Subseasonal Forecast Skill Improvement From Strongly Coupled Data Assimilation With a Linear Inverse Model

Gregory J. Hakim[1] ![ORCID], **Chris Snyder[2]** ![ORCID], **Stephen G. Penny[3,4]** ![ORCID], **and Matthew Newman[4]** ![ORCID]

[1]University of Washington, Seattle, WA, USA, [2]National Center for Atmospheric Research, Boulder, CO, USA, [3]Sofar Ocean Technologies, San Francisco, CA, USA, [4]CIRES, University of Colorado Boulder & NOAA Physical Sciences Laboratory, Boulder, CO, USA

Abstract Strongly coupled data assimilation (SCDA), such as using atmospheric observations to update ocean analyses, is critical for properly initializing Earth System models to predict subseasonal to decadal timescales. We show that a Kalman filter with a linear emulator of the coupled dynamics can be used to efficiently assimilate observations with SCDA. A linear inverse model (LIM), trained on 25 years of Climate Forecast System Reanalysis gridded data, is used to assimilate observations daily during an independent 7-year period. SCDA sea-surface temperature (SST) analysis errors are reduced over 20% in global-mean mean-squared error relative to a control experiment where only SST observations are assimilated with an SST LIM. The analysis improvements enhance forecast skill for leads of at least 50 days. In contrast, extratropical Northern Hemisphere 2 m air temperature forecast errors increase for coupled data assimilation in these experiments, despite reduction during the training period.

Plain Language Summary Using observations to consistently initialize a forecast is very difficult with coupled Earth system models due to their enormous computational demand. Here we show that a simplified model can be used to improve forecasts through this initialization process. In particular, we show that this approach allows observations of the atmosphere to be used to estimate sea-surface temperature (SST) and to improve forecasts of SST when compared to an approach that uses only observations of SST. This suggests the potential to improve SST forecasts one to 2 months in the future using such an approach.

1. Introduction

Many problems in weather and climate, such as subseasonal and decadal forecasting, involve initializing coupled Earth System models to predict future states (e.g., Laloyaux et al., 2016; Lea et al., 2015; Penny & Hamill, 2017; Penny et al., 2017; Yang et al., 2013; Zhang et al., 2005). Initialization typically involves using observations of the system to improve an existing estimate, such as a forecast from an earlier time, using data assimilation (DA). In addition to a prior estimate of the state, DA requires an estimate of the errors, which significantly elevates the demands on an already computationally intensive set of calculations. This burden has slowed experimentation and algorithm development in coupled DA (S. G. Penny et al., 2019), and promotes implementations where coupled DA is approximated. Building on recent work in coupled paleoclimate DA (Perkins & Hakim, 2021), we test the hypothesis that a stochastically forced linear emulator of the coupled atmosphere–ocean dynamics can be used for coupled DA to improve subseasonal forecasts. We provide a concise proof-of-concept calculation using a Kalman filter to process observations drawn from reanalysis data to assess the impact of coupled DA on analysis and forecast errors.

In the context of atmosphere–ocean coupled forecasting, strongly coupled DA (SCDA) involves assimilating observations to apply updates across the domain interface. This allows, for example, relatively more abundant atmospheric observations to inform the ocean analysis. To work well, SCDA requires an accurate estimate of the covariance between the atmospheric observations and the ocean variables (e.g., Sluka et al., 2016), which contributes to the aforementioned computational demands. A range of weakly coupled DA (WCDA) approaches approximate SCDA. A common one, which we adopt here, is to use separate DA systems for the atmosphere and the ocean, in order to generate independent analyses in each domain using only the observations in that domain (S. Penny et al., 2019). For WCDA, coupling occurs during the forecast step, when a coupled model evolves the state from the analysis to the next assimilation time. A significant drawback of WCDA is that observations of one

domain do not directly affect the other, which could lead to imbalances between the atmosphere and ocean during the forecast step (Mulholland et al., 2015; Zhang, 2011).

In this study, we use SCDA to generate initial conditions for subseasonal forecasts. To break the computational bottleneck for SCDA, we use a linear inverse model (LIM; Albers & Newman, 2019; Newman et al., 2003; Penland & Sardeshmukh, 1995) for coupled atmosphere–ocean prediction. Although LIMs have long been used for subseasonal prediction (e.g., Albers & Newman, 2019; Newman et al., 2003), we believe the application to SCDA is novel. The LIM is trained on Climate Forecast System Reanalysis (CFSR) data (Saha et al., 2014), and applied to DA experiments using a Kalman filter on observations of both atmosphere and ocean variables. Two novel aspects of this approach, not currently possible with operational subseasonal coupled forecasts systems, are: (a) use of the full covariance matrix, rather than an estimate derived from a small ensemble; and (b) SCDA is performed with a full Kalman filter, rather than an approximation. This proof-of-concept calculation may also be useful for developers of SCDA systems, since LIMs trained on the output of complex models are straightforward to develop for prototype applications and hypothesis testing. For example, since the LIM forecasts are skillful beyond 50 days, this approach may prove useful for improving operational forecasts on subseasonal to seasonal timescales.

2. Methods and Data

Here we describe the data (Section 2.1), linear inverse modeling (Section 2.2), Kalman filtering (Section 2.3), observations and their errors (Section 2.4), and validation and error quantification (Section 2.5).

2.1. Data

All data for this study are taken from the CFSR (Saha et al., 2014), which is defined on global lat–lon grids. Calculations related to training pertain to the period 1979–2003; validation is performed over 2004–2010. Gridded CFSR data is available every 6 hr, which we first average to daily, and then to a running 5-day mean. The seasonal cycle is then defined at each grid point by the leading three Fourier harmonics of the annual cycle averaged over the training period, and removed from the 5-day-average data. The LIM state vector is defined by the following variables: 2 m air temperature (T_{2m}), sea-surface temperature (SST), u and v wind components at 850 hPa (u_{850} and v_{850}), and outgoing longwave radiation (OLR). Each field is truncated to the leading 30 empirical orthogonal functions (EOFs), which are computed from grid point values area weighted by $\cos(\phi)^{1/2}$, where ϕ is latitude. This truncation retains about 50%–60% of the variance for all variables but OLR (33%) (Table S1 in Supporting Information S1). Each variable is standardized after truncation to unit variance. Increasing the number of EOFs leads to overfitting the LIM, although results are insensitive to the exact truncation value.

2.2. Linear Inverse Model

LIMs capture the linear dynamics of anomalies about a chosen mean state, where the anomalies have zero mean, with zero-lag time-mean covariance matrix $\mathbf{C}_0$. For state vector $\mathbf{x}$, a LIM is defined by (e.g., Penland & Sardeshmukh, 1995)

$$\frac{d\mathbf{x}}{dt} = \mathbf{L}\mathbf{x} + \xi, \tag{1}$$

where t is time, $\mathbf{L}$ is a matrix containing the deterministic dynamics, and ξ is a random vector that is uncorrelated in time (but has correlations in the state dimension). The first integral of (1) yields a mapping from any initial condition to the forecast at lag δt,

$$\mathbf{x}(t + \delta t) = \mathbf{G}_{\delta t}\mathbf{x}(t) + \mathbf{n}, \tag{2}$$

where $\mathbf{G}_{\delta t}$ is related to $\mathbf{L}$ by

$$\mathbf{G}_{\delta t} = \exp(\mathbf{L}\delta t) \tag{3}$$

and $\mathbf{n}$ is the integrated contribution from the random vector, ξ. Assuming that the state and the error are uncorrelated, (2) gives a forecast equation for the error covariance

$$\text{cov}(\mathbf{x}(t + \delta t), \mathbf{x}(t + \delta t)) = \mathbf{G}_{\delta t}\text{cov}(\mathbf{x}(t), \mathbf{x}(t))\mathbf{G}_{\delta t}^T + \mathbf{N}_{\delta t} \tag{4}$$

where $\mathbf{N}_{\delta t}$ is the stochastic error covariance specific to the time lag δt. Since (4) applies to any initial condition, including $\mathbf{C}_0$, and the time-mean covariance statistics are stationary, we may solve for $\mathbf{N}_{\delta t}$ algebraically (Penland, 1989, Equation 11):

$$\mathbf{N}_{\delta t} = \mathbf{C}_0 - \mathbf{G}_{\delta t}\mathbf{C}_0\mathbf{G}_{\delta t}^T. \tag{5}$$

We calibrate the LIM using CFSR data during the training period by determining the least squares solution to (2) for $\delta t = 5$ days, and then recover $\mathbf{L}$ from (3). For any lead time δt, $\mathbf{G}_{\delta t}$ is defined from $\mathbf{L}$ by (3) and $\mathbf{N}_{\delta t}$ by (5).

2.3. Kalman Filter

Given a set of observations and a prior estimate of the state mean and covariance, the Kalman filter gives the best linear unbiased estimator for the analysis mean

$$\mathbf{x}_a = \mathbf{x}_f + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}_f), \tag{6}$$

and covariance

$$\mathbf{P}_a = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}_f. \tag{7}$$

The Kalman gain matrix is given by

$$\mathbf{K} = \mathbf{P}_f\mathbf{H}^T\left[\mathbf{H}\mathbf{P}_f\mathbf{H}^T + \mathbf{R}\right]^{-1}, \tag{8}$$

where $\mathbf{x}_f$ and $\mathbf{P}_f$ are the prior mean and covariance, respectively, $\mathbf{x}_a$ and $\mathbf{P}_a$ are the analysis mean and covariance, respectively, $\mathbf{y}$ is a vector of observations having error covariance $\mathbf{R}$, and $\mathbf{H}$ is the observation operator, which maps from the state to the observations. Having solved for $\mathbf{x}_a$ and $\mathbf{P}_a$ at one time, (2) (with $\mathbf{n} = 0$) and (4), respectively, are solved for $\mathbf{x}_f$ and $\mathbf{P}_f$ at the next time that observations are available, which allows (6–7) to be solved again; this process is repeated 1 day at a time for the entire validation period.

2.4. Observations

Observations are drawn once each day for every variable directly from the CFSR gridded data, and estimated from the truncated EOF basis of the LIM,

$$\mathbf{y} = \hat{\mathbf{H}}\hat{\mathbf{x}} = \mathbf{H}\mathbf{x} + \epsilon, \tag{9}$$

where $\hat{\mathbf{x}}$ is the CFSR lat–lon gridded data and $\mathbf{x}$ is the truncation of $\hat{\mathbf{x}}$ to the EOF basis for the variable. The observation operator in the LIM basis, $\mathbf{H}$, is related to the observation operator on the lat–lon grid, $\hat{\mathbf{H}}$, by $\mathbf{H} = \hat{\mathbf{H}}\mathbf{U}$, where the (30) columns of $\mathbf{U}$ are the EOFs. Observations are defined on a regular grid every 20° latitude, and in longitude every $20°/\cos(\phi)$, where ϕ is latitude; observation locations are shown on Figure 3.

Since observations are drawn directly from the CFSR analysis grids by $\hat{\mathbf{H}}$, the observation error covariance in (8), $\mathbf{R}$, is determined completely by representativeness error from the EOF truncation: $\mathbf{R} = \text{cov}(\epsilon, \epsilon)$. However, the truncation error for each variable is not independent of the resolved components of the other variables, so we define a second observation operator by removing a linear estimate of the dependent part of the truncation error,

$$\epsilon = \mathbf{A}\mathbf{H}\mathbf{x} + \mathbf{e}. \tag{10}$$

Matrix $\mathbf{A}$ is found by least-squares regression to minimize var($\mathbf{e}$) during the training period, and the observation error covariance matrix in this case is given by $\mathbf{R} = \text{cov}(\mathbf{e}, \mathbf{e})$. We shall refer to use of this observation operator and observation-error covariance as "regression-$\mathbf{R}$," and use of (9) and $\mathbf{R} = \text{cov}(\epsilon, \epsilon)$ as "control-$\mathbf{R}$." For both cases, the correlated portion of the observation errors results from the spatial dependence of the EOF truncation, and these errors are assumed to be uncorrelated in time.

2.5. Error Quantification and Validation

Errors are measured in observation space and on the full lat–lon grid. The misfit to observations of the 1-day forecast is defined as

$$\mathbf{d} = \mathbf{y} - \mathbf{H}\,\mathbf{x}_f. \tag{11}$$

If all errors are consistently specified, the covariance of the misfits averaged over all validation times should approximately equal the innovation covariance from (8) (e.g., Houtekamer & Mitchell, 1998),

$$\text{cov}(\mathbf{d}, \mathbf{d}) \approx \mathbf{H}\mathbf{P}_f\mathbf{H}^T + \mathbf{R}. \tag{12}$$

We note that, because $\mathbf{H}$, $\mathbf{R}$, $\mathbf{G}$, and $\mathbf{N}$ are constant, (4) and (7) reveal that $\mathbf{P}_f$ and $\mathbf{P}_a$ are as well, so that the right-side of (12) has a single value.

Errors on the lat–lon grid are defined with respect to the truncated EOF basis. We compute the mean-squared error (MSE) of the analyses and forecasts to 50 days lead time during the validation period at all grid points, and area weight the global mean of these values. When comparing different experiments, we compute the percentage change in global-mean MSE. For brevity, we limit the analysis of atmospheric variables to T_{2m}; results for the others are qualitatively similar.

3. Results

We present results for WCDA and SCDA experiments, compared to separate control experiments for the atmosphere and ocean. The atmospheric control experiment consists of cycling DA on atmospheric observations only, and the model used in the forecast step ((2) with $\mathbf{n} = 0$, and (4)) is limited to the atmospheric elements of $\mathbf{L}$. Similarly, the oceanic control experiment consists of cycling DA on SST observations only, and the model used in the forecast step is limited to predicting SST from SST alone. We note that the control experiments are not directly comparable to operational systems for the atmosphere and ocean, since the LIM has no explicit boundary conditions, and atmosphere–ocean coupling is implicit from the training of the LIM variables. The weakly coupled DA (WCDA) experiment is defined by a coupled forecast using the full LIM, with separate DA in the atmosphere and ocean, and no "cross-domain" covariances. Specifically, WCDA sets to zero those entries of $\mathbf{P}_f$ corresponding to covariances between SST and the atmospheric variables. The SCDA experiment is defined by a coupled forecast using the full LIM, and full coupling in the analysis, so that atmospheric observations affect the SST analysis and vice versa. Observations are assimilated once per day to produce an analysis, from which a single forecast is made to 50 days using (2) (with $\mathbf{n} = 0$, representing the average over an infinitely large ensemble). This results in 2556 daily analyses and forecasts during the validation period.

Looking first at the fit of the 1-day forecast to observations, from (12) we compute for each observation i the calibration ratio

$$\text{cov}(\mathbf{d}, \mathbf{d})_{ii} \big/ \left[\mathbf{H}\mathbf{P}_f\mathbf{H}^T + \mathbf{R}\right]_{ii} \tag{13}$$

and summarize the distribution over all observations in violin plots (Figure 1). For the control-$\mathbf{R}$ case, where observation errors are defined simply by truncation error, results for all experiments show smaller errors in the mean analysis than expected from the innovation variance. This over-dispersion is largely corrected when using regression-$\mathbf{R}$, with particularly good calibration for SCDA and all experiments for SST. For T_{2m}, WCDA and the control experiment show over-dispersion for regression-$\mathbf{R}$, with longer tails toward larger values. For the remainder of the results we adopt regression-$\mathbf{R}$.

For T_{2m} forecasts over the 7-year validation period, the global-mean MSE increases rapidly with lead time up to about 10 days, and then increases more slowly before saturating around 20–30 days (Figure 2, top panel). As an illustration of the spatial distribution of errors, Figure 3 shows that at a lag of 10 days, errors are largest over extratropical land masses and the Antarctic coastline. Errors grow more slowly in SST (Figure 2, top panel), and are not near saturation at 50 days. Spatially, SST errors are localized in the eastern tropical Pacific Ocean, and in the midlatitudes (Figure 3).
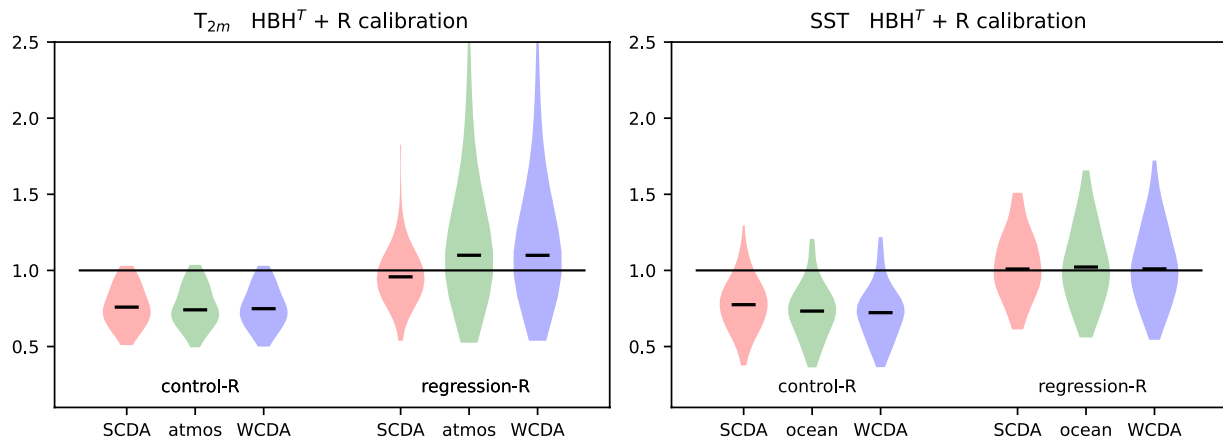
Figure 1. Violin plots of calibration ratio ((13); unitless) over all T_{2m} observations (left) and sea-surface temperature observations (right). Results labeled "control-**R**" apply to the observation error covariance defined exclusively by truncation error (9); those labeled "regression-**R**" remove the component of error linearly predictable from the truncated state using (10). Data assimilation (DA) experiments are strongly coupled data assimilation, weakly coupled DA, ocean only, and atmosphere only. Bold black lines denote the median of each distribution.

Comparing the coupled-DA experiments against the respective control experiments in the atmosphere and ocean reveals that, in the global-mean MSE, WCDA errors are larger than the control case (Figure 2, bottom panel), although the spatial distribution of errors at 10-day lead time reveals improvements in the tropical Pacific (Figure 4, upper left). For SCDA, the global-mean MSE shows little change from the control case in the analysis, and a slow increase in error during the first 20 days before leveling off at about 2% (Figure 2, bottom panel). The error increase derives mainly from the Northern Hemisphere extratropics, with improvements to 20 days in the tropics and Southern Hemisphere (Figure S1 in Supporting Information S1). Improvements in T_{2m} cover a larger fraction of the tropics compared to WCDA, and portions of the Southern Hemisphere, with larger increases in error relative to the control over large areas of the Northern Hemisphere (Figure 4, lower left). For SST, SCDA shows large improvement relative to the control case, over 20% in the global-mean MSE at short leads (Figure 2, bottom panel), and about 40% over large areas (Figure 4, lower right). Our SST improvements are qualitatively
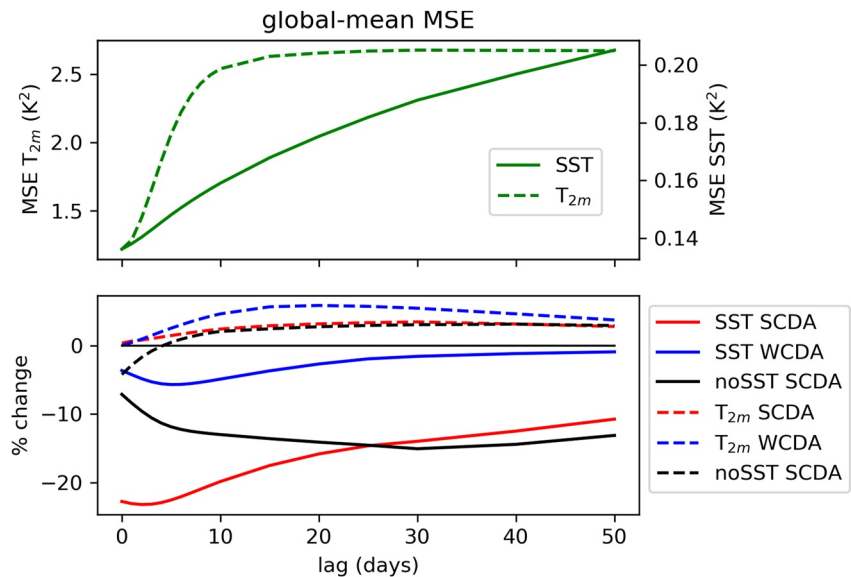

Figure 2. (Top) Global-mean mean-squared error over the validation period for the control experiments where only atmospheric observations are assimilated using the atmospheric component of the linear inverse model (LIM) (dashed line shows T_{2m}) and sea-surface temperature (SST) observations are assimilated using the SST components of the LIM (solid line). (bottom) % change of Strongly coupled data assimilation (red) and weakly coupled data assimilation (blue) from the control experiments for T_{2m} (dashed) and SST (solid).
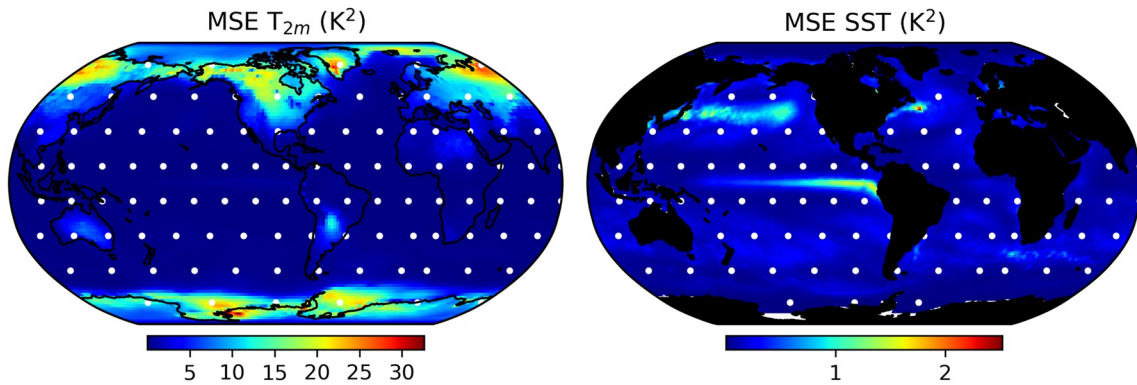
Figure 3. Spatial pattern of mean-squared error over the validation period for 10-day forecasts of (left) T_{2m} and (right) sea-surface temperature. Observation locations are denoted by white dots.

similar to those of Sluka et al. (2016) in the tropical Pacific and near the midlatitudes, although they find larger improvements over the sub-Arctic North Atlantic.

An additional experiment, aimed at exploring the degradation in T_{2m} forecasts for SCDA, is the same as the SCDA case, but excludes SST observations from assimilation (black lines in Figure 2, bottom panel). In this case, SCDA improves upon the control for T_{2m} analyses and forecasts up to about 5 days lead, before slowly converging upon the previous SCDA results at long leads. SST analysis improvements are much smaller than for the SCDA case, but the improvements increase during the forecast, such that they become about as large as the main SCDA case after about 25 days. Further insight is provided by Figures S2 and S3 in Supporting Information S1, which show that SCDA substantially improves both SST and T_{2m} forecasts globally during the training period. While the atmosphere-only case during the training period produces similar results to the validation period (cf. Figure S2 in Supporting Information S1, top panel with Figure 2, top panel), the SST-only case has smaller errors during the training period, suggesting that low-frequency SST variability (e.g., ENSO) is not well sampled during the LIM training period.
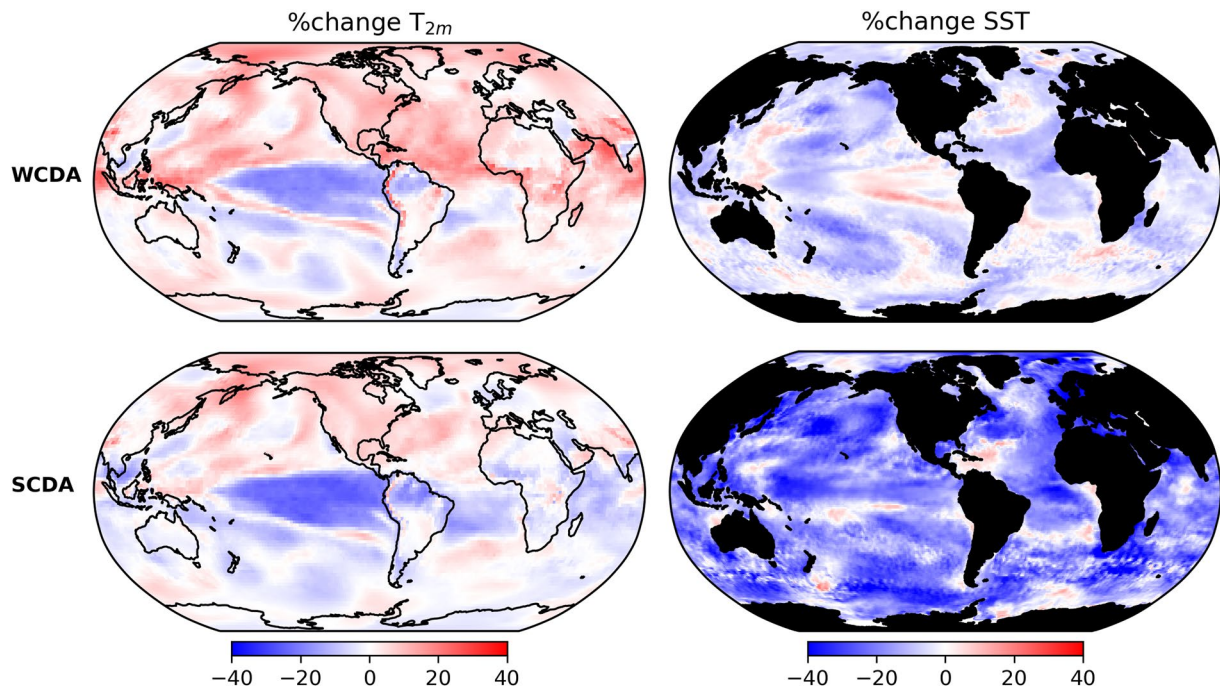

Figure 4. Spatial pattern of % change in mean-squared error during the validation period for 10-day forecasts of (top) weakly coupled data assimilation and (bottom) strongly coupled data assimilation. (left) T_{2m} and (right) sea-surface temperature.

4. Concluding Summary

Strongly coupled DA is essential for Earth System analysis and prediction because it allows observations to consistently influence components of the system other than those directly measured. This is particularly important when the dynamics and/or observing density are very different across components, as is the case for the atmosphere and ocean. Despite this promise, progress on strongly coupled DA has been slowed due to the enormous computational burden of both simulation and assimilation, which has promoted the use of weakly coupled approximations where assimilation is performed independently in each domain. Here we used an empirical model of coupled atmosphere–ocean dynamics, the LIM, as a low-dimensional Earth System emulator to test two approaches to coupled DA. The LIM is skillful to at least 50 days, and highly computationally efficient, so that when integrated with a Kalman filter, it allows us to compare the performance of SCDA to WCDA and single-domain control experiments.

For a LIM calibrated on CFSR data, and DA experiments on observations drawn from CFSR during a 7-year validation period, we find that SCDA produces large improvements to the ocean control case, which has just the ocean component of the LIM and SST observations. SCDA SST analysis errors are reduced by over 20% in the global-mean compared to the control case, and 40% over local regions. In contrast, WCDA SST analysis errors are reduced by only about 5% in the global mean compared to the control, and error reduction nearly vanishes by 50 days into the forecast. Forecasts derived from SCDA analyses show that the SST analysis error reduction persists through the forecast with over 10% improvement compared to the control at 50-day lead. Outside of the tropics, atmospheric forecasts are degraded in WCDA and SCDA relative to the control cases, although SCDA has smaller error than WCDA at all forecast leads. While we show that a portion of this change in performance is due to the assimilation of SST observations, comparison with results during the training period suggests that the degradation appears to result mainly from non-stationary statistics.

We also show that SCDA has the best probabilistic reliability when comparing 1-day forecasts with observations. This result depends on consistent specification of observation error statistics, which in our experiments is determined entirely by representativeness error due to truncation onto the LIM EOF basis. We find that truncation error correlates with the LIM basis, and removing this relationship in the forward operator (**H**) results in SCDA analyses that are well calibrated.

This work provides a proof-of-concept demonstration that low-dimensional Earth System emulators are useful for testing approaches to coupled DA. As such they may provide an important tool for rapidly prototyping experiments before deployment in the full modeling system. Moreover, since the forecasts are skillful beyond 50 days, future research may explore how to use this approach for operational forecasting on subseasonal to seasonal timescales. This work was also limited by the sample size afforded by the CFSR data set, and future work may also explore the role of sample size in training the LIM to resolve coupled atmosphere-ocean variability for DA applications.

Data Availability Statement

Climate Forecast System Reanalysis data may be found at: https://cfs.ncep.noaa.gov/cfsr/. Software developed for this research is available at https://doi.org/10.5281/zenodo.6513266.

References

Albers, J. R., & Newman, M. (2019). A priori identification of skillful extratropical subseasonal forecasts. *Geophysical Research Letters*, *46*(21), 12527–12536. https://doi.org/10.1029/2019GL085270

Houtekamer, P. L., & Mitchell, H. L. (1998). Data assimilation using an ensemble Kalman filter technique. *Monthly Weather Review*, *126*(3), 796–811. https://doi.org/10.1175/1520-0493(1998)126<0796:dauaek>2.0.co;2

Laloyaux, P., Thépaut, J.-N., & Dee, D. (2016). Impact of scatterometer surface wind data in the ECMWF coupled assimilation system. *Monthly Weather Review*, *144*(3), 1203–1217. https://doi.org/10.1175/mwr-d-15-0084.1

Lea, D., Mirouze, I., Martin, M., King, R., Hines, A., Walters, D., & Thurlow, M. (2015). Assessing a new coupled data assimilation system based on the Met Office coupled atmosphere–land–ocean–sea ice model. *Monthly Weather Review*, *143*(11), 4678–4694. https://doi.org/10.1175/mwr-d-15-0174.1

Mulholland, D. P., Laloyaux, P., Haines, K., & Balmaseda, M. A. (2015). Origin and impact of initialization shocks in coupled atmosphere–ocean forecasts. *Monthly Weather Review*, *143*(11), 4631–4644. https://doi.org/10.1175/mwr-d-15-0076.1

Newman, M., Sardeshmukh, P. D., Winkler, C. R., & Whitaker, J. S. (2003). A study of subseasonal predictability. *Monthly Weather Review*, *131*(8), 1715–1732. https://doi.org/10.1175/2558.1

Penland, C. (1989). Random forcing and forecasting using principal oscillation pattern analysis. *Monthly Weather Review*, *117*(10), 2165–2185. https://doi.org/10.1175/1520-0493(1989)117<2165:rfafup>2.0.co;2

Penland, C., & Sardeshmukh, P. D. (1995). The optimal growth of tropical sea surface temperature anomalies. *Journal of Climate*, *8*(8), 1999–2024. https://doi.org/10.1175/1520-0442(1995)008<1999:togots>2.0.co;2

Penny, S., Bach, E., Bhargava, K., Chang, C.-C., Da, C., Sun, L., & Yoshida, T. (2019). Strongly coupled data assimilation in multiscale media: Experiments using a quasi-geostrophic coupled model. *Journal of Advances in Modeling Earth Systems*, *11*(6), 1803–1829. https://doi.org/10.1029/2019ms001652

Penny, S. G., Akella, S., Alves, O., Bishop, C., Buehner, M., Chevallier, M., et al. (2017). *Coupled data assimilation for integrated earth system analysis and prediction: Goals, challenges, and recommendations (Tech. Rep.).* Retrieved from https://library.wmo.int/doc_num.php?explnum_id=10830

Penny, S. G., Akella, S., Balmaseda, M. A., Browne, P., Carton, J. A., Chevallier, M., et al. (2019). Observational needs for improving ocean and coupled reanalysis, S2S prediction, and decadal prediction. *Frontiers in Marine Science*, *6*, 391. https://doi.org/10.3389/fmars.2019.00391

Penny, S. G., & Hamill, T. M. (2017). Coupled data assimilation for integrated Earth system analysis and prediction. *Bulletin of the American Meteorological Society*, *98*(7), ES169–ES172. https://doi.org/10.1175/bams-d-17-0036.1

Perkins, W., & Hakim, G. (2021). Coupled atmosphere–ocean reconstruction of the last millennium using online data assimilation. *Paleoceanography and Paleoclimatology*, *36*(5), e2020PA003959. https://doi.org/10.1029/2020pa003959

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., et al. (2014). The NCEP climate forecast system version 2. *Journal of Climate*, *27*(6), 2185–2208. https://doi.org/10.1175/jcli-d-12-00823.1

Sluka, 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/10.1002/2015gl067238

Yang, X., Rosati, A., Zhang, S., Delworth, T. L., Gudgel, R. G., Zhang, R., et al. (2013). A predictable AMO-like pattern in the GFDL fully coupled ensemble initialization and decadal forecasting system. *Journal of Climate*, *26*(2), 650–661. https://doi.org/10.1175/jcli-d-12-00231.1

Zhang, S. (2011). A study of impacts of coupled model initial shocks and state–parameter optimization on climate predictions using a simple pycnocline prediction model. *Journal of Climate*, *24*(23), 6210–6226. https://doi.org/10.1175/jcli-d-10-05003.1

Zhang, S., Harrison, M., Wittenberg, A., Rosati, A., Anderson, J., & Balaji, V. (2005). Initialization of an ENSO forecast system using a parallelized ensemble filter. *Monthly Weather Review*, *133*(11), 3176–3201. https://doi.org/10.1175/mwr3024.1