

ABSTRACT

Title of dissertation: COVARIANCE LOCALIZATION IN
STRONGLY COUPLED DATA ASSIMILATION

Takuma Yoshida, Doctor of Philosophy, 2019

Dissertation directed by: Professor Eugenia Kalnay
Dept. of Atmospheric and Oceanic Science

The recent development of accurate coupled models of the Earth system and enhanced computation power have enabled numerical prediction with the coupled models in weather, sub-seasonal, seasonal, and interannual time scales as well as climate projection. In the shorter timescales, the initial condition, or the estimate of the present state of the system, is essential for accurate prediction. Coupled data assimilation (DA) based on an ensemble of forecasts seems to be a promising approach for this state estimate due to its inherent ability to estimate flow-dependent error covariance.

Strongly coupled DA tries to incorporate more observations of the other subsystems into an analysis (e.g., ocean observations into the atmospheric analysis) using the coupled error covariances; the covariance is estimated with a finite ensemble, and spurious covariance must be eliminated by localization. Because the coupling strength between subsystems of the Earth is not a simple function of a distance, we develop a better localization strategy than the distance-dependent localization.

Based on the estimated benefit of each observation into each analysis variable,

we first propose the correlation-cutoff method, where localization of strongly coupled DA is guided by ensemble correlations of an offline DA cycle. The method achieves improved analysis accuracy when tested with a simple coupled model of the atmosphere and ocean.

As a related topic, error growth and predictability of a coupled dynamical system with multiple timescales are explored using a simple chaotic model of the atmosphere and ocean. A discontinuous response of the attractor's characteristics to the coupling strength is reported.

The characteristic of global atmosphere-ocean coupled error correlation is investigated using two sets of ensemble DA systems. This knowledge is essential for effectively implementing global strongly coupled atmosphere-ocean DA. We report and discuss common and uncommon features, and the importance of ocean model resolution is stressed.

Finally, the correlation-cutoff method is realized for global atmosphere-ocean strongly coupled DA with neural networks. The combination of static information provided by the neural networks and flow-dependent error covariance estimated by the ensemble improves the atmospheric analysis in our proof-of-concept experiment. The neural networks' ability to reproduce the error statistics, computation cost in a DA system, as well as analysis quality are evaluated.

COVARIANCE LOCALIZATION IN
STRONGLY COUPLED DATA ASSIMILATION

by

Takuma Yoshida

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2019

Advisory Committee:
Professor Eugenia Kalnay, Chair/Advisor
Professor Stephen G. Penny, Co-Advisor
Professor James A. Carton
Professor Brian R. Hunt
Dr. Safa Motesharrei

© Copyright by
Takuma Yoshida
2019

To Rie, for her dedicated support

Acknowledgments

First, I want to express my immense gratitude to my advisor, Prof. Eugenia Kalnay. She has always been the best advisor possible; she allowed me enormous freedom to explore the field following my curiosity, and at the same time, she gave me insightful guidance as the most knowledgeable scientist in the field. My study could not have been more exciting with any other advisors.

I also thank Prof. Steve Penny, who provided many thoughtful ideas, including the use of machine learning technique, as a co-advisor and a leading scientist in the ocean and coupled data assimilation. I also enjoyed his two courses, both of which brought me relevant and enlightening ideas. The other committee members, Prof. James Carton, Prof. Brian Hunt, and Dr. Safa Motesharrei, also gave me valuable guidance and feedback on my project based on their unique expertise.

Many other people supported this project. Ms. Tammy Hendershot has always managed to fit my irregular scheduling constraints into the Department and University curricula; without her timely support, this project should have been delayed indefinitely. Dr. Yun Liu, Dr. Sreenivas Pentakota, Dr. Travis Sluka, Cheng Da, and Eviatar Bach provided computer programs and data essential for this project. Discussion with Prof. Kayo Ide, Prof. Jonathan Poterjoy, Dr. Kriti Bhargava, Dr. Tse-Chun Chen, Chu-Chun Chang, and Luyu Sun as well as people listed above deepened my understanding of broad topics. Also, starting my study at the University of Maryland was impossible without pioneering achievements by Dr. Takemasa Miyoshi and Dr. Daisuke Hotta.

I appreciate the financial support from the Japanese Government (the National Personnel Authority, the Ministry of Land, Infrastructure, Transportation and Tourism, and the Japan Meteorological Agency), Burgers Program, and Ann Wylie Green Fund. The Deepthought2 supercomputer from the Department of Information Technology and various open-source computer programs were essential for this project. The tips shared in *Deep Work* by Cal Newport enhanced my productivity.

My personal life in the United States as a foreigner was enormously supported by Mika Ebisawa, Jim and Michiko Masutani-Purser, Masahiro and Chikako Sawada, and Teppei Kinami as well as my former roommates Kevin Schneider, David Deisenroth, and Rana Karimpour.

Lastly, I acknowledge my family. My wife Rie has perpetually encouraged me and supported my family. I could immerse myself to this project only with her substantial support and patience. I also thank my son Itsuki for providing us peace of minds (and occasional distractions).

Table of Contents

Dedication	ii
Acknowledgments	iii
Table of Contents	v
List of Tables	viii
List of Figures	ix
List of Abbreviations	xvii
List of Symbols	xix
1 Introduction	1
1.1 Coupled problems	1
1.2 Previous efforts on coupled DA	4
1.2.1 Coupled data assimilation products	6
1.2.2 Other global coupled DA efforts	8
1.2.3 Methodological advances and small model experiments	11
1.2.4 Coupled data assimilation not for atmosphere-ocean state	14
1.3 Terminology: background error correlation and other	15
1.4 Theory and approaches of localization	17
1.4.1 Localization in the model space	18
1.4.2 Localization in the observation space	19
1.4.3 Effect of localization	20
1.4.4 Choice of localization length scale	24
1.4.5 Advanced localization methods	28
1.4.6 Localization for applications other than EnKF	30
1.5 Problem statement and outline of this thesis	32

2	The correlation-cutoff method — experiments with a nine-variable coupled model	34
2.1	Introduction	34
2.2	Theoretical analysis	35
2.3	Methods	38
2.3.1	Model	38
2.3.2	Data assimilation method	40
2.3.3	Experimental settings	42
2.3.4	Offline experiment and error statistics	43
2.3.5	Covariance localization	45
2.4	Results	47
2.5	Summary	49
3	Sudden and major changes of dynamics observed in coupled chaotic systems	52
3.1	Background	52
3.2	Brief introduction to Lyapunov exponents and vectors	54
3.2.1	Tangent linear operator and its adjoint	54
3.2.2	The multiplicative ergodic theorem of Oseledets: the existence of Lyapunov exponents and covariant Lyapunov vectors	55
3.2.3	Forward and backward Lyapunov vectors	57
3.2.4	Numerical methods for obtaining Lyapunov exponents and vectors	59
3.3	Methods	62
3.3.1	The coupled model and decoupling	62
3.3.2	Experimental settings	64
3.4	Results	65
3.5	Discussion and remarks	72
4	Implementation of FOAM-LETKF	74
4.1	Fast Ocean Atmosphere Model (FOAM)	74
4.1.1	FOAM specifications	74
4.1.2	FOAM characteristics	79
4.2	Observation network	84
4.3	Data assimilation system: FOAM-LETKF	90
5	Background error statistics of atmosphere-ocean systems	96
5.1	Introduction	96
5.2	Previous efforts on estimating and modeling error correlations	98
5.2.1	Background error correlations of the atmosphere	99
5.2.2	Background error correlations of the ocean	103
5.3	Methods and data sets	104
5.3.1	Ensemble-based estimates of background error correlations	106
5.3.2	FOAM-LETKF WCDA system/data	107
5.3.3	CFS-IITM system/data	108
5.4	Results: background error correlations of FOAM-LETKF	110

5.4.1	Pointwise surface error correlations (FOAM-LETKF)	110
5.4.2	Mean response to single observation assimilation (FOAM-LETKF)	113
5.5	Results: background error correlations of CFS-IITM	117
5.5.1	Pointwise surface error correlations (CFS-IITM)	117
5.5.2	Mean response to single observation assimilation	120
5.6	Discussion	120
5.7	Summary	126
6	Localization modeling with neural networks and assimilation experiments with FOAM	129
6.1	Introduction	129
6.2	Methodological concept: correlation-cutoff with neural networks	130
6.2.1	Overview of the correlation-cutoff method application	130
6.2.2	Mathematical validity as a localization function	133
6.3	Preliminary neural network experiments with toy data	135
6.3.1	Toy correlation models from geostrophic theory	136
6.3.2	Specification of neural networks and training algorithm	138
6.3.3	Explanatory variables for fitting toy data	141
6.3.4	Results and discussion — fitting toy correlation models	142
6.4	Fitting to FOAM ensemble correlations	144
6.4.1	Generation of training data	144
6.4.2	Neural network and training configurations	146
6.4.3	Explanatory variables for regression of FOAM ensemble correlation	147
6.4.4	Regression results and discussion	148
6.5	Assimilation experiments with correlation-cutoff	155
6.5.1	Choice of cutoff function	155
6.5.2	One-year forecast-analysis cycle experiment	157
6.5.3	Computation cost	165
6.6	Concluding remarks and future directions	168
7	Concluding summary	172
7.1	Summary of this thesis	172
7.2	Future direction	174
A	Speedup observation lookup with Octree	176
	Bibliography	183

List of Tables

4.1	Specification of the Fast Ocean Atmosphere Model (FOAM), version 1.5.	75
4.2	Observation networks simulated and assimilated. For vertical profile observations (radiosondes, radiance, and Argo floats), each model level is directly observed; i.e., the observation operator only incorporates horizontal interpolation.	88
4.3	Vertical coordinate of the FOAM-LETKF analysis system.	93
5.1	Settings of the weakly coupled FOAM-LETKF observation system simulation experiment (OSSE) from which ensemble statistics are sampled.	107
5.2	Settings of CFS-IITM reanalysis from which ensemble statistics are sampled.	109
6.1	Subjective comparison of regression methodologies. Here, <i>curse of dimensionality</i> refers to the exponential growth of nonlinear combinations of input variables, which will affect necessary memory space, training samples needed, and evaluation cost. An <i>analytical</i> training means that the regression parameters are obtained by a weighted average of training data without iterative minimization.	132
6.2	Settings of strongly coupled assimilation experiments <i>Control</i> and <i>Cutoff</i> . See text for the differences between <i>Cutoff</i> , <i>Cutoff-nodeep</i> , and <i>Cutoff-nodeep-ocnhalf</i> experiments.	159
A.1	Time spent on observation lookup with and without Octree. The computation time for the slowest process and the mean of all 128 processes are shown. Each number is the average of three experiments. For comparison, the computation time for LETKF core calculation (roughly equivalent to lines 19 to 32 of Algorithm 2) is also shown.	182

List of Figures

1.1	Growing complexity of coupled models used for climate projection. Figure courtesy IPCC WG1 (2014).	2
1.2	An early idea of coupled DA for initialization of seasonal forecasts (originally called as a <i>comprehensive DA system</i>). In the design, a coupled model is employed ($-t_a \leq t \leq 0$) with its ocean part spun up for an extra period ($-t_b \leq t \leq -t_a$). Figure courtesy Miyakoda (1986)	6
1.3	Example of probability density functions (PDFs). The grey shading shows the <i>climatological</i> PDF $p(\mathbf{x})$ of the Lorenz (1963) attractor. The red shading shows an <i>analysis</i> PDF $p(\mathbf{x}_t \mathbf{y}_t)$ that have assimilated two scalar observations shown with the plus sign ($x = 10$ and $z = 20$ at time t ; with uncorrelated error standard deviation of 3 implied) using the climatological distribution as the prior. The blue shading is a <i>forecast</i> PDF $p(\mathbf{x}_{t+0.08} \mathbf{y}_t)$ (0.08 time units after the analysis). All the PDFs shown are two-dimensional projections of three-dimensional density numerically estimated.	15
1.4	Inseparability of insufficient rank and spurious correlation problems. Panel (a) shows an identity matrix \mathbf{I}_{20} . Each of the panels (b-d) shows one of its r -rank approximations \mathbf{C}_r . For r -rank approximation, its matrix square root $\mathbf{X}_r \in \mathbb{R}^{20 \times r}$ is randomly initialized and optimized numerically minimizing the Frobenius norm $\ \mathbf{I}_{20} - \mathbf{C}_r\ _F$. The diagonal elements of \mathbf{C}_r are constrained to unity. Panels (e-h) are the same as panels (a-d) except that the target correlation matrix (e) is more spatially correlated.	22
1.5	Schematic representation of model-space covariance localization. Panel (a) shows an example of a matrix of localization weights \mathbf{C} . Panel (b) shows the original background error covariance matrix \mathbf{P} , and panel (c) shows the localized background error covariance matrix $\mathbf{C} \circ \mathbf{P}$. The matrices are for a 100-variable model with the one-dimensional spatial extent with periodic boundary conditions. Figure courtesy Petrie and Dance (2010).	23

1.6	An example of an atmosphere-ocean coupled localization function (Eq. 1.6). The localization length scales are different in the atmosphere-like fluid ($L_{x A} = 2$ and $L_{z A} = 2000$) and in the ocean-like fluid ($L_{x O} = 0.4$ and $L_{z O} = 200$). The red lines show the $r = 0.5$ contours. Figure courtesy Frolov et al. (2016).	27
1.7	The relative error of ensemble covariance as a function of the true correlation ρ and ensemble size n . The larger the true correlation and the larger the ensemble size, the ensemble correlation becomes more reliable. Figure courtesy Hamill et al. (2001)	29
2.1	(a) Temporal mean of the squared background error correlation for each pair of variables, obtained by an offline LETKF run. (b) Temporal mean of the squared background error correlation for all 81 ordered pairs of variables, in descending order. Note that the correlation drops off after the first 45 pairs of variables, which indicates that <i>ENSO-coupling</i> is optimal.	44
2.2	Covariance localization patterns tested. Allowed background error covariances between components are indicated by shading. Since only the y variables are observed, only the background error covariances indicated by “+” signs are actually used in our experiments.	46
2.3	Temporal mean analysis root-mean-square error (RMSE) for each experiment. The shading indicates the covariance localization pattern used. The errors in the extratropical atmosphere, tropical atmosphere, and ocean are separately shown in each panel. Horizontal lines show the observation errors σ_o^{atm} and σ_o^{ocn} for comparison. Each panel is the result of experiments with (a) 4 members (b) 6 members, and (c) 10 members. Note that the filter diverged in the 4-member <i>Full</i> experiment.	47
3.1	Lyapunov exponents of the 6-variable model. (a): entire spectrum. (b): magnification of the first five exponents. For each value of α , a red dot represents the largest positive exponent of the system; orange, yellow, green, blue, and purple follows.	66
3.2	Sample time series of the models with different α values (written in each panel). The ordinate ranges shown are $x_t, y_t \in [-25, 25]$; $z_t \in [0, 60]$; $X, Y \in [-100, 100]$; and $Z \in [-70, 130]$	69
3.3	Two-dimensional projections of attractors $\mathbf{x}(t)$ for $5,000 \leq t \leq 10,000$ TU with different α values (written in each panel). The color shows the relative frequency of each pair of values on a log scale. Superimposed numbers show mutual information between the two variables in bits. The mutual information is calculated after dividing each variable’s range into 100 bins with regular intervals. The abscissa and ordinate ranges are automatically adjusted for each panel and not shown for conciseness; see Figure 3.2 instead for obtaining a sense of variabilities.	70

3.4	Same as Figures 3.2 and 3.3 but for $\alpha = 0.56$	71
4.1	Monthly temporal anomaly of Niño3.4 regional (5°S-5°N and 170-120°W) average sea surface temperature. The anomaly is defined as the difference from the average of 250 instances on the same day of a year. For visualization purpose, a five-month running mean is applied to the anomaly.	80
4.2	One hundred year average of the ocean temperature. Within the 100-year period, Total 1200 instances on the first day of each month are averaged. Top: sea surface temperature. Bottom: ocean temperature cross-section at the equator. The y-axis of the bottom panel shows depth in meters, and the color bar shows the temperature in degrees Celsius.	81
4.3	Same as Figure 4.1 but only the last 100 years is shown; the anomaly is defined against the 100-year average.	82
4.4	The first two empirical orthogonal functions (EOFs) calculated from the monthly, 100-year time series of the sea surface temperature. Top row: the first and second EOFs normalized by the variance of each grid, i.e., each grid value's temporal correlation to the principal component. Bottom row: The time series of the first and second principal components (PCs).	83
4.5	One-hundred-year average of bottom-level wind fields of different seasons. The bottom model level roughly corresponds to 992.5 hPa. The arrows' length represent the wind speed, with the arrow at the bottom being 15 m/s. Top: boreal winter (January 1st). Bottom: boreal summer (July 1st).	85
4.6	Horizontal distribution of assimilated observation networks listed in Table 4.2. The ship and Argo observations are randomly generated and vary with time. The ship locations are sampled from historical Voluntary Observing Ship program (VOS) locations to simulate its nonuniform distribution while Argo float locations are uniformly random over the model ocean.	87
4.7	The quasi-uniform observation network used for some experiments. (a): atmospheric variables (T, Q, U, V, P_s) are observed in the bottom 12 levels with observation error standard deviations of (1 K, 1 g/kg, 1 m/s, 1 m/s, 1 hPa), respectively. (b): oceanic variables (T, S, U, V) are observed in the top 12 model levels with observation error standard deviations of (0.1 K, 0.01 PSU, 1 cm/s, 1 cm/s), respectively.	89

4.8	Abstract call trees of FOAM after implementation of incremental analysis update (IAU). PCCM3 and OM3 are the names of the component models, and ABL stands for atmospheric boundary layer (parameterization scheme). In FOAM, the ocean model (OM3) is called as one of the physical components of the atmospheric model (PCCM3), and they are sequentially coupled through the exchange of SST and fluxes. In the same level of a tree, tasks are generally processed from left to right. The atmospheric call tree is a modification of Jacob (1997).	95
5.1	Innovation (observation minus background; a proxy to the background error) correlation between (a): height and height at the 500 hPa level and (b): height and transversal (or tangential) wind at the 850 hPa level. The observations are radiosonde over North America, and the background is 6-h forecast by their global model with horizontal resolution T63. Figure courtesy Lönnberg and Hollingsworth (1986).	100
5.2	Increment caused by the assimilation of a positive temperature innovation at 45°N, 100°W at layer 5. Increment of (a): temperature at layer 5, (b): zonal wind at layer 7, and (c): meridional wind at layer 7. Figure courtesy Parrish and Derber (1992).	101
5.3	Modeled vertical background error correlation length when the mixed layer depth is 97 m. Figure courtesy Waters et al. (2015).	104
5.4	Effect of empirical orthogonal functions as background error covariance in the ocean temperature data assimilation. The top and bottom rows represent assimilation experiment of dense and sparse observation networks, respectively. Left: assimilated observations. Center: analysis increment with EOF-based background error covariance. Right: analysis increment with diffusion-based background error covariance. Figure courtesy Weaver et al. (2017).	105
5.5	Pointwise time-mean surface background ensemble correlations. (a): between surface air temperature and sea surface temperature (SST). (b): between surface specific humidity and SST. (c): between precipitation and SST. (d): between precipitation and sea surface salinity. Background precipitation is the value accumulated over the preceding 24-hour window.	111
5.6	Same as Figure 5.5 but for wind and currents. Top and bottom rows show zonal and meridional winds, and left and right columns show zonal and meridional currents, respectively.	112
5.7	Mean pointwise background ensemble correlations of surface variables without direct interactions. (a): air temperature and zonal current. (b): meridional wind and air temperature (both variables in the atmosphere). (c): meridional wind and zonal current (same as Figure 5.6c).	114

5.8	Examples that strongly coupled DA can provide useful information to the other subsystem. Each panel shows mean (RMS) background ensemble correlations in a meridional cross-section by shading (contours) to an observation background (white crosses). (a): atmospheric and oceanic temperature cross-section to surface air temperature at 0°N 120°E, (b): atmospheric and oceanic temperature cross-section to sea surface temperature at 40°S 80°E, (c): atmospheric meridional wind and oceanic zonal current cross-section to surface zonal current at 40°S 80°E, and (d): atmospheric zonal wind and oceanic meridional current cross-section to surface zonal wind at 40°N 180°E.	115
5.9	Same as Figure 5.8, but for examples that strongly coupled DA will not provide additional constraints to the other subsystem. (a): wind speed and oceanic temperature cross-section to surface wind speed at 40°N 180°E, (b): atmospheric temperature and salinity cross-section to the atmospheric temperature at 40°S 80°E.	116
5.10	Same as Figure 5.5 but from CFS-IITM data. Background precipitation is the value accumulated over the preceding 6-hour window.	118
5.11	Same as Figure 5.6 but from CFS-IITM data.	119
5.12	Time-mean background ensemble correlations to a point from CFS-IITM data. Each panel shows mean (RMS) background ensemble correlations in a meridional cross-section by shading (contours) to an observation background. (a and b): temperature fields to atmospheric temperature at 0°N, 120°W and 40°S, 80°E; (c): meridional wind and zonal current fields to zonal current at 40°S, 80°E; and (d): zonal wind and salinity fields to zonal wind at 55°S, 30°W.	121
5.13	Sea surface height (SSH) anomaly in meters from (a): FOAM and (b): CFS. Note the former is from a nature run, while the latter is the 6-hour forecast (background) used in the analysis cycle. Also note these panels show different seasons.	124
5.14	Temporal evolution and anomaly correlation of coupled forced oscillators (Eq. 5.1). Solid and dotted lines represent displacements x_A and x_O . Values of C_O are written in each panel, and other parameters $k_A = k_O = 5$, $f_A = 1$, $f_O = 0.4$, $C_A = 1$ are constant.	125
6.1	Schematic of the correlation-cutoff localization method applied to analysis of coupled geophysical systems, where physical distance alone cannot explain the relevance between an analysis variable and an observable.	131
6.2	Four toy correlation models and their generating functions ($r = \sqrt{x^2 + y^2}$). Shading shows the value z at each horizontal location (x, y) . Contours showing the same quantity are added for visual clarity.	136

6.3	Schematic illustration of the multivariate error correlation functions in northern hemisphere mid-latitude atmosphere. These are generated assuming the perfect geostrophic relation between wind (U, V) and geopotential (Φ). Another assumption is the isotropic and Gaussian-shaped spatial correlation of geopotential errors shown in the upper left panel. Figure courtesy Kalnay (2003).	137
6.4	Schematic of a two-layer feedforward neural network. The green arrows show how input signals propagate to the output. Modified Bishop (2006).	139
6.5	Regression by neural networks to toy correlation models. The left-most column shows the error-free generating function, and the other columns show the regression field with a different set of explanatory variables. Superimposed to each panel are root-mean-squared (RMS) regression errors to the independent validation dataset.	142
6.6	Same to Figure 6.5 but for squared generating functions.	143
6.7	Panels (a, b, d): RMS regression error to validation data set of neural networks to squared ensemble correlations with different set of explanatory variables: $(r, z_{\text{anl}}, z_{\text{obs}})$, $(r, z_{\text{anl}}, z_{\text{obs}}, \phi_{\text{anl}})$, and $(r, z_{\text{anl}}, z_{\text{obs}}, \phi_{\text{anl}}, \cos \theta, \sin \theta)$, respectively. Superimposed numbers are RMS regression error multiplied by 1000 (redundant to color). Panels (c, e): relative regression errors, (b) divided by (a) and (d) divided by (b), respectively. Blue color means improved regression by additional explanatory variable(s). Superimposed numbers are the ratio minus one multiplied by 1000 (redundant to color).	150
6.8	Latitudinal sensitivity of vertical error correlation of U -wind. Panels (a) vertical background ensemble auto-correlation of U -wind reproduced by neural networks; correlations are to the 8th level of FOAM (approximately 500 hPa). Panel (b): vertical error auto-correlation of U -wind to the 11th level of another model (approximately 500 hPa), as a function of latitude, estimated by forecast difference method (the NMC method) in January 1999; figure courtesy Ingleby (2001).	151
6.9	Background ensemble correlation reproduced by a neural network, between surface pressure observation and V -wind field at 60°S . Panels (a) and (b) are regressions to the raw and squared background ensemble correlations, respectively.	152

6.10	Capability of neural networks to reproduce the time-mean background ensemble correlation between the atmosphere and the ocean. Top: time-average of background ensemble correlation, between a hypothetical observation of surface meridional wind at 40°S 80°E (southern Indian ocean) and meridional wind and zonal current meridional cross-sections along 80°E. Bottom: mean background ensemble correlation reproduced by a neural network, between a hypothetical surface meridional wind observation at 40°S and meridional wind and zonal current field, as a function of horizontal distance and the analysis levels. Note that the former is a meridional cross-section, but the latter is a radially symmetric function.	153
6.11	Global RMSE of Control (black) and Cutoff (red) experiments as a function of model levels. Solid and dotted lines show the background (i.e., 24-hour forecast) and analysis errors to the known truth. Panels (a-d) show atmospheric temperature (K), specific humidity (g/kg), zonal wind (m/s), and meridional wind (m/s), respectively. Panels (e-h) show oceanic temperature (K), salinity (PSU), zonal current (cm/s), and meridional current (cm/s), respectively. Out of 360 days, the first 30 days are excluded from averaging.	158
6.12	Same as Figure 6.11 but for Control (black) and Cutoff-nodeep (red) experiments.	160
6.13	Same as Figure 6.11 but for Control (black) and Cutoff-nodeep-ocnhalf (red) experiments.	161
6.14	Atmospheric background (24-hour forecast) RMSE of the Cutoff-nodeep-ocnhalf experiment relative to the Control experiment ($\text{RMSE}_{\text{Cutoff-nodeep-ocnhalf}} - \text{RMSE}_{\text{Control}}$). Blue (red) colors show improvement (degradation) from the Control experiment. Panels (a-e) show surface pressure, temperature, specific humidity, zonal wind, and meridional wind, respectively. Panels (f) and (g) show zonal and meridional winds at 250 hPa (11th model level). Units are written in each panel. Out of 360 days, the first 30 days are excluded from averaging.	163
6.15	Same as Figure 6.14 but for oceanic variables. Panels (a-e) show sea surface height, sea surface temperature, sea surface salinity, surface zonal current, and surface meridional current, respectively.	164

A.1 Horizontal schematic of Octree application on the spherical coordinate. The pink box is one of the three-time-divided boxes, which is 45 degrees in longitude and 22.5 degrees in latitude. The green box is one of the four-time-divided boxes. The green cross, dots, and circle show the *center* of the green box (at 39.375°N 78.75°W), hypothetical observations in the green box, and the *radius* of the box after adding the observations (approximately 943 km), respectively. Note that the observations (green dots) exist only in the intersection of the green box and circle. Queries like the blue (red) circles intersect (does not intersect) with the green circle and may (may not) find observations in the green box. 179

List of Abbreviations

3D-Var	Three-dimensional variational data assimilation
4D-Var	Four-dimensional variational data assimilation
ABL	Atmospheric boundary layer
AMIP	Atmospheric Model Intercomparison Project
BUFR	Binary universal form for the representation of meteorological data
CCM3	Community Climate Model, third version
CERA	Coupled ECMWF Re-Analysis
CERA-20C	Coupled ECMWF Re-Analysis of the twentieth century
CERA-SAT	Coupled ECMWF Re-Analysis of satellite era
CESM	Community Earth System Model
CFS	Coupled Forecast System
CFSR	Climate Forecast System Reanalysis
CPU	Central processing units
DA	Data assimilation
DART	Data Assimilation Research Testbed
EAKF	Ensemble adjustment Kalman filter
ECCO	Estimating the Circulation and Climate of the Ocean
ECMWF	European Center for Medium-range Weather Forecast
ECO-RAP	Ensemble correlations raised to a power (method)
EDA	Ensemble of Data Assimilation
EFSO	Ensemble forecast sensitivity to observations
EnKF	Ensemble Kalman filter
EnOI	Ensemble optimal interpolation
ENSO	El Niño Southern Oscillation
EnSRF	Ensemble square root filter
EnVar	Ensemble-based variational data assimilation
EOF	Empirical orthogonal function
ERA-20C	ECMWF Reanalysis of the twentieth century
ETKF	Ensemble transform Kalman filter
FGAT	First guess at appropriate time
FOAM	Fast Ocean Atmosphere Model
GCM	General circulation model
GFDL	Geophysical Fluid Dynamics Laboratory
GFS	Global Forecast System
GPU	Graphics processing units
IAU	Incremental analysis update
IITM	Indian Institute of Tropical Meteorology
k D-tree	k -dimensional tree (data structure)
LETKF	Local ensemble transform Kalman filter

LETLM	Local ensemble tangent linear model
MPI	Message Passing Interface
MOM	Modular Ocean Model
NCAR	National Center for Atmospheric Research
NCEP	National Centers for Environmental Prediction
NEMO	Nucleus for European Modelling of Ocean
NEMOVAR	Variational data assimilation for NEMO
NMC	National Meteorological Center
OM3	Ocean Model (at University of Wisconsin-Madison) version three
ORA-20C	Ocean Reanalysis of the twentieth century
OSSE	Observation system simulation experiment
PC	Principal component
PCCM3	Parallel implementation of CCM3
PDF	Probability density function
PREPBUFR	Prepared (or quality controlled) observations in BUFR format
PSU	Practical salinity unit
RMS	Root mean squared
RMSE	Root mean squared error
RTPP	Relaxation to prior perturbation (method)
RTPS	Relaxation to prior spread (method)
SCDA	Strongly coupled data assimilation
SSH	Sea surface height
SST	Sea surface temperature
STDV	Standard deviation
TU	(Nondimensional) time unit
UTC	Coordinated Universal Time
VOS	Voluntary Observing Ship (program)
WCDA	Weakly coupled data assimilation

List of Symbols

The list below only contains important symbols. Some symbols may have multiple meanings, which will be clarified in the text. Following the convention, this notation roughly follows the suggestion of Ide et al. (1997).

Operations and qualifiers

$\langle \cdot \rangle$	Temporal mean
\cdot'	Deviation from the temporal/ensemble mean
$\bar{\cdot}$	Ensemble mean
$\tilde{\cdot}$	Variable in ensemble space
\circ	Schur (elementwise) product of matrices
a	Analysis
A, atm	Atmosphere
b	Background (forecast used for data assimilation)
o	Observed
O, ocn	Ocean
t	True
$(\cdot)^T$	Matrix transpose

Integers

D	Number of predictors for a neural network
k	Number of different Lyapunov exponents
K	Ensemble size
L	Number of outputs for a neural network
M	Number of hidden units of a two-layer neural network
n	Number of model variables
p	Number of observations

Real scalars

P_s	Surface pressure
P_{top}	Ocean pressure anomaly at the surface
Q	Specific humidity
S	Salinity
T, t, τ	Time, time index (can be an integer)
T	Temperature
U, u, V, v	Zonal and meridional wind and current
x, y, z	Cartesian coordinate in longitudinal, latitudinal, and vertical directions
λ	Lyapunov exponents

Vectors

b	Bias parameters of a neural network
x	Model state (\mathbb{R}^n), predictors of a neural network (\mathbb{R}^D)
y	Observation (including background/analysis interpolated to the observation space; \mathbb{R}^p), independent variables of a neural network (\mathbb{R}^L)

Matrices

A	Analysis error covariance matrix ($\mathbb{R}^{n \times n}$)
B	Background error covariance matrix ($\mathbb{R}^{n \times n}$)
E	(Covariant) Lyapunov splitting spanned by covariant Lyapunov vectors
H	Linearized observation operator ($\mathbb{R}^{p \times n}$)
I	Identity matrix (shape may be clarified with a subscript)
K	Kalman gain matrix ($\mathbb{R}^{n \times p}$)
M	Tangent linear model ($\mathbb{R}^{n \times n}$)
P	Error covariance matrix in model space ($\mathbb{R}^{n \times n}$)
R	Observation error covariance matrix ($\mathbb{R}^{p \times p}$)
U	Orthogonal eigenspace of an Oseledets matrix, spanned by forward/backward Lyapunov vectors
W	Linear combination matrix in a neural network
X	Ensemble perturbation matrix ($\mathbb{R}^{n \times K}$)
Y	Ensemble perturbation matrix in observation space ($\mathbb{R}^{p \times K}$)
Γ	(Nested) Oseledets subspace
Ξ	Oseledets matrix ($\mathbb{R}^{n \times n}$)
ρ	Localization matrix (also for localization function)

Functions

$f(\cdot)$	Tendency function ($\mathbb{R}^n \rightarrow \mathbb{R}^n$)
$F^{(\cdot)}(\cdot)$	Forward integration operator ($\mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}^n$)
$h(\cdot)$	Forward observation operator ($\mathbb{R}^n \rightarrow \mathbb{R}^p$), activation function of a neural network ($\mathbb{R}^M \rightarrow \mathbb{R}^M$)
$M(\cdot)$	Forward integration operator ($\mathbb{R}^n \rightarrow \mathbb{R}^n$)
$p(\cdot)$	Probability density function ($\mathbb{R}^n \rightarrow \mathbb{R}$)
$P(\cdot)$	Probability measure (sets on $\mathbb{R}^n \rightarrow \mathbb{R}$)

Other

C^k	k times continuously differentiable functions
\mathcal{J}	A set of observation indices
$\mathcal{N}(\mu, \sigma^2)$	Normal distribution with mean μ and variance σ^2
\mathcal{O}	Order of computation cost
ϕ	Empty set

Chapter 1: Introduction

1.1 Coupled problems

Numerical prediction models, in their successful 70-year history, have been incorporating more and more components of the Earth. The first successful numerical weather prediction by [Charney et al. \(1950\)](#) started with filtered quasi-geostrophic equations with only one type of prognostic variable, potential vorticity. Since then, numerical prediction models became to solve primitive equations of the atmosphere, with parameterized physical processes such as radiation and convection. Furthermore, they are gradually internalizing other components: land surface and vegetation, ocean, aerosols, atmospheric chemistry, and sea and land ice ([Figure 1.1](#)). These components, historically approximated to be steady boundary conditions of the atmosphere, are expected to provide extra predictability exceeding the two-week limit previously supposed (e.g., [Shukla, 1998](#)). This trend of internalization may even incorporate the human system, which has now dominant influences onto the Earth system and should be bidirectionally coupled with the Earth system ([Motesharrei et al., 2016](#)).

