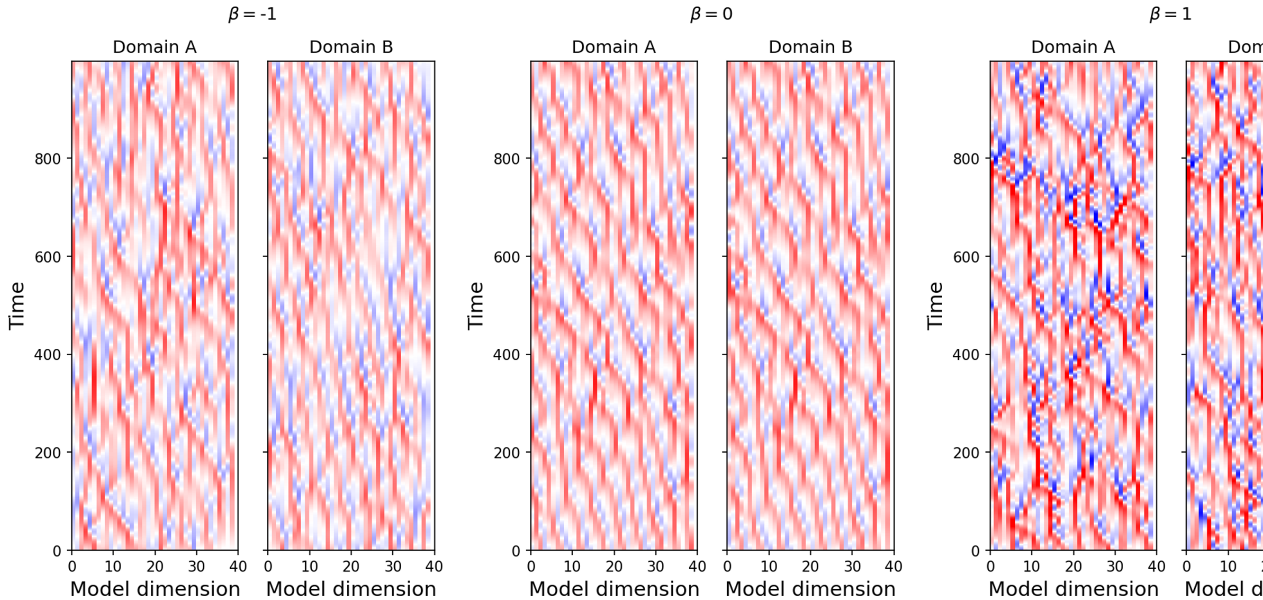
Exp. ID	Experiment	Coupled-DA method used	Model settings
-4		wCDA, sCDA	Symmetric coupling with different temporal scales: = , -1.4 1.4, c=0.5, 0.3, 0.2
	Bias in model parameters	wCDA, sCDA	Symmetric coupling: i) bias in inter-domain parameters: $\alpha_{\text{true}} = \beta_{\text{true}} = 0.5$ ii) bias in inter-domain parameters: $\alpha_{\text{true}} = \beta_{\text{true}} = -1.0$ iii) bias in intra-domain parameters: $F_{\text{true}} = 8.0$
	Estimate interaction (spatiotemporally uniform)	wCDA, sCDA	Symmetric coupling: = , -1.4 1.4 (We know that interaction is spatiotemporally uniform)

*Note: for all types of DA experiments, we conducted 640 DA run with four types of localization in domain A (3, 5, 7 and 9), localization in domain B (3, 5, 7 and 9), four types of covariance inflation (1.02, 1.06, 1.10 and 1.14 for other than Experiment 2 and 1.06, 1.10, 1.14 and 1.18 for Experiment 2), two types of ensembles (20 and 60), and five types of initial conditions.

sCDA			Observation used	
			Domain A	Domain B
			$x_1 \dots x_{40}$	$y_1 \dots y_{40}$
Analyzed grid	Domain A	$x_1 \dots x_{40}$	+	+
	Domain B	$y_1 \dots y_{40}$	+	+

3. Results

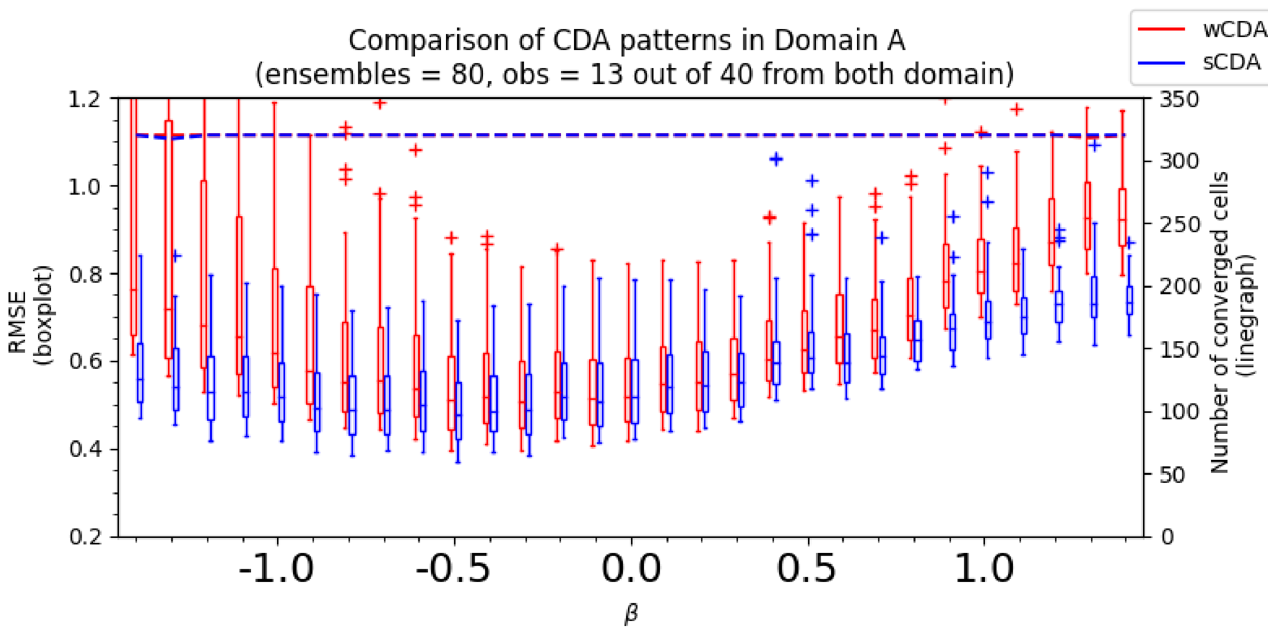
3.1 Basic characteristics of the joint-Lorenz96 model

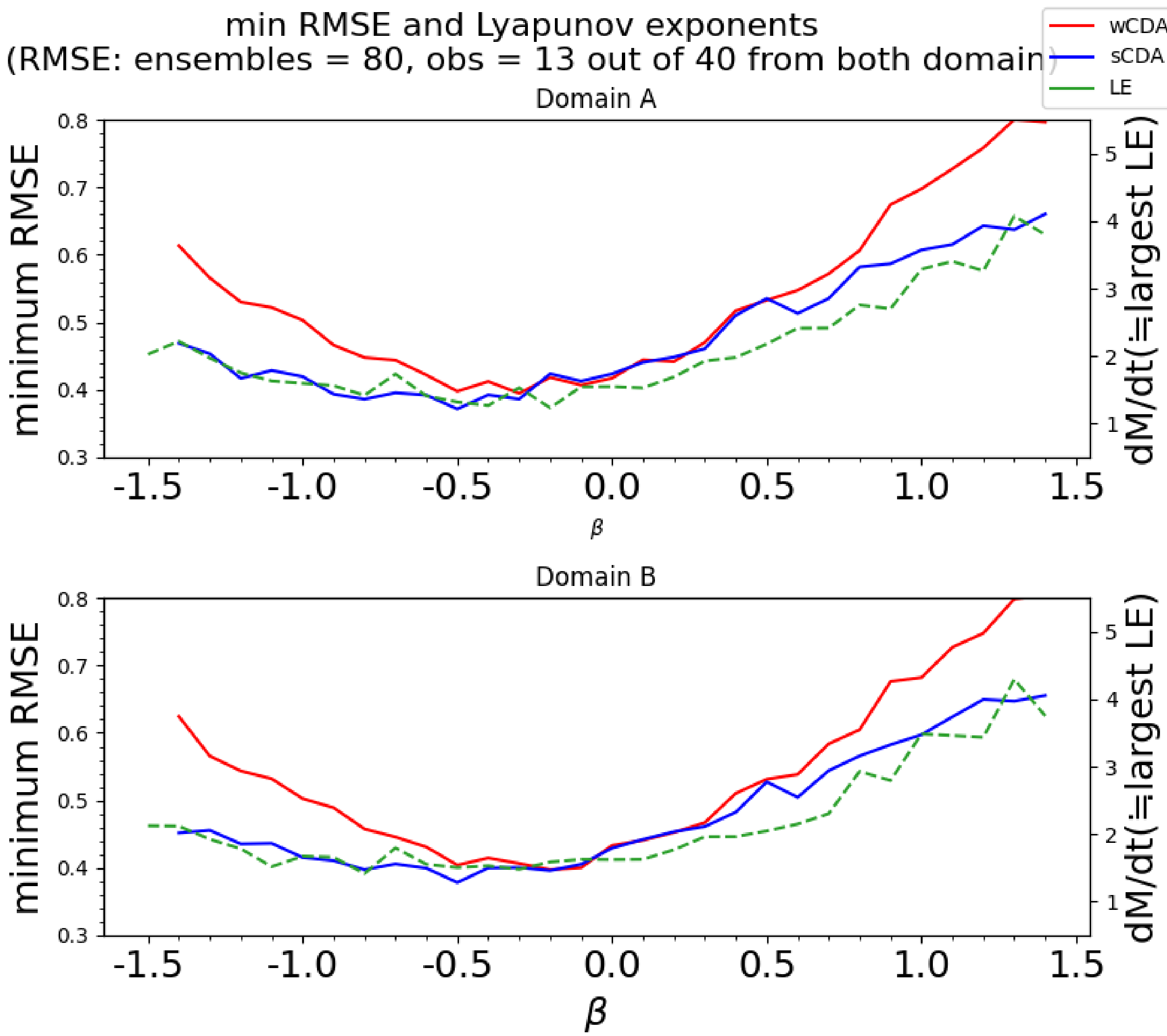


the time evolution of the joint-Lorenz96 model is shown in Figure 2 ($c = 1$, $\alpha = \beta$). The values oscillate largely when the absolute value of β is large, which implies the relative instability of the system. The Covariant Lyapunov Vectors (CLVs) are computed. Lyapunov vectors indicate the direction where system's error grows rapidly. When the Lyapunov Exponents (LEs, which indicate the speed of error growth on the direction of corresponding CLVs) are positive, the system is unstable in that direction and small error grows exponentially. Figure 3 indicates the number of positive LEs. When β is negative, LEs are not related to the absolute value of β, while when β is positive, the number of positive LEs increases when β increases. The maximum number of positive LEs is 30, which implies 40 ensemble members (consequently 80 ensembles as well) are sufficient to run ensemble based Kalman Filters in this model.

3.2 Perfect model experiments

Figure 4 shows the result of the experiment 1 with the ensemble size of (a) 20 and (b) 80. The dashed lines indicate the number of converged experiments out of 320 trials (64 hyperparameter settings * 5 initial conditions). When the cross-domain interaction is too large, the systems is dominantly driven by the cross-domain interaction term. The results indicate that the system is more likely to diverge when the absolute value of β is large, especially in the case of small ensembles. The boxplot indicates the stability of LETKF to the hyperparameters. As there are 64 patterns of hyperparameters in LETKF, we can calculate $\overline{\mathrm{tRMSE}}$ for each setting, which are used to draw each boxplot. The horizontal axis shows cross-domain interaction, and the vertical axis shows $\overline{\mathrm{tRMSE}}$. Firstly, when the ensemble size is large (Figure 4-a), sCDA stably scores better than wCDA. Even when the cross-domain interaction is zero (and thus no cross-domain covariance), sCDA can accurately provide the information of "having no cross-correlation", so that sCDA does not degrade the score. Besides, sCDA is more stable than wCDA especially when the cross-domain interaction is large. Meanwhile, when the ensemble size is small (Figure 4-b), the sCDA scores better when β increases under optimal hyperparameter settings. However, when the domain interaction is small, wCDA scores better than sCDA, which indicates that the strong cross-domain interaction is essential for the superiority of sCDA. In addition, there are large variances of $\overline{\text{tRMSE}}$ with each β in the case of sCDA , which indicates that $\overline{\text{tRMSE}}$ significantly changes when the different set of hyperparameters is given. Subsequently, the performance of sCDA is quite sensitive to hyperparameters. Hence, the calibration of hyperparameters is essential for the accurate and stable prediction in sCDA.





5 compares CDA scores and chaoticity of the coupled system. The minimum $\overline{\text{tRMSE}}$ of wCDA and sCDA is extracted from the boxplots in Figure 5 in red and blue solid lines, respectively. In addition, the green dotted line shows the value of dM/dt. It is a logarithm of the error expansion speed when we added small perturbations to the initial conditions on attractor, and thus corresponds to the chaoticity of the system. For detailed explanation of dM/dt, see Gutiérrez et al. (2008). Figure 5 shows that the minimum $\overline{\text{tRMSE}}$ is well correlated with dM/dt, which implies that minimum $\overline{\text{tRMSE}}$ can be explained by the impact of the cross-domain interaction on the chaoticity of the system. In addition, when β is negative (and its absolute value is large), sCDA scores

better than wCDA even with small ensembles. In other words, sCDA is the most effective when cross-domain interaction is large, albeit it does not contribute to the chaoticity of the system. This is one of the important findings regarding the conditions under which sCDA works better than wCDA.

Figure 6 shows the average $\overline{\text{tRMSE}}$ of domain A as a function of domain A's localization scales and the inflation magnitudes. The result is averaged over the four localization patterns in domain B. Overall, the poor tuning of inflation may degrade the average $\overline{\text{tRMSE}}$. In other words, the range of inflation which provides better estimates depends on the number of ensembles. Small inflation provides poor estimates when the ensembles is small, while large inflation provides poor estimates when the ensemble is large. The reason may be that the small ensemble causes the underestimation of background error covariance, which requires larger covariance inflation for appropriate filtering. In terms of localization, the second and fourth graphs in the bottom row show that the unstable result is caused also by the inappropriate choice of localization scales in sCDA with small number of ensembles. These results occur irrespective to β.

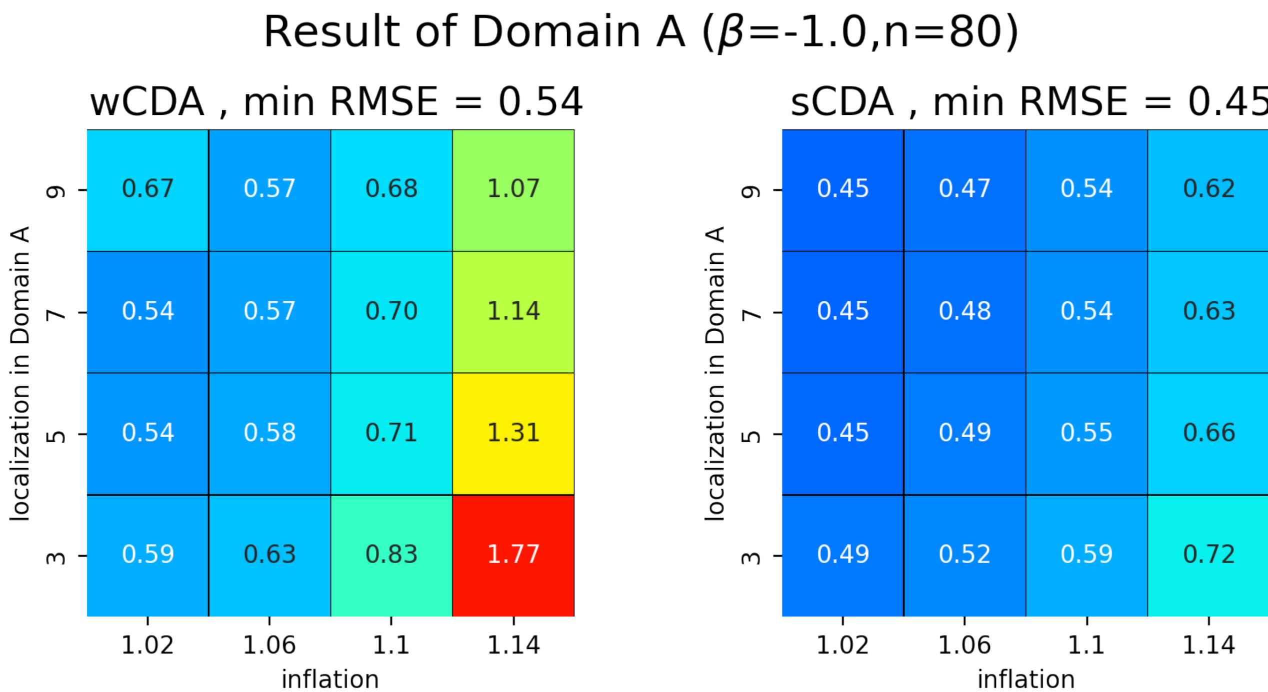
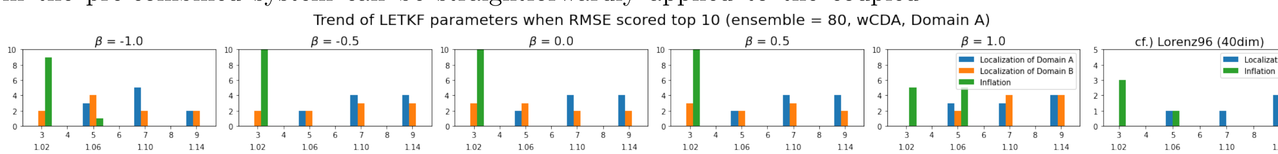


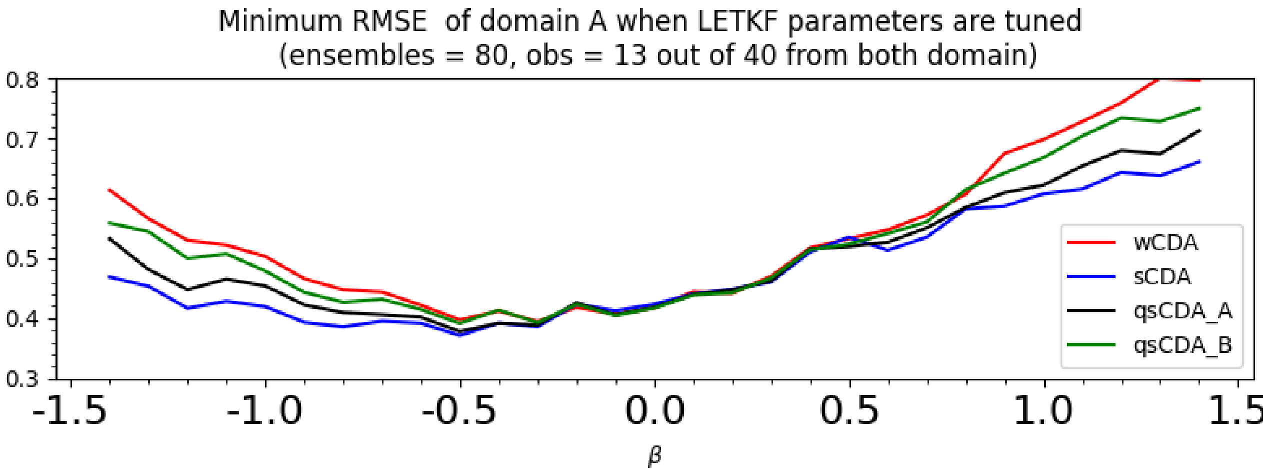
Figure 7 shows the hyperparameter settings with top 10 $\overline{\text{tRMSE}}$ score in domain A among the 64 patterns of LETKF hyperparameter settings. In general, the tendency of the hyperparameters that show good scores strongly depends on ensemble size, which is a universal property for both sCDA and wCDA. The settings that work well for the small ensemble can be applied to the large ensemble for neither wCDA nor sCDA. In other words, it is difficult to initially compute using small ensembles with the cheap computational cost to find the optimal hyperparameters and then apply them to large ensembles. As optimal parameter settings greatly depend on the domain interaction and the number of ensembles, it is essential to tune the hyperparameters from time to time. The optimal inflation tends to be small when the ensemble is large. Although there is no clear trend for localization, it seems that it is better to take a longer localization scale and pick up weak correlations in large ensembles. The rightmost column shows the hyperparameter characteristics of the uncoupled Lorenz96 model. If the parameter trends for the uncoupled model are the same as those for the coupled models, the optimal settings in the pre-combined system can be straightforwardly applied to the coupled



However, the optimal settings in the uncoupled Lorenz96 system can be used only when the ensemble is sufficiently large and when 0 (i.e., the domain interaction does not affect the chaoticity of the system). To summarize, it can be concluded that (1) the strategy of ”firstly estimating the hyperparameters in the uncoupled system, and then applying the obtained optimal settings to the large-ensemble sCDA” does not work well, and (2) CDA is quite sensitive to the hyperparameters of LETKF.

Figure 8 shows the results when quasi-strongly coupled data assimilation is introduced; we swept the hyperparameters (inflation, domain A’s localization, and domain B’s localization), and the shown RMSE is the result with optimal hyperparameters. When the ensemble size is set to sufficiently large, sCDA scores the best, qsCDA_A (sCDA only in domain A) and qsCDA_B (sCDA only in domain B) follow, and wCDA is the worst. Even when only domain B is strongly coupled, domain A can be improved to some extent because improved estimates of domain B have a positive effect on the estimation of domain A through the cross-domain interaction in the forecast step. When the ensemble size is small, wCDA scores better unless the interaction term is large.

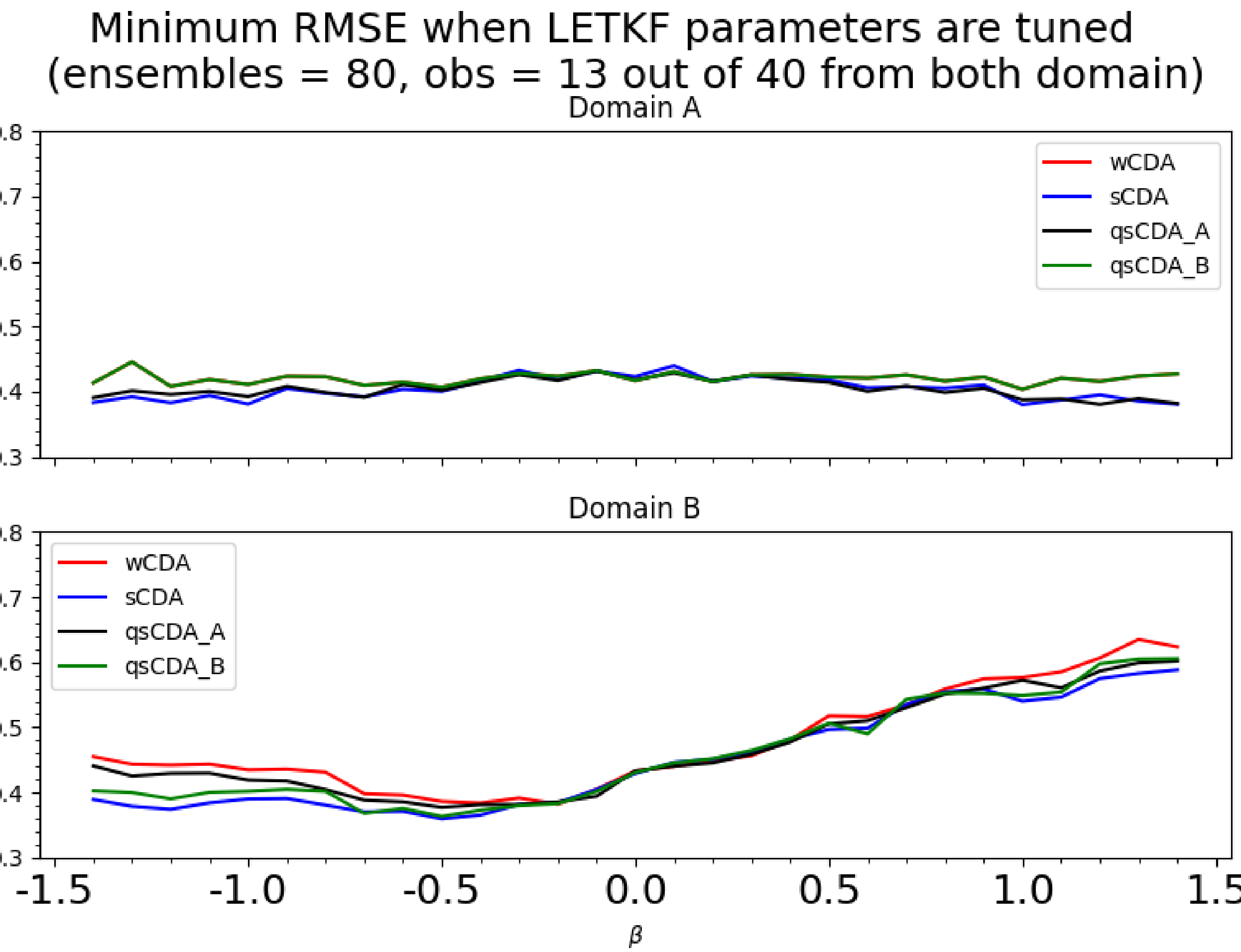
Next, we provide the results of the one-way coupled system (experiment ID: 1-2). Figure 9 shows the results (hyperparameters are swept, and the results with optimal hyperparameters are shown) in the one-way coupled system (only domain A influences domain B). Since there is no reverse impact (namely domain A to domain B), it is theoretically impossible that the effect of strongly coupled B spills over to A and make A’s estimates better. Therefore, when only one of them can be strongly coupled due to constraints such as a development cost, it is better to choose the domain that drives the entire coupled system. Besides, when β takes large negative value, sCDA outperforms wCDA even when ensemble is small. However, compared to the difference between ”wCDA” and ”qsCDA_B” in Figure 8-a, the difference of them in Figure 9 is small. Compared to the one-way coupled system, the interdependent coupled system is more likely to be improved by sCDA.

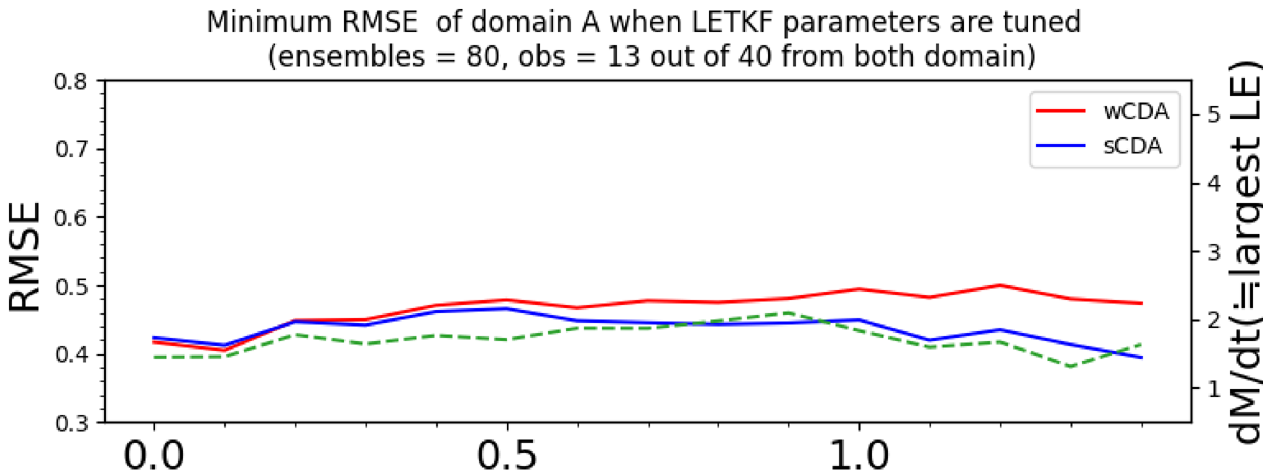


same experiment was conducted for the case where the signs of and were different, which hereafter called inverse coupling (experiment ID: 1-3, Figure 10). The results stayed the same: 1) for small ensembles, wCDA is better when the inter-domain interaction is small, and sCDA is better only when the inter-domain interaction becomes strong, and 2) for large ensembles, when the inter-domain interaction is small, the scores are similar, but when the inter-domain interactions are large, sCDA is better. This also suggests that, at least for coupled systems with equal velocities, the magnitude of the domain interaction is important, and it does not substantially depend on the direction of domain interactions (whether it is a two-way, one-way, or inverse).

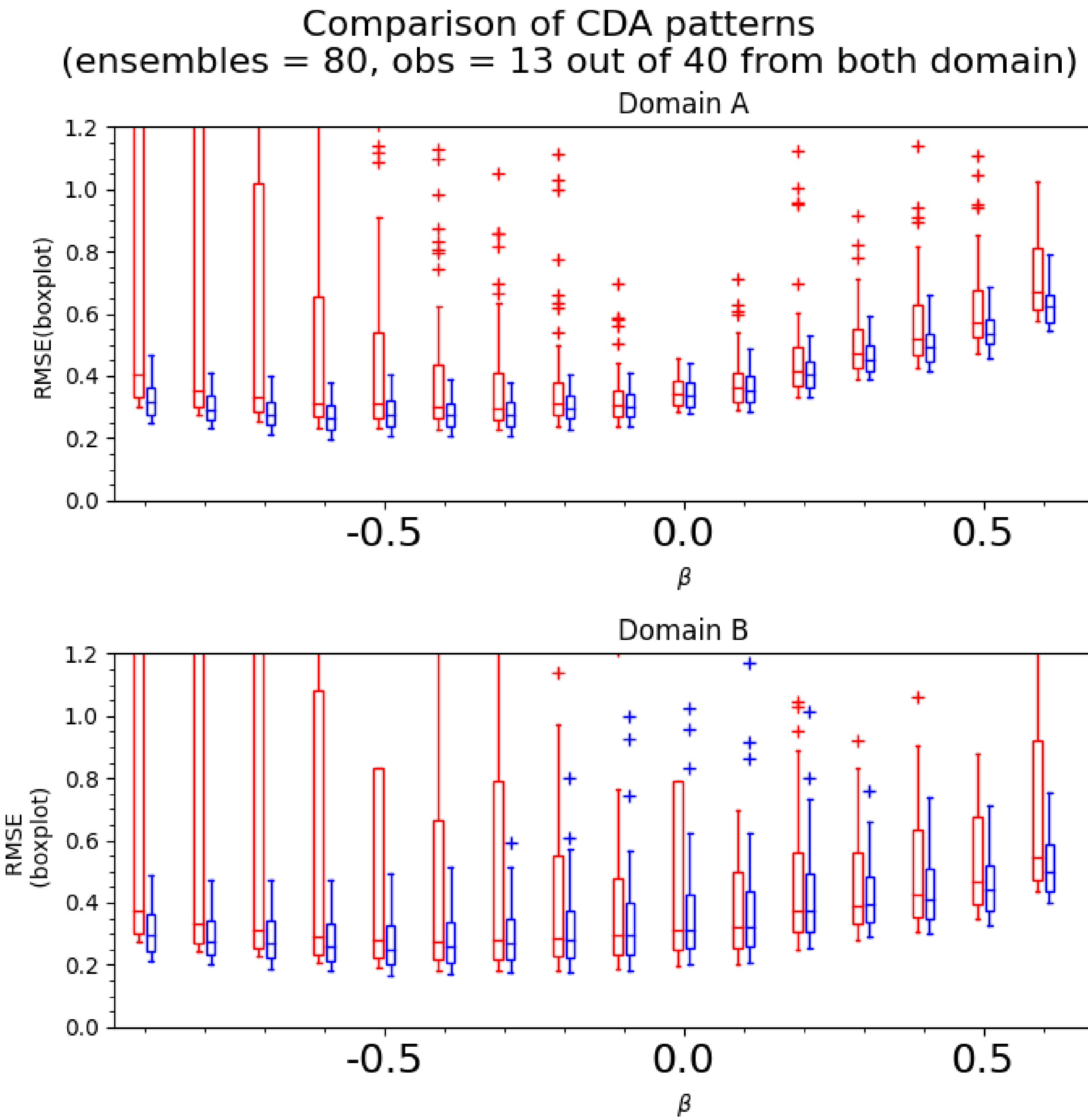
Figure 11 shows the results when there exists time-scale difference between the domains (experiment ID: 1-4). When domain B is twice as fast as domain A (Figure 11-a: c = 0.5), the main result remains the same as those in the previous experiments: sCDA outperforms wCDA when the cross-domain interaction is large (especially large negative). In addition, wCDA starts to outperform sCDA especially in the slow domain when the time-scale difference becomes larger (Figure 11-b, and Figure 11-c). The result implies that when the time scale difference becomes more significant, sCDA may not be the appropriate choice.

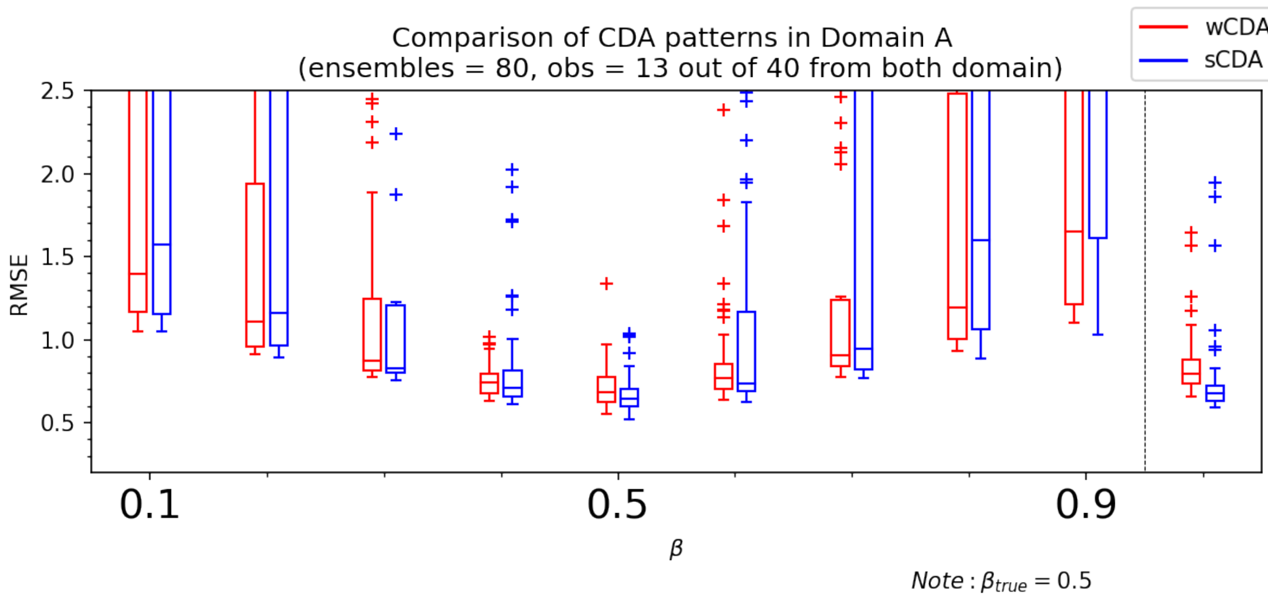
In terms of hyperparameters, although the calibration of localization parameters is important in small ensembles, the score primarily depends on an inflation parameter. In addition, the range of inflation with better estimates depends on the number of ensembles. The tendency is the same as experiment 1-1 (shown in Figure 6): Small inflation provides poor estimates when the ensembles is small, while large inflation provides poor estimates when the ensemble is large. The only difference is that the extent to which the result degrades more severely with inappropriate inflation, which in turn appears that CDA is more sensitive to hyperparameters in the coupled systems with larger time scale differences. Besides, the instability occurs when the slower system is weakly coupled (not shown); namely, the observation in the fast domain is not assimilated to the slower domain. It implies that observing and assimilating the domain with large scale/perturbation can constrain the whole system, even though the background error covariance is overestimated. Although overestimation of error covariance may contribute to degrading the score, the use of observation from both domains may be able to cancel such a drawback, and thus sCDA's score does not depend so heavily on hyperparameters as wCDA.





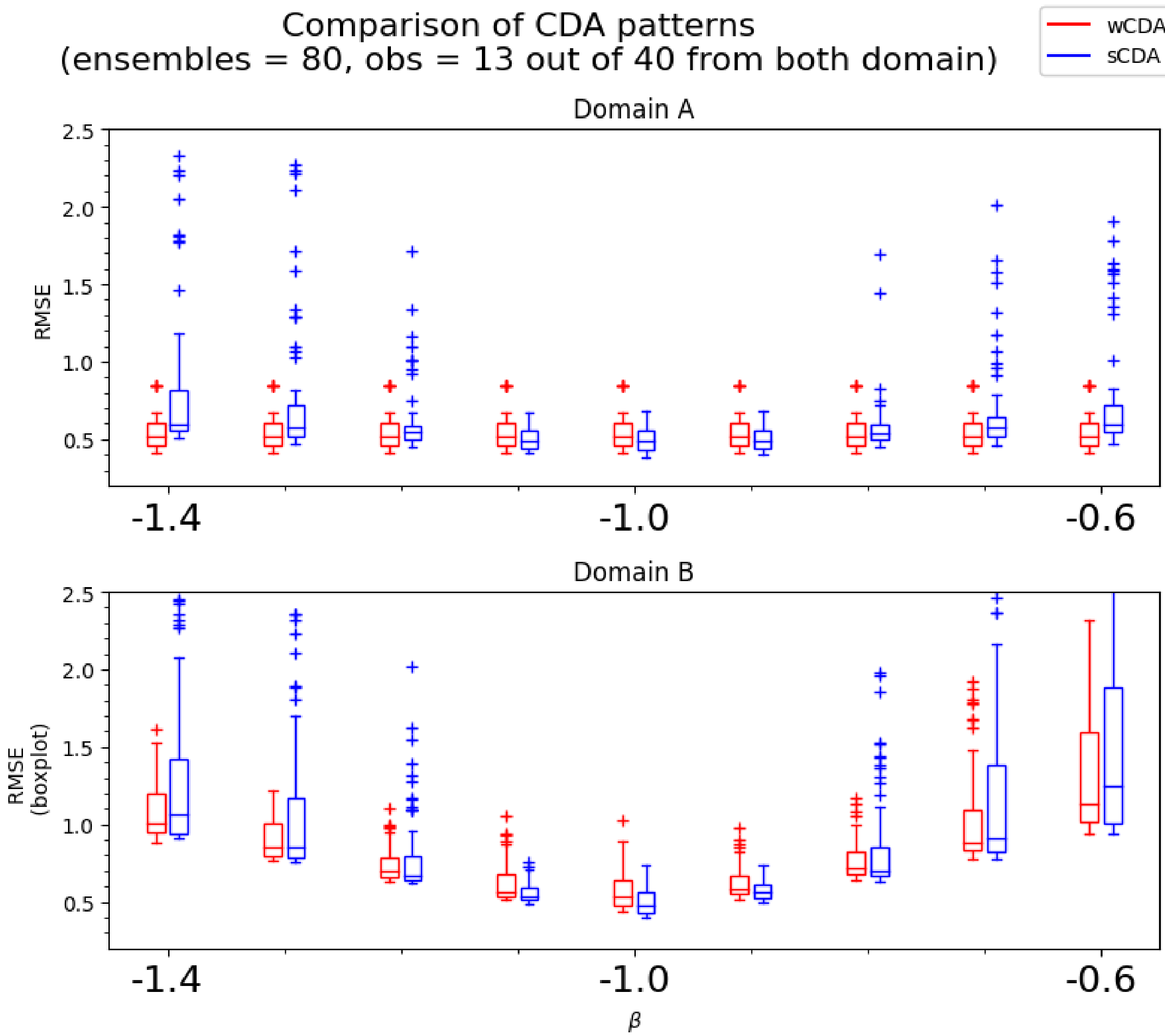
Minimum RMSE of domain A when LETKF parameters are tuned
(ensembles = 80, obs = 13 out of 40 from both domain)
RMSE
dM/dt(≒largest LE)
wCDA
sCDA
0.8
0.7
0.6
0.5
0.4
0.3
0.0
0.5
1.0
5
4
3
2
1


Comparison of CDA patterns
(ensembles = 80, obs = 13 out of 40 from both domain)
Domain A
RMSE(boxplot)
1.2
1.0
0.8
0.6
0.4
0.2
0.0
-0.5
0.0
0.5
β
Domain B
RMSE
(boxplot)
1.2
1.0
0.8
0.6
0.4
0.2
0.0
-0.5
0.0
0.5
β



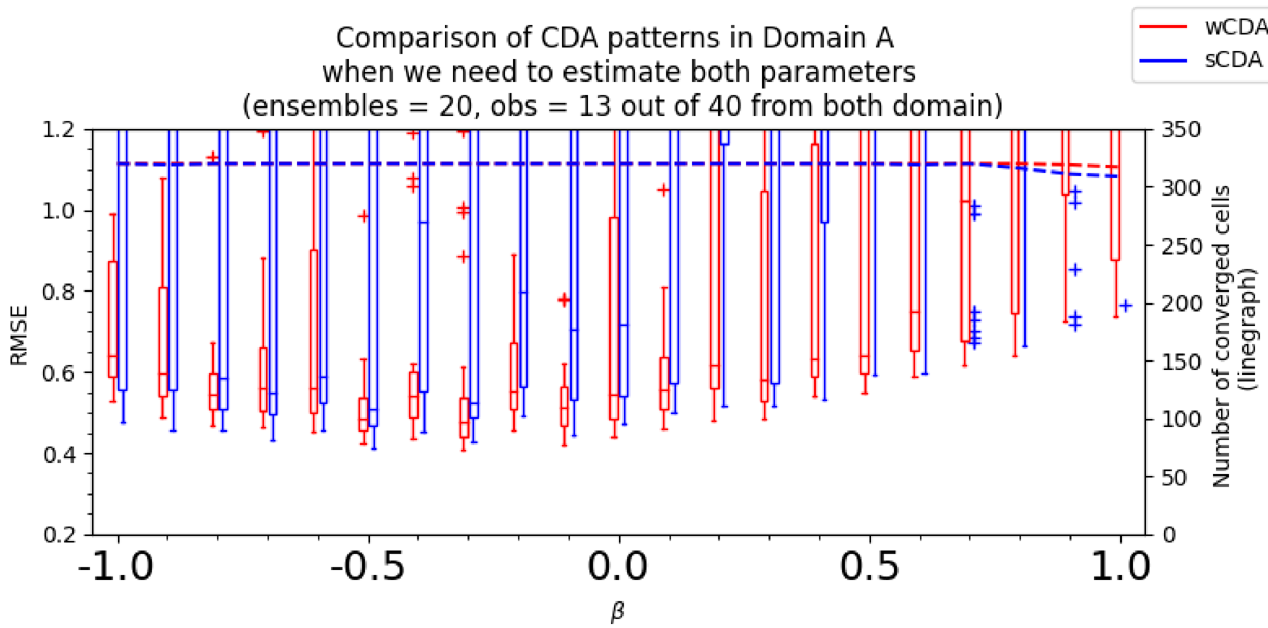
3.3 Imperfect model experiments

In the imperfect model experiments, we assumed that some of the model parameters are biased. In Figure 12, we assumed symmetrical domain interactions (i.e., interaction term $=$). In Figures 12-a to 12-b, the true model parameters were set to $\alpha = \beta = 0.5$. However, in the forecast phase (the time evolution phase) of DA, we used the biased interaction terms, ranging from 0.1 to 0.9. The horizontal axis shows biased , and the vertical axis shows $\overline{\text{tRMSE}}$. While sCDA performs better than wCDA in terms of the optimal scores, sCDA is more sensitive to hyperparameters than wCDA (namely, the range of the boxplot is long) especially with the small ensemble size. When the interaction is relatively weak, the cross-domain error covariance, which is inherently weak and hard to estimate, becomes even harder to be estimated under imperfect model settings. Figure 12-c shows the case where $\alpha = \beta = -1.0$. Our perfect model experiments showed that sCDA is effective in this domain interaction term. However, in this imperfect model experiment, sCDA only slightly outperforms wCDA (see also Figure 4). Such a decrease in the effectiveness of sCDA may be due to the disadvantage of using the disturbed background error covariance by the model's imperfectness. sCDA substantially degrades the score when there are large biases in the process model. Figure 12-d shows the results for the one-way coupling case where $\alpha = 0$ and $\beta = -1.0$. When the bias is large (i.e., the biased is 0.1 or 0.9), wCDA outperforms sCDA even in the case of the large ensemble (80 ensembles). It seems to contradict with the result that sCDA was more effective where is significantly negative. This can be explained as follows: in one-way coupled system, when domain A were weakly coupled, it would continue to be accurate as it is not affected by the bias. However, if domain A is strongly coupled, the bias in the interaction term deteriorates the estimation of domain A. This will spill over to domain B, resulting in a worse score than with wCDA. In other words, in a one-way coupled system, because the presence of bias can disrupt the score of the whole system, wCDA's ability to prevent these disruptions (see the upper boxplots in Figure 12-d) by not updating domain A's model forecast via biased dynamics, must be taken into consideration. In addition, we also tested the case where the bias exists in the intra-domain dynamics. Figure 12-e shows the results when we assume a bias in the external force term F (true value is 8.0) in the domain for symmetric coupling case ($\alpha = \beta = -1.0$). Even in the case of 80 ensembles, a 20% bias in F makes wCDA better. It turns out that when intra-domain bias prevents the accurate estimation of the intra-domain covariance, it is even more difficult to correctly estimate the cross-domain covariance. In other words, it is strongly suggested that sCDA may not contribute to the better estimation of the system when the dynamics of each domain is not accurately described by the process models. When we compare the threshold of the bias when wCDA outperforms sCDA, it is approximately 40% in domain interaction term (in figure 12-d, wCDA is better when the biased β is 0.7), while it is approximately 20% in the intra-domain dynamics (in figure 12-e, wCDA is better when the biased F is 9.6). Therefore, the biases in both intra-domain dynamics and inter-domain dynamics deteriorate sCDA score. As the origin of the bias may not be clearly identifiable in many real-world applications, the poor skill of sCDA when bias is assumed is an important implication for the real-world applications.



Next, we provide the results of imperfect model experiments, in which we assumed that the domain interaction terms were unknown and thus jointly estimated with the state variables.(Experiment ID: 3). The rightmost columns in Figures. 12-a to 12-c show the results of the estimation of the interaction term β by state-augmentation. Note that we have a strong hypothesis here: we know that the interaction is symmetric and uniform in all 40 dimensions. In such cases, as a lot of information can be used for parameter estimation and the bias in the model can be corrected by data assimilation.

Figure 13 shows the results of the joint state-parameter estimation with various β value. Figures 13-a and 13-b shows the boxplot of the $\overline{\text{tRMSE}}$ for ensembles 20 and 80, respectively. The result is similar to that of the perfect model experiment; sCDA works well 1) when the cross-domain interaction β is large (especially largely negative) 2) and when the ensemble number is large. Considering that the deviation of the estimated β from the true value is approximately 0.05 in this experiment (not shown), it can be said that the present experiment is equivalent to the case of the complete model. It suggests that sCDA performs well under imperfect model scenarios when we can utilize DA to mitigate/lessen model biases. In addition, we conducted OSSE where β varies spatially and temporarily (not shown). The overall results were same: sCDA outperforms wCDA when we have large number of ensembles, or the cross-domain terms is large (especially negative).



4, Discussion

The extensive analysis has revealed the CDA's characteristics with various cross-domain interaction. First, the estimation skills (explained by the minimum $\overline{\text{tRMSE}}$ of both sCDA and wCDA) corresponds to dM/dt (chaos intensity index in Gutiérrez et al., 2008). It is reasonable that the chaoticity determines the DA skills. When comparing sCDA and wCDA, the large number of ensembles and the large magnitude of the domain interaction term are necessary for sCDA to be more effective than wCDA. Although the previous studies have found the importance of the large ensembles in sCDA (e.g., Raboudi et al., 2021), the efficiency of sCDA with regard to the chaoticity has not been discussed.

The advantage of sCDA is most pronounced in systems with small dM/dt and large cross-domain interactions. In other words, sCDA efficiently works in the systems where the cross-domain interaction is large but does not enlarge the chaoticity of the whole system. This finding is useful to identify the systems in which sCDA can be efficiently applied. For example, in the field of tropical disturbance where atmosphere-ocean interaction may suppress the evolution of the system, introducing sCDA may efficiently improve the predictability of the whole system.

Second, in a one-way coupled system, both master and slave systems can be improved by strongly coupling the master system that drives the whole system, which is consistent with the previous works (Z. Liu et al., 2013; Sawada et al., 2018). In addition, we suggest that the degree of effectiveness increases with the coupling strength, which is reasonable considering that dM/dt is hardly changed by the coupling strength in a one-way coupled system.

Third, the intensive calibration of hyperparameters of covariance inflation and localization is necessary in sCDA. The results of our study indicate that sCDA outperforms wCDA with the use of adequate hyperparameters. In terms of inflation parameter, the results of Han et al. (2013), which used fixed inflation value and showed sCDA does not easily outperform wCDA with seemingly sufficient ensemble size, implies that the inflation may have played the key role. In terms of localization, the poor tuning of localization deteriorated sCDA's score in this study. Therefore, the cross-domain localization, as well as the intra-domain localization, is also an important perspective in sCDA as discussed in Stanley et al. (2021), which discussed the localization methods with various localization functions. In addition, we also found that the hyperparameter settings that works well in a small ensemble cannot be directly transferred to a large ensemble. Efficient hyperparameter tuning methods are necessary to maximize the potential of sCDA.

Fourth, sCDA is vulnerable to both intra-model and inter-model biases because the biased model causes the inaccurate estimates of the background cross-covariance. In our case, 20% bias in intra-domain terms made wCDA better than sCDA, while it was 40% bias in the inter-domain case. Therefore, sCDA cannot be used for fields where the model structures or model parameters are highly uncertain such as socio-hydrological models (see Sawada and Hanazaki 2020, for instance).

Fifth, when there is a time scale difference in the system, the main result remains the same: sCDA outperforms wCDA when the cross-domain interaction is largely negative. In addition, the improvement of sCDA's score drastically decreased especially in the slow domain when the time scale difference becomes more significant. As suggested in previous studies, the difficulty of coupled data assimilation lies in the fact that timescales differ across domains (Yoshida & Kalnay, 2018). This result suggests that data assimilation that simply exploits cross-domain covariance may not work. Recently, it has been found that coupled systems have a unique mode that was not present in the single domain

(Carlu et al., 2019). To make coupled data assimilation more effective, it may be necessary to maximize the potential of sCDA by extracting information which is specific to the coupled system, such as focusing on the stable subspace (Quinn et al., 2020).

Finally, we would like to discuss possible future strategies for the study of more generalized CDA. Since we attempted to analyze the sensitivity of LETKF hyperparameters in this study, we do not consider automatic/dynamic tuning of hyperparameters such as localization and covariance inflation. However, it is impractical in terms of computational cost to conduct hyperparameter sweep in the real-world applications. Hence, it is essential to investigate whether the scores of dynamic localization and covariance inflation schemes (Miyoshi, 2011; Yoshida & Kalnay, 2018) depend on domain interactions.

5.Conclusion

In this study, we clarified the characteristics of coupled data assimilation in chaotic systems. OSSE with the joint-Lorenz96 model was performed in both complete and incomplete model settings, and two conditions for the success of sCDA were found: sufficient ensemble sizes and sufficient magnitude of interactions. The impact of sCDA was particularly pronounced in the case where the interaction term is large but does not contribute to the chaoticity of the system. In the absence of a sufficient ensemble size, 1) sCDA's score strongly depends on the LETKF hyperparameters, i.e., highly unstable, and 2) sCDA outperformed wCDA only in cases where the interaction can damp the dynamics of other domains, and the bias of the model is small. Experiments with various domain interaction schemes, such as unidirectional, symmetric, and inverse, revealed that the above properties basically depend on the magnitude of the interaction, and not so much on the direction of the interaction.

Acknowledgement

This work was supported by the JST FOREST program (grant no. JPMJFR205Q).

Open Research

The data used in this study is available upon request. LETKF execution code, which is the basis of sCDA and wCDA, is available at https://github.com/takemasa-miyoshi/letkf.s

Author Contribution Statement (may not be necessary for JAMES)

The authors confirm contributions to the paper as follows: study conception and design: Miwa and Sawada; data collection and preparation: Miwa; analysis and interpretation of results: Miwa, and Sawada; draft manuscript preparation: Miwa; manuscript revision: Sawada. All authors reviewed the results and approved the final version of the manuscript.

References

Anderson, J. L., & Anderson, S. L. (1999). A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Review*, *127*(12), 2741–2758. https://doi.org/10.1175/1520-0493(1999)127<2741:AMCIOT>2.0.CO;2Bakhshaii, A., & Johnson, E. A. (2019). A review of a new generation of wildfire–atmosphere modeling. *Canadian Journal of Forest Research*, *49*(6), 565–574. https://doi.org/10.1139/cjfr-2018-0138Ballabrera-Poy, J., Kalnay, E., & Yang, S. C. (2009). Data assimilation in a system with two scales - Combining two initialization techniques. *Tellus, Series A: Dynamic Meteorology and Oceanography*, *61*(4), 539–549. https://doi.org/10.1111/j.1600-0870.2009.00400.xDe Cruz, L., Demaeyer, J., & Vannitsem, S. (2016). The Modular Arbitrary-Order Ocean-Atmosphere Model: MAOOAM v1.0. *Geoscientific Model Development*, *9*(8), 2793–2808. https://doi.org/10.5194/gmd-9-2793-2016Di Baldassarre, G., Viglione, A., Carr, G., Kuil, L., Salinas, J. L., & Blöschl, G. (2013). Socio-hydrology: Conceptualising human-flood interactions. *Hydrology and Earth System Sciences*, *17*(8), 3295–3303. https://doi.org/10.5194/hess-17-3295-2013Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, *99*(C5). https://doi.org/10.1029/94jc00572Gutiérrez, J. M., Primo, C., Rodríguez, M. A., & Fernández, J. (2008). Spatiotemporal characterization of Ensemble Prediction Systems the Mean-Variance of Logarithms (MVL) diagram. *Nonlinear Processes in Geophysics*, *15*(1), 109–114. https://doi.org/10.5194/npg-15-109-2008Han, G., Wu, X., Zhang, S., Liu, Z., & Li, W. (2013). Error covariance estimation for coupled data assimilation using a lorenz atmosphere and a simple pycnocline ocean model. *Journal of Climate*, *26*(24), 10218–10231. https://doi.org/10.1175/JCLI-D-13-00236.1Hunt, B. R., Kostelich, E. J., & Szunyogh, I. (2007). Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D: Nonlinear Phenomena*, *230*(1–2), 112–126. https://doi.org/10.1016/j.physd.2006.11.008Jia, B., Zhou, J., Zhang, Y., Tian, M., He, Z., & Ding, X. (2021). System dynamics model for the coevolution of coupled water supply–power generation–environment systems: Upper Yangtze river Basin, China. *Journal of Hydrology*, *593*(December 2020). https://doi.org/10.1016/j.jhydrol.2020.125892Kalnay, E., Li, H., Miyoshi, T., Yang, S. C., & Ballabrera-Poy, J. (2007). 4-D-Var or ensemble Kalman filter? *Tellus, Series A: Dynamic Meteorology and Oceanography*, *59*(5), 758–773. https://doi.org/10.1111/j.1600-0870.2007.00261.xLin, L. F., & Pu, Z. (2020). Improving near-surface short-range weather forecasts using strongly coupled land-atmosphere data assimilation with gsi-enkf. *Monthly Weather Review*, *148*(7), 2863–2888. https://doi.org/10.1175/MWR-D-19-0370.1Liu, X., & Lim, S. (2018). An agent-based evacuation model for the 2011 Brisbane City-scale riverine flood. *Natural Hazards*, *94*(1), 53–70. https://doi.org/10.1007/s11069-018-3373-1Liu, Z., Wu, S., Zhang, S., Liu, Y., & Rong, X. (2013). Ensemble data assimilation in a simple coupled climate model: The role of ocean-atmosphere interaction. *Advances in Atmospheric Sciences*, *30*(5), 1235–1248. https://doi.org/10.1007/s00376-013-2268-zLorenz, E. N. (1995). Predictabiity: A problem partly solved. *Seminar on Predictability*. https://www.ecmwf.int/node/10829%0ALorenz, E. N., & Emanuel, K. A. (1998). Optimal sites for supplementary weather observations: Simulation with a small model. *Journal of the Atmospheric Sciences*, *55*(3), 399–414. https://doi.org/10.1175/1520-0469(1998)055<0399:OSFSWO>2.0.CO;2Miyoshi, T., & Yamane, S. (2007). Local ensemble transform Kalman filtering with an AGCM at a T159/L48 resolution. *Monthly Weather Review*, *135*(11), 3841–3861. https://doi.org/10.1175/2007MWR1873.1Miyoshi, T., Yamane, S., & Enomoto, T. (2007). Localizing the Error Covariance by Physical Distances within a Local Ensemble Transform Kalman Filter (LETKF). *Sola*, *3*(1), 89–92. https://doi.org/10.2151/sola.2007-023Ng, G. H. C., McLaughlin, D., Entekhabi, D., & Ahanin, A. (2011). The role of model dynamics in ensemble Kalman filter performance for chaotic systems. *Tellus, Series A: Dynamic Meteorology and Oceanography*, *63*(5), 958–977. https://doi.org/10.1111/j.1600-0870.2011.00539.xPenny, S. G. (2014). The hybrid local ensemble transform Kalman filter. *Monthly Weather Review*, *142*(6), 2139–2149. https://doi.org/10.1175/MWR-D-13-00131.1Raboudi, N. F., Ait-El-Fquih, B., & Hoteit, I. (2018). Ensemble Kalman filtering with one-step-ahead smoothing. *Monthly Weather Review*, *146*(2), 561–581. https://doi.org/10.1175/MWR-D-17-0175.1Raboudi, N. F., Ait-El-Fquih, B., Subramanian, A. C., & Hoteit, I. (2021). Enhancing ensemble data assimilation into one-way-coupled models with one-step-ahead smoothing. *Quarterly Journal of the Royal Meteorological Society*, *147*(734), 249–272. https://doi.org/10.1002/qj.3916Ravier, F., Jarvinen, H., Klinker, E., Mahfouf, J. ., & Simmons, A. (2000). The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics. *Quarterly Journal of the Royal Meteorological Society*, *126*, 1143–1170.Sakov, P., & Oke, P. R. (2008). Implications of the form of the ensemble transformation in the ensemble square root filters. *Monthly Weather Review*, *136*(3), 1042–1053. https://doi.org/10.1175/2007MWR2021.1Sawada, Y., Nakaegawa, T., & Miyoshi, T. (2018). Hydrometeorology as an Inversion Problem: Can River Discharge Observations Improve the Atmosphere by Ensemble Data Assimilation? *Journal of Geophysical Research: Atmospheres*, *123*(2), 848–860. https://doi.org/10.1002/2017JD027531Sluka, T. C., Penny, S. G., Kalnay, E., & Miyoshi, T. (2016). Assimilating atmospheric observations into the ocean using strongly coupled ensemble data assimilation. *Geophysical Research Letters*, *43*(2), 752–759. https://doi.org/10.1002/2015GL067238Stanley, Z., Grooms, I., & Kleiber, W. (2021). Multivariate localization functions for strongly coupled data assimilation in the bivariate Lorenz 96 system. *Nonlinear Processes in Geophysics*, *28*(4), 565–583. https://doi.org/10.5194/npg-28-565-2021Sun, T. A., & Hilker, F. M. (2021). Comparison between best-response dynamics and replicator dynamics in a social-ecological model of lake eutrophication. *Journal of Theoretical Biology*, *509*, 110491.

https://doi.org/10.1016/j.jtbi.2020.110491Suzuki, K., Zupanski, M., & Zupanski, D. (2017). A case study involving single observation experiments performed over snowy Siberia using a coupled atmosphere-land modelling system. *Atmospheric Science Letters*, *18*(3), 106–111. https://doi.org/10.1002/asl.730Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, *93*(4), 485–498. https://doi.org/10.1175/BAMS-D-11-00094.1Tondeur, M., Carrassi, A., Vannitsem, S., & Bocquet, M. (2020). On Temporal Scale Separation in Coupled Data Assimilation with the Ensemble Kalman Filter. *Journal of Statistical Physics*, *179*(5–6), 1161–1185. https://doi.org/10.1007/s10955-020-02525-zTrevisan, A., D'Isidoro, M., & Talagrand, O. (2010). Four-dimensional variational assimilation in the unstable subspace and the optimal subspace dimension. *Quarterly Journal of the Royal Meteorological Society*, *136*(647), 487–496. https://doi.org/10.1002/qj.571Wang, X., & Bishop, C. H. (2003). A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. *Journal of the Atmospheric Sciences*, *60*(9), 1140–1158. https://doi.org/10.1175/1520-0469(2003)060<1140:ACOBAE>2.0.CO;2Whitaker, J. S., & Hamill, T. M. (2012). Evaluating methods to account for system errors in ensemble data assimilation. *Monthly Weather Review*, *140*(9), 3078–3089. https://doi.org/10.1175/MWR-D-11-00276.1Yoshida, T., & Kalnay, E. (2018). Correlation-cutoff method for covariance localization in strongly coupled data assimilation. *Monthly Weather Review*, *146*(9), 2881–2889. https://doi.org/10.1175/MWR-D-17-0365.1Zhang, F., Zhang, M., & Hansen, J. A. (2009). Coupling ensemble Kalman filter with four-dimensional variational data assimilation. *Advances in Atmospheric Sciences*, *26*(1), 1–8. https://doi.org/10.1007/s00376-009-0001-8