To maximize prediction skills, it is essential to have both accurate initial conditions that account for all the past observations, and the physical laws coded in the

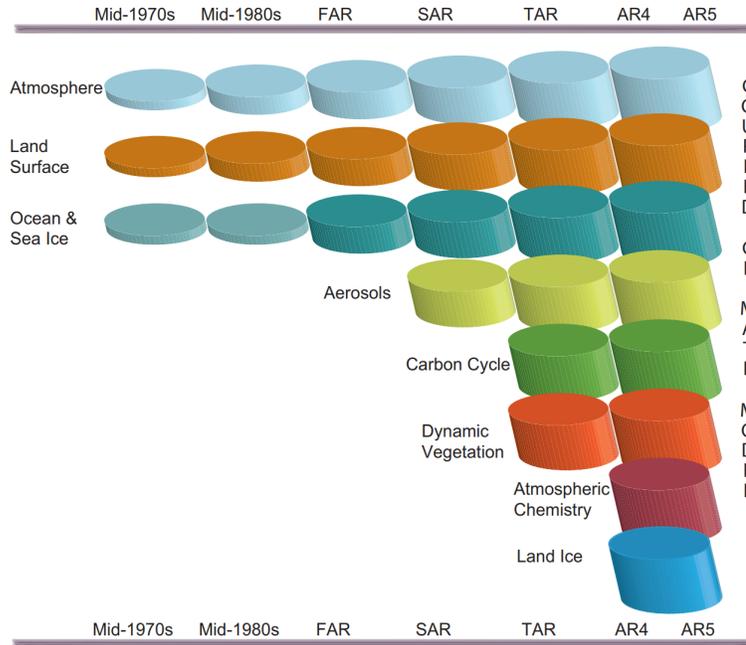


Figure 1.1: Growing complexity of coupled models used for climate projection. Figure courtesy [IPCC WG1 \(2014\)](#).

numerical models. Data assimilation (DA) is a broad class of practices to prepare such initial conditions given a numerical model and observations based on statistical assumptions. At the operational numerical prediction centers, data assimilation (or objective analysis) has been done separately for each Earth’s component (e.g., atmospheric analysis, oceanic analysis, and land analysis). Those uncoupled analyses are disregarding mutual constraints otherwise available in coupled models, possibly undermining the analysis accuracy and consistency. *Coupled data assimilation* refers to a class of methodologies to provide those coupled Earth system models with accurate and consistent initial conditions (e.g., [Penny et al., 2017](#)).

Most straightforward, coupled DA is expected to provide more accurate estimates of the truth because of additional constraints available, especially in under-

observed components (e.g., [Sluka et al., 2016](#)). Another expectation is that coupled DA provides more balanced initial condition (i.e., analysis state well settled onto the coupled model’s attractor) to the coupled model. Although quantitative estimates of degradation caused by coupled imbalance are limited (e.g., [Mulholland et al., 2015](#)), in some cases a combination of uncoupled atmospheric and land initial conditions causes rapid heating of the lower atmosphere and subsequent excitation of external gravity waves (Kenta Ochi, 2015, personal communication). Such fast and imbalanced modes, although they hardly interact with slow modes of interest and are finally dumped by adjustment processes, can cause a spurious observation innovation when the short forecast is used for the background of the subsequent analysis (Section 6.3 of [Daley, 1993](#)).

In addition to providing initial conditions for numerical (weather, seasonal, or interannual) predictions, those objective analyses provide dynamically consistent and uniform estimate of the Earth’s state for academic and industrial purposes. *Reanalysis* is an objective analysis conducted retrospectively with minimum system changes dedicated to these purposes. Therefore, reanalysis products with appropriate coupled DA methodology will provide insights into coupled phenomena, including those caused by complex, non-local interactions. One prominent example is El Niño Southern Oscillation (ENSO), where air-sea flux exchange, atmospheric large scale circulation, and oceanic basin-scale vacillation play a combined role.

However, coupled DA has many practical difficulties. First, the spatiotemporal scales of physical processes are vastly different from each other. The variety of spatiotemporal scales increase the computation cost (e.g., the broad range needs

to be solved at fine resolution) and numerical inaccuracy/instability (e.g., inverting ill-conditioned matrices). It also worsens the signal-to-noise ratio included in the observations through representation error (e.g., [Janjić et al., 2018](#)). Second, the increased complexity often causes superlinear growth of computation cost. For example, for an n -element system, the number of covariances grows at $\mathcal{O}(n^2)$. This inherent complexity makes the problem infeasible even with powerful supercomputers. Third, coupled models often have biased model climatology to the correct climate due to model and parameter errors, which has been obscured by using prescribed boundary conditions.

Therefore, we are going to explore a methodology to overcome the dilemma between the desire for more exhaustive use of information achieved and increasing unreliable information faced in coupled DA. This will be achieved by the improved practice of localization, one of the key element of ensemble-based DA systems ([Section 1.4](#)).

1.2 Previous efforts on coupled DA

Efforts before 2016 are summarized in a white paper based on an International Workshop on Coupled DA ([Penny et al., 2017](#); [Penny and Hamill, 2017](#)), part of which overlaps the review below. Furthermore, we review increasing efforts on understanding and implementing coupled DA after the Workshop.

Although there exist a spectrum of coupling strengths possible in coupled state estimate, we will follow the convention and roughly classify coupled DA algorithms

as follows:

- Uncoupled DA: background state is propagated using uncoupled dynamic models (with prescribed boundary conditions). The analysis is done independently in each subsystem. A pair of atmospheric and oceanic reanalysis products with one-way forcing is also classified as uncoupled (e.g., in the ERA-20C atmospheric reanalysis and the ORA-20C oceanic reanalysis, the latter uses surface fluxes estimated by the former).
- Weakly coupled data assimilation (WCDA): the background state is propagated with a coupled model. However, the analysis is done independently in each subsystem.
- Strongly coupled data assimilation (SCDA): the background state is propagated with a coupled model. The analysis also is done in a coupled way using the coupled background error covariances; atmospheric observations can be used to correct the oceanic background and vice versa. A variety of approximations to partially use the coupled background error covariance information are emerging (quasi-SCDA).

Although the maturity of coupled modeling has enabled recent study and implementation of coupled DA, the idea of coupled DA is, in fact, not new. For example, [Miyakoda \(1986\)](#), in a Global Atmospheric Research Programme publication, suggested using a coupled model with continuous data injection scheme (e.g., nudging) for the initialization of seasonal forecasts (Figure 1.2). The primary motivation

Data Assimilation in an Air-Sea Coupled Model

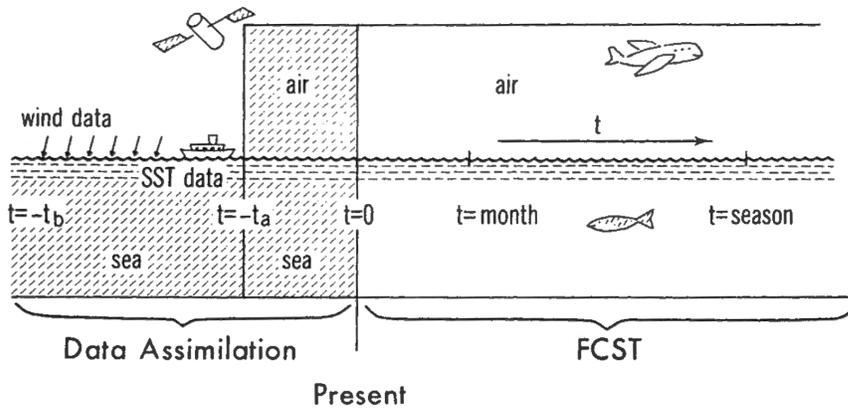


Figure 1.2: An early idea of coupled DA for initialization of seasonal forecasts (originally called as a *comprehensive DA system*). In the design, a coupled model is employed ($-t_a \leq t \leq 0$) with its ocean part spun up for an extra period ($-t_b \leq t \leq -t_a$). Figure courtesy Miyakoda (1986)

for the design is a potentially improved estimate of the tropical sea surface temperature (SST) and atmospheric sub-seasonal modes, and the design is equivalent to the WCDA paradigm in today’s classifications.

1.2.1 Coupled data assimilation products

The world’s leading modeling centers are increasingly implementing atmosphere-ocean coupled DA products.

The National Centers for Environmental Prediction (NCEP) of the United States has been producing the Climate Forecast System Reanalysis (CFSR) and climate forecast based on the analysis (Saha et al., 2010, 2014). The CFSR is classified as a weakly coupled DA system; the atmospheric and oceanic background state is predicted by a coupled model, whereas the analyses are conducted in in-

dividual 3D-Var systems every 6 hours. One of its superiorities over the previous, uncoupled atmospheric reanalysis is that the lead-lag temporal correlations between precipitation and SST closely follow the observed relationship. With the presence of atmosphere-to-ocean feedback in the forward model, the excessive simultaneous correlation observed in the previous analysis was largely corrected¹.

The Met Office of the United Kingdom also implemented a weakly coupled data assimilation system (Lea et al., 2015). With a common window of 6 hours, the atmosphere is updated by a 4D-Var system, the ocean and sea ice are analyzed with a 3D-Var-first-guess-at-appropriate-time (FGAT) system. Its analysis quality is comparable to the uncoupled DA, possibly undermined by the inaccurate diurnal cycle of the ocean model and runoff from their river model, which are prescribed in their uncoupled control.

The European Center for Medium-range Weather Forecast (ECMWF) is producing a few coupled reanalysis products based on their Coupled ECMWF Re-Analysis (CERA) system (Laloyaux et al., 2016). The CERA system combines an atmospheric weakly-constrained 4D-Var and an oceanic 3D-Var-FGAT system with a nonlinear coupled model used in each iteration of the outer loop (quasi-SCDA). A common 24-hour window is used for the atmosphere and the ocean. Even in the absence of coupled background error covariance at the beginning of the window, an observation innovation can correct both the atmospheric and oceanic states with their outer-loop coupling. With a long enough 24-hour window, their outer-loop cou-

¹This improved relationship, however, is also attributed to the SST fields used for the verification (Kumar et al., 2013).

pling is found as skillful as if the coupled background error covariance was explicitly represented by a small ensemble (Laloyaux et al., 2018a). One of their products is CERA-20C, which covers the twentieth century by only assimilating surface pressure and marine observations. The CERA-20C reanalysis showed an improved representation of coupled features such as ocean heat budget and the tropical instability waves (Laloyaux et al., 2018b). The other product is CERA-SAT, where the coupled state from 2008 to 2016 is analyzed using both conventional and satellite-based observations. With an eddy-permitting quarter-degree ocean model, the CERA-SAT reanalysis shows improvement in the tropical analysis relative to the uncoupled control. It also detects unrealistic ocean heat budget primarily constrained by the analysis increments and the SST relaxation rather than the surface heat fluxes, a defect of the coupled model (Schepers et al., 2018). Both of these products provide uncertainty estimate based on the ensemble of data assimilation (EDA) generated by perturbed model physics and observations (10 members each). The coupled forecasts initialized by the CERA system produce smaller initial adjustments than the uncoupled counterparts (Mulholland et al., 2015).

1.2.2 Other global coupled DA efforts

The ocean reanalysis of the Estimating the Circulation and Climate of the Ocean (ECCO) Consortium (Stammer et al., 2004) takes a unique approach, which has something in common with the concept of strongly coupled DA. They use an adjoint-based, strong-constrained 4D-Var approach with a very long (approximately

10 years) analysis window. Its analysis control variables are three-dimensional initial potential temperature and salinity (defined only at the beginning of the 10-year window), as well as daily surface forcing of net heat, net freshwater, and momentum fluxes. An underlying assumption is that the dominant uncertainty resides in these initial and boundary conditions, not the model deficiencies or oceanic internal error growth. Based on a comparison to independent estimates, they find that the global error of net heat flux estimate from an atmospheric reanalysis reaches tens of watts per square meter. They also argue that the estimate of wind stress (whose first guess derives from the atmospheric reanalysis) adjusted by the ocean observations has a smaller bias to wind measurements than that of the original atmospheric reanalysis.

[Sugiura et al. \(2008\)](#) adopted a similar approach to ECCO, using spatially and temporally variable multiplicative adjustment factors of surface fluxes (latent heat, sensible heat, and momentum) as well as initial ocean conditions as control variables of 9-month 4D-Var windows. In their strongly coupled 4D-Var system, both oceanic and atmospheric observations are assimilated after 10-day average, to capture seasonal to interannual variabilities while filtering day-to-day variabilities. They demonstrated that their analysis has predictive skill for the 1997-1998 El Niño for a 1.5-year lead time.

With an ensemble method, [Zhang et al. \(2007\)](#) implemented a weakly coupled ensemble adjustment Kalman filter (EAKF) of the global atmosphere and ocean and explored the importance of multivariate analysis within each fluid. They showed that the use of the atmospheric geostrophic relationship and oceanic temperature-salinity relationship represented by the ensemble are necessary to obtain balanced

and accurate analyses.

[Sluka et al. \(2016\)](#) conducted a strongly coupled atmosphere-ocean observation system simulation experiment (OSSE) with a low-resolution global atmosphere-ocean model and the local ensemble transform Kalman filter (LETKF). Their experiment showed that the assimilation of atmospheric observations, in the absence of ocean observations, reduces the ocean analysis error. [Sluka \(2018\)](#), in Chapter 3, further applied the same methodology to a more realistic coupled model and assimilation of both atmospheric and oceanic observations. They pointed out several challenges to be overcome before strongly coupled DA brings substantial improvements over weakly coupled DA: better representation of diurnal processes at the surface, bias correction of observations, and appropriate localization. They also discussed the directional benefit of SCDA relative to WCDA that the observations of the downwind (i.e., dynamically driven) system have a greater positive impact when directly assimilated to the upwind (i.e., dynamically driving) system than the opposite. They explained that the observations of the upwind system can readily correct the downwind system with WCDA through the dynamical coupling, and there remains less room for improvement by direct assimilation (i.e., SCDA).

[Karspeck et al. \(2018\)](#) developed a weakly coupled EAKF system for the Community Earth System Model (CESM) and the Data Assimilation Research Testbed (DART) at the National Center for Atmospheric Research (NCAR). Although they disclaim their results are not a reanalysis product, their 30-member coupled reanalysis is shown to have skills close to major reanalysis products in several metrics.

[O’Kane et al. \(2019\)](#) focused on the initialization of coupled ensemble forecast.

They showed that their strongly coupled ensemble transform Kalman filter (ETKF) assimilation of oceanic observations improves the ensemble mean analysis compared to ensemble optimal interpolation (EnOI). For perturbations to initialize ensemble forecast, bred vectors with a rescaling interval and an error norm targeting ENSO perturbation outperformed the perturbations generated by ETKF.

1.2.3 Methodological advances and small model experiments

A majority of coupled DA methodologies utilize ensemble-based filtering, where the coupled background error covariance is readily estimated. However, a variety of approaches to strongly coupled variational DA is also emerging.

With a simple coupled model of the atmosphere and the ocean, [Singleton \(2011\)](#) explored several strongly coupled DA methodologies, including ensemble Kalman filter (EnKF), 4D-Var, and ECCO-like 4D-Var. They showed the EnKF's (4D-Var's) preference of shorter (longer) analysis windows. An optimal analysis window is especially relevant to coupled DA, where subsystems with different time scales are analyzed together.

[Han et al. \(2013\)](#) investigated single-media (i.e., weakly coupled) and multiple-media (i.e., strongly coupled) assimilation with a simple atmosphere-ocean coupled model. They showed that the very different timescales between the subsystems render the ensemble-based estimate of error covariance inaccurate, and without thousands of ensemble members, strongly coupled DA degrades the analysis accuracy compared to weakly coupled DA². They also showed a directional preference that

²This result mildly contradicts with recent understanding of deterministic EnSRFs that the

the faster subsystem is more difficult to be improved by the assimilation of observations of slower subsystems.

[Shen et al. \(2018\)](#) conducted strongly coupled EAKF experiments with the two-scale [Lorenz \(2005\)](#) model, showing that the observations of the small-scale subsystem, when directly assimilated to the large-scale subsystem, have more positive impacts. This result supports the claim of [Han et al. \(2013\)](#) with a slightly more realistic model with spatial extent. They also showed that the cross-domain localization is a key for accurate strongly coupled analysis, with a cross-domain localization function derived from the localization function of the subsystem where the observation resides.

[Huntley and Hakim \(2010\)](#) proposed assimilation of time-averaged observations into an atmospheric model to reduce the analysis cost without much degrading the analysis of long-term variabilities. [Tardif et al. \(2014, 2015\)](#) and [Lu et al. \(2015a,b\)](#) further studied assimilation of time-averaged observations for coupled atmosphere-ocean data assimilation with EnKFs. They showed that assimilation of time-averaged atmospheric observations improves the ocean analysis by the reduced detrimental effect of “weather noise” onto the ocean. Their approach is analogous to the assimilation of superobservations, the spatial average of observation innovations used to counteract the representation error (e.g., [Janjić et al., 2018](#)).

The interface solver of [Frolov et al. \(2016\)](#) enables a small subset of atmospheric (oceanic) observations to be assimilated into the ocean (atmosphere) with

analysis accuracy for a deterministic-chaotic model is almost insensitive to the ensemble size if it exceeds the dimension of the system’s unstable subspace.

small additional costs in the variational framework. They showed with a regional atmosphere-ocean coupled model that their interface solver can produce analyses of higher accuracy than exhaustive strongly coupled DA. They also proposed how to incorporate different length scales into a localization function, which will be explained later in Subsection 1.4.4.

Bishop et al. (2017) pointed out that the exact numerical differentiation of a grid-based nonlinear model can be obtained if the ensemble size exceeds the number of variables influenced by each model variable. Their local ensemble tangent linear model (LETLM) can be used for strongly coupled 4D-Var, where the coupled tangent linear model is unavailable in an analytical form to the date.

The convergence of variational DA minimization is sensitive to the conditioning of the error covariance matrices to be inverted. Smith et al. (2018) pointed out that the background error covariance matrix of an atmosphere-ocean coupled system has a large conditional number. They suggested inflating smallest eigenvalues of the background error correlation matrix and applying the model space localization (Subsection 1.4.1) to condition the matrix.

Storto et al. (2018) augmented their ocean 3D-Var cost function with a balance operator, the tangent linearized bulk formulas between the atmospheric mixed layer and ocean variables. The tangent linear approximation is found to outperform statistical regression based on monthly anomaly covariance.

A few researchers try to estimate the background error covariance between the atmosphere and ocean, an integral part of variational data assimilation (Smith et al., 2017; Feng et al., 2018). Those papers will be discussed in Section 5.1.

All the above methodologies focus on either model (\mathbf{x}_b , M , or \mathbf{M}) or background error covariance (\mathbf{B}) coupling. An additional benefit of coupled DA can come from the observation operator (h or \mathbf{H}) coupling. [Geer et al. \(2018\)](#) discussed how today’s uncoupled DA hinders the appropriate use of near-surface radiance observations. They showed that only strongly coupled DA with accurate observation operators (i.e., radiative transfer models) can correctly handle uncertainties.

1.2.4 Coupled data assimilation not for atmosphere-ocean state

Aside from the atmosphere-ocean coupled system, the same methodology can also be applied to other coupled systems. [Lin and Pu \(2018\)](#) thoroughly examined the background error covariance of the atmosphere-land coupled model over the contiguous United States, estimated by the National Meteorological Center (NMC) method ([Parrish and Derber, 1992](#)). They found negative error correlation between near-surface air temperature and soil moisture (-0.20 when spatially and temporarily averaged for July) and positive error correlation between near-surface humidity and soil moisture (0.15 when similarly averaged). They also found that the error correlations are stronger during local summer and daytime and concluded that strongly coupled assimilation of soil moisture observations could improve the near-surface atmospheric state estimate.

The same methodology works for parameter estimation as well as the state estimation problem. [Zhang et al. \(2012\)](#) and [Liu et al. \(2014\)](#) investigated ensemble-based parameter estimation simultaneous to the weakly coupled state estimate for

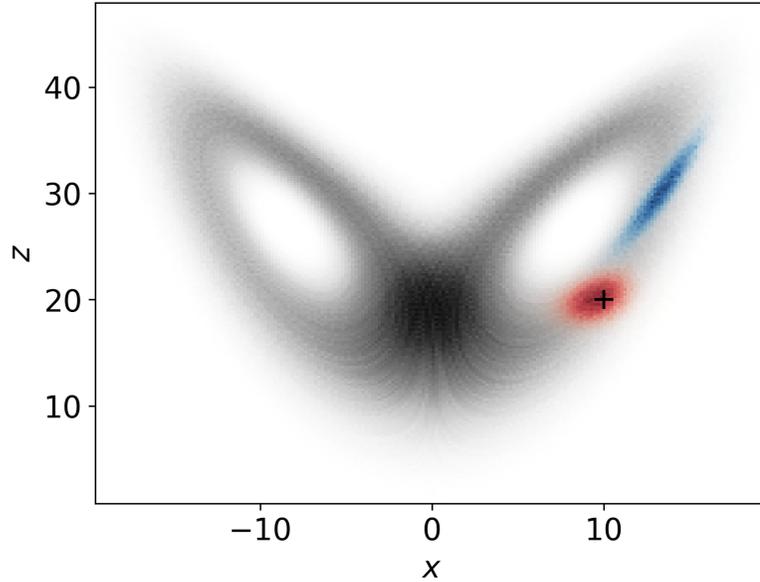


Figure 1.3: Example of probability density functions (PDFs). The grey shading shows the *climatological* PDF $p(\mathbf{x})$ of the [Lorenz \(1963\)](#) attractor. The red shading shows an *analysis* PDF $p(\mathbf{x}_t|\mathbf{y}_t)$ that have assimilated two scalar observations shown with the plus sign ($x = 10$ and $z = 20$ at time t ; with uncorrelated error standard deviation of 3 implied) using the climatological distribution as the prior. The blue shading is a *forecast* PDF $p(\mathbf{x}_{t+0.08}|\mathbf{y}_t)$ (0.08 time units after the analysis). All the PDFs shown are two-dimensional projections of three-dimensional density numerically estimated.

toy and realistic atmosphere-ocean coupled models, respectively. With parameter estimation activated after state estimation error reaching quasi-steady-state, they showed an improved estimate of uncertain parameters dictating atmosphere-ocean coupling.

1.3 Terminology: background error correlation and other

In order to clarify relevant terminologies for this thesis, we show a schematic of climatological, analysis, and forecast probability density functions (PDFs) in [Figure 1.3](#). We summarize important points, some of which are apparent from the figure:

- The *climatological* PDF (grey) is invariant if the dynamical system is autonomous. For autonomous and chaotic dynamical systems, we expect a long time-average of some function of the state to converge to its expectation over this PDF (ergodicity). Therefore, *anomaly* and *temporal* correlations refer to the correlations calculated over this PDF.
- Given an analysis PDF (red; conditioned by past observations), we can obtain the PDF of the negative analysis *error* (truth minus mean analysis) by translating the analysis PDF so that its mean corresponds to the origin; the truth is one realization of the distribution from the Bayesian perspective, and the mean analysis is our best point estimate. The forecast error PDF is similarly obtained by translating the forecast PDF (blue).
- The analysis/forecast error PDFs are generally time-dependent, or flow-dependent. However, we can also consider time-averaged error distributions as an approximation to flow-dependent error distribution (e.g., static background error covariance matrix used in variational methods).
- The PDFs of analysis and short forecast are generally more compact than the climatological PDF. As a result, perturbations (or errors) in the analysis/forecast PDF are less affected by the nonlinearities of the forward model, and the analysis/forecast PDF is closer to a multivariate Gaussian than the climatological PDF. This is why we can handle the errors as if they behave linearly on a tangent space of the attractor; the extended Kalman filter and incremental variational methods are examples.

- The forecast PDF is generally more elongated than the analysis PDF because the error has grown in some directions and decayed in other directions.
- Since the (short) forecast often serves as the background of the subsequent analysis, the term *background* is used as a synonym for the short forecast unless otherwise noted.
- In ensemble-based forecast-analysis systems, the analysis/background PDFs are represented by an ensemble (i.e., a set of realizations of state). Therefore, we use the terms “background *error* correlation” and “background *ensemble* correlation” almost interchangeably to denote the correlation based on the blue density; terms are similarly used for covariance and analyses. However, to be more specific, we sometimes use the term *error* for the (true and unknown) underlying distribution, and the term *ensemble* for its finite-sample approximation.

1.4 Theory and approaches of localization

In the field of data assimilation, *localization* refers to a family of methods to enhance the signal-to-noise ratio by ignoring unphysical background error covariance between distant variables (e.g., [Hamill et al., 2001](#)). As we will shortly see, there exist various approaches to achieve localization. Those approaches originate from different concepts, and they have different characteristics.

1.4.1 Localization in the model space

The non-localized EnKF gain is

$$\mathbf{K} = \mathbf{P}^b \mathbf{H}^T [\mathbf{H} \mathbf{P}^b \mathbf{H}^T + \mathbf{R}]^{-1}, \quad (1.1)$$

where $\mathbf{P}^b = (K - 1)^{-1} \mathbf{X}^b \mathbf{X}^{bT}$ is the background error covariance for background ensemble perturbations \mathbf{X}^b and the ensemble size K , \mathbf{H} is the linearized observation operator, and \mathbf{R} is the observation error covariance matrix (Evensen, 1994).

The straightforward formulation of covariance localization is the model space localization given by

$$\mathbf{K} = (\rho_m \circ \mathbf{P}^b) \mathbf{H}^T [\mathbf{H} (\rho_m \circ \mathbf{P}^b) \mathbf{H}^T + \mathbf{R}]^{-1}, \quad (1.2)$$

where $\rho_m \in \mathbb{R}^{n \times n}$ is a covariance localization matrix in the model space, which is chosen to be positive semi-definite³ so that the Schur product $\rho_m \circ \mathbf{P}^b$ is a valid covariance matrix (e.g., Gaspari and Cohn, 1999; Houtekamer and Mitchell, 2001). The elements of ρ_m , the localization weight between each pair of model variables, is usually in a range $[0, 1]$, and its less-than-unity elements discount the non-diagonal elements of the background error covariance \mathbf{P}^b . However, model space localization is expensive because the convolution of the localization and observation functions is necessary.

³See also Subsection 6.2.2.

1.4.2 Localization in the observation space

The other form of localization, observation space localization, is further divided into two forms, namely, B-localization and R-localization (Shlyueva et al., 2018).

With the B-localization in the observation space, the gain matrix reads

$$\mathbf{K} = [\rho_{o1} \circ (\mathbf{P}^b \mathbf{H}^T)] [\rho_{o2} \circ (\mathbf{H} \mathbf{P}^b \mathbf{H}^T) + \mathbf{R}]^{-1}, \quad (1.3)$$

where $\rho_{o1} \in \mathbb{R}^{n \times p}$ and $\rho_{o2} \in \mathbb{R}^{p \times p}$ are covariance localization matrices in the observation space (Houtekamer and Mitchell, 2001; Lei and Whitaker, 2015). The elements of ρ_{o1} represent the localization weights between each pair of an observation and a model variable, and the elements of ρ_{o2} represent the localization weights between each pair of observations. The B-localization in the observation space often takes the sequential form, where the term $\rho_{o2} \circ (\mathbf{H} \mathbf{P}^b \mathbf{H}^T)$ is represented by using \mathbf{P} that has been updated by the assimilation of previous nearby observations (e.g., Hamill et al., 2001; Whitaker and Hamill, 2002; Anderson, 2003). This sequential assimilation renders the localization weight between observations ρ_{o2} unnecessary, and the B-localization problem in the observation space reduces to defining localization weights between each observation and each model variable (ρ_{o1} , hereafter $\rho \in \mathbb{R}^{n \times p}$).

An early prototype of the other observation-space localization, R-localization, can be found in Houtekamer and Mitchell (1998), who parallelly analyzed each model variable using a subset of observations within a horizontal cutoff radius of each analysis grid. By assimilating only observations which have accurate background error covariances to the analysis variable, this approach has a similar effect

to B-localization. The R-localization for deterministic EnKF is later developed to enhance the degree of parallelism of the algorithm (Ott et al., 2004). Hunt et al. (2007) further proposed multiplying elements of the inverse observation error covariance matrix (\mathbf{R}^{-1}) used in each local analysis by factors between zero and one, so that the influence of each observation smoothly decays from one (nearby observations) to zero (faraway observations). Therefore, the problem of R-localization also reduces to defining localization weights between each observation and each model variable (i.e., $\rho \in \mathbb{R}^{n \times p}$). Note that with the same localization weight matrix in the observation space, the final analysis by the B-localization and the R-localization can differ. Generally, the B-localization further reduces the impact of observation (i.e., the B-localization is tighter with the same ρ ; Greybush et al., 2011; Nerger, 2015).

1.4.3 Effect of localization

The localization has two major effects: to increase the rank of the space to which analysis increment belongs and to suppress the spurious ensemble correlation.

Without localization, an EnKF can find an analysis increment in the space spanned by the ensemble perturbations, which is at most $K - 1$ dimension for a K -member ensemble.

Ng et al. (2011) and Trevisan and Palatella (2011) showed that the number of non-negative Lyapunov exponents, equal to the dimensionality the unstable tangent space of an attractor, is the minimum requirement for the number of independent ensemble perturbations (i.e., $K - 1 \geq D$ is a necessary condition for a deterministic

ensemble filter for the system with D non-negative Lyapunov exponents to converge). It turned out that all the growing error modes need to be constrained by the analysis so that all the conditional Lyapunov exponents of the forecast-analysis system become negative (Penny, 2017).

Indeed, the low-rank and spurious correlation problems are inseparable. Figure 1.4 (a-d) shows the 20-dimensional identity matrix \mathbf{I}_{20} (panel a) and its low-rank approximations numerically obtained. If the rank r is lower than the rank of the underlying correlation matrix (20), spurious correlations must appear as the off-diagonal elements (panels b and c). When the rank is full (20), we can reproduce the target matrix almost error-free with the same method (panel d). If the underlying distribution is more spatially correlated (i.e., if its effective rank is smaller), smaller rank approximations can have considerably smaller spurious correlations (panels e-h). Therefore, the rank problems discussed by Ng et al. (2011) and Trevisan and Palatella (2011) and the spurious correlations apparent in a finite-size ensemble (e.g., Whitaker and Hamill, 2002; Miyoshi et al., 2014) are likely to be different aspects of the same problem.

The spurious background error correlations do not only act as noise to the ensemble mean analysis but also excessively reduce the analysis ensemble variance (i.e., spread; Whitaker and Hamill, 2002). This underestimated variance is harmful to the DA cycle because it leads to an overconfident analysis, which eventually diverges from the truth ignoring the observations. This overconfident analysis ensemble must be compensated with artificial inflation of error covariance (Anderson and Anderson, 1999; Mitchell and Houtekamer, 2000; Zhang et al., 2004; Whitaker

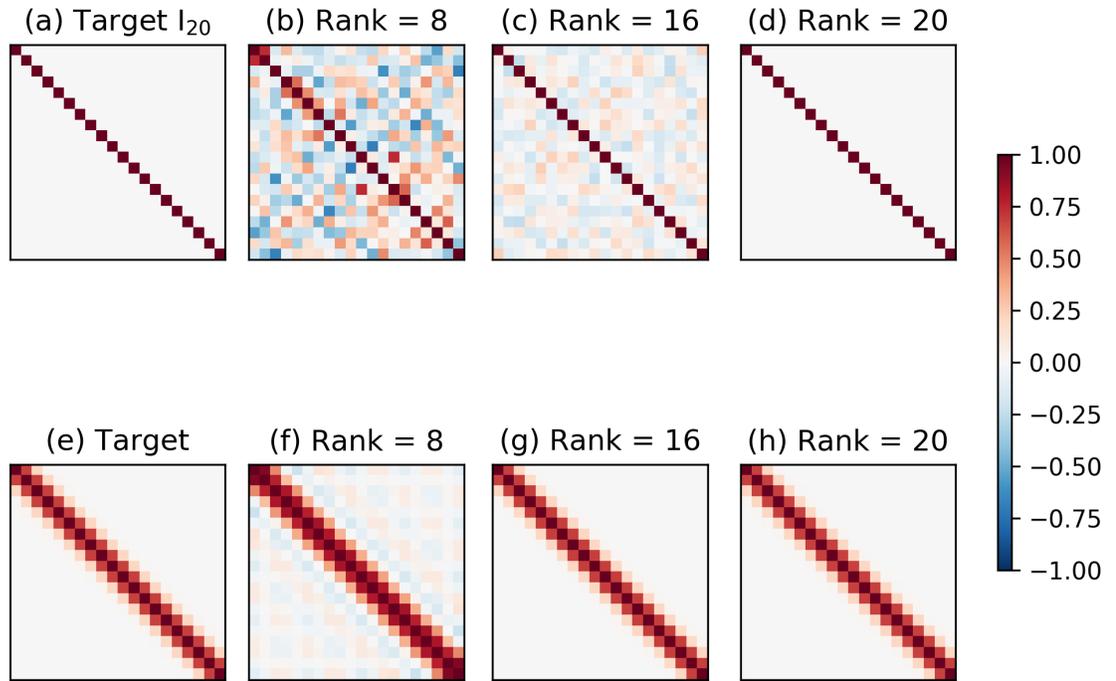


Figure 1.4: Inseparability of insufficient rank and spurious correlation problems. Panel (a) shows an identity matrix \mathbf{I}_{20} . Each of the panels (b-d) shows one of its r -rank approximations \mathbf{C}_r . For r -rank approximation, its matrix square root $\mathbf{X}_r \in \mathbb{R}^{20 \times r}$ is randomly initialized and optimized numerically minimizing the Frobenius norm $\|\mathbf{I}_{20} - \mathbf{C}_r\|_F$. The diagonal elements of \mathbf{C}_r are constrained to unity. Panels (e-h) are the same as panels (a-d) except that the target correlation matrix (e) is more spatially correlated.

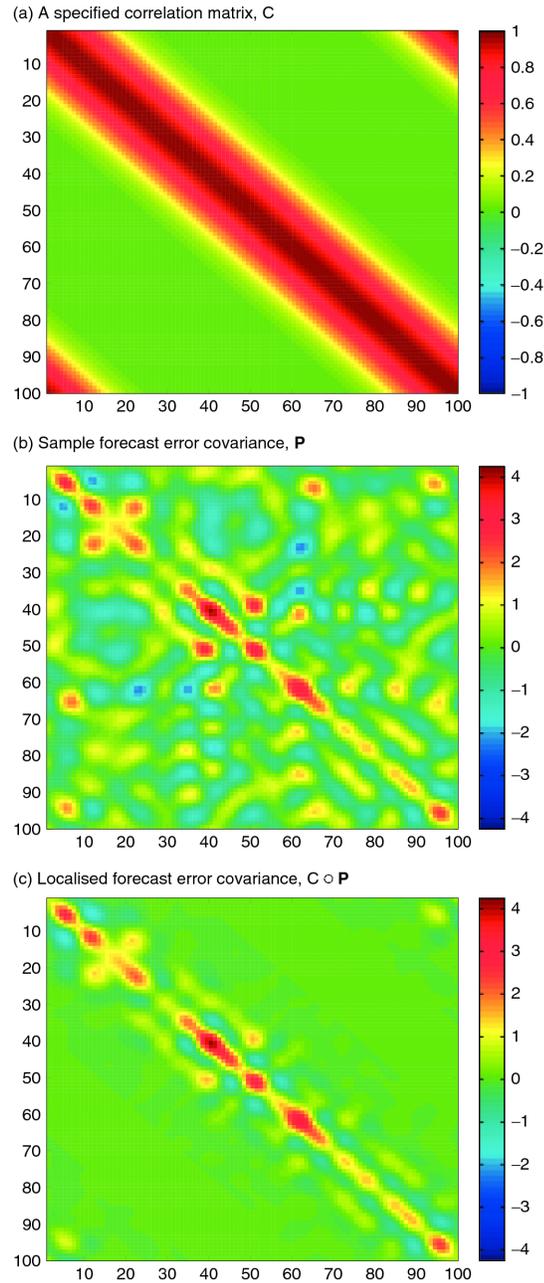


Figure 1.5: Schematic representation of model-space covariance localization. Panel (a) shows an example of a matrix of localization weights \mathbf{C} . Panel (b) shows the original background error covariance matrix \mathbf{P} , and panel (c) shows the localized background error covariance matrix $\mathbf{C} \circ \mathbf{P}$. The matrices are for a 100-variable model with the one-dimensional spatial extent with periodic boundary conditions. Figure courtesy [Petrie and Dance \(2010\)](#).

and Hamill, 2012).

To counteract those problems, global atmospheric analysis with EnKF almost always needs localization. Figure 1.5 schematically shows the effect of model-space localization, where most of spurious, off-diagonal error covariances are filtered out by a Schur product. The 10,240-member LETKF experiment by Miyoshi et al. (2014) is a rare exception with which they captured long-range, non-isotropic background error correlations like a wave train.

1.4.4 Choice of localization length scale

For most atmospheric applications, the localization weight ρ is a static function of the distance between the observation and analyzed variable:

$$\rho_{ij} = \rho(\text{distance between the } i\text{th analysis variable and the } j\text{th observation}). \quad (1.4)$$

A popular choice of this function is the compact support, piecewise polynomial, positive definite function of Gaspari and Cohn (1999):

$$\rho(r) = \begin{cases} -\frac{1}{4}r^5 + \frac{1}{2}r^4 + \frac{5}{8}r^3 - \frac{5}{3}r^2 + 1 & (0 \leq r \leq 1) \\ \frac{1}{12}r^5 - \frac{1}{2}r^4 + \frac{5}{8}r^3 + \frac{5}{3}r^2 - 5r + 4 - \frac{2}{3}r^{-1} & (1 < r \leq 2) \\ 0 & (2 < r), \end{cases} \quad (1.5)$$

where $r = d/c$ is the normalized distance for the distance d and the cut-off parameter c , and $L = c\sqrt{0.3}$ is the localization length scale. We can similarly apply three-dimensional localization by combining the distances in each direction as $r = \sqrt{(d_x/L_x)^2 + (d_y/L_y)^2 + (d_z/L_z)^2}$ where the subscripts x , y , and z represent the

longitudinal, latitudinal, and vertical directions. For most atmospheric application, the function is horizontally isotropic ($L_x = L_y$), but for some ocean applications, the zonal length scale is set a few times larger than the meridional length scale ($L_x > L_y$) to account for the zonally prolonged error correlations.

The choice of localization length scale is affected by multiple factors. [Ying et al. \(2018\)](#) is a comprehensive study on the selection of localization length scale, who showed

- The optimal localization length can be associated with the spatial scale (i.e., correlation length) of the dynamics. On the other hand, the optimal localization length is rather insensitive to the model resolution as long as the key dynamical processes (e.g., synoptic weather) are well represented.
- When multiple scales of dynamics are resolved and analyzed simultaneously, the smaller scale features are better analyzed with smaller localization length, and vice versa.
- With a larger ensemble size, the optimal localization length becomes larger.
- Denser observation network results in shorter correlation length in background error and prefers shorter localization length.

For a coupled atmosphere-ocean problem, where the atmosphere exist above the ocean, the localization function of [Frolov et al. \(2016\)](#) is a natural extension of this distance-dependent localization for a case that the observation and the analysis variable locate in different subsystem (i.e., one is in the atmosphere, subscripts A,

and the other is in the ocean, subscripts O):

$$r = \sqrt{\left[\frac{x_A - x_O}{0.5(L_{x|A} + L_{x|O})}\right]^2 + \left[\frac{y_A - y_O}{0.5(L_{y|A} + L_{y|O})}\right]^2 + \left[\frac{z_A - z_O}{L_{z|A} - L_{z|O}}\right]^2 + \left[\frac{t_A - t_O}{0.5(L_{t|A} + L_{t|O})}\right]^2}, \quad (1.6)$$

where x , y , z , and t represent the longitude, latitude, vertical, and time of the observation or the analysis variable, and L with subscripts represents the localization length scale (for example, $L_{x|A}$ is the localization length scale in the x -direction in the atmosphere). This function allows the existence of different scales in the subsystems, where horizontal localization length scale $L_{\{x,y\}}$ generally takes a smaller value in the ocean than in the atmosphere (Figure 1.6). A limitation of this function is that it assigns 100% weight to the adjacent variables (e.g., subsurface current and surface wind), which will be shown to be an inaccurate representation (Chapter 5).

[Laloyaux et al. \(2018a\)](#), in order to evaluate their outer-loop coupling method, compared the implicit background error correlation represented by outer-loop coupling to the ensemble-based background error correlation. For that purpose, they used the adaptive localization of [Ménétrier et al. \(2015a,b\)](#) to estimate the localization weight from the instantaneous ensemble covariance. However, the limitation of flow-dependent localization methods (to be reviewed) is that they are based on the same instantaneous ensemble information to be localized and are generally noisy due to the small available samples (generally tens of members). Therefore, a localization method that combines the climatological information with the flow-dependent ensemble information is needed to estimate the localization weight more robustly.

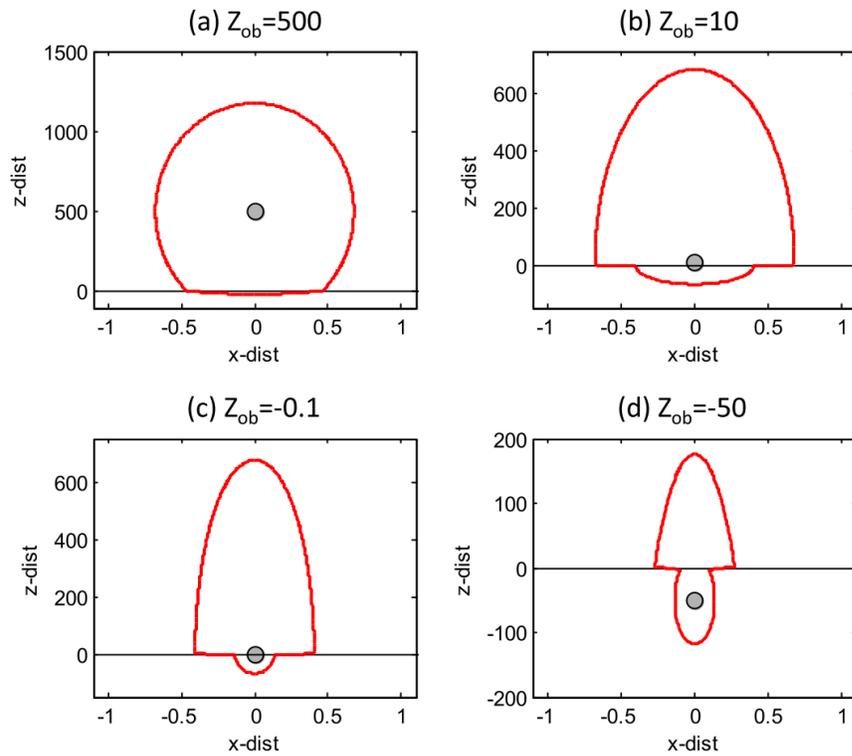


Figure 1.6: An example of an atmosphere-ocean coupled localization function (Eq. 1.6). The localization length scales are different in the atmosphere-like fluid ($L_{x|A} = 2$ and $L_{z|A} = 2000$) and in the ocean-like fluid ($L_{x|O} = 0.4$ and $L_{z|O} = 200$). The red lines show the $r = 0.5$ contours. Figure courtesy [Frolov et al. \(2016\)](#).

1.4.5 Advanced localization methods

In contrast to the distance-dependent localization, there are several localization methods, some of which are adaptive and flow-dependent.

[Anderson \(2007\)](#) proposed a flow-dependent localization method (hierarchical filter) based on regression confidence factor; the localization weight is a function of the degree of coincidence of ensemble covariances between sub-ensembles. This method can deal with a localization function far from a bell shape, and also the temporal localization. [Gasperoni and Wang \(2015\)](#) applied the method to the forecast sensitivity problem.

[Bishop and Hodyss \(2009\)](#) developed the ensemble correlations raised to a power (ECO-RAP) method, another flow-dependent localization. The ECO-RAP method derives a localization weight from a convolution of a smoothing kernel and an ensemble correlation matrix raised to high power (e.g., 6th power). This method is based on an observation that the reliable ensemble correlation is close to be unity, and an assumption that the ensemble correlation near the strong ensemble correlation also is reliable.

The sampling error correlation method of [Anderson \(2012\)](#) converts the absolute value of ensemble correlation and the ensemble size to the localization weight by a look-up table obtained by an offline Monte Carlo simulation.

Those flow-dependent localizations, as well as our correlation-cutoff method (Chapter 2), uses the fact that the stronger ensemble correlations generally are more reliable (Figure 1.7) and also beneficial for reducing the uncertainty.

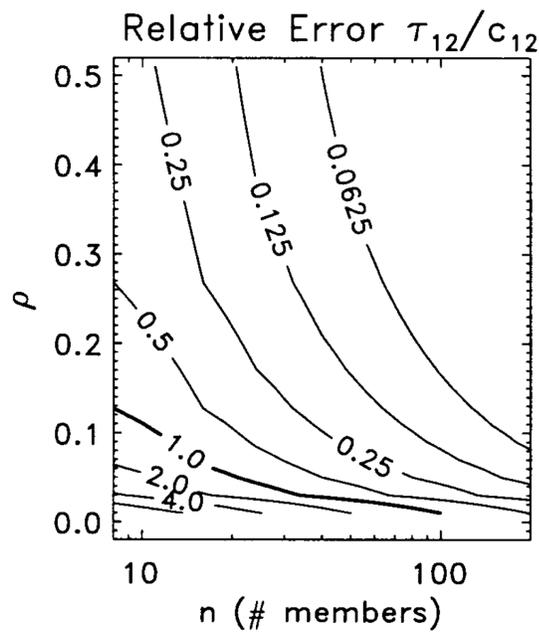


Figure 1.7: The relative error of ensemble covariance as a function of the true correlation ρ and ensemble size n . The larger the true correlation and the larger the ensemble size, the ensemble correlation becomes more reliable. Figure courtesy [Hamill et al. \(2001\)](#)

[Anderson and Lei \(2013\)](#) proposed a method to derive a static localization function ρ for binned pair of observation and analysis grid locations by minimization of the analysis error in an OSSE. Their method is tested with the 40-variable model of [Lorenz \(2005\)](#) to achieve a comparable analysis accuracy to the optimally tuned [Gaspari and Cohn \(1999\)](#) localization.

[Miyoshi and Kondo \(2013\)](#) proposed a multi-scale localization, where the small-scale features are analyzed with shorter localization length while the large-scale features are analyzed with longer localization length, where the scale of features is decomposed by spectral truncation. The multi-scale localization is shown to improve the analysis.

[Kang et al. \(2011\)](#) studied localization between different types of variables, who showed that the analysis of CO₂ concentration and flux should be independent of the observation of temperature, humidity, and surface pressure. Their approach “variable localization” is based on the observation that the CO₂ variables and those dynamical variables are hardly interacting.

1.4.6 Localization for applications other than EnKF

Localization is also beneficial for geophysical analysis and inverse methods other than EnKFs. Some of the important applications are discussed below.

[Fukumori \(2002\)](#) applied the partitioned Kalman filter and smoother to oceanic data assimilation, which ignore the background error cross-covariances between the different partitioned regions. By partitioning, the background error covariance ma-

trix \mathbf{B} reduces from an $n \times n$ matrix to the l sub-matrices of the shape $(n/l) \times (n/l)$ for model dimension n and number of regional partitions l . This partition reduces the overall computation and memory costs by l -fold.

[Penny and Miyoshi \(2015\)](#) and [Poterjoy \(2016\)](#) independently developed localized particle filters, where the resampling weight of each particle is computed only from the information within local patches. The localized particle filters alleviate the exponential growth of computational cost required by particle filters by limiting the dimensionality of the problem while retaining the advantages of particle filters that allows non-Gaussian probability distributions that derive from nonlinear dynamics and observation operators.

Other ensemble-based, data assimilation methods also require localization to suppress spurious correlation and increase the rank of the analysis. [Liu et al. \(2009\)](#) and [Yokota et al. \(2016\)](#) respectively applied model-space and observation-space localizations to the ensemble-based variational data assimilation (EnVar). [Liu and Kalnay \(2008\)](#) developed the ensemble forecast sensitivity to observations (EFSO) method, which uses the ensemble regression between the analysis time and the forecast time to obtain each observation's impact to a forecast at a particular time. Localization is also necessary for EFSO to obtain accurate regression with limited ensemble sizes.

1.5 Problem statement and outline of this thesis

We have reviewed efforts toward strongly coupled DA and the approaches and effects of localization. Most of these methodologies, however, are intended for univariate or tightly coupled multivariate fluids (e.g., horizontal wind and geopotential are tightly coupled in the shallow-water equations), where relevance of an observation to an analysis variable is well approximated by their physical distance. However, as is clear from [Kang et al. \(2011\)](#)'s example, this approximation does not hold for more loosely coupled Earth's subsystems. Furthermore, for Earth system models with growing complexity, it will be unrealistic to experimentally obtain a good compromise between inclusive (strongly coupled) and exclusive (tightly localized or univariate) data assimilation configurations⁴. Although we can employ physical intuitions to obtain plausible configurations, the nonlinear interactions modeled into the numerical models are diverse and complex.

Therefore, the goal of this thesis is to find a localization method that incorporates the variable type dependency as well as spatial separation, in a natural, accurate, and effortless (data-driven) way. Although our experiments focus on the atmosphere-ocean coupled system, we try to find a methodology that is also applicable to other coupled systems.

This thesis is organized as follows. In Chapter 2, we introduce the correlation-cutoff method, an objective criterion of covariance localization. The method is tested

⁴With l_{obs} types of observations and l_{anal} types of analysis variables, there exist $2^{l_{\text{obs}}l_{\text{anal}}}$ ways of variable localization.

with a nine-variable coupled model of the fast atmosphere and the slow ocean. Chapter 3 explores the effect of coupling on chaotic models with very different timescales by examining the Lyapunov exponents and other characteristics of the attractors. Chapter 4 explains our strongly coupled assimilation system for the global atmosphere and ocean that will be used in the following two chapters. In Chapter 5, we investigate the background error correlation of atmosphere-ocean coupled system, an essential quantity for the strongly coupled DA. Two ensemble-based weakly coupled DA systems with different configurations reveal its strong sensitivity to configurations, and suggestions for strongly coupled DA is made. We extend the correlation-cutoff method (Chapter 2) in Chapter 6, by a combination of neural networks and data from the previous chapter. For proof-of-concept experiments, computational feasibility of our method is tested with the global atmosphere-ocean coupled model, and its accuracy is compared to that of a strongly coupled DA experiment with distance-dependent localization. Finally, in Chapter 7, we summarize our findings and make suggestions for future directions.

Chapter 2: The correlation-cutoff method — experiments with a nine-variable coupled model

2.1 Introduction

We have reviewed efforts toward SCDA in Section 1.2, where the relative accuracy of ensemble-based SCDA over WCDA is mixed; some study (e.g., [Sluka et al., 2016](#)) showed improvements, while other (e.g., [Han et al., 2013](#); [Kang et al., 2011](#)) showed that a straightforward implementation of SCDA with finite ensemble degrades the analysis.

These apparently contradicting results raise an important question: *Under what conditions does SCDA provide a better analysis than WCDA?* Therefore, in this chapter, we address this question and propose an offline method to determine which observations should be assimilated into which variables during the analysis update. This methodology, the correlation-cutoff method, is tested with a nine-variable coupled chaotic model of the atmosphere and the ocean.

2.2 Theoretical analysis

In this section, we derive an expression that estimates the analysis uncertainty reduction by the assimilation of each observation.

Here, we assume that only a single observation is assimilated at a time. This is not a strong assumption because, in both Gaussian and Bayesian frameworks, theoretical analysis shows that the observations can be assimilated sequentially without changing the resulting analysis if they have mutually independent error distributions (Houtekamer and Mitchell, 2001; Anderson, 2003). Furthermore, Anderson (2003) pointed out that the observations with correlated errors can be transformed into ones with uncorrelated errors by performing a singular value decomposition on the observation error covariance matrix \mathbf{R} . Note that when considering a second or later observation in sequential assimilation, the background error covariance \mathbf{B} in the following derivation should be replaced with the one used for the assimilation of the observation of interest (i.e., the *analysis* error covariance after assimilating all the previous scalar observations).

We start our derivation from the state-update equations of the Kalman filter (Kalman, 1960). Assuming that the background error covariance \mathbf{B} and observation error covariance \mathbf{R} are correctly specified, and that the observation errors are not correlated with the background errors, the analysis error covariance \mathbf{A} is given by

$$\mathbf{A} = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{B} \tag{2.1}$$

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}, \tag{2.2}$$

where \mathbf{K} is the Kalman gain, \mathbf{I} is the identity matrix, and \mathbf{H} is a linearized observation operator (e.g., Gelb et al., 1974).

Consider the analysis error variance of the i th model variable ($1 \leq i \leq n$),

$$A_{ii} = B_{ii} - \sum_{k=1}^n \sum_{l=1}^p K_{il} H_{lk} B_{ki}, \quad (2.3)$$

where n is the size of the state vector, p is the number of observations, and capital scalars with a subscript denote corresponding matrix elements (e.g., K_{il} is the (i, l) element of \mathbf{K}). The fractional decrease of the uncertainty of the i th model variable is given by

$$\frac{B_{ii} - A_{ii}}{B_{ii}} = \frac{\sum_{k=1}^n \sum_{l=1}^p K_{il} H_{lk} B_{ki}}{B_{ii}}. \quad (2.4)$$

Assuming that there is only one observation ($p = 1$), the observation error variance can be expressed by a scalar as $\mathbf{R} = R = \sigma_{y_0}^2$. With this assumption, we can reduce Eq. (2.4) to

$$\frac{B_{ii} - A_{ii}}{B_{ii}} = \frac{[(\mathbf{HB})_{1i}]^2}{(\mathbf{HBH}^T + R)B_{ii}}, \quad (2.5)$$

where we have used the Kalman gain (Eq. 2.2) and the single observation assumption repeatedly (note that \mathbf{HB} and \mathbf{HBH}^T are a $1 \times n$ matrix and a scalar, respectively). We then rewrite the covariance between the background errors of the observable (δy_b) and the i th model variable (δx_{bi}) as a product of their correlation and standard deviations ($\sigma_{y_b} = \sqrt{\mathbf{HBH}^T}$ and $\sigma_{bi} = \sqrt{B_{ii}}$ for the observable and the i th model variable, respectively) as $(\mathbf{HB})_{1i} = \sigma_{bi} \sigma_{y_b} \text{corr}(\delta x_{bi}, \delta y_b)$. We finally obtain

$$\frac{\sigma_{bi}^2 - \sigma_{ai}^2}{\sigma_{bi}^2} = \frac{\sigma_{y_b}^2}{\sigma_{y_b}^2 + \sigma_{y_0}^2} \text{corr}^2(\delta x_{bi}, \delta y_b), \quad (2.6)$$

where $\sigma_{ai} = \sqrt{A_{ii}}$ is the standard deviation of the analysis error of the i th model variable. A similar derivation for a two-variable example is provided in Hamill et al. (2001). It is informative to compare this equation with the analysis uncertainty reduction in the univariate analysis, in which a single state variable is directly observed by a single observation,

$$\frac{\sigma_b^2 - \sigma_a^2}{\sigma_b^2} = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_o^2}, \quad (2.7)$$

where σ_b^2 , σ_a^2 , and σ_o^2 are the error variances for the background, analysis, and observation, respectively. Equation (2.6) is similar to Eq. (2.7), except that the right-hand side is multiplied by the square of the correlation between the background errors of the analyzed and observed variables.

Equation (2.6) indicates that the relative improvement of the estimate of the state of each model variable by the assimilation of an observation is the product of two quantities: (i) the ratio of the background and total error variances at the observation location (which is large when observations are precise relative to the background) and (ii) the square of the background error correlation between the analyzed and observed variables. This equation also provides a quantitative estimate of the analysis error reduction by SCDA using estimates of the background error covariances between different components.

We hypothesize that in an EnKF, the assimilation of “irrelevant” observations in SCDA degrades the analysis if the detrimental effect of spurious correlations from the limited ensemble size exceeds the expected error reduction from the Kalman filter. Based on this hypothesis, we propose a correlation-cutoff method to localize

strongly coupled EnKFs, in which we only allow strongly coupled assimilation between variables that have strong background error correlations.

2.3 Methods

Local EnKFs such as the LETKF allows us to assimilate different subsets of observations for each model variable. Therefore, we can define a “localization pattern”, in which we select observations to be assimilated into each model variable depending on which component the observation and the model variable are located in (see details in Subsection 2.3.5 and Figure 2.2). In this section, the optimal localization pattern for a simple coupled model is sought by estimating the strength of background error correlation using an offline analysis cycle of the LETKF. Then, the localization pattern is tested in independent LETKF cycles with various ensemble sizes, and the accuracy of the resulting analysis is compared to those obtained with other localization patterns.

2.3.1 Model

We test the correlation-cutoff method with a nine-variable, multi-timescale coupled model proposed by [Peña and Kalnay \(2004\)](#). The governing equations of

the model are as follows:

$$\begin{aligned}
\dot{x}_e &= \sigma(y_e - x_e) - c_e(Sx_t + k_1) \\
\dot{y}_e &= rx_e - y_e - x_ey_e + c_e(Sy_t + k_1) \\
\dot{z}_e &= x_ey_e - bz_e \\
\dot{x}_t &= \sigma(y_t - x_t) - c(SX + k_2) - c_e(Sx_e + k_1) \\
\dot{y}_t &= rx_t - y_t - x_tz_t + c(SY + k_2) + c_e(Sy_e + k_1) \\
\dot{z}_t &= x_ty_t - bz_t + c_zZ \\
\dot{X} &= \tau\sigma(Y - X) - c(x_t + k_2) \\
\dot{Y} &= \tau rX - \tau Y - \tau SXZ + c(y_t + k_2) \\
\dot{Z} &= \tau SXY - \tau bZ - c_zz_t.
\end{aligned} \tag{2.8}$$

This coupled model consists of three components: a fast “extratropical atmosphere” (x_e, y_e, z_e) , a fast “tropical atmosphere” (x_t, y_t, z_t) , and a slow “(tropical) ocean” (X, Y, Z) . Each component is the [Lorenz \(1963\)](#) three-variable model, and they are coupled by coefficients c , c_z , and c_e . The “ocean” is slowed down by a factor of 10 through τ to mimic the slower variations of the ocean. The extratropical atmosphere is only loosely coupled ($c_e = 0.08$) with the tropical atmosphere, and the tropical atmosphere is tightly coupled ($c = c_z = 1$) with the ocean. There is no direct interaction between the extratropical atmosphere and the ocean. The parameters are kept the same as in [Peña and Kalnay \(2004\)](#): $(\sigma, r, b, \tau, c, c_z, c_e, S, k_1, k_2) = (10, 28, 8/3, 0.1, 1, 1, 0.08, 1, 10, -11)$. The model is integrated using the fourth order Runge-Kutta scheme with time steps $\Delta t = 0.01$ non-dimensional time units.

Despite its extreme simplicity, the multi-timescale coupled model shares several important characteristics with the real atmosphere-ocean system and is an excellent testbed for testing ideas for coupled DA problems. The model shows a chaotic behavior with two distinct regimes: the coupled “tropical atmosphere” and “ocean” cycle into a random number of “normal years” (between 2 and 7), interrupted by an “El Niño year” with large negative anomaly in X , before returning to “normal years” (see Figure 2 of [Peña and Kalnay, 2004](#)). Since this asymmetric oscillation neither occurs in the uncoupled “tropical atmosphere” nor “ocean”, it is regarded as an intrinsically coupled instability. Therefore, the model developers called the coupled “tropical atmosphere” and “ocean” as an ENSO-like coupled system. The “extratropical atmosphere”, on the other hand, behaves almost like an individual chaotic system due to its weak coupling with the other components. [Norwood et al. \(2013\)](#) examined the properties of this coupled model and showed that it has two positive, five negative, and two near-zero Lyapunov exponents.

2.3.2 Data assimilation method

We use the LETKF ([Hunt et al., 2007](#)), one of the deterministic implementations of the ensemble Kalman filters classified as ensemble square root filters (EnSRFs; [Tippett et al., 2003](#)). The LETKF allows us to assimilate only a subset of the observations into the analysis of each variable.

According to [Ng et al. \(2011\)](#) and [Trevisan and Palatella \(2011\)](#), the dimension of the subspace spanned by perturbations is at most $K - 1$ for a K -member ensemble.

ble, and this dimension should be equal or larger than the number of non-negative Lyapunov exponents. Hence our coupled model needs at least 5 ensemble members so that the perturbations span the unstable subspace given its 4 non-negative Lyapunov exponents. Therefore, we conduct our experiments with $K = 4, 6$, and 10 ensemble members. With 10 members, the ensemble perturbations can span the entire 9-dimensional model space, so it represents a case with sufficient members. We expect the 4-member experiment to represent a case with insufficient members, and the 6-member experiment to represent an intermediate situation.

Although we have also conducted some experiments with 100 members, the resulting temporal mean analysis RMS error is not qualitatively different, and only the results with $K = 4, 6$, and 10 are shown below. Note that the LETKF is designed to provide the same analysis mean and analysis error covariance matrix as those of the extended Kalman filter for linear forward operators if ensemble size is sufficient to factorize the background error covariance matrix (Hunt et al., 2007). However, the premises may be violated if the model is biased or stochastic, or if the nonlinearity is significant (i.e., the errors are too large to neglect the second and higher order terms in the Taylor expansions of the nonlinear forward operators). In these difficult situations, a larger ensemble size will be beneficial because the sampling of the stochastic or nonlinear error growth becomes more accurate. The insensitivity of EnSRF’s averaged analysis error to excessive ensemble size is thoroughly discussed in Sakov and Oke (2008).

For covariance inflation, we use the adaptive multiplicative inflation of Wang and Bishop (2003). The diagnosed inflation factor Δ° is first limited within $0.9 \leq$

$\Delta^\circ \leq 1.2$ and then temporarily smoothed with a forgetting factor $\kappa = 1.01$ (we follow the notation of [Li et al., 2009](#)).

2.3.3 Experimental settings

We test our method by performing identical twin experiments. The model (Eq. 2.8) is started from random initial conditions and spun up for 25,000 time steps before saving the subsequent 75,000 time steps as the truth. Observations are produced by adding Gaussian noise to the truth with a mean of zero and standard deviation of $\sigma_o^{\text{atm}} = 1.0$ and $\sigma_o^{\text{ocn}} = 5.0$. The observations are available once every 8 time steps, and only one variable in each component (y_e, y_t, Y) is observed to simulate a sparse observation network. We use the y -variables here because the observations of y are the most informative when assimilated in the three-variable Lorenz model ([Yang et al., 2006](#)).

The ensemble members are initialized with random numbers (with different random seeds from the one used for the truth) and spun up for 25,000 time steps before starting the analysis cycle so that the background ensemble members for the first analysis are random samples on the model's attractor. Analysis experiments are conducted for the subsequent 75,000 time steps, the same period as the one for which we have saved the truth and the observations. The analysis is updated every 8 time steps, and therefore, the observations are only available at the end of each window. Within the 75,000 time steps (9,375 analysis windows), only the last 50,000 time steps (6,250 analysis windows) are used for calculating the background error

correlation and the analysis error in the following subsections because we are only interested in the filter performance after its initial transient.

2.3.4 Offline experiment and error statistics

We first conduct an *offline experiment* to obtain the error statistics of the model. For this purpose, we use the same analysis system as discussed in the previous subsection but with the truth, observations, and initial ensemble members independent from the main experiments. We use the fully coupled ETKF (*Full* pattern in the following subsection) with $K = 10$ members for this offline run.

For each pair of model variables x_i and x_j , we first calculate an instantaneous background ensemble correlation at each analysis time t :

$$\text{corr}_{ij}(t) = \frac{\sum_{k=1}^K [x_{ki}(t) - \bar{x}_i(t)][x_{kj}(t) - \bar{x}_j(t)]}{\sqrt{\sum_{k=1}^K [x_{ki}(t) - \bar{x}_i(t)]^2} \sqrt{\sum_{k=1}^K [x_{kj}(t) - \bar{x}_j(t)]^2}}, \quad (2.9)$$

where $x_{ki}(t)$ is the i th model variable of the k th ensemble member at time t , and $\bar{x}_i(t)$ is the ensemble mean of the i th model variable at time t . Then we obtain the temporal mean the squared background error correlation for each pair (x_i, x_j) :

$$\langle \text{corr}_{ij}^2 \rangle = \frac{1}{T} \sum_{t=1}^T \text{corr}_{ij}^2(t), \quad (2.10)$$

where $T = 6,250$ is the number of assimilation windows used to estimate the error statistics.

Figure 2.1 shows the mean of squared background error correlation for each pair of variables. In this model there are only weak error correlations ($\langle \text{corr}^2 \rangle < 0.03$) between the “extratropical atmosphere” and the other components, whereas the errors in the “tropical atmosphere” and the “ocean” are more strongly correlated

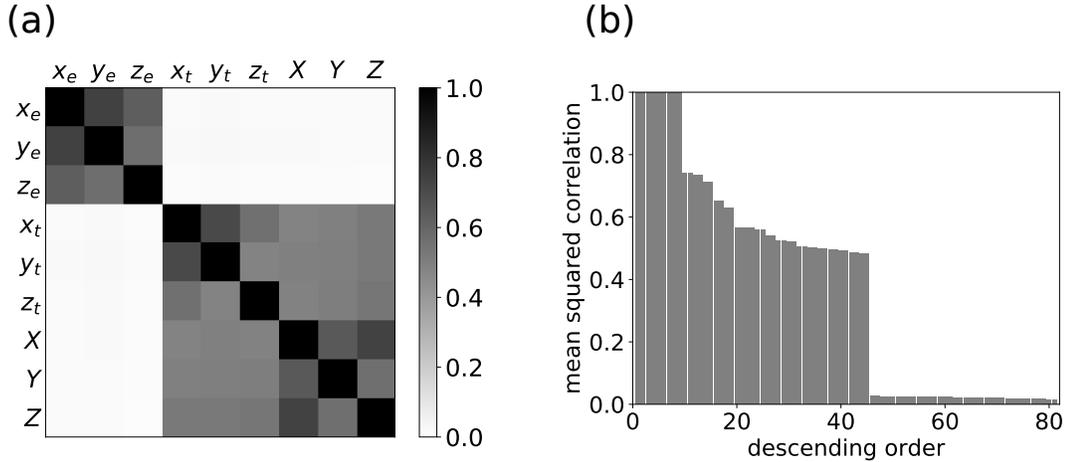


Figure 2.1: (a) Temporal mean of the squared background error correlation for each pair of variables, obtained by an offline LETKF run. (b) Temporal mean of the squared background error correlation for all 81 ordered pairs of variables, in descending order. Note that the correlation drops off after the first 45 pairs of variables, which indicates that *ENSO-coupling* is optimal.

($\langle \text{corr}^2 \rangle \sim 0.5$). These offline statistics suggest ignoring the background error covariance between the extratropical atmosphere and the other components and using only the background error covariance between the tropical atmosphere and the ocean when performing data assimilation.

The use of time-mean *squared* error correlation instead of *time-mean* error correlation is supported by a thought experiment. See the blue background PDF of Figure 1.3, which shows a large positive error correlation between x and z , two state variables of the Lorenz (1963) model. If we take time-mean of background error correlation, both on the right and left lobes, the mean will be near-zero because of the symmetry against the $x = 0$ plane. However, even if the time-mean background error correlation between x and z is near-zero, the background errors of these two variables are still relevant to each other; they are sometimes positively correlated

and sometimes negatively correlated. This average strength of error coupling is captured by the mean squared error correlation.

Let us conduct another thought experiment for a more realistic system. Imagine a point vortex with uncertain strength exists on a two-dimensional domain, and its location is accurately known as a function of time. Then we think of error correlation of zonal and meridional winds at the origin. Due to the strength uncertainty, the zonal and meridional wind errors are negatively correlated when the vortex is in the first or third quadrant; the errors are positively correlated when the vortex is in the second or fourth quadrant. If the vortex location (over a period) is symmetrically distributed around the origin, then the time-mean error correlation of zonal and meridional winds at the origin tends to zero. However, these wind errors are still relevant to each other. Again, mean squared error correlation will be able to apprehend that these errors are correlated at each instance.

2.3.5 Covariance localization

We test the five covariance localization patterns shown in Figure 2.2.

- *Full* is the standard SCDA in which every observation is assimilated into the analysis of every state variable.
- *Adjacent* uses the background error covariance only between directly interacting components. The background error covariances between the extratropical atmosphere and ocean are therefore ignored.

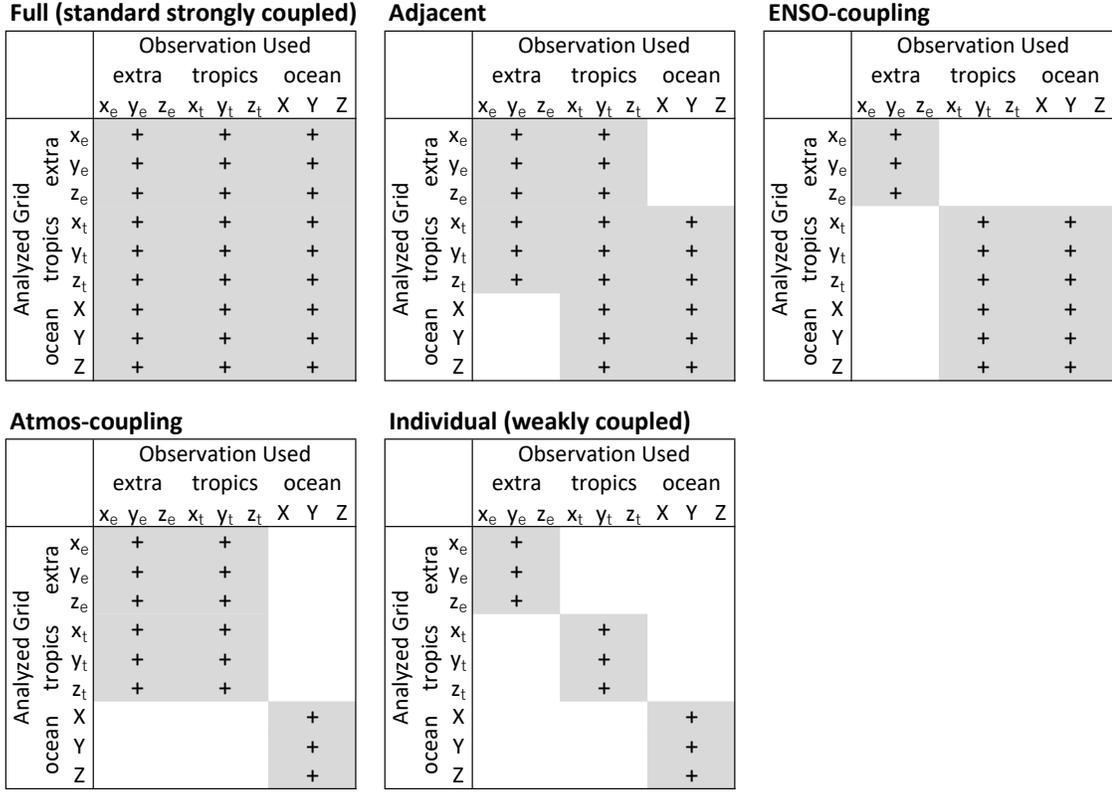


Figure 2.2: Covariance localization patterns tested. Allowed background error covariances between components are indicated by shading. Since only the y variables are observed, only the background error covariances indicated by “+” signs are actually used in our experiments.

- *ENSO-coupling* is the pattern suggested by our theoretical analysis and the offline experiment. The observations of the ENSO-like coupled system (i.e., the tropical atmosphere and the ocean) are mutually assimilated, but the extratropical atmosphere is analyzed individually.
- *Atmos-coupling* analyzes the extratropical and tropical atmosphere together but the ocean separately. This pattern separately analyzes the fast and slow components. This is same as the *subsystem localization* tested by Singleton (2011).

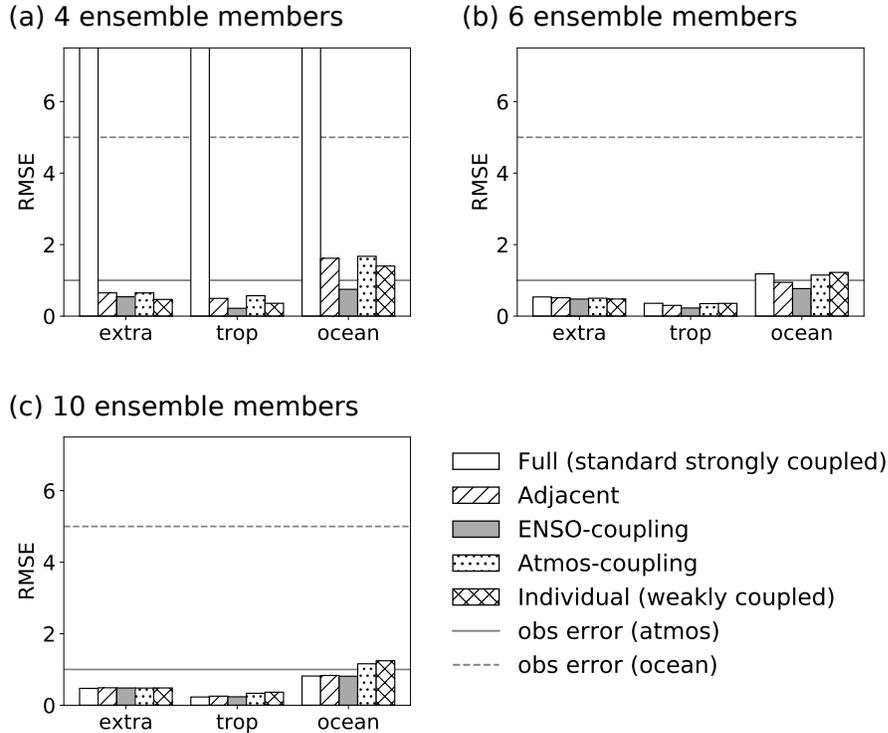


Figure 2.3: Temporal mean analysis root-mean-square error (RMSE) for each experiment. The shading indicates the covariance localization pattern used. The errors in the extratropical atmosphere, tropical atmosphere, and ocean are separately shown in each panel. Horizontal lines show the observation errors σ_o^{atm} and σ_o^{ocn} for comparison. Each panel is the result of experiments with (a) 4 members (b) 6 members, and (c) 10 members. Note that the filter diverged in the 4-member *Full* experiment.

- *Individual* analyzes each component individually. The background is updated by the coupled model, but the analysis step is individually implemented for each component, which is equivalent to WCDA for this three-component model.

2.4 Results

The resulting analysis errors are plotted in Figure 2.3.

Full (standard SCDA) performs worse than *Individual* (WCDA) when the ensemble size is small ($K = 4, 6$). The negative impact of SCDA in this case is expected given the rank deficiency and the resulting spurious correlations. As the ensemble becomes larger, *Full* gradually becomes better, whereas the analysis accuracy of *Individual* is not so sensitive to the ensemble size. This result suggests the importance of a larger ensemble for successful implementation of the strongly coupled LETKF.

As Eq. (2.6) indicates, the assimilation of any type of observations with the Kalman filter will, on average, not increase the analysis uncertainty if the background and observation error covariance matrices are accurately specified. When the ensemble size is sufficient, the assimilation of observations whose background error is uncorrelated with that of an analysis variable will be neither beneficial nor harmful since the ETKF converges to the Kalman filter. The number of ensemble members needed for successful implementation of SCDA will be highly model dependent and may be affected by other factors like the use of covariance inflation.

The *ENSO-coupling* pattern suggested by the correlation-cutoff method performs best in essentially all experiments, as we expected. In comparison to *Individual* (WCDA) and *Atmos-coupling*, *ENSO-coupling* is superior regardless of the ensemble size. The inferior performance of *Individual* and *Atmos-coupling* is noticeable in the tropical atmosphere and the ocean, between which these inferior patterns ignore strong background error covariances. This comparison shows the importance of including the background error covariances between the tropical atmosphere and the ocean in this model. In contrast to *Full* (standard SCDA) and *Adjacent*, *ENSO-coupling* performed well with the smaller ensembles ($K = 4, 6$), though all patterns

using the background error covariance between the tropical atmosphere and the ocean performed similarly well when the ensemble size was large enough ($K = 10$). The inferior performance of *Full* and *Adjacent* with insufficient ensemble size is seen in all components and can be attributed to the use of spurious correlations between the extratropical atmosphere and the other components.

These comparisons support the *ENSO-coupling* pattern, or the decision of ignoring the weak background error covariance between the “extratropical atmosphere” and the other components while considering the strong covariance between the “tropical atmosphere” and the “ocean”, as suggested by Figure 2.1.

2.5 Summary

We first derived a simplified equation for the expected analysis error reduction when assimilating an observation into the analysis of each model variable. The experimental results with five different covariance localization patterns support the intuitive idea that SCDA benefits only when the variables of different components have strong background error correlations.

We then experimentally showed that the use of background error covariance in the LETKF could be detrimental when the ensemble size is too small. This supports the claim of Han et al. (2013) that a large ensemble is needed to improve the analysis using the full background error covariance. With a limited number of ensemble members, localizing the background error covariance is essential to obtain an accurate analysis. We proposed the correlation-cutoff method: first, estimate the

mean squared background error correlation with an offline DA cycle, then, uncouple the data assimilation if the background error correlation between the analyzed and the observed variables is weak. In our experiments with the nine-variable coupled model of [Peña and Kalnay \(2004\)](#), the correlation-cutoff method, intermediate to the standard SCDA and WCDA approaches, results in the best analysis and is the most robust to the choice of ensemble size.

Covariance localization guided by the correlation-cutoff method is a general idea to increase the signal-to-noise ratio of data assimilation. This method, however, is particularly important for the SCDA, where the correlation strength between different model components cannot be summarized by a simple function of distance, as represented by the carbon-dynamics data assimilation of [Kang et al. \(2011\)](#). Although the distance-dependent localization ([Hamill et al., 2001](#)) showed great success in atmospheric and oceanic DA, it cannot deal with characteristics of the dynamics that are distance-independent. On the other hand, the squared ensemble correlation is a nondimensional quantity between 0 and 1, which can be measured between any pair of observation and model variables. Furthermore, the method is also applicable before the implementation of SCDA; if a weakly coupled EnKF system has been already implemented, by measuring the squared ensemble correlations, one can assess the variance reduction that could be achieved by implementing the SCDA in advance. With these two characteristics, the correlation-cutoff method can be particularly useful for coupled EnKF applications.

In the toy model we used, there was a clear distinction between strongly and weakly correlated pairs of variables ([Figure 2.1](#)), and therefore, it was clear where to

stop the coupled data assimilation. The application of the correlation-cutoff method to more complex and realistic system will be explored in Chapter 6.

Most contents of this chapter are first published in [Yoshida and Kalnay \(2018\)](#), whose copyright is reserved by the American Meteorological Society.

Chapter 3: Sudden and major changes of dynamics observed in coupled chaotic systems

3.1 Background

For a dynamical system, its Lyapunov exponents characterize, to first order, the growth of perturbations in the state vector. That is, if a dynamical system has positive Lyapunov exponents, small differences in the initial conditions will exponentially grow so that it imposes a finite limit to the practical predictability of the system (Lorenz, 1963). Also, in the data assimilation context, the number of non-negative Lyapunov exponents is associated to the required minimum necessary ensemble size such that a non-localized deterministic EnKF cycle converges (Ng et al., 2011; Trevisan and Palatella, 2011). Furthermore, if we regard the forecast-analysis cycle as a non-autonomous dynamical system (forced by observation information from the true trajectory), the absence of non-negative conditional Lyapunov exponents is directly related to the convergence of analysis and synchronization of the analysis solution to the truth (Penny, 2017). Therefore, the Lyapunov exponents and their associated subspaces (Lyapunov splitting/vectors) of the atmosphere and the other dynamical systems has been an active area of research (e.g., Legras and

Vautard, 1996).

For coupled geophysical dynamical systems such as the atmosphere-ocean system, there exist several studies that try to associate each Lyapunov exponent of a coupled dynamical system with its subsystem by projecting its associated Lyapunov vector to each subsystem and understand its dynamical origin (e.g., Norwood et al., 2013; Vannitsem and Lucarini, 2016). The Lyapunov spectrum of the multiple-timescale coupled systems is characterized by many near-zero Lyapunov exponents associated to slow or neutral modes and geometrically degenerated covariant Lyapunov vectors¹ (Vannitsem and Lucarini, 2016; Penny et al., 2019). In general, the Lyapunov exponents of a dynamical system cannot always be associated with one of its subsystems, especially if the coupling is tight. For an extreme example, the geopotential and wind fields of the atmosphere are tightly coupled, and the baroclinic instability resides neither within the geopotential nor wind field.

In this chapter, we do not try to associate each of Lyapunov exponents of the coupled system deterministically. Instead, we try to answer the following questions:

- When can we associate each Lyapunov exponent of a coupled system with one of its subsystems?
- How does the coupling strength affect the Lyapunov spectrum if a series of dynamical systems with incremental coupling strength are examined?

For these purposes, we gradually change the coupling strength between zero (uncoupled) to the original (coupled) and see how the coupling strength affects the

¹That is, given some norm, covariant Lyapunov vectors have small angles with each other.

Lyapunov spectrum and the attractor’s characteristics.

For this purpose, we use the ENSO-type 6-variable model introduced by [Peña and Kalnay \(2004\)](#). Despite its simple construction, the 6-variable model exhibits a realistic chaotic oscillation in its “tropical atmosphere” and “ocean” parts.

3.2 Brief introduction to Lyapunov exponents and vectors

In this section, we briefly introduce Lyapunov exponents and vectors of an ergodic and nonlinear dynamical system, which have many implications on the system’s predictability and stability of data assimilation cycles. The goal of this section is to provide a feasible algorithm to obtain Lyapunov exponents and vectors given a nonlinear forward operator of a dynamical system. We do not provide proof of existence here, and readers are referred to more rigorous materials where necessary.

3.2.1 Tangent linear operator and its adjoint

Assume that an n -dimensional, autonomous, continuous-time dynamical system whose tendency is described by a continuous mapping $f(\mathbf{x})$ for a state vector $\mathbf{x} \in \mathbb{R}^n$ ($f : \mathbb{R}^n \rightarrow \mathbb{R}^n$). We then define a once-differentiable mapping $F^{(\cdot)}(\cdot) : \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}^n$, which satisfies $\partial F^t(\mathbf{x})/\partial t|_{t=0} = f(\mathbf{x})$, $F^0(\mathbf{x}) = \mathbf{x}$, and $F^s[F^t(\mathbf{x})] = F^{s+t}(\mathbf{x})$ for times $t, s \in \mathbb{R}$. The mapping F is called a nonlinear forward operator in the numerical weather prediction context; that is, $\mathbf{x}(t_0 + t) = F^t[\mathbf{x}(t_0)]$.

For a window length t and a basic state \mathbf{x} defined at the beginning of the

window, we can then define a *tangent linear operator* $\mathbf{M}_{\mathbf{x}}^t$ as

$$\mathbf{M}_{\mathbf{x}}^t = \left. \frac{\partial F^t(\mathbf{x}')}{\partial \mathbf{x}'} \right|_{\mathbf{x}'=\mathbf{x}}, \quad (3.1)$$

which can be expressed as an $n \times n$ matrix. We also write $\mathbf{M}_{\mathbf{x}(t_0)}^t$ as $\mathbf{M}_{t_0 \rightarrow t_0+t}$ if the trajectory $\mathbf{x}(t_0)$ is obvious from the context², or we may completely drop the annotation. By definition, a tangent linear operator $\mathbf{M}_{\mathbf{x}}^t$ describes how an infinitesimal perturbation $\delta \mathbf{x}$ added to \mathbf{x} will be at the end of the window of length t :

$$\delta \mathbf{x}_t \equiv F^t(\mathbf{x} + \delta \mathbf{x}) - F^t(\mathbf{x}) \quad (3.2)$$

$$\frac{\delta \mathbf{x}_t}{\|\delta \mathbf{x}\|} \xrightarrow{\|\delta \mathbf{x}\| \rightarrow 0} \frac{\mathbf{M}_{\mathbf{x}}^t \delta \mathbf{x}}{\|\delta \mathbf{x}\|}. \quad (3.3)$$

It is also useful to introduce an *adjoint* \mathbf{M}^* of a tangent linear operator \mathbf{M} , which is characterized by the property $\langle \mathbf{x}, \mathbf{M}\mathbf{y} \rangle = \langle \mathbf{M}^*\mathbf{x}, \mathbf{y} \rangle$ for an appropriate norm $\langle \cdot \rangle$ and $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. For our \mathbb{R}^n formulation and the Euclidean norm, the adjoint \mathbf{M}^* corresponds to the matrix transpose \mathbf{M}^T (Kalnay, 2003).

3.2.2 The multiplicative ergodic theorem of Oseledets: the existence of Lyapunov exponents and covariant Lyapunov vectors

We next introduce the Lyapunov exponents and covariant Lyapunov vectors guided by Bochi (2008).

We first assume the mapping F^t to be invertible. If $P(A)$ is either 0 or 1 for all $t \in \mathbb{R}$ and all attractors $A \subset \mathbb{R}^n$ invariant to F^t (i.e., $F^{-t}(A) = A$), then a

²By definition, $\mathbf{M}_{t_0 \rightarrow t_2} = \mathbf{M}_{t_1 \rightarrow t_2} \mathbf{M}_{t_0 \rightarrow t_1}$.

probability measure P defined on \mathbb{R}^n is said to be *ergodic* with respect to F^3 . With an ergodic probability measure P of a dynamical system, ergodic theorems prove that a time-average of some function of the state vector converges to the function's space average on the attractor, which is independent of the initial condition and the length of integration.

Of ergodic theorems, we are interested in the multiplicative ergodic theorem of Oseledets (Oseledets, 1968; Bochi, 2008). The theorem assures the existence of Lyapunov exponents and covariant Lyapunov vectors for an ergodic dynamical system. Although the exponents and vectors will be defined using a specific norm, the theorem proves that they are independent of the norm under general conditions.

The *Lyapunov exponents* are k ($1 \leq k \leq n$) distinct numbers unique to the dynamical system

$$\lambda_1 > \lambda_2 > \dots > \lambda_k \tag{3.4}$$

with positive integer multiplicities d_i ($i = 1, 2, \dots, k$)⁴, which satisfy $\sum_{i=1}^k d_i = n$.

³Intuitively, this condition claims that one and only one well-mixed attractor with nonzero probability measure exists. If two disjoint attractors $A, B \subset \mathbb{R}^n$ that are invariant to F exist and if both A and B have nonzero probability measures, the condition is not satisfied. When two attractors A and B of the mapping F exist, then by re-defining probability measures P_A and P_B that are vanishing on the other attractor, we can separately apply ergodic theorems for each attractor and discuss the system's characteristics in each basin of attraction.

⁴We will consistently use i for the index of distinct Lyapunov exponents throughout this section ($i = 1, \dots, k$). We will later use d (without subscript) for the index of Lyapunov exponents counted with multiplicity ($d = 1, \dots, n$). Both indices are in descending order (i.e., the first Lyapunov exponent is the largest positive one).

Note that the Lyapunov exponents depend on the mapping F and the probability measure P of the attractor but are independent of the state vector \mathbf{x} .

The *Lyapunov splitting* is a series of linear subspaces (represented by matrices whose column vectors constitute each subspace's basis)

$$\mathbf{E}_{\mathbf{x}}^1 \oplus \mathbf{E}_{\mathbf{x}}^2 \oplus \dots \oplus \mathbf{E}_{\mathbf{x}}^k = \mathbb{R}^n, \quad (3.5)$$

defined for P -almost-every $\mathbf{x} \in \mathbb{R}^n$, where the dimension of the subspace (i.e., number of columns of the matrix) $\mathbf{E}_{\mathbf{x}}^i$ is d_i ; each subspace is *covariant* in a sense $\mathbf{M}_{\mathbf{x}}^t \mathbf{E}_{\mathbf{x}}^i = \mathbf{E}_{F^t(\mathbf{x})}^i$, that is, the i -th subspace of the Lyapunov splitting propagated with the tangent linear operator remains the i -th subspace of the Lyapunov splitting forever. A non-orthogonal basis of each subspace is called *covariant Lyapunov vectors*; for an exponent without multiplicity, there exists a unique covariant Lyapunov vector up to scalar multiplication.

The Lyapunov exponents and covariant Lyapunov vectors are related by

$$\lim_{t \rightarrow \pm\infty} \frac{1}{t} \log \|\mathbf{M}_{\mathbf{x}}^t \mathbf{v}_{\mathbf{x}}^i\| = \lambda_i \quad \text{for all } \mathbf{v}_{\mathbf{x}}^i \in \mathbf{E}_{\mathbf{x}}^i, \|\mathbf{v}_{\mathbf{x}}^i\| = 1, \quad (3.6)$$

that is, a Lyapunov exponent is the average long-term growth rate of its associated covariant Lyapunov vectors.

3.2.3 Forward and backward Lyapunov vectors

As a corollary of the multiplicative ergodic theorem of Oseledets, we can also define forward and backward Lyapunov vectors. The discussion below follows [Legras and Vautard \(1996\)](#) and [Ginelli et al. \(2013\)](#).

First, we obtain matrix limits called forward (+) and backward (-) *Oseledets matrices*

$$\begin{aligned}\Xi_{\mathbf{x}}^+ &= \lim_{t \rightarrow +\infty} \frac{1}{2t} \log (\mathbf{M}_{\mathbf{x}}^{t\mathbf{T}} \mathbf{M}_{\mathbf{x}}^t) \\ \Xi_{\mathbf{x}}^- &= \lim_{t \rightarrow +\infty} \frac{1}{2t} \log (\mathbf{M}_{F^{-t}(\mathbf{x})}^t \mathbf{M}_{F^{-t}(\mathbf{x})}^{t\mathbf{T}}),\end{aligned}\tag{3.7}$$

which are symmetric and proved to exist under the same hypothesis of the multiplicative ergodic theorem of Oseledets. These two matrices share the same eigenvalues $\lambda_1 > \lambda_2 > \dots > \lambda_k$ independent of \mathbf{x} , which correspond to the Lyapunov exponents defined by Eqs. (3.4) and (3.6).

Then we obtain orthogonal eigenspaces $\mathbf{U}_{\mathbf{x}}^{\pm(i)} \in \mathbb{R}^{n \times d_i}$ ($i = 1, \dots, k$) of the Oseledets matrices $\Xi_{\mathbf{x}}^{\pm}$ by eigendecomposition. The i -th eigenspaces associated with eigenvalue λ_i have dimension d_i , the multiplicity of the Lyapunov exponent. Orthogonal bases of these eigenspaces are called *forward* and *backward Lyapunov vectors*. For the future discussion, we define $n \times n$ orthogonal matrices $\mathbf{U}_{\mathbf{x}}^{\pm}$ whose columns are forward/backward Lyapunov vectors in descending order. Note the forward and backward Lyapunov vectors are norm-dependent and not covariant.

We then construct two sequences of embedded subspaces called *Oseledets subspaces* defined by

$$\begin{aligned}\Gamma_{\mathbf{x}}^{+(i)} &= \bigoplus_{j=i}^k \mathbf{U}_{\mathbf{x}}^{+(j)} \\ \Gamma_{\mathbf{x}}^{-(i)} &= \bigoplus_{j=1}^i \mathbf{U}_{\mathbf{x}}^{-(j)},\end{aligned}\tag{3.8}$$

which satisfy

$$\begin{aligned}\mathbb{R}^n &= \Gamma_{\mathbf{x}}^{+(1)} \supset \dots \supset \Gamma_{\mathbf{x}}^{+(k)} \supset \Gamma_{\mathbf{x}}^{+(k+1)} \equiv \phi \\ \mathbb{R}^n &= \Gamma_{\mathbf{x}}^{-(k)} \supset \dots \supset \Gamma_{\mathbf{x}}^{-(1)} \supset \Gamma_{\mathbf{x}}^{-(0)} \equiv \phi\end{aligned}\tag{3.9}$$

and

$$\mathbf{E}_x^i = \Gamma_x^{+(i)} \cap \Gamma_x^{-(i)} \quad (3.10)$$

for $i = 1, \dots, k$, where (covariant) Lyapunov splitting \mathbf{E}_x^i is defined in Eq. (3.5). We note that almost every⁵ random vector in the i -th forward Lyapunov splitting $\Gamma_x^{+(i)}$ will grow exponentially at average speed λ_i in the future. Similarly, almost every⁶ random vector in the i -th backward Lyapunov splitting $\Gamma_x^{-(i)}$ has exponentially grown at average speed λ_i in the past.

3.2.4 Numerical methods for obtaining Lyapunov exponents and vectors

So far, we have obtained constructive definitions of the Lyapunov exponents and covariant/forward/backward Lyapunov vectors by Eqs. (3.1) and (3.7) - (3.10) for a given nonlinear forward operator F . However, practically we cannot obtain the exponents and vectors directly from the definition because the evaluation of Eq. (3.7) requires evaluation of exponential functions of t . Computers can only handle real numbers with finite precision and range, and the exponential growth leads to computational overflow/underflow even with reasonably small t .

To overcome this issue, we exploit Eqs. (3.6), (3.9), and (3.10). Assume that we can uniquely define a space spanned by the first d ($d = 1, \dots, n$) Lyapunov vectors (i.e., there exist $1 \leq i_d \leq k$ such that $\sum_{i=1}^{i_d} d_i = d$). With $t \rightarrow +\infty$, a random

⁵Exceptions are vectors in $\Gamma_x^{+(i+1)}$. Thanks to the relationship (Eq. 3.9) and the dimensionality of each subspace, this probability is zero for uniformly random vectors in $\Gamma_x^{+(i)}$.

⁶Exceptions are vectors in $\Gamma_x^{-(i-1)}$.

d -dimensional linear subspace of \mathbb{R}^n defined at time $-t$ (i.e., far past) almost surely converges to the d -dimensional subspace spanned by the first d backward Lyapunov vectors (at time 0) if it is propagated with a tangent linear operator $\mathbf{M}_{-t \rightarrow 0}$. Similarly, with $t \rightarrow +\infty$, a random d -dimensional linear subspace of \mathbb{R}^n defined at time t (i.e., far future) almost surely converges to the d -dimensional subspace spanned by the first d forward Lyapunov vectors (at time 0) if propagated back with an adjoint operator $\mathbf{M}_{0 \rightarrow t}^T$. Therefore, without explicitly calculating Oseledets matrices $\Xi_{\mathbf{x}}^{\pm}$, we can estimate the orthogonal spaces spanned by first d forward/backward Lyapunov vectors. When propagating a space with the tangent linear operator and its adjoint, we can divide t into many small windows and regularly orthonormalize the space's basis to avoid overflow/underflow; see footnote 2. At the same time, we can estimate the Lyapunov exponents by calculating the mean logarithm of the growth rate of the orthogonal vectors. Algorithm 1 summarizes the procedure described in this paragraph. This algorithm is also applicable for obtaining Lyapunov exponents and vectors even if the Lyapunov spectrum has multiplicity; the resulting vectors are one realization of non-unique basis vectors.

After obtaining estimated forward/backward Lyapunov vectors, covariant Lyapunov vectors for each time are obtained by Eqs. (3.8) and (3.10). The intersection of two subspaces can be obtained using the singular value decomposition.

Although there is another computational issue, namely that the computation and memory costs grow at $\mathcal{O}(n^4t)$ and $\mathcal{O}(n^2t)$, the present algorithm is practical for our purposes with small models. If we are only interested in obtaining few leading Lyapunov exponents and vectors of a large dynamical system, there exist a few

Algorithm 1: Gram-Schmidt method to obtain Lyapunov exponents and forward/backward Lyapunov vectors. Note that backward Lyapunov vectors near the beginning of the procedure and forward Lyapunov vectors near the end of the procedure can be inaccurate due to spin-up.

Symbols for non-local variables:

- \mathbf{x}_0 : State vector at time 0
- Δt : Orthonormalization interval (short enough to avoid overflow/underflow)
- m : Number of orthonormalization windows
(as long as we afford; results are calculated from time 0 to $m\Delta t$)
- j : Time index ($0, 1, \dots, m$)
- $\widehat{\boldsymbol{\lambda}}$: Estimate of Lyapunov exponents in descending order
(n -vector; counted with multiplicity)
- $\widehat{\mathbf{U}}_t^\pm$: Estimate of forward/backward Lyapunov vectors $\mathbf{U}_{\mathbf{x}(t)}^\pm$
- $\mathbf{QR}(\cdot)$: QR decomposition (returns two matrices)

Other symbols (n , F , and \mathbf{M}) as defined in the text

```

function Gram-Schmidt-Lyapunov( $\mathbf{x}_0$ ,  $\Delta t$ ,  $m$ )
  ▷ Get tangent linear operators
1  for  $j$  in  $1, \dots, m$  do
2    Compute and store  $\mathbf{M}_{(j-1)\Delta t \rightarrow j\Delta t}$  using  $\mathbf{x}_{(j-1)\Delta t}$ 
3     $\mathbf{x}_{j\Delta t} \leftarrow F^{\Delta t} [\mathbf{x}_{(j-1)\Delta t}]$ 
4  end
5
  ▷ Get backward Lyapunov vectors and exponents
6   $\widehat{\mathbf{U}}_0^- \leftarrow \mathbf{Q}$  of  $\mathbf{QR}$ (a random  $n \times n$  matrix)
7  for  $j$  in  $1, \dots, m$  do
8     $\widehat{\mathbf{U}}_{j\Delta t}^-, \mathbf{R}_{(j-1)\Delta t \rightarrow j\Delta t} \leftarrow \mathbf{QR}(\mathbf{M}_{(j-1)\Delta t \rightarrow j\Delta t} \widehat{\mathbf{U}}_{(j-1)\Delta t}^-)$ 
9     $\gamma_{(j-1)\Delta t \rightarrow j\Delta t} \leftarrow$  diagonal of  $\mathbf{R}_{(j-1)\Delta t \rightarrow j\Delta t}$       ▷ No sort by values
10 end
11  $\widehat{\boldsymbol{\lambda}} \leftarrow \frac{1}{m\Delta t} \sum_{j=1}^m \ln \gamma_{(j-1)\Delta t \rightarrow j\Delta t}$ 
12
  ▷ Get forward Lyapunov vectors
13  $\widehat{\mathbf{U}}_{m\Delta t}^+ \leftarrow \mathbf{Q}$  of  $\mathbf{QR}$ (a random  $n \times n$  matrix)
14 for  $j$  in  $m, m-1, \dots, 1$  do
15    $\widehat{\mathbf{U}}_{(j-1)\Delta t}^+, \mathbf{R}_{j\Delta t \rightarrow (j-1)\Delta t} \leftarrow \mathbf{QR}(\mathbf{M}_{(j-1)\Delta t \rightarrow j\Delta t}^\top \widehat{\mathbf{U}}_{j\Delta t}^+)$ 
16 end
17
18 return  $\widehat{\boldsymbol{\lambda}}$ ,  $\{\widehat{\mathbf{U}}_{j\Delta t}^-\}$ ,  $\{\widehat{\mathbf{U}}_{j\Delta t}^+\}$ 
end function

```

algorithms to alleviate the cost; see [Ginelli et al. \(2013\)](#) and the references therein.

3.3 Methods

3.3.1 The coupled model and decoupling

We will now observe how the Lyapunov spectrum and other characteristics of the attractor of a coupled model vary when the coupling strength is continuously varied from the zero (uncoupled) to original (coupled).

For a simple but realistic model of a coupled chaotic dynamical system, we use the 6-variable, “ENSO” type model of [Peña and Kalnay \(2004\)](#). The prognostic equations of the model are as follows:

$$\begin{aligned}
 \dot{x}_t &= \sigma(y_t - x_t) - \alpha c(SX + k_2) \\
 \dot{y}_t &= rx_t - y_t - x_t z_t + \alpha c(SY + k_2) \\
 \dot{z}_t &= x_t y_t - bz_t + \alpha c_z Z \\
 \dot{X} &= \tau \sigma(Y - X) - \alpha c(x_t + k_2) \\
 \dot{Y} &= \tau r X - \tau Y - \tau S X Z + \alpha c(y_t + k_2) \\
 \dot{Z} &= \tau S X Y - \tau b Z - \alpha c_z z_t.
 \end{aligned} \tag{3.11}$$

Here, $\mathbf{x} \equiv (x_t, y_t, z_t, X, Y, Z)^T$ represents state variables, a coupling strength parameter $\alpha = 1$ is introduced (and will be modified), and other parameters are kept the same as the original values, namely, $(\sigma, r, b, \tau, c, c_z, S, k_2) = (10, 28, 8/3, 0.1, 1, 1, 1, -11)$.

The dots above the state variables represent their temporal derivative. In comparison to the 9-variable model used in the previous chapter, this 6-variable model does

not have the “extratropical atmosphere” for further simplicity. Note that the terms including α are not diffusive and do not contain the time derivative. For example, if the third equation was instead $\dot{z}_t = x_t y_t - b z_t + \alpha c_z (Z - z_t)$, increasing α would increase the diffusivity and stabilize the system.

This model is integrated using the fourth-order Runge-Kutta method with a timestep of 0.01 nondimensional time unit (TU). This single (external) integration step can be denoted as $F^{0.01\text{TU}}$ in the previous section’s notation, and hereafter we only think time intervals that are multiples of 0.01 TU. The tangent linear operator $\mathbf{M}_x^{0.01\text{TU}}$ is obtained by numerical differentiation of $F^{0.01\text{TU}}$ around the basic state \mathbf{x} .

The original six-variable coupled model ($\alpha = 1$) has Lyapunov exponents of (0.318, 0, -0.47, -0.794, -1.811, -12.276) in our previous experiment; the model has one positive, one near-zero, and four negative exponents.

On the other hand, the three Lyapunov exponents of the [Lorenz \(1963\)](#) model with the proposed parameters are numerically known to be (0.906, 0, -14.572) ([Sprott, 1997](#)). Our tropical atmosphere (the first three equations of Eq. [3.11](#)) is the same as the Lorenz model when uncoupled ($\alpha = 0$). In addition, the uncoupled ocean (the last three equations of Eq. [3.11](#) with $\alpha = 0$) should have Lyapunov exponents of (0.0906, 0, -1.4572); these values are one-tenth of the ones for uncoupled tropical atmosphere because the entire tendency equation is multiplied by $\tau = 0.1$, and the dynamics is unchanged except that it is ten times slower. With these combined, the six-variable uncoupled ($\alpha = 0$) model should have Lyapunov exponents of (0.906, 0.0906, 0, 0, -1.4572, -14.572), which is numerically validated

(the leftmost values of Figure 3.1 later).

The uncoupled three-variable model is designed to represent a hydrodynamic flow, by trigonometric decomposition of two-dimensional (one horizontal and one vertical dimensions) convective dynamics forced by vertical temperature gradient; variable X represents the convective motion, and variables Y and Z represent the horizontal and vertical temperature gradient perturbations, respectively (Lorenz, 1963). Therefore, the coupled model (Eq. 3.11) couples the momentum terms (x_t, X) and two pairs of thermal terms (y_t, Y) and (z_t, Z) , respectively. The k_2 terms are introduced as an “uncentering” parameter, which can be thought as the mean state felt by the other subsystem. As a result of coupling, the model exhibits an asymmetric coupled oscillation with chaotic periods like ENSO; for example, the ocean variable X experiences occasional large negative anomaly (analogous to an El Niño), followed by several cycles of “normal years” (see Figure 3.2d later).

3.3.2 Experimental settings

Our main experiments vary the coupling strength parameter α within a range $[0, 1]$ by an interval of 0.005 and observe the system’s Lyapunov exponents and the attractor’s other characteristics. For each value of α , the following procedure is repeated to correct data:

1. We randomly initialize an initial state vector $\mathbf{x}(t = 0) \in \mathcal{N}(0, 100^2)$. The initial variance of 100^2 is determined to roughly cover the attractor. The random initial state is independent for each value of α so that the results

represent the entire basin of attraction.

2. We integrate the model for 5,000 TU to obtain $\mathbf{x}(t = 5,000 \text{ TU})$. This state vector $\mathbf{x}(t = 5,000 \text{ TU})$ is expected to be a random sample from the attractor due to ergodicity.
3. Using Algorithm 1 with arguments $\mathbf{x}(t = 5,000 \text{ TU})$, $\Delta t = 0.01 \text{ TU}$, and $m = 500,000$, we obtain the estimate of Lyapunov exponents and vectors. The trajectory $\mathbf{x}(t)$ for $5,000 \leq t \leq 10,000 \text{ TU}$ is also saved for later examinations. The random matrices to initialize $\widehat{\mathbf{U}}_0^-$ and $\widehat{\mathbf{U}}_{m\Delta t}^+$ are sampled from the multidimensional normal distribution so that each column vector's direction before orthonormalization is uniformly random in the \mathbb{R}^6 space. The random vectors are independently sampled for each α .

3.4 Results

First, we investigate Figure 3.1, which shows that the Lyapunov exponents are not continuously dependent on the coupling strength. The most prominent change occurs between $\alpha = 0.22$ and $\alpha = 0.225$. We define this threshold as α_1 . When we change from $\alpha < \alpha_1$ to $\alpha > \alpha_1$, the largest positive exponent disappears; at the same time, another neutral mode (i.e., a near-zero exponent) appears. We also notice that with $\alpha < \alpha_1$, two neutral modes co-exist. However, it is well known that a continuous-time, autonomous, and ergodic dynamical system has at least one

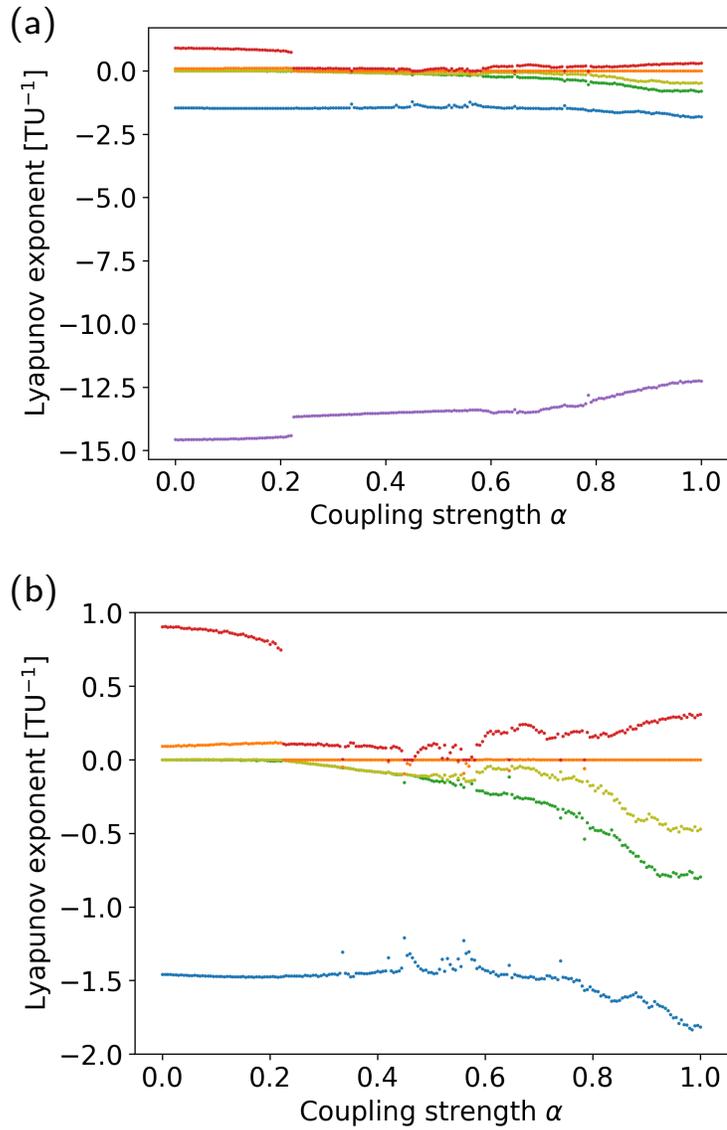


Figure 3.1: Lyapunov exponents of the 6-variable model. (a): entire spectrum. (b): magnification of the first five exponents. For each value of α , a red dot represents the largest positive exponent of the system; orange, yellow, green, blue, and purple follows.

neutral exponent corresponding to temporal translation⁷. Similarly, if a dynamical system has two non-interacting subsystems, then the system should have at least two null Lyapunov exponents (counted with multiplicity), whose associated covariant subspace contains the temporal translations in the two subsystems. Note that its converse does not hold; a counterexample is a harmonic oscillator with two degrees of freedom, which has two null Lyapunov exponents, but the two degrees of freedom are interacting. We may interpret that the two temporal translation modes, each originating from each uncoupled subsystem, are intact until we increase α to exceed α_1 . Overall, the models with $\alpha < \alpha_1$ are, from the Lyapunov spectrum perspective, qualitatively same with a dynamical system with two uncoupled subsystems.

We can also find for some values of $\alpha > \alpha_1$, no positive Lyapunov exponent exist (e.g., $\alpha \sim 0.56$). These values of α are called periodic windows (e.g., Chapter 10 of [Strogatz, 2015](#)) and later examined in detail. Each model with $\alpha > \alpha_1$ has a single unstable mode (positive exponent) and a single null exponent except for the periodic windows.

Let us look at other characteristics of the models with different coupling parameter α , especially those with values close to α_1 .

Figure 3.2 shows time series $\mathbf{x}(t)$ ($9,900 \leq t \leq 10,000$ TU) with different values

⁷For the mapping F , $\mathbf{x} \in \mathbb{R}^n$, and $\delta t \in \mathbb{R}$, a nonlinear perturbation $F^{\delta t}(\mathbf{x}) - \mathbf{x}$ will be $F^t[F^{\delta t}(\mathbf{x})] - F^t[\mathbf{x}] = F^{\delta t}[F^t(\mathbf{x})] - F^t[\mathbf{x}]$ after time t . However, if \mathbf{x} is randomly sampled from the attractor, due to ergodicity, random variables \mathbf{x} and $F^t(\mathbf{x})$ share the same probabilistic measure for any t . Therefore, in particular, $\langle \|F^{\delta t}(\mathbf{x}) - \mathbf{x}\| \rangle = \langle \|F^{\delta t}[F^t(\mathbf{x})] - F^t[\mathbf{x}]\| \rangle$ is satisfied for any δt .

of α . Note that the first two panels correspond to $\alpha < \alpha_1$, and the last two panels correspond to $\alpha > \alpha_1$. First, with $\alpha = 0.0$ (panel a), the two systems are completely independent. Then with $\alpha = 0.2$ (panel b), we notice that the state evolution is very similar to the previous case. When we increase α slightly beyond α_1 (panel c), the fast atmospheric oscillation suddenly disappears. Although it is hard to see from the panel, the atmosphere is synchronized with the ocean oscillation with small amplitude. By increasing α further toward unity (panel d), the amplitude and the frequency of both the atmosphere and the ocean variations increase. This change between panels (c) and (d) is rather gradual (not shown). Most of these features do not contradict with what we can infer from the change of the Lyapunov spectrum.

To further investigate how parameter α affects the dynamics, we show pairwise scatter diagrams of attractors in Figure 3.3. First, with $\alpha = 0.0$ (panel a), we notice that cross-relationships between (x_t, y_t, z_t) and (X, Y, Z) shown in the nine lower left subpanels are obviously independent. Here, two scalar random variables A and B are called independent if for all a and $b \in \mathbb{R}$, $P\{A \leq a, B \leq b\} = P\{A \leq a\}P\{B \leq b\}$. This independence is also supported by near-zero mutual information

$$\int_B \int_A p(a, b) \log_2 \left[\frac{p(a, b)}{p(a)p(b)} \right] da db \quad (3.12)$$

between these variables, where p denotes the probability density function. If mutual information is large between two variables, having access to climatological information, we can infer one variable's value with good accuracy given the other variable's value. The nearly independent behavior between (x_t, y_t, z_t) and (X, Y, Z) persists even if we increase α to 0.2 (panel b) although the slightly increased mutual in-

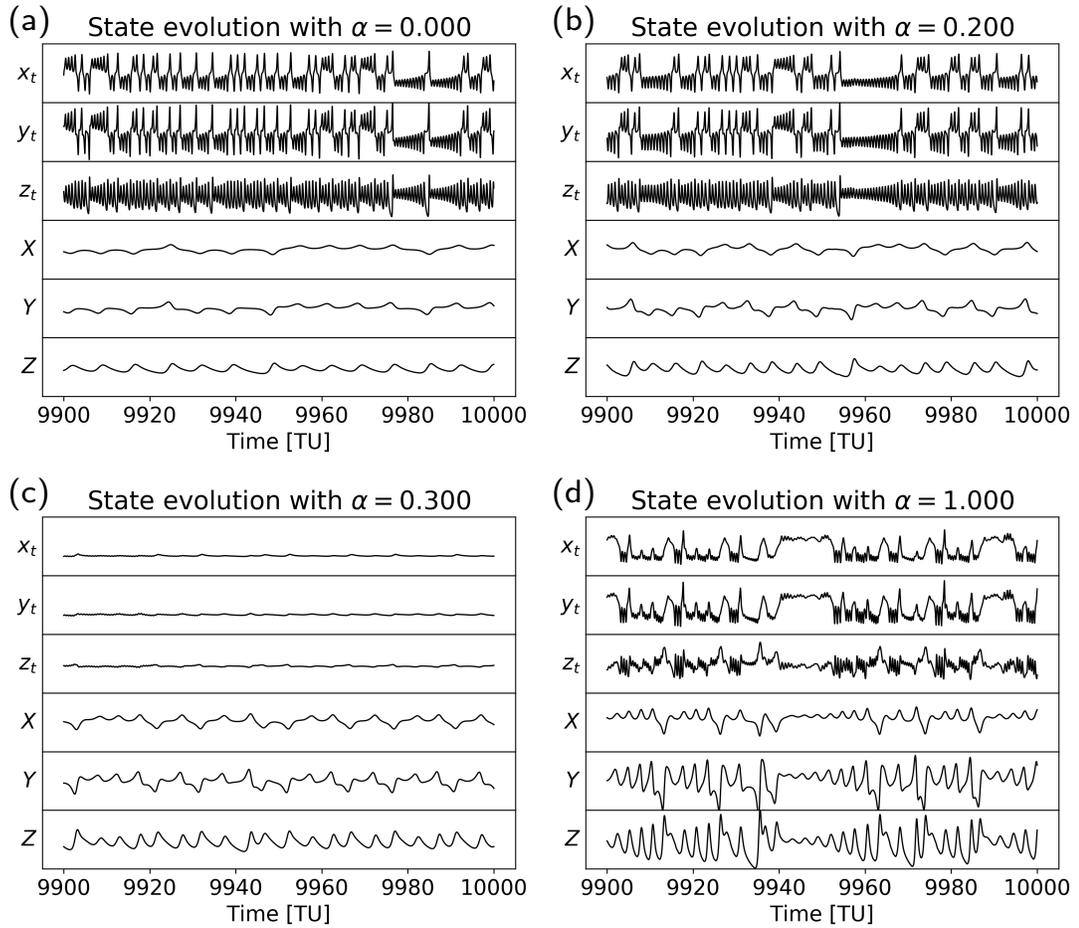


Figure 3.2: Sample time series of the models with different α values (written in each panel). The ordinate ranges shown are $x_t, y_t \in [-25, 25]$; $z_t \in [0, 60]$; $X, Y \in [-100, 100]$; and $Z \in [-70, 130]$.

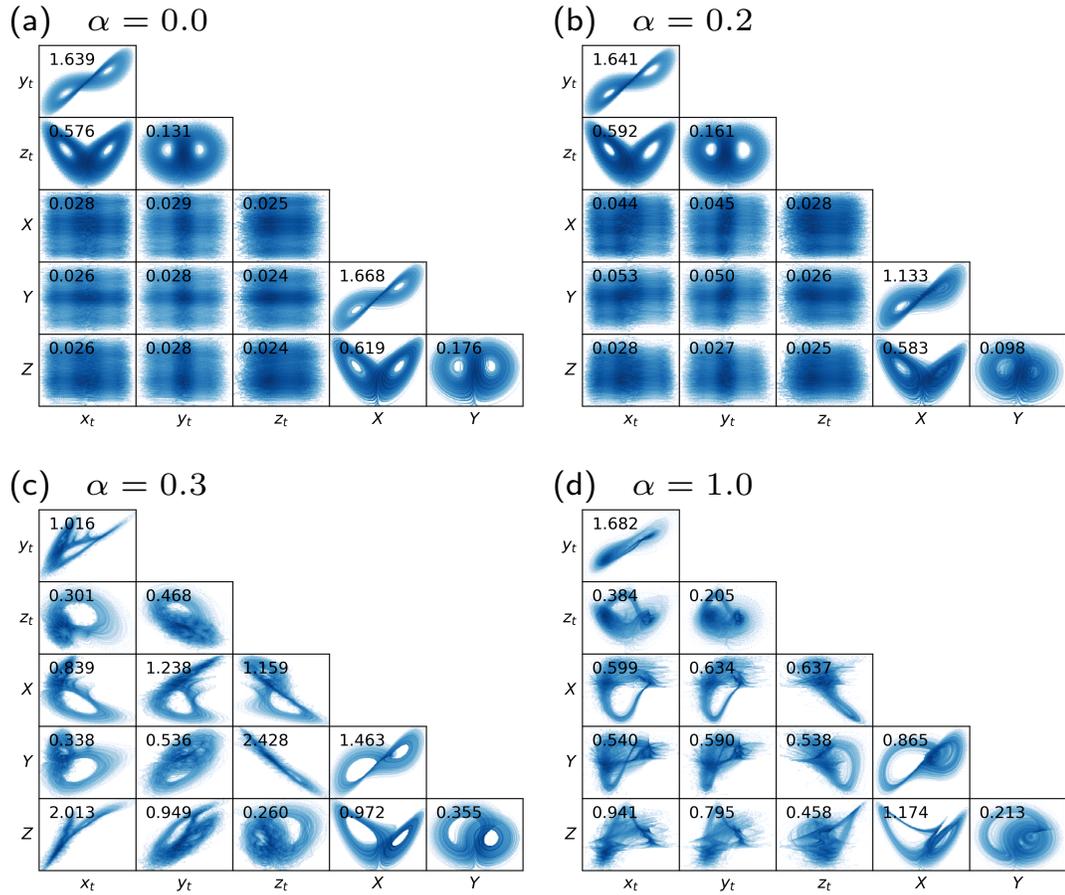


Figure 3.3: Two-dimensional projections of attractors $\mathbf{x}(t)$ for $5,000 \leq t \leq 10,000$ TU with different α values (written in each panel). The color shows the relative frequency of each pair of values on a log scale. Superimposed numbers show mutual information between the two variables in bits. The mutual information is calculated after dividing each variable's range into 100 bins with regular intervals. The abscissa and ordinate ranges are automatically adjusted for each panel and not shown for conciseness; see Figure 3.2 instead for obtaining a sense of variabilities.

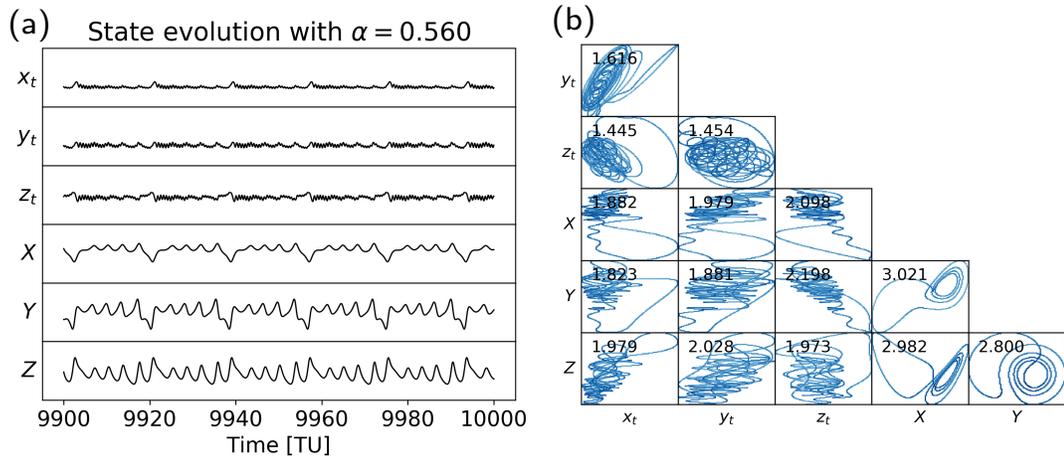


Figure 3.4: Same as Figures 3.2 and 3.3 but for $\alpha = 0.56$.

formation between (x_t, y_t) and (X, Y) indicate weak mutual dependence. The two-dimensional projections of the attractor look totally different when α is increased beyond α_1 (panels c and d). We can see distinct structures in the scatter diagrams, which show that these pairs of variables are closely interacting. Values of mutual information also support the tight coupling of these variables.

We have seen that the model has no positive Lyapunov exponents for some values of α (periodic windows; Figure 3.1). Figure 3.4 shows a time series and two-dimensional projection of the attractor for such value of α . First, we notice that the trajectory is a closed loop in the panel (b), most obviously in the relation between Z and Y . An autonomous system with a closed loop must be periodic, as we can see in panel (a) that there exists a period of ~ 18 TU. As expected from the Lyapunov spectrum, the 6-variable model has a periodic attractor with these α values. This periodic orbit is examined to be attracting by repeating experiments with a few different initial conditions.

3.5 Discussion and remarks

We have shown that the Lyapunov spectrum of two chaotic subsystems with very different timescales experiences a qualitative and sudden change when the subsystems are coupled gradually.

When the chaotic subsystems have the same or similar dynamics and the timescales, a behavior known as synchronization of chaos (Pecora and Carroll, 1990; Pecora et al., 1997) is observed; for one-way coupling, this is characterized by all the conditional Lyapunov exponents of the response system gradually becomes negative as increasing coupling strength (e.g., Rosenblum et al., 1996; Liu et al., 2003). Our study showed a similar, abrupt change of dynamics in a coupled dynamical system with very different timescales.

The original coupled model of Peña and Kalnay (2004) ($\alpha = 1$) is constructed mimicking the strong coupling between chaotic tropical atmospheric weather with the slow ocean, which also is chaotic. Each subsystem is designed to represent a hydrodynamic flow by trigonometric decomposition of two-dimensional convective dynamics forced by vertical temperature gradient. With $\alpha > 0$, the two subsystems are coupled both kinetically and thermally. Therefore, the model developers called this coupled model as “ENSO type”, indicating that this coupling process closely resembles that of the tropical atmosphere-ocean system, where the anomalous wind stress drives the SST anomaly, which in turn drives anomalous wind circulation (Bjerknes feedback). It is therefore interesting to see that these two subsystems originating from fluid mechanics do not just affect each other but also modify the

other subsystem's dynamics dramatically through thermal and mechanical coupling.

Such discontinuous sensitivity to the coupling strength may also apply to more realistic coupled geophysical subsystems, whose coupling strength can vary geologically and dynamically. For example, the momentum coupling strength between the atmosphere (say top of the boundary layer) and the ocean (say a few meters depth) will be affected by wave height and by vertical mixing coefficient within each fluid, which in turn are affected by wind speed and vertical stability. For some regimes, these subsystems may be tightly coupled and mutually provide a good amount of information, while for another regime, these subsystems may behave as if they were irrelevant to each other.

Another implication of our results is that a slight misspecification of the model coupling parameters may cause severe over or under-estimation of uncertainty by its model-based estimate such as ensemble forecast. In our example, if our best estimate of α is greater than α_1 but the true α is less than α_1 , then our estimate of coupled predictability will be too optimistic (and vice versa). The nonlinear sensitivity will also render parameter estimation difficult, where observable quantities of the attractor discontinuously depend on the model parameter.

Further investigation is needed for more realistic coupled models of the atmosphere and ocean, especially those with multiple positive Lyapunov exponents (i.e., hyperchaotic models) and abilities to reproduce coupled instabilities like ENSO.

Chapter 4: Implementation of FOAM-LETKF

In this short chapter, we describe a FOAM-LETKF system, that will be commonly used for the two subsequent chapters.

4.1 Fast Ocean Atmosphere Model (FOAM)

4.1.1 FOAM specifications

To test our coupled data assimilation methodologies, we employ a more realistic coupled model of the atmosphere and ocean, the Fast Ocean Atmosphere Model (FOAM; [Jacob, 1997](#)). We acknowledge that the model program and guidance on implementation of an analysis system are kindly provided by Dr. Yun Liu. The model is a coupled general circulation model (GCM) efficiently implemented for multi-processing and represents the atmosphere and the ocean as well as simple parameterized processes of sea ice, land, and river runoff. [Table 4.1](#) summarizes FOAM's general specifications.

The atmospheric component of the model, PCCM3, originates from the third version of the NCAR Community Climate Model (CCM3) with modifications for parallel computers. The dry dynamics of PCCM3 uses the spectral method for

Table 4.1: Specification of the Fast Ocean Atmosphere Model (FOAM), version 1.5.

Atmospheric model (PCCM3)	
Horizontal resolution	R15 spectral (40 latitudes \times 48 longitudes)
Vertical resolution	18 levels (hybrid σ - p)
Integration timestep	30 minutes 1 hour for radiation
Parameterized schemes	Convection Cloud Radiation Surface physics Vertical diffusion Gravity wave drag
Oceanic model (OM3)	
Horizontal resolution	128 latitudes \times 128 longitudes (polar grid)
Vertical resolution	24 levels (z -coordinate)
Integration timestep	6 hours
Vertical mixing scheme	Bulk scheme (based on Richardson number)
Sea ice (CSIM 2.2.6)	
Horizontal resolution	128 latitudes \times 128 longitudes (same as ocean)
Integration timestep	30 minutes
Modeled processes	Formation/melting Thermal conduction Snow on top Radiation
Land, Hydrology, and River Runoff	
Horizontal resolution	128 latitudes \times 128 longitudes (same as ocean)
Integration timestep	30 minutes

solving the primitive equations on the sphere, but the moisture advection is computed with the semi-Lagrangian scheme. The model has 18 hybrid sigma-pressure coordinate levels and resolves the spectral modes up to 15 zonal and meridional wavenumbers (rhomboidal truncation; R15). This horizontal configuration corresponds to 40 grids in latitudinal direction (approximately 500 km) and 48 grids in the longitudinal direction (approximately 830 km at the equator). The subgrid, moist, and radiative physics are parameterized. The prognostic variables of the model are surface pressure, temperature, specific humidity, and zonal/meridional components of the wind.

The ocean component of the model (Ocean Model version three developed at the University of Wisconsin-Madison; OM3) uses the finite-difference representation of the primitive equations with hydrostatic and Boussinesq approximations. The ocean model has 24 vertical levels in z -coordinate with a free surface. The global ocean is divided into 128×128 horizontal (latitude-longitude) grid boxes. Numerical instability in the Arctic is avoided by application of a spatial filter. The oceanic prognostic variables are temperature, salinity, zonal/meridional components of current, and the surface pressure anomaly. The vertical mixing is parameterized, with the mixing length depending on the Richardson number (i.e., it depends on the mechanical shear and static stability).

Coupling between the atmosphere and the ocean is implemented as follows. First, the ocean model provides an SST field to the coupler, and similarly, the sea ice/land models provide surface moisture, temperature, roughness, and albedo. The coupler is a subroutine of the atmospheric model and computes the heat, momentum,

and freshwater fluxes; the atmospheric state is updated consistently. The total heat flux TH (in W/m^2) is

$$TH = SW + LW + SH + LH, \quad (4.1)$$

where SW and LW are shortwave and longwave radiations, and SH and LH are sensible and latent heat fluxes (positive downward). The total freshwater flux FW (in $\text{kg}/\text{m}^2\text{s}$) is

$$FW = P - E + Rnf + SI, \quad (4.2)$$

where P is the precipitation, E is the evaporation, Rnf is the runoff from the river model, and SI is the water gain from sea ice melting (negative for sea ice formation). Then, the fluxes are accumulated over a 6-hour window before passed to the ocean model. Finally, the fluxes are used to update the ocean state at the top layer:

$$\begin{aligned} \frac{\partial T}{\partial t} &= \frac{TH}{\rho h C_p^{\text{ocn}}} \\ \frac{\partial \ln S}{\partial t} &= -\frac{FW}{\rho h} \\ \frac{\partial u}{\partial t} &= \frac{\tau_x}{\rho h} \\ \frac{\partial v}{\partial t} &= \frac{\tau_y}{\rho h}. \end{aligned} \quad (4.3)$$

Here, (T, S, u, v) are the oceanic prognostic variables (temperature, salinity, zonal current, and meridional current, respectively), ρ is the mean density of sea water, h is the thickness of the top ocean level, (τ_x, τ_y) are the momentum fluxes (in N/m^2), and C_p^{ocn} is the specific heat of sea water.

The flux calculations in the coupler follow the CCM3 parameterizations (Kiehl et al., 1996). Here, we review only important flux calculations over the ocean. The

sensible heat flux SH , evaporation E (proportional to the latent heat flux LH), and momentum fluxes (τ_x, τ_y) follow the bulk formulae:

$$\begin{aligned}
SH &= \rho_A C_p C_H U_0 (\theta_0 - SST) \\
E &= \rho_A C_E U_0 [q_0 - q_{\text{sat}}(SST)] \\
\tau_x &= \rho_A C_M U_0 u_0 \\
\tau_y &= \rho_A C_M U_0 v_0.
\end{aligned}
\tag{4.4}$$

Here, (C_H, C_E, C_M) are the transfer coefficients for sensible heat, evaporation, and momentum (which depend on the atmospheric vertical stratification and the wind speed), C_p is the specific heat of the air at constant pressure, ρ_A is the density of the air, $(\theta_0, q_0, U_0 = \sqrt{u_0^2 + v_0^2})$ are the potential temperature, the specific humidity, and the wind speed at the bottom level, (u_0, v_0) are the wind velocity in the zonal and meridional directions at the bottom level, and q_{sat} is a saturation specific humidity over the ocean as an exponential function of SST (SST).

The land and sea ice components of the model are computed at the same horizontal resolution as the ocean and interchange the fluxes with the atmosphere through the coupler. The sea ice model predicts sea ice thickness, fraction, temperature, as well as snow depth on it. Lateral and vertical formation/melting, thermal conduction, and radiation in the sea ice are modeled (Bettge et al., 1996). The existence of sea ice modifies surface roughness and albedo.

From a software perspective, the non-atmosphere components of the coupled model are implemented as the subprocesses of the atmospheric model and serially coupled with the other components. The FOAM utilizes collocated two-dimensional

decomposition for parallel computing (i.e., each process computes the same horizontal portion of the atmosphere, ocean, and the other components). The model source code is compiled into a single executable, and the message passing interface (MPI) is employed for the parallel implementation. When 16 processors of the Deepthought2 cluster at the University of Maryland (Ivy Bridge, 2.8 GHz) are employed, the model finishes a 50-model-year integration within three hours. The computational efficiency of the model is essential for an early-stage study of coupled DA since EnKFs need to run tens of ensemble members in parallel. Besides, a coupled model has longer spin-up time than an atmospheric model, which means we need more extended experiments to evaluate the method accurately.

4.1.2 FOAM characteristics

We briefly examine and review the relevant characteristics of the model in this subsection. We focus on the model’s inherent ability to reproduce internal variabilities, especially those originating from the atmosphere-ocean interaction (e.g., ENSO). However, quantitative comparisons of the model climate to the real climate is omitted because all of the subsequent assimilation experiments are conducted in a perfect-model scenario.

The model is integrated for 250 years from the sample initial condition provided by the model developers¹. With this long integration, we can examine the model’s climatology as well as its seasonal and interannual variabilities.

We first examine the discrepancy between the initial condition and the model’s

¹<ftp://ftp.mcs.anl.gov/champp/foam>, last accessed June 24, 2019.

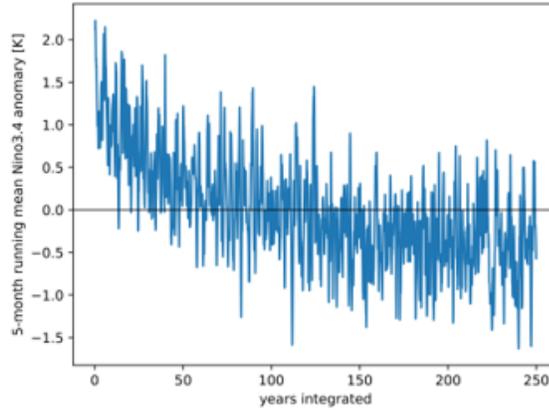


Figure 4.1: Monthly temporal anomaly of Niño3.4 regional (5°S - 5°N and 170 - 120°W) average sea surface temperature. The anomaly is defined as the difference from the average of 250 instances on the same day of a year. For visualization purpose, a five-month running mean is applied to the anomaly.

attractor. Figure 4.1 shows the temporal anomaly of Niño3.4 regional average sea surface temperature (SST); the variable is chosen to represent the prominent interannual variabilities. We notice that the Niño3.4 SST has a downward drift of roughly 2 Kelvins in the first 150 years of the integration. In contrast, the last 100 years show almost no signs of drift. Although some variables including the temperature of the deep ocean continue to drift in the timescale of thousands of years (examined in an independent experiment; not shown), for our purpose of real-time analysis and reanalysis experiments, 150-year spin-up is long enough for the model to settle on its attractor. Therefore, later in this subsection, we exclusively examine the last 100 years of this integration as the model’s climatological behavior.

Next, we check the annual mean state to see if the model’s climatology is a reasonable representation of the real atmosphere-ocean system. Figure 4.2 shows the annual mean climatology of SST and the ocean temperature at the equator. In the top panel, we can see that the warm pool extends over the Indian Ocean and

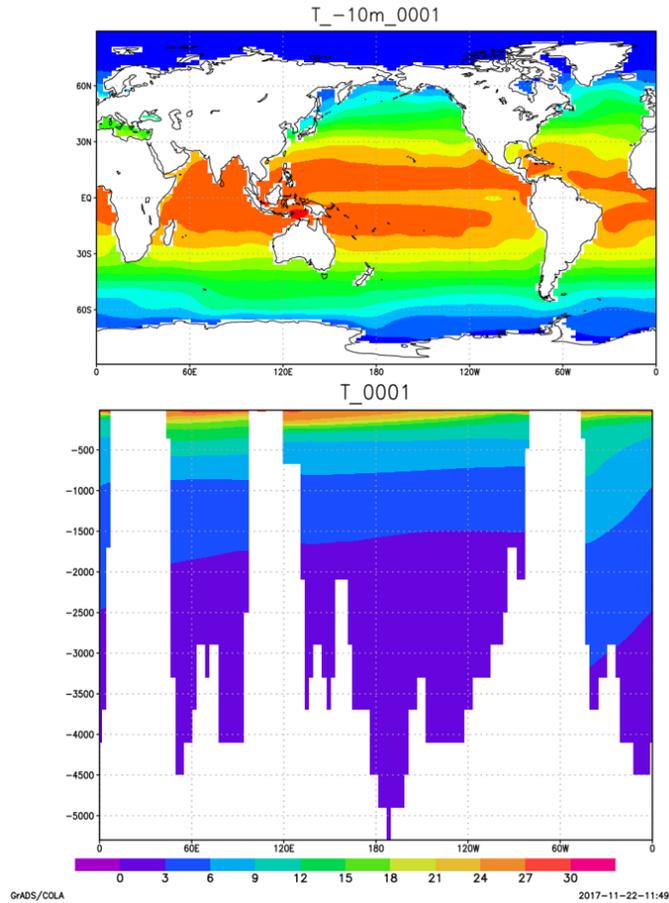


Figure 4.2: One hundred year average of the ocean temperature. Within the 100-year period, Total 1200 instances on the first day of each month are averaged. Top: sea surface temperature. Bottom: ocean temperature cross-section at the equator. The y-axis of the bottom panel shows depth in meters, and the color bar shows the temperature in degrees Celsius.

the Maritime Continents. In comparison to the observed sea surface temperature (e.g., Figure 7.14 and 8.9 of [Hartmann, 2016](#)), the tropical and equatorial sea surface temperature in the model is colder by a few Kelvins. However, the model successfully reproduces 3-5 K East-West temperature gradient at the equatorial Pacific, which should be the result of the atmospheric Walker circulation coupled to the oceanic model.

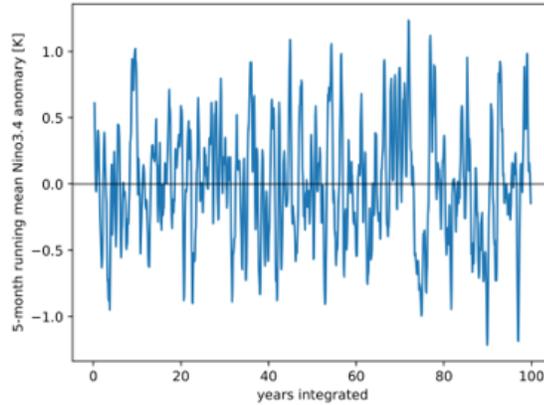


Figure 4.3: Same as Figure 4.1 but only the last 100 years is shown; the anomaly is defined against the 100-year average.

One of the primary expectations on the seasonal prediction by the atmosphere-ocean coupled models is to reproduce the El Niño Southern Oscillation (ENSO) variability, which is thought to be an intrinsically coupled mode explained by the Bjerknes feedback. Figure 4.3 shows the temporal anomaly of the Niño3.4 regional mean SST. Although the amplitude (~ 1 K) is insufficient and the period (a few years) is too short, the model exhibits anomalous SST at the region at the inter-annual time scale. Since the El Niño variability is the largest variability of the sea surface temperature in seasonal to interannual timescales, the variability is expected to appear in the empirical orthogonal function (EOF) analysis. Figure 4.4 shows the result of the EOF analysis. The EOF analysis is applied to the monthly sea surface temperature field over the 100 years. Since monthly data is used, high-frequency variabilities with period shorter than one month are implicitly filtered. In panel (a), we can see that the EOF1 corresponds well to the known El Niño variability. Comparing Figure 4.4(c) with Figure 4.3, we see that the variation of Niño3.4 explains well the first principal component (PC1) time series. Hence the El

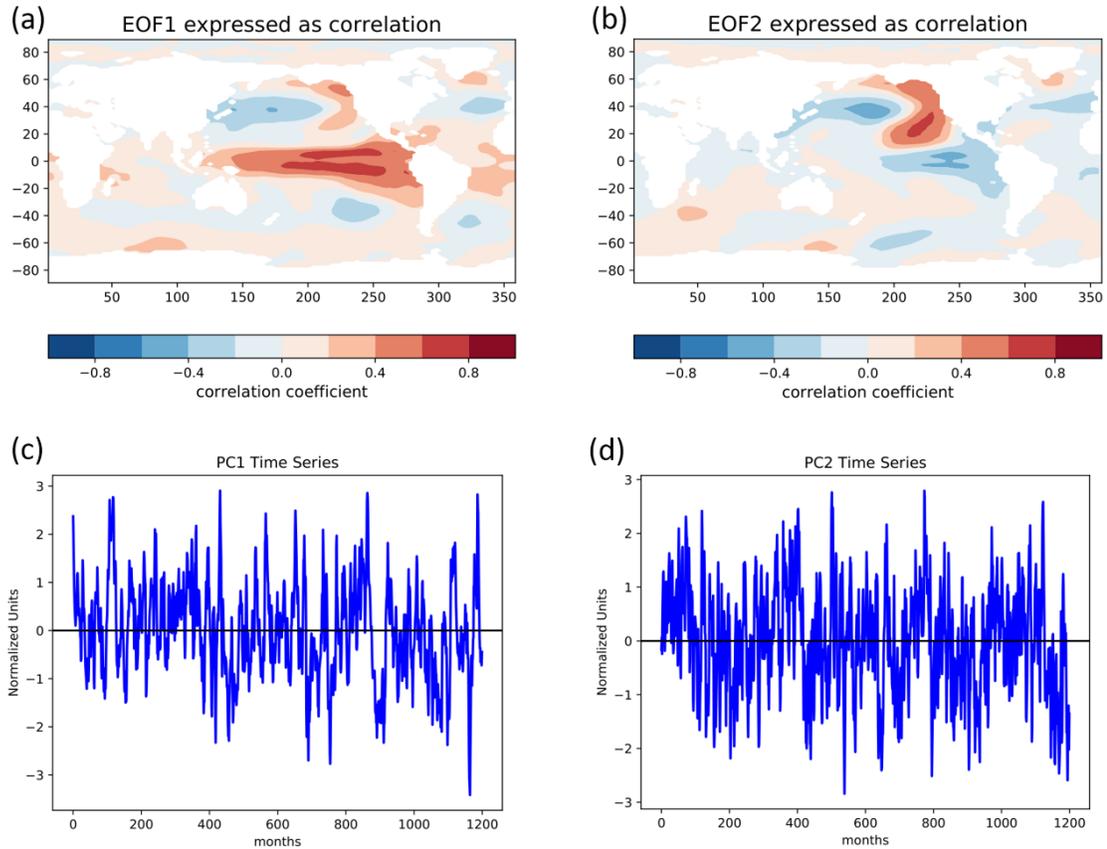


Figure 4.4: The first two empirical orthogonal functions (EOFs) calculated from the monthly, 100-year time series of the sea surface temperature. Top row: the first and second EOFs normalized by the variance of each grid, i.e., each grid value’s temporal correlation to the principal component. Bottom row: The time series of the first and second principal components (PCs).

Niño naturally appears as the largest internal SST variability of this model. The second EOF shown in panel (b) resembles the north Pacific oscillation (Deser and Blackmon, 1995; Hartmann, 2015) whose typical timescale is slightly shorter than that of ENSO (panel d).

Finally, we examine the seasonal variation of the atmospheric circulation, including the Asian Monsoon. Figure 4.5 shows the 100-year average wind field at the bottom model level on January 1st and July 1st. From the figure, we can first no-

tice some features prevalent throughout the year; the most prominent is the strong westerly wind at 45°S surrounding Antarctica. Some other features are seasonally varying. For example, in the boreal winter (upper panel), we can see cyclonic circulation at the northernmost Pacific, which corresponds to the Aleutian low. We can also see that the northeasterly wind is prevalent over India and Southeast Asia. In the boreal summer (lower panel), the dominant wind direction over India and Southeast Asia is southwesterly, which advects moisture from the ocean to the continent. In the northern Atlantic, anti-cyclonic circulation corresponding to the Bermuda high is also apparent. These continental-scale, seasonally dependent surface wind circulations closely resemble that of the real atmosphere (e.g., Figure 1.19 of [Wallace and Hobbs, 2006](#)).

From these observations, we can conclude that this affordable coupled model has good overall abilities to represent the global variabilities observed in the atmosphere and the ocean, and it is suitable for our purpose of exploring coupled data assimilation methodologies.

4.2 Observation network

In order to evaluate DA methodologies in idealized settings, we will conduct observation system simulation experiments (OSSEs), which mimic the data assimilation applications. In an OSSE, the true atmosphere-ocean state to be estimated is simulated by an integration of a numerical prediction model, which is referred to as the “nature run” or the “simulated truth”. Then, the state of the nature

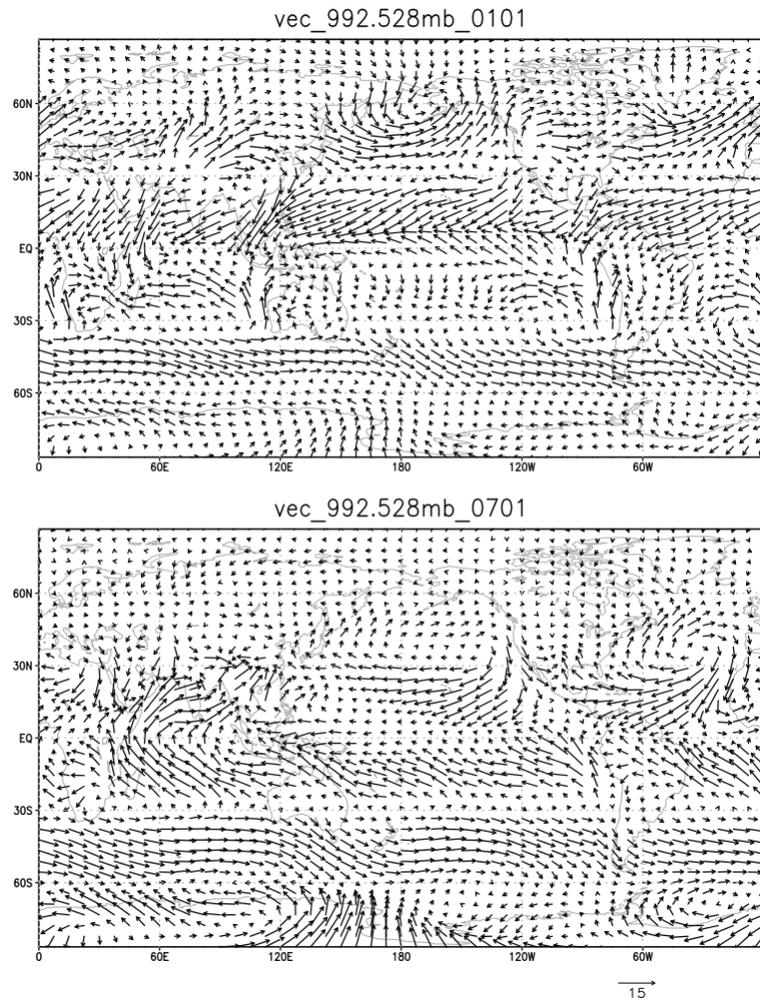


Figure 4.5: One-hundred-year average of bottom-level wind fields of different seasons. The bottom model level roughly corresponds to 992.5 hPa. The arrows' length represent the wind speed, with the arrow at the bottom being 15 m/s. Top: boreal winter (January 1st). Bottom: boreal summer (July 1st).

run is imperfectly observed by simulated observations, which are functions of the simulated true state plus an observation error:

$$\mathbf{y}^o = h(\mathbf{x}^{\text{truth}}) + \boldsymbol{\epsilon}^o, \quad (4.5)$$

where h and $\mathbf{x}^{\text{truth}}$ are the observation operator and the simulated true state, and $\boldsymbol{\epsilon}^o \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$ is the observation error for the observation error covariance matrix \mathbf{R} . It is impossible to perfectly know the true state from these observations because the number of observations available p is much smaller than the number of unknown state variables n , and the observations are erroneous. This setting simulates the actual data assimilation applications, where we can only infer the true state through imperfect observations that only partially cover the domain.

Observation system simulation experiments have certain advantages and disadvantages compared to experiments with real observations. One major advantage is that we can evaluate the analysis/forecast accuracy by comparing it to the “truth”, which is unknown in the real applications. It also eliminates the imperfections in the dynamical model M , observation operator h , and observation error covariance \mathbf{R} if these imperfections are not of interest. This ideal setting enables us to isolate the differences caused by DA methods, and effectively investigate DA methodology.

In designing an OSSE, it is important to simulate the observability of the real system sufficiently well. The real observing network is highly nonuniform. For example, conventional atmospheric observations such as radiosondes, are mostly available over land and thus more densely distributed over the northern hemisphere. The observation networks simulated and assimilated are shown in Table 4.2 and

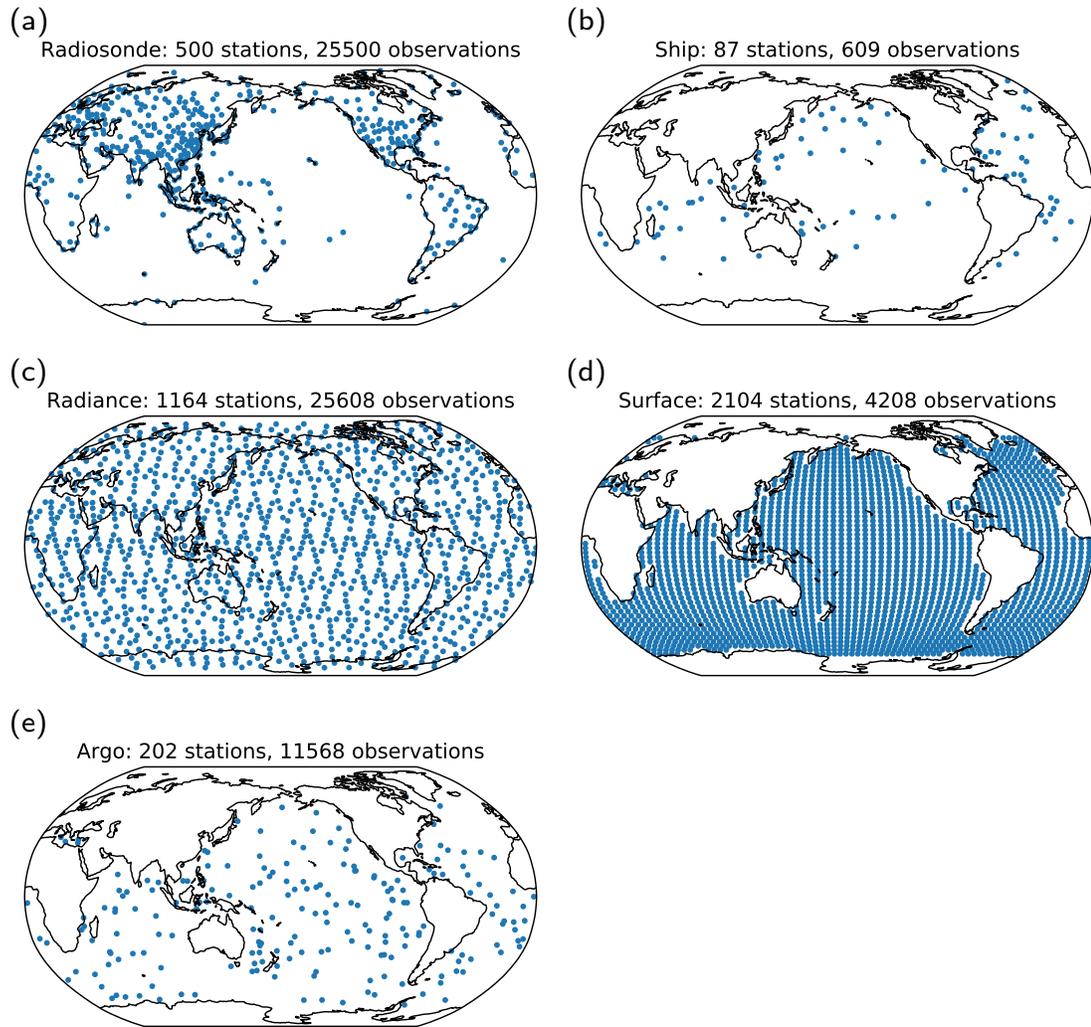


Figure 4.6: Horizontal distribution of assimilated observation networks listed in Table 4.2. The ship and Argo observations are randomly generated and vary with time. The ship locations are sampled from historical Voluntary Observing Ship program (VOS) locations to simulate its nonuniform distribution while Argo float locations are uniformly random over the model ocean.

Figure 4.6. For some experiments, the quasi-uniform observation network of Figure 4.7 is also used.

Table 4.2: Observation networks simulated and assimilated. For vertical profile observations (radiosondes, radiance, and Argo floats), each model level is directly observed; i.e., the observation operator only incorporates horizontal interpolation.

Observation type	Observed element (error STDV)	Approx. horizontal number of stations	Vertical availability
(Atmosphere)			
Radiosondes	U, V (1 m/s) T (1 K) Q (0.1 g/kg) Ps (1 hPa)	500	Up to 100 hPa (500 hPa for Q)
Satellite radiance	T (2 K) Q (0.2 g/kg)	1000	Up to 100 hPa (500 hPa for Q)
Ship	U, V (1 m/s) T (1 K) Q (0.1 g/kg) Ps (1 hPa)	100	Surface (bottom level)
(Ocean)			
Ship	T (1 K) S (0.1 PSU)	100	Surface (top level)
Satellite SST, SSH	T (1 K) SSH (1 cm)	2000	Surface (top level)
Argo floats	U, V (1 cm/s) T (1 K) S (0.1 PSU)	200	Down to 2000 m

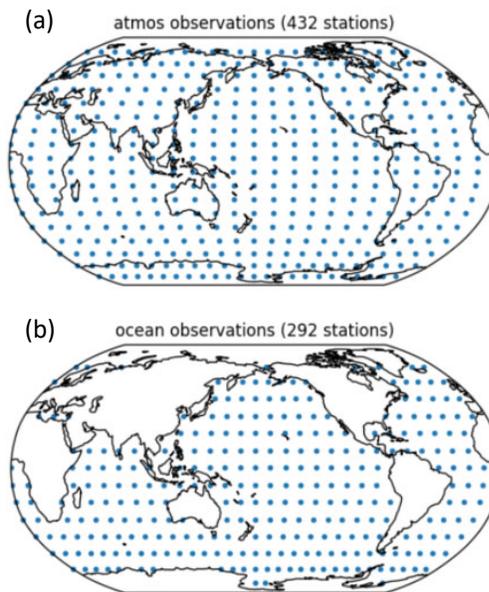


Figure 4.7: The quasi-uniform observation network used for some experiments. (a): atmospheric variables (T, Q, U, V, P_s) are observed in the bottom 12 levels with observation error standard deviations of (1 K, 1 g/kg, 1 m/s, 1 m/s, 1 hPa), respectively. (b): oceanic variables (T, S, U, V) are observed in the top 12 model levels with observation error standard deviations of (0.1 K, 0.01 PSU, 1 cm/s, 1 cm/s), respectively.

4.3 Data assimilation system: FOAM-LETKF

We have implemented a strongly-coupled atmosphere-ocean LETKF system for our experiments (FOAM-LETKF).

The analysis variables are the same as the prognostic variables of the model. For the atmosphere, temperature (T), specific humidity (Q), the zonal wind (U), and the meridional wind (V) are analyzed at every 3-D grid point, and surface pressure (P_s) is analyzed at every 2-D grid point. These add up to 140,160 analysis variables in the atmosphere. For the ocean, temperature (T), salinity (S), the zonal current (U), and the meridional current (V) are analyzed at every 3-D grid point, and pressure anomaly of the top layer (P_{top}) is analyzed at every 2-D grid point. These add up to 781,334 analysis variables in the ocean.

Although we may extend the same methodology to all the subsystems of the coupled model, we focus on the analysis of the atmosphere and the ocean. Therefore, the prognostic variables of land, sea ice, and river models are unchanged in the analysis steps.

The analysis update is based on the local ensemble transform Kalman filter (LETKF; [Hunt et al., 2007](#)), as summarized in Algorithm 2. The LETKF enables each analysis variable to be updated concurrently and is implemented in parallel using the Message Passing Interface (MPI).

Unless otherwise noted, spatial localization weight (Line 4 of Algorithm 2) is given by the piecewise polynomial function (Eq. 1.5) based on the horizontal and vertical distances between the analysis variable and the observation. In the analysis

Algorithm 2: LETKF

```
1  $n$  : Number of model variables
2  $K$  : Number of ensemble members
3  $p$  : Number of observations
4  $weight(i, j)$  : Localization weight defined between  $i$ th model variable and  $j$ th
   observation
5  $\bar{\mathbf{x}}_g^b$  : Background ensemble mean model state  $\triangleright (n \times 1)$ 
6  $\mathbf{X}_g^b$  : Background perturbation model state  $\triangleright (n \times K)$ 
7  $\mathbf{y}_g^o$  : Observations  $\triangleright (p \times 1)$ 
8  $\bar{\mathbf{y}}_g^b$  : Background ensemble mean observations  $\triangleright (p \times 1)$ 
9  $\mathbf{Y}_g^b$  : Background perturbation observations  $\triangleright (p \times K)$ 
10  $\mathbf{R}_g$  : Observation error covariance matrix  $\triangleright (p \times p)$ 
11
12 function  $LETKF(\bar{\mathbf{x}}_g^b, \mathbf{X}_g^b, \mathbf{y}_g^o, \bar{\mathbf{y}}_g^b, \mathbf{Y}_g^b, \mathbf{R}_g)$ 
13    $\bar{\mathbf{x}}_g^a \leftarrow$  Empty  $n$ -column vector
14    $\mathbf{X}_g^a \leftarrow$  Empty  $n \times K$  matrix
15   for  $i \in [1, n]$  do
16      $\bar{\mathbf{x}}^b \leftarrow$   $i$ th row of  $\bar{\mathbf{x}}_g^b$ 
17      $\mathbf{X}^b \leftarrow$   $i$ th row of  $\mathbf{X}_g^b$ 
18      $\mathcal{J} \leftarrow \{j | j \in [1, p], weight(i, j) > 0\}$ 
19      $\mathbf{y}^o \leftarrow$  Rows  $\mathcal{J}$  of  $\mathbf{y}_g^o$ 
20      $\bar{\mathbf{y}}^b \leftarrow$  Rows  $\mathcal{J}$  of  $\bar{\mathbf{y}}_g^b$ 
21      $\mathbf{Y}^b \leftarrow$  Rows  $\mathcal{J}$  of  $\mathbf{Y}_g^b$ 
22      $\mathbf{R} \leftarrow$  Rows  $\mathcal{J}$ , columns  $\mathcal{J}$  of  $\mathbf{R}_g$ 
23     for  $j \in \mathcal{J}$  do
24        $j' \leftarrow$  Local index of  $\mathbf{y}^o$  which corresponds to  $j$ 
25       Multiply  $j'$ th column of  $\mathbf{R}^{-1}$  by  $\sqrt{weight(i, j)}$ 
26       Multiply  $j'$ th row of  $\mathbf{R}^{-1}$  by  $\sqrt{weight(i, j)}$ 
27     end
28      $\tilde{\mathbf{P}} \leftarrow [(K - 1)\mathbf{I} + \mathbf{Y}^{bT}\mathbf{R}^{-1}\mathbf{Y}^b]^{-1}$ 
29      $\mathbf{W}^a \leftarrow [(\tilde{\mathbf{P}})^{1/2}]$ 
30      $\bar{\mathbf{w}}^a \leftarrow \tilde{\mathbf{P}}\mathbf{Y}^{bT}\mathbf{R}^{-1}(\mathbf{y}^o - \bar{\mathbf{y}}^b)$ 
31      $i$ th row of  $\bar{\mathbf{x}}_g^a \leftarrow \bar{\mathbf{x}}^b + \mathbf{X}^b\bar{\mathbf{w}}^a$ 
32      $i$ th row of  $\mathbf{X}_g^a \leftarrow \mathbf{X}^b\mathbf{W}^a$ 
33   end
34   return  $\bar{\mathbf{x}}_g^a, \mathbf{X}_g^a$ 
35 end function
```

system, integer vertical coordinates on Table 4.3 is used for vertical localization. An important aspect of implementation is observation lookup, which is explained in Appendix A.

To counteract the under-dispersive analysis ensemble, the relaxation to prior perturbation method (RTPP; Zhang et al., 2004) is employed. The RTPP method relaxes the analysis perturbations back to the prior perturbations:

$$\mathbf{X}^a \leftarrow (1 - \alpha)\mathbf{X}^a + \alpha\mathbf{X}^b, \quad (4.6)$$

where $\alpha \in [0, 1]$ is a relaxation parameter, \mathbf{X}^a and \mathbf{X}^b are the analysis and background ensemble perturbations.

After applying the LETKF and RTPP updates, the positive sign of specific humidity analysis is ensured by rounding values up to 10^{-3} g/kg, independently for each member and each grid point. This positive sign for the specific humidity is required to avoid problems in running the subsequent forecast. The threshold used is roughly equal to the minimum value found in the free integration of the model.

In order to enhance the stability of the DA cycle, incremental analysis update (IAU; Bloom et al., 1996) is implemented into the model. Incremental analysis update distributes the analysis increment estimated at certain time, into a finite time window, so that the model state can gradually incorporate the analysis increment. Incremental analysis update is implemented in both atmospheric and oceanic parts of FOAM.

The original IAU by Bloom et al. (1996) distributes the analysis increment into a window centered at the analysis time. For example, if the analysis interval is 24

Table 4.3: Vertical coordinate of the FOAM-LETKF analysis system.

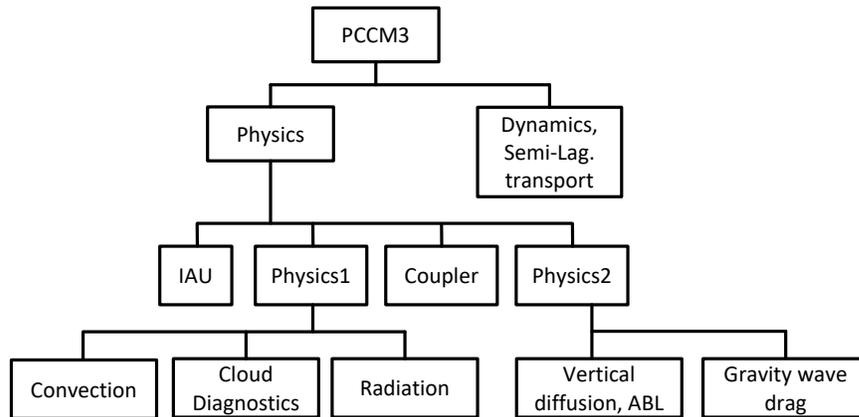
Level	Representative pressure/depth	Variables defined
18	4.81 hPa	
17	13.07 hPa	
16	32.56 hPa	
15	63.95 hPa	
14	99.04 hPa	
13	138.71 hPa	
12	189.19 hPa	
11	252.14 hPa	
10	324.85 hPa	T, Q, U, V
9	408.96 hPa	(atmos)
8	501.28 hPa	
7	598.25 hPa	
6	695.17 hPa	
5	786.51 hPa	
4	866.41 hPa	
3	929.28 hPa	
2	970.45 hPa	
1	992.53 hPa	
0	-	P _s (atmos), P _{top} (ocean)
-1	10 m	
-2	30 m	
-3	51 m	
-4	74 m	
-5	99.5 m	
-6	129 m	
-7	165 m	
-8	210 m	
-9	271 m	
-10	357 m	
-11	484 m	
-12	675 m	T, S, U, V
-13	952.5 m	(ocean)
-14	1307.5 m	
-15	1700 m	
-16	2100 m	
-17	2500 m	
-18	2900 m	
-19	3300 m	
-20	3700 m	
-21	4100 m	
-22	4500 m	
-23	4900 m	
-24	5300 m	

hours, the analysis increment valid at 0 UTC, January 2 is distributed to a 24-hour window from 12 UTC, January 1 to 12 UTC, January 2. [Yan et al. \(2014\)](#) named this original method as IAU 50, as 50% of the window overlaps the pre-existing background integration. However, we may also distribute the analysis increment to a window not centered at the analysis time. [Yan et al. \(2014\)](#) tested three different IAU implementations (IAU 0, IAU 50, and IAU 100; defined by how many percents of the window overlaps the pre-existing background integration) with a simple non-linear ocean model. They showed that the difference between these IAU schemes on the global scale is limited, and they recommended the IAU 0 scheme when the computational cost is taken into account. Following their recommendation, the IAU 0 scheme is adopted in our implementation for FOAM-LETKF; an analysis increment valid at 0 UTC, January 2 is distributed to the subsequent 24-hour window from 0 UTC, January 2 to 0 UTC, January 3.

Incremental analysis update is implemented as if it was one of the physical forcings in the model. [Figure 4.8](#) shows the schematic call trees of the atmosphere and ocean components of FOAM, implemented with IAU.

From a software perspective, the analysis system is implemented as follows: the truth and background states are read from the restart files of FOAM, noisy observations are generated within the LETKF executable, and the analysis ensemble states are written into restart files. When IAU is enabled, the analysis increments are written into independent files; then the coupled model reads analysis increment files as well as the background restart files to compute the subsequent background states. In this sense, FOAM-LETKF is an offline assimilation system that is implemented

(a) Atmosphere model



(b) Ocean model

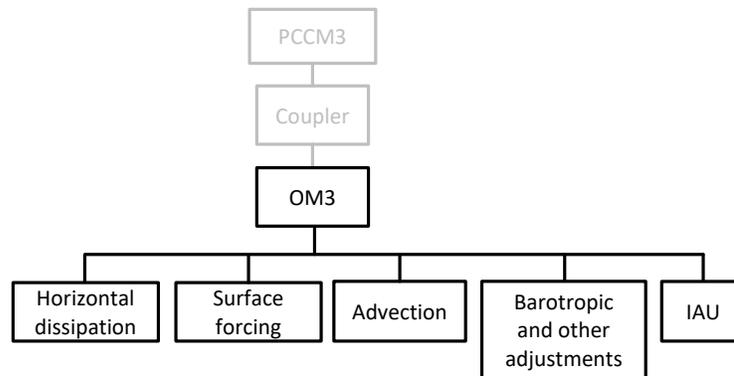


Figure 4.8: Abstract call trees of FOAM after implementation of incremental analysis update (IAU). PCCM3 and OM3 are the names of the component models, and ABL stands for atmospheric boundary layer (parameterization scheme). In FOAM, the ocean model (OM3) is called as one of the physical components of the atmospheric model (PCCM3), and they are sequentially coupled through the exchange of SST and fluxes. In the same level of a tree, tasks are generally processed from left to right. The atmospheric call tree is a modification of [Jacob \(1997\)](#).

separately from the model executable ([Liu et al., 2017](#)).

Chapter 5: Background error statistics of atmosphere-ocean systems

5.1 Introduction

In most methods of data assimilation including variational methods and Kalman filters, estimation of error statistics is an essential problem. An underlying assumption for these analysis methods is that we know better about the long-term statistics of states or errors than the evolving state itself.

Of error statistics, we are particularly interested in the background error covariance (**B** matrix). The **B** matrix should reflect the dynamic processes, including the growing error mode, that is, the direction (in the phase space) in which forecast error tend to grow rapidly.

For variational methods, the **B** matrix must be specified *a priori*. For ensemble Kalman filters (EnKFs), although they can estimate the flow-dependent background error covariances, the covariance estimate can be improved by appropriate localization based on offline statistics (Chapter 2). In other words, the EnKFs can also benefit from the prior knowledge of background error covariances.

Correlations are covariances normalized by variances, and therefore useful for multivariate systems where each variable has different units (e.g., hPa for surface pressure and PSU for salinity). For the atmosphere-ocean coupled system, there

exist two pilot studies that have examined the background or analysis error correlations. [Smith et al. \(2017\)](#) investigated the vertical aspects of error correlations between the atmosphere and ocean, using a strongly coupled ensemble of a single-column model at a particular point in the northwestern Pacific. They have shown strong diurnal and seasonal variabilities of error structures, as well as its vertical distribution, mostly bounded to the atmospheric boundary layer and the oceanic mixed layer. [Feng et al. \(2018\)](#) examined the analysis errors of surface temperatures (surface air temperature and SST), which is expected to be one of the dominant mechanisms of coupling. They used an ensemble of coupled variational data assimilations (CERA-20C by ECMWF) to examine the analysis errors of the surface temperatures. Both of these pilot studies provided substantial knowledge for the coupled data assimilation community. However, they only examined the vertical or pointwise features of coupled errors, which is insufficient for three-dimensional data assimilation system. The influence of an observation must also be horizontally distributed to provide smooth analysis increments.

Therefore, our study aims to extend their investigation in a few points:

- As examined by [Feng et al. \(2018\)](#), the thermal coupling through radiative processes and exchange of sensible and latent heat fluxes is known to have major impacts. However, the atmosphere and ocean also exchange momentum (mostly the wind driving the currents) and freshwater (through precipitation and evaporation). These processes also need attention because they play a crucial role in, for example, the genesis of western boundary currents, El Nino

Southern Oscillation, and Atlantic Meridional Overturning Circulation.

- Horizontal dynamics of ocean is absent in the investigation by [Smith et al. \(2017\)](#), which may change the error structures.
- Atmosphere and ocean are known to have different length scales of error correlations, vertically and horizontally. However, in the existence of coupled processes, the horizontal correlation length scales near the surface should affect each other. The consistent picture of horizontal error correlations near the surface will be sought.
- The error growth of a coupled system is known to be affected by the model resolutions, especially that of the ocean model (e.g., [Hallberg, 2013](#); [De Cruz et al., 2018](#)). We briefly compare the error statistics of two global coupled models with very different resolutions.
- Finally, the time-mean error correlation does not include the flow-dependent portion of error correlation. By examining the mean *squared* error correlation, we obtain average strength of error correlation, which can be used for localization of strongly coupled EnKFs (also see [Section 5.3](#) later).

5.2 Previous efforts on estimating and modeling error correlations

Before estimating the coupled error structure between the atmosphere and the ocean, we quickly review known characteristics and previous efforts on estimating and modeling error correlations in the atmosphere and ocean. This review will help

to interpret our results.

5.2.1 Background error correlations of the atmosphere

The estimation of background error statistics in the atmosphere given observation networks, has been attempted with several approaches.

An innovation-based estimation of background error statistics is proposed by [Hollingsworth and Lönnberg \(1986\)](#). [Daley \(1993\)](#), in Chapters 4 and 5, summarizes univariate and multivariate error covariances estimated by their method. The errors are reported to be largely geostrophic and nondivergent. For example, [Figure 5.1](#) shows the innovation correlation within the height field (panel a) and between the height and wind fields (panel b). The height innovation is positively correlated over a thousand kilometers with correlation decreasing with the distance. The transversal wind (i.e., rotational wind centering the height observation) innovation has peak correlation to the height innovation at the distance of hundreds of kilometers, reflecting the quasi-geostrophic and quasi-nondivergent wind structures.

To obtain the error statistics in the model space, [Parrish and Derber \(1992\)](#) proposed the NMC method; the difference between two forecasts initialized at different times and valid at the same time serves as a proxy to the background error. The sampled forecast differences are then modeled into a product of sparse matrices so that it can be stored in computers. They first transformed the model's prognostic variables to analysis variables with mutually uncorrelated errors by decomposing them into balanced and unbalanced parts. Further, vertical error correlation is rep-

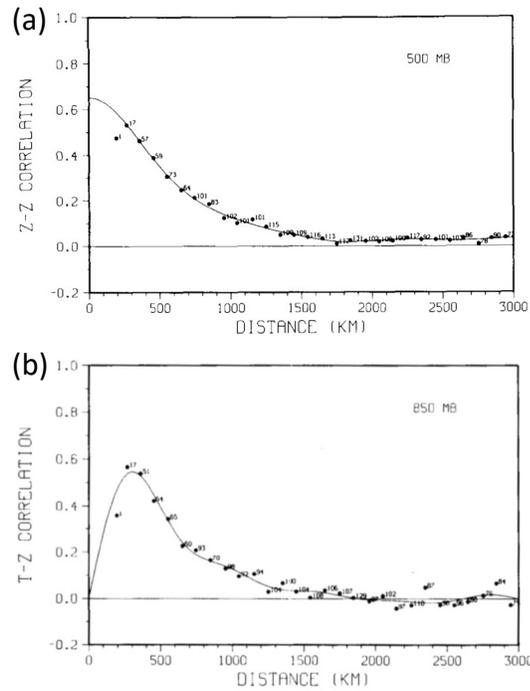


Figure 5.1: Innovation (observation minus background; a proxy to the background error) correlation between (a): height and height at the 500 hPa level and (b): height and transversal (or tangential) wind at the 850 hPa level. The observations are radiosonde over North America, and the background is 6-h forecast by their global model with horizontal resolution T63. Figure courtesy [Lönnerberg and Hollingsworth \(1986\)](#).

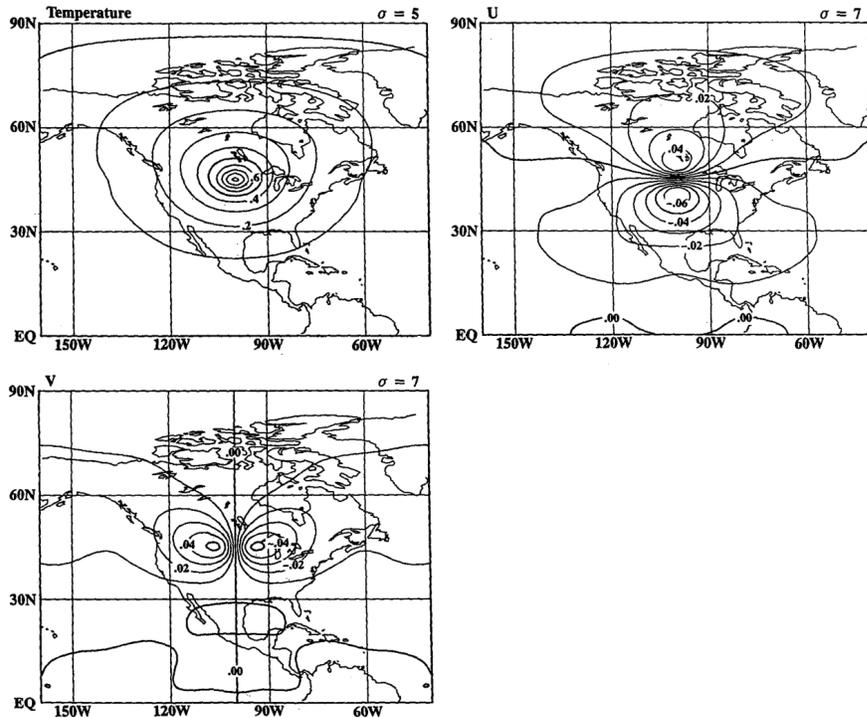


Figure 5.2: Increment caused by the assimilation of a positive temperature innovation at 45°N , 100°W at layer 5. Increment of (a): temperature at layer 5, (b): zonal wind at layer 7, and (c): meridional wind at layer 7. Figure courtesy Parrish and Derber (1992).

represented by empirical orthogonal functions for each analysis variable. They further assume that background error of each analysis variable is horizontally homogeneous and isotropic so that the background error covariance is diagonal in the spectral space. By this way, spatial error correlation is represented by a sum of spectral modes, and multivariate error correlation is represented by the balanced part of analysis variables (Figure 5.2).

Two primary forms of grid-space correlation modeling are the diffusion operator (Derber and Rosati, 1989) and the higher-order recursive filter (Purser et al., 2003), both of which can incorporate inhomogeneity and anisotropy. The latter

is more computationally efficient for the same accuracy, while the former is more flexible for complex boundary conditions and topographies.

[Ingleby \(2001\)](#) provided a broad picture of background errors based on the NMC method. He noted several features: vertical correlation scale increases with smaller horizontal wavenumber; horizontal error correlation length is larger in the stratosphere and the tropics consistently to the Rossby deformation radius; latitudinal dependency of correlation length; and multivariate error correlations between pressure and temperature.

[Fisher \(2003\)](#) introduced an ensemble of (variational) analysis method for estimation of background error covariance and discussed its advantage over the NMC method. The ensemble of analysis method is superior because the forecast length is the same as the one used for background, and unlike the NMC method, the ensemble method can capture large background errors where observations are sparse. They further introduced wavelet representation of error modes, which enabled the spatially and spectrally variable error structures.

[Bannister \(2008a,b\)](#) are review papers summarizing the atmospheric error covariance structure and how modeling centers model the covariance into a product of sparse matrices. In addition to the efforts mentioned above, an important characteristic of background error covariance models for variational applications is that its matrix square root form is available for the efficient minimization (control variable transform).

5.2.2 Background error correlations of the ocean

For ocean models, due to topographic constraints, covariance modeling is limited to spatial representations.

The diffusion operator of [Derber and Rosati \(1989\)](#) was first developed for the global ocean data assimilation at the Geophysical Fluid Dynamics Laboratory (GFDL). In their univariate analysis of global ocean temperature with 3D-Var, the Laplacian operator is repeatedly applied so that the resulting horizontal error correlation approximates a Gaussian-shaped function. They further modeled the shorter correlation length in higher latitudes and zonally elongated correlation near the equator following the predominant dynamical scales.

[Waters et al. \(2015\)](#) discusses recent efforts at the United Kingdom Met Office for the NEMOVAR system. Their system is set up with a global quarter-degree ocean model, which may be referred to be eddy-permitting. The background covariance model includes the transform of variables to the balanced and unbalanced variables, spatially variable horizontal and vertical correlation lengths via application of diffusion operators, and inclusion of flow dependencies using the background mixed layer depth and vertical temperature gradient. [Figure 5.3](#) shows that the correlation model can represent the well-correlated mixed layer, short correlation length in the seasonal thermocline, and the longer correlation length in the deeper ocean.

[Weaver et al. \(2017\)](#) explains another background error formulation in the hybrid NEMOVAR system. Their implementation includes a approach unique for

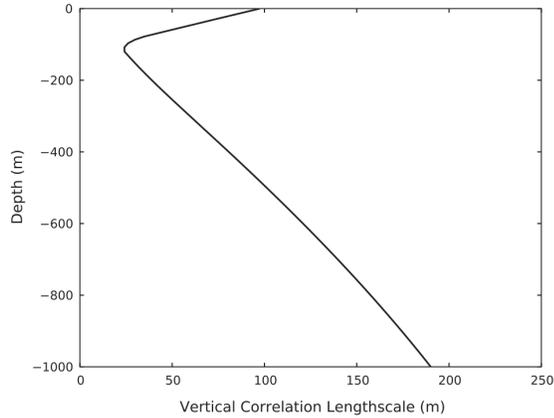


Figure 5.3: Modeled vertical background error correlation length when the mixed layer depth is 97 m. Figure courtesy [Waters et al. \(2015\)](#).

the ocean error modeling, the use of multivariate empirical orthogonal functions (EOFs). Here, empirical orthogonal functions are leading modes of *anomaly* (or temporal) covariance, which explains the large-scale error correlations ([Fujii and M. Kamachi, 2003](#)). Figure 5.4 shows how the EOF-based background error covariance model can spread the observation information to the analysis field better than the diffusion-based background error covariance model. The use of anomaly covariance as the background error covariance is justified especially with sparse observation network or during the spin-up period; the background (or forecast) PDF will fall back to the climatological PDF when no observational constraints are available.

5.3 Methods and data sets

In this section, we explain the two data sets of global atmosphere-ocean weakly coupled EnKF systems and important terminologies.

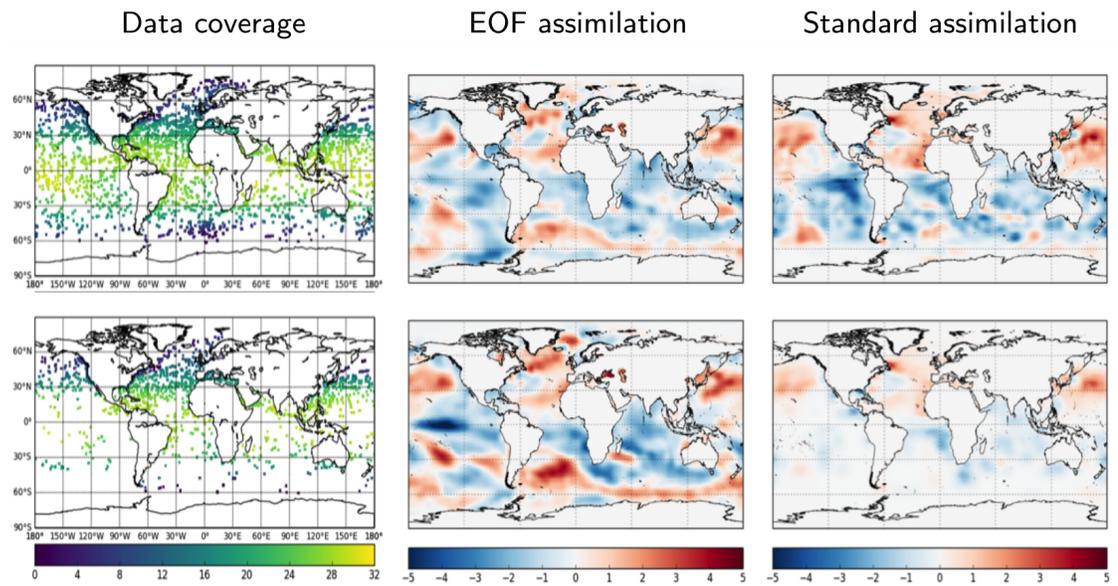


Figure 5.4: Effect of empirical orthogonal functions as background error covariance in the ocean temperature data assimilation. The top and bottom rows represent assimilation experiment of dense and sparse observation networks, respectively. Left: assimilated observations. Center: analysis increment with EOF-based background error covariance. Right: analysis increment with diffusion-based background error covariance. Figure courtesy [Weaver et al. \(2017\)](#).

5.3.1 Ensemble-based estimates of background error correlations

In our investigation, ensemble covariance of EnKF serves as a proxy to the background error covariance. The ensemble perturbations can contain multiple sources of uncertainties by using perturbed initial conditions (e.g., [Hoffman and Kalnay, 1983](#)) and perturbed observations (e.g., [Houtekamer and Mitchell, 1998](#)). In the deterministic EnSRFs, the observation error uncertainty is implicitly considered without perturbing observations by matching the analysis ensemble covariance to that of the extended Kalman filter ([Tippett et al., 2003](#)). One advantage of ensemble-based covariance estimation is that it provides flow-dependent estimates of error covariance represented by $\mathcal{O}(100)$ of ensemble perturbations.

We stress the difference between two distinct quantities that are thoroughly discussed:

- **Time-mean background ensemble correlations** (or *mean* ensemble correlations for short): $\langle \rho \rangle = \frac{1}{T} \sum_{t=1}^T \rho(t)$ for some time-dependent ensemble correlation $\rho(t)$ and the number of assimilation windows T . This will be the static background ensemble correlation used in the variational methods, with which we expect predominant coupling mechanisms to be explained.
- **Time-mean squared background ensemble correlations** (or *mean squared* ensemble correlations): $\langle \rho^2 \rangle = \frac{1}{T} \sum_{t=1}^T [\rho(t)]^2 = \langle \rho \rangle^2 + \langle [\rho'(t)]^2 \rangle$, where $\rho'(t) = \rho(t) - \langle \rho \rangle$ is the temporal fluctuation of the ensemble correlation. As shown by definition, this quantity includes the flow-dependent ensemble correlations as

Table 5.1: Settings of the weakly coupled FOAM-LETKF observation system simulation experiment (OSSE) from which ensemble statistics are sampled.

Experiment period	One model year (an arbitrary year from January 1; one model year of FOAM is constituted by 360 days)
Observation network	Quasi-uniform network available at the end of each window
Analysis method	Weakly coupled LETKF
Horizontal localization	700 km (atmosphere) and 300 km (ocean)
Vertical localization	1.5 model levels (atmosphere) and 1 model levels (ocean)
Analysis variables	T, Q, U, V, P_s (atmosphere), $T, S, U, V, P_{\text{top}}$ (ocean)
Analysis interval	24 hours
Ensemble	64 members; initial ensemble members and the true state are sampled from a 100-year free integration (the same date of a year)
Covariance inflation	RTPP; 50% prior perturbations (atmosphere) and 99% prior perturbations (ocean)

well as the time-mean ensemble correlation. Therefore, this quantity will be useful to guide the localization of ensemble-based data assimilation methods (Chapters 2 and 6). In some diagrams its square root, the *root-mean-square (RMS)* ensemble correlation is shown to provide a direct comparison to time-mean ensemble correlations.

5.3.2 FOAM-LETKF WCDA system/data

The first dataset is an observation system simulation experiment (OSSE) of the FOAM-LETKF system described in the previous chapter. The setting of the data assimilation cycle is listed in Table 5.1. We will explain a few important characteristics of the model and experimental setups.

The LETKF cycle is weakly coupled, that is, the forward evolution is calculated with the coupled model, but the atmosphere (ocean) model only assimilate the atmosphere (ocean) observations. Although it is unclear how strongly coupled analysis changes the error correlation structure, the background ensemble is expected to reflect the coupled error growth; preceding experiments with a coupled toy model showed that the error correlation structure obtained with weakly coupled EnKF is qualitatively similar with the one obtained with strongly coupled EnKF.

Note that the model is deterministic and uncertainty in model structure or parameters is not represented. Therefore, the background ensemble perturbations only depend on the initial condition uncertainties (i.e., analysis ensemble perturbations) and their growth in the coupled forward model; the analysis ensemble perturbations, in turn, depend on the previous background ensemble perturbations and observation uncertainties.

5.3.3 CFS-IITM system/data

To support the characteristics of coupled background error correlations found with FOAM-LETKF, we repeat some analyses with a more realistic coupled assimilation system, namely, CFS-IITM¹.

¹CFS stands for the Coupled Forecasting System, the atmosphere-ocean coupled model developed by the National Oceanic and Atmospheric Administration (version two, [Saha et al., 2010](#), [2014](#)). The LETKF system has been prepared by Dr. Travis Sluka ([Sluka, 2018](#)). IITM stands for Indian Institute of Tropical Meteorology, where CFS-IITM reanalysis is carried out. The data is provided by Dr. Sreenivas Pentakota and his collaborators.

Table 5.2: Settings of CFS-IITM reanalysis from which ensemble statistics are sampled.

Atmosphere model (GFS)	
Horizontal resolution	T126 (190 × 384 grids)
Vertical resolution	64 levels (hybrid $\sigma - p$ coordinate)
Ocean model (MOM4)	
Horizontal resolution	0.5 degrees (410 × 720 grids)
Vertical resolution	40 levels (z -coordinate)
Data assimilation system	
Analysis scheme	Weakly coupled 4D-LETKF
Analysis interval	6 hours
Ensemble size	40 members
Horizontal localization	1000 km (atmosphere) 200-720 km (ocean)
Vertical localization	0.4 lnP (atmosphere) none (ocean)
Covariance inflation	Relaxation to prior spread at 95%
Assimilated observations	PREPBUFR (See Figure 3.1 and Table 3.1 of Sluka, 2018)
Other	
Coupling frequency of the model	30 minutes
Period of analysis	2000-
Investigated period	June 2006 (out of 120 six-hourly analyses, six instances are excluded due to partial data loss)

Table 5.2 summarizes the important characteristics of this reanalysis product. We note three major differences from the FOAM-LETKF system. First, the system assimilates real observations of the atmosphere and the ocean, unlike the identical-twin setting of FOAM-LETKF. Second, the atmosphere and ocean models have

higher horizontal and vertical resolutions. Third, the system produces four times more frequent analysis; the six-hour analysis window is the same as major global atmospheric analyses. Overall, the system should represent operational atmosphere-ocean DA systems more closely, and some observed difference will be associated with these systematic differences.

5.4 Results: background error correlations of FOAM-LETKF

In this section, the background error structures of FOAM-LETKF is shown. The results of this section will serve as a basis for the correlation-cutoff experiments in the next chapter.

5.4.1 Pointwise surface error correlations (FOAM-LETKF)

We first examine the pointwise background error correlations of surface variables, which is expected to have the largest correlations. Here, an ensemble of atmospheric state vectors is bilinearly interpolated to the higher-resolution oceanic grid before background ensemble correlations are calculated.

Figure 5.5 shows the surface error correlations mainly related to the thermal coupling, namely, the exchange of sensible heat, latent heat, shortwave radiation, and longwave radiations. In addition to the exchange of thermal fluxes (Eqs. 4.3 and 4.4), the temperature tendency in each subsystem includes horizontal and vertical advectations and diffusion as well as other diabatic physical processes. Most straightforward, the surface temperature errors of the atmosphere and ocean are positively

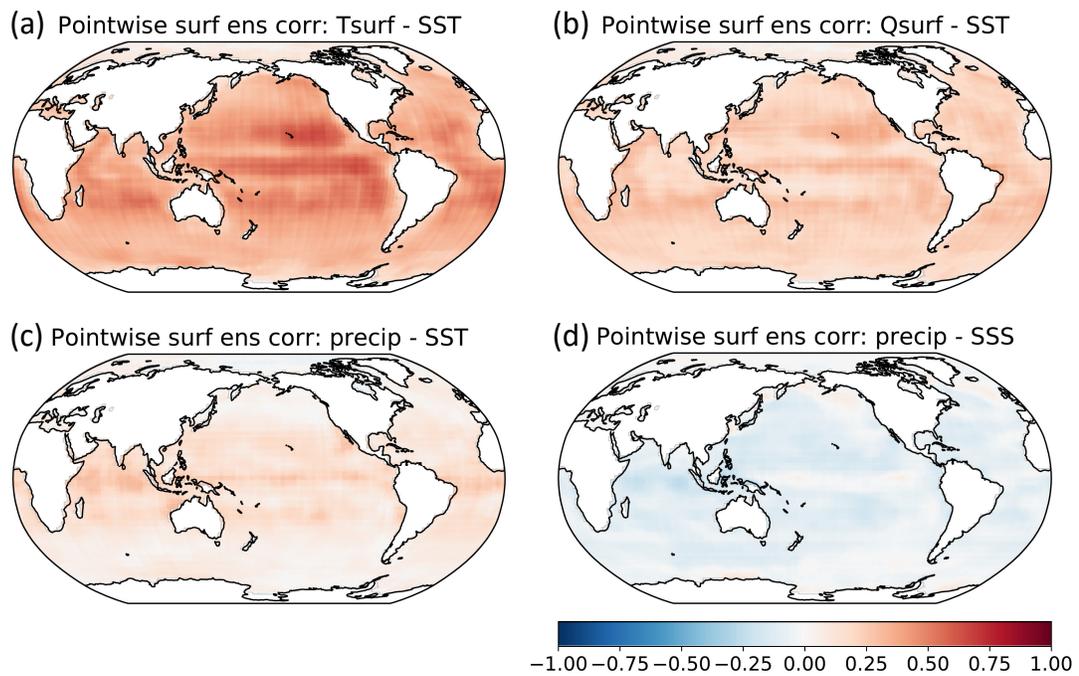


Figure 5.5: Pointwise time-mean surface background ensemble correlations. (a): between surface air temperature and sea surface temperature (SST). (b): between surface specific humidity and SST. (c): between precipitation and SST. (d): between precipitation and sea surface salinity. Background precipitation is the value accumulated over the preceding 24-hour window.

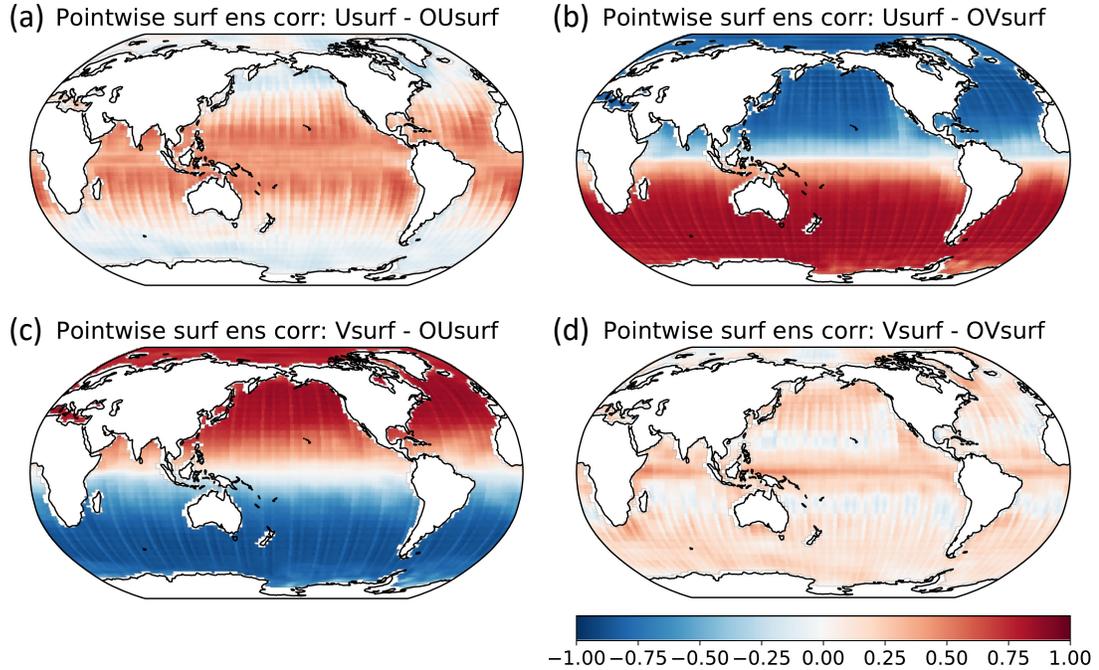


Figure 5.6: Same as Figure 5.5 but for wind and currents. Top and bottom rows show zonal and meridional winds, and left and right columns show zonal and meridional currents, respectively.

correlated almost everywhere, with slightly larger correlation over the equator and subtropics (panel a). Errors of specific humidity (Q) also is positively correlated to the SST error throughout the globe with weaker strength (panel b). Precipitation, which may be enhanced by the warmer SST, has positively correlated background error to the SST (panel c). Precipitation, in turn, is negatively driving the surface salinity error, by supplying freshwater (panel d).

Figure 5.6 shows the surface error correlations between the wind and currents. Within the FOAM-LETKF background, the strongest error correlation exists between the midlatitude wind and current, especially in the perpendicular directions (panels b and c). These relationships can be explained by the linearized Ekman

layer dynamics, where the subsurface water is transported to the right (left) of the wind direction in the northern (southern) hemisphere; at linear limit, the same relationship will be applicable to the error fields. The wind and current errors in the parallel directions (panels a and d) are correlated positively but less strongly (panels b and c). The meridional stripes present in those figures are most likely to be associated with FOAM's coupling strategy, primarily designed for simplicity and conservation of fluxes (Figure 4 of [Jacob, 1997](#)).

Errors of the other variables are correlated with weaker amplitude, consistently with the internal and cross-correlations of other variables. For example, surface air temperature and zonal subsurface current have positively correlated error in the midlatitude (Figure 5.7a). Errors of both variables are strongly correlated with the meridional wind error through thermal advection (panel b) and the Ekman transport (panel c); symbolically, the panel (a) can be understood by the product of panels (b) and (c).

5.4.2 Mean response to single observation assimilation (FOAM-LETKF)

We are also interested in how those error correlations penetrate the upper atmosphere and deeper ocean, which dictates how deep an observation can provide information by SCDA. Since background ensemble correlation serves as the proxy to the normalized increment incurred by a single observation innovation, we will see cross-sections of error correlation to hypothetical observation backgrounds.

Figure 5.8 shows examples that an ocean (atmosphere) observation can pro-

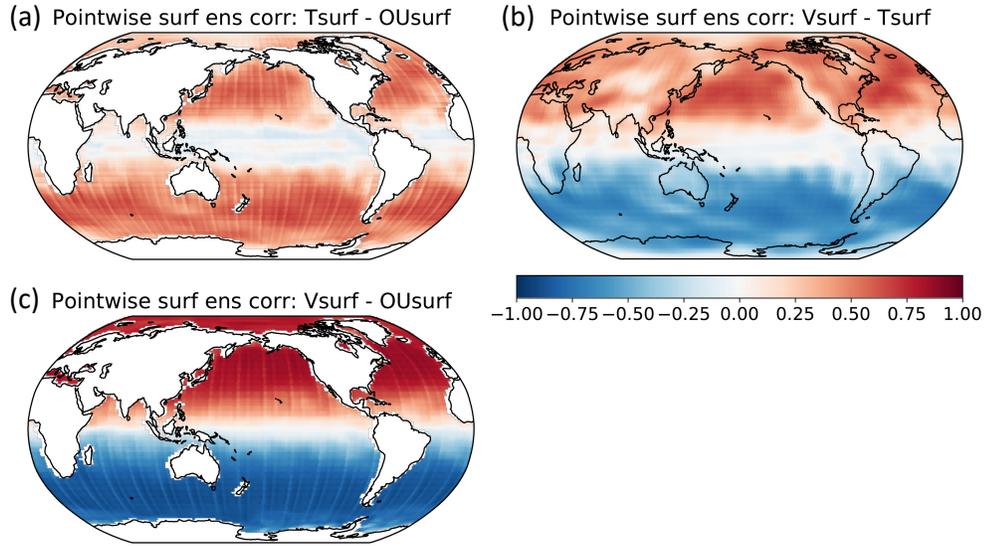


Figure 5.7: Mean pointwise background ensemble correlations of surface variables without direct interactions. (a): air temperature and zonal current. (b): meridional wind and air temperature (both variables in the atmosphere). (c): meridional wind and zonal current (same as Figure 5.6c).

vide useful information to the atmosphere (ocean) analysis. Panels (a) and (b) show the background ensemble correlation between air and sea temperatures. Both panels show that the near-surface observations can, on average, provide useful information to the analysis of the other subsystem. Although the strongest correlation exists near the surface, the impact can penetrate a few hundred meters (ocean) or a few hundreds of hectopascals (atmosphere). The correlations found in the very deep ocean seem spurious; the one-year assimilation cycle is too short for those variables to provide sufficiently independent samples. Panels (c) and (d) show the background ensemble correlations between wind and subsurface currents. Those panels show, respectively, surface current and wind observations can provide useful information into the other subsystems. In the midlatitude atmosphere, the wind error is vertically well correlated, showing quasi-barotropic nature of wind errors.

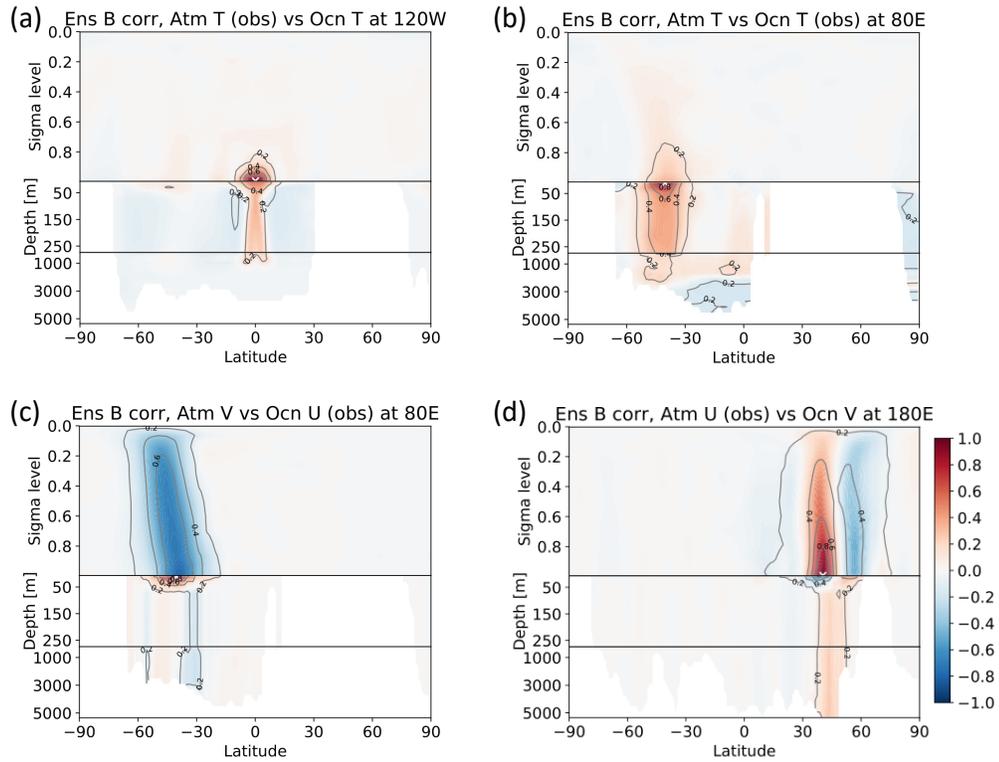


Figure 5.8: Examples that strongly coupled DA can provide useful information to the other subsystem. Each panel shows mean (RMS) background ensemble correlations in a meridional cross-section by shading (contours) to an observation background (white crosses). (a): atmospheric and oceanic temperature cross-section to surface air temperature at 0°N 120°E , (b): atmospheric and oceanic temperature cross-section to sea surface temperature at 40°S 80°E , (c): atmospheric meridional wind and oceanic zonal current cross-section to surface zonal current at 40°S 80°E , and (d): atmospheric zonal wind and oceanic meridional current cross-section to surface zonal wind at 40°N 180°E .

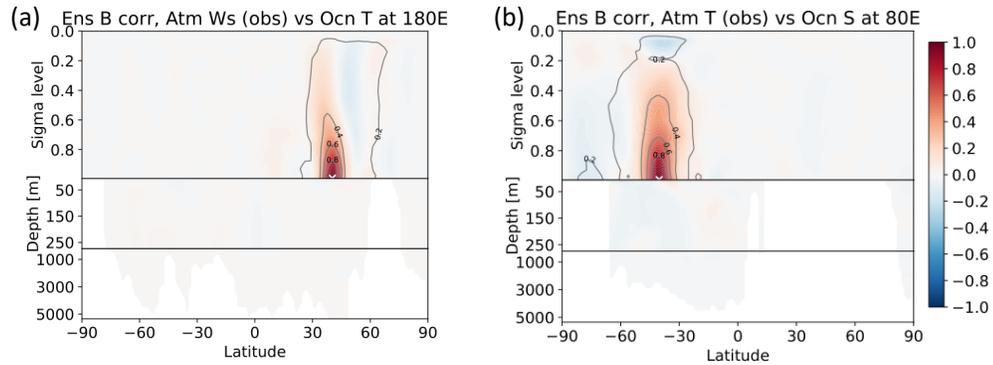


Figure 5.9: Same as Figure 5.8, but for examples that strongly coupled DA will not provide additional constraints to the other subsystem. (a): wind speed and oceanic temperature cross-section to surface wind speed at 40°N 180°E , (b): atmospheric temperature and salinity cross-section to the atmospheric temperature at 40°S 80°E .

However, there exist a larger number of examples where strongly coupled DA will hardly provide any additional benefits by assimilation of near-surface observations. A limited number of examples are shown in Figure 5.9. Panels (a) and (b) show that a wind speed observation will provide almost no constraints to ocean temperature, and an atmospheric temperature observation will provide almost no constraints to salinity. Indeed, the background error correlation between the atmosphere and ocean is fairly sparse in our global OSSE, supporting the importance of appropriate variable and spatial localizations in SCDA.

Those various strengths of background error correlation, especially among different types of variables, should be taken into account so that we obtain accurate analysis from the strongly coupled DA (Chapter 6).

5.5 Results: background error correlations of CFS-IITM

In this section, the investigations in the previous paragraph is repeated for another data set so that we can discuss the fidelity of our observations and the dependency of error structures on the coupled DA configurations.

5.5.1 Pointwise surface error correlations (CFS-IITM)

Figure 5.10 shows a global map of pointwise surface background error correlation, which is comparable to Figure 5.5 of FOAM-LETKF. Panel (a) show the error correlation between the atmospheric and oceanic surface temperatures, which is almost everywhere positive. The strongest positive correlation resides near the equatorial upwelling of the ocean and near Greenland, which is likely to be associated with the sea ice boundary. The average correlation is almost vanishing near the Maritime Continents and northern Indian ocean, where the strongest convective activity is observed. Feng et al. (2018) associated the weak ensemble correlations between temperature fields in this region to cooling by re-evaporating precipitation, which is internal variability of the atmosphere nearly independent of the SST. Those geographical features of strong and weak error correlations are not found in Figure 5.5 with FOAM-LETKF. One most noticeable difference found near Greenland can be associated with the sea ice schemes of the two models. The sea ice model of FOAM (Section 4.1) does not represent horizontal advection. On the other hand, the sea ice model of CFS accounts for the advection as well as latent heat of fusion. Therefore, the existence and concentration of sea ice can be a large source

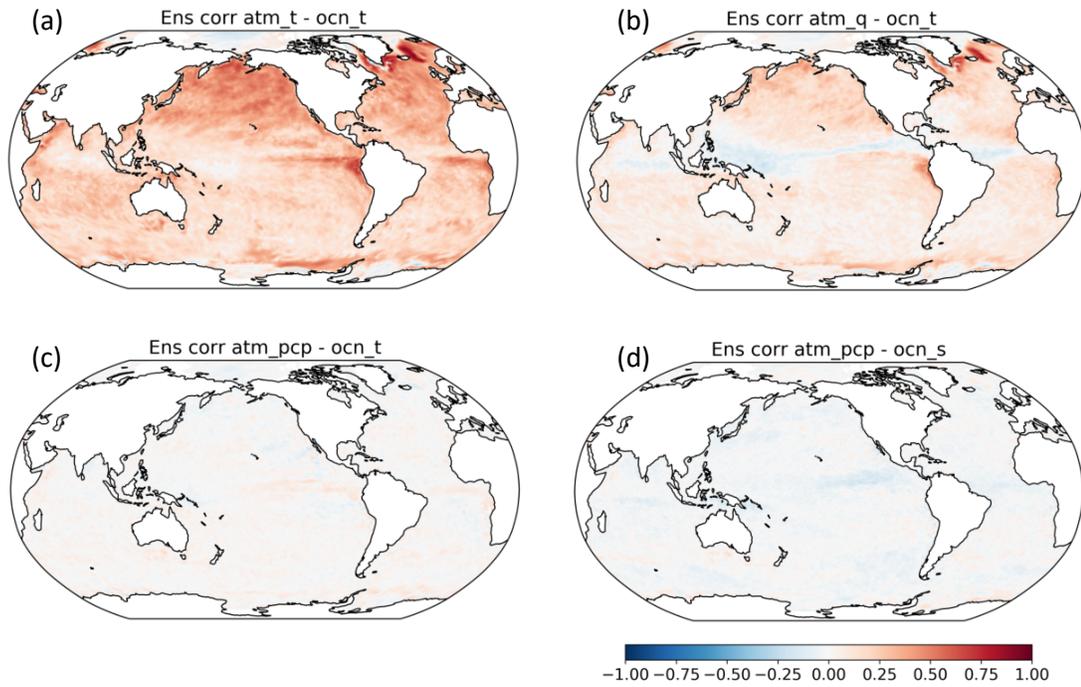


Figure 5.10: Same as Figure 5.5 but from CFS-IITM data. Background precipitation is the value accumulated over the preceding 6-hour window.

of temperature uncertainty, strong enough to dictate the background errors of atmosphere and ocean temperatures. This interpretation is supported by the salinity error, which is positively correlated to the temperature error showing that the melting sea ice provides freshwater as well as negative latent heat of fusion (not shown). The error correlation between specific humidity and SST is weakly positive over the globe but slightly negative over the intertropical convergence zone (ITCZ). Here, the dominant error-coupling mechanism should be the enhanced convection reduces the insolation (and vice versa).

Another significant difference is found in the momentum error correlations. Figure 5.11 shows the pointwise error correlations of surface wind and currents. The error correlations explained by parallel drag, between zonal wind and zonal

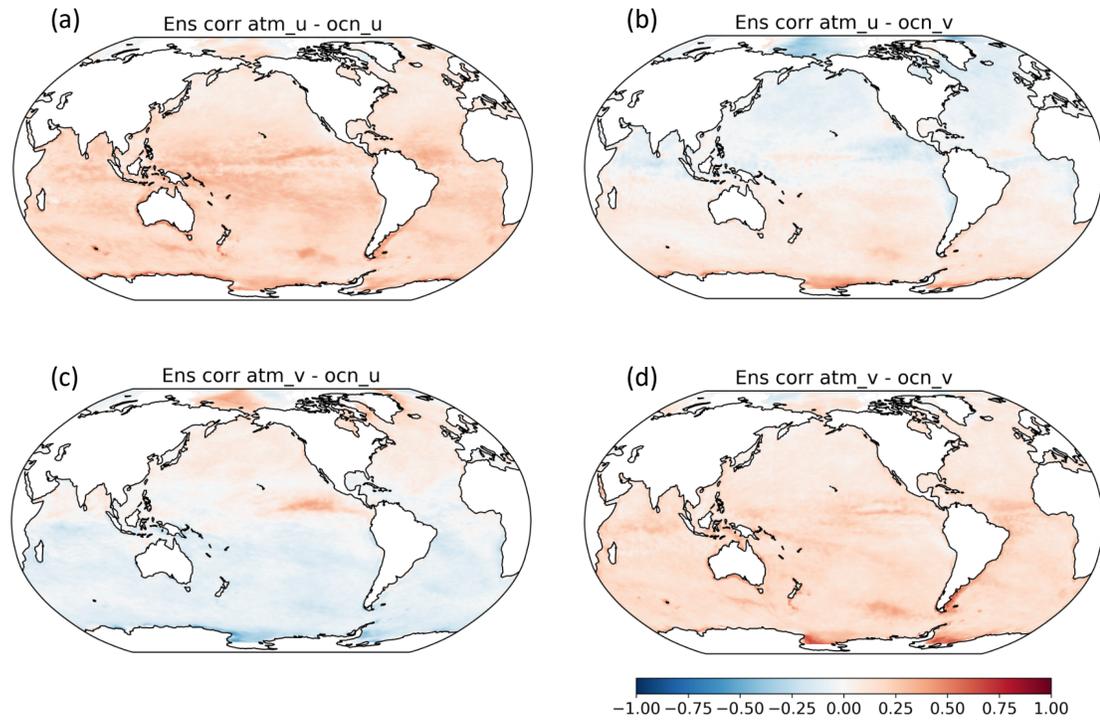


Figure 5.11: Same as Figure 5.6 but from CFS-IITM data.

current (panel a), or meridional wind and meridional current (panel d) are globally slightly weaker than those of FOAM-LETKF (Figure 5.6). The error correlations in the perpendicular directions, namely, between zonal wind and meridional current (panel b), and meridional wind and zonal current (panel c) are much weaker than that of FOAM-LETKF. The error correlations in perpendicular direction in the second and third ocean layers (which represent similar physical quantities to the FOAM's top ocean layer with 20 m thickness) is not much stronger than the surface layer (not shown). The much weaker wind-current error correlations in CFS-IITM dataset will be discussed later.

5.5.2 Mean response to single observation assimilation

Figure 5.12 shows meridional cross-sections of background error correlation, which is comparable to Figure 5.8. In general, the horizontal error correlation length is smaller both in the atmosphere and the ocean, but the difference is most significant in the ocean. This shorter correlation length will be the combined effect of denser observation network, shorter analysis interval, and the higher model resolutions. The panels (a) and (b) show that the atmospheric temperature observations can provide some useful information to the ocean temperature analysis. As can be expected from Figures 5.10 and 5.11, the smaller background error correlations between wind and current render the mutual assimilation between these variables almost irrelevant (panels c and d).

Similarly to FOAM-LETKF (Figure 5.9), the background error of CFS-IITM is almost uncorrelated between most of the variables (such as salinity and wind; not shown). Overall, the importance of variable and spatial localization of cross-update is further stressed by the CFS-IITM data.

5.6 Discussion

From the dynamical viewpoint, the higher resolutions of CFS-IITM may either strengthen or weaken the coupled error correlations. The discussion here is qualitative and will apply to both anomaly and background-error cross-correlations.

First and most straightforward, higher vertical resolution can increase the cross-correlations between the surface layers because the adjacent layers become

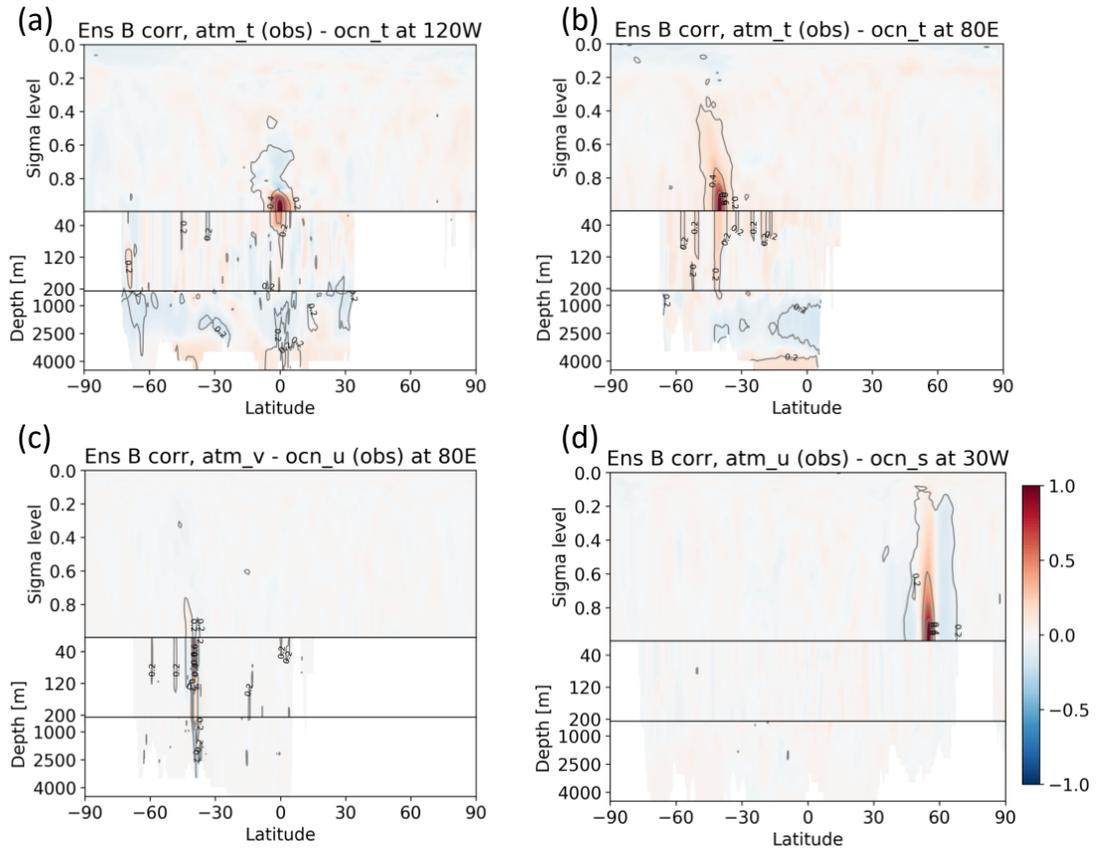


Figure 5.12: Time-mean background ensemble correlations to a point from CFS-IITM data. Each panel shows mean (RMS) background ensemble correlations in a meridional cross-section by shading (contours) to an observation background. (a and b): temperature fields to atmospheric temperature at 0°N , 120°W and 40°S , 80°E ; (c): meridional wind and zonal current fields to zonal current at 40°S , 80°E ; and (d): zonal wind and salinity fields to zonal wind at 55°S , 30°W .

physically closer. For example, the bottom atmospheric layer of FOAM (18 layers) represents an average over $0.985 \leq \sigma \leq 1$, whereas the bottom atmospheric layer of CFS (64 layers) represents an average over $0.995 \leq \sigma \leq 1$. Here, $\sigma = p/p_s$ is the vertical coordinate. Therefore, these “surface layers” represent different physical quantities, and the latter is expected to be more strongly affected by the boundary condition at $\sigma = 1$. In other words, the thinner surface layer has smaller mass and heat capacity so that its response for given flux will be larger (see Eq. 4.3, for example). For ocean models, it is known that surface processes such as diurnal cycles of SST and freshwater lenses following precipitation can only be represented with higher vertical resolutions (e.g., [Miller et al., 2017](#)).

The horizontal resolution of each model and the coupler will also affect how well the model can react to the small-scale phenomena resolved in the other subsystem. For example, [Minobe et al. \(2008\)](#) showed spatial correspondence between the Laplacian of SST, atmospheric low-level wind convergence, and precipitation fields using an atmospheric GCM (T239, ~ 50 km resolution) driven by observed SST fields. They showed that if the SST field used in the GCM is smoothed to a lower resolution (tens of degrees), the GCM no longer reproduces the strong precipitation observed over the warm flank of the Gulf stream. Therefore, if the atmospheric model has an insufficient horizontal resolution, it will not be able to respond to the SST variabilities with small horizontal scales and therefore can lower the cross-correlations.

In addition, different resolutions lead to different dynamical processes reproduced *within* each subsystem. An ocean model is commonly referred to as “eddy-

permitting” if its horizontal grid spacing is less than a half of the first baroclinic deformation radius, which in turn depends on latitude, stratification, and the ocean depth (e.g., [Hallberg, 2013](#)). The CFS ocean (0.25-0.5 degrees) has high enough horizontal resolution to permit the oceanic baroclinic instabilities in the lower latitudes while the FOAM ocean (1.4-2.8 degrees) does not. Therefore, we may assume that the CFS ocean is more internally chaotic than the FOAM ocean. This assumption is supported by apparently different horizontal dynamics of these ocean models (Figure 5.13), where we can see meandering along with the western boundary currents and Antarctic circumpolar current only in the CFS. On the other hand, we may assume that both CFS and FOAM atmospheres are resolving the chaotic quasi-geostrophic dynamics of the atmosphere, which have deformation radius of thousands of kilometers. Former experiments with an atmosphere-ocean coupled quasi-geostrophic model showed that the Lyapunov dimension (\sim the number of chaotic modes) increases with higher ocean resolution while it is insensitive to the atmosphere resolution ([De Cruz et al., 2018](#)). Upon these considerations, the change of oceanic horizontal resolution can work both ways on the cross-correlations. When we change the ocean’s resolution (internal dynamics) from low (inert) to high (highly chaotic), we will observe:

1. If no oceanic internal oscillation exists, then the ocean will behave like a “slave” of the atmosphere.
2. If the oceanic internal oscillation has moderate amplitude, the atmosphere and the ocean will follow their internal modes somewhat independently.

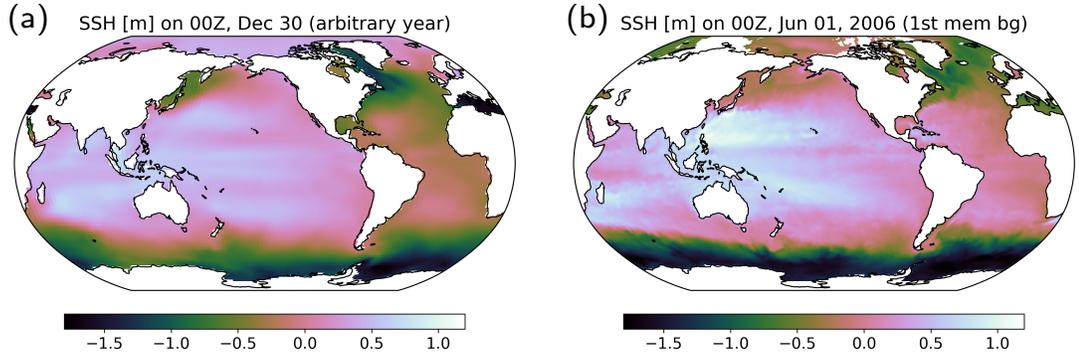


Figure 5.13: Sea surface height (SSH) anomaly in meters from (a): FOAM and (b): CFS. Note the former is from a nature run, while the latter is the 6-hour forecast (background) used in the analysis cycle. Also note these panels show different seasons.

3. If the oceanic internal oscillation is very strong, then the atmosphere will behave like the “slave” of the ocean.

Going from regime 1 to 2 will decrease the cross-correlations while going from regime 2 to 3 will increase the cross-correlations.

To support the idea, we think two highly-damped forced oscillators representing the atmosphere (A) and the ocean (O) that are mutually coupled:

$$\begin{aligned}
 k_A x_A(t) &= C_A \cos(f_A t) + x_O(t) \\
 k_O x_O(t) &= C_O \cos(f_O t) + x_A(t),
 \end{aligned}
 \tag{5.1}$$

with non-dimensional displacements $x(t)$, non-dimensional restoring coefficients k , forcing frequencies f , and non-dimensional forcing amplitudes C . The forcings are analog of chaotic oscillations inherent to each subsystem. For $f_A \neq f_O$ and $k_A, k_O > 1$, there exist analytical solutions of temporal anomaly correlation between x_A and

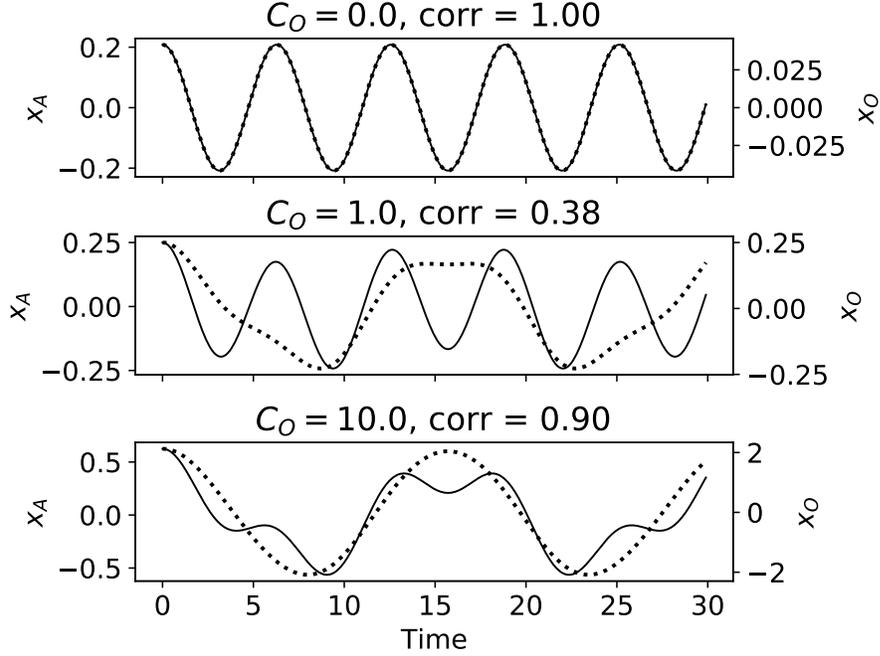


Figure 5.14: Temporal evolution and anomaly correlation of coupled forced oscillators (Eq. 5.1). Solid and dotted lines represent displacements x_A and x_O . Values of C_O are written in each panel, and other parameters $k_A = k_O = 5$, $f_A = 1$, $f_O = 0.4$, $C_A = 1$ are constant.

x_O :

$$\text{corr}(x_A, x_O) = \frac{k_O C_A^2 + k_A C_O^2}{\sqrt{k_O^2 C_A^2 + C_O^2} \sqrt{C_A^2 + k_A^2 C_O^2}}, \quad (5.2)$$

which, as a function of C_O , has a minimum in $(0, \infty)$. Figure 5.14 shows how temporal anomaly correlations of the two variables first decrease and then increase when amplitude C_O is increased, supporting the discussion in the previous paragraph. Therefore, the smaller wind-current error cross-correlations observed in the CFS-IITM may be explained by the more chaotic ocean in CFS and the transition from regime 1 to 2. On the other hand, the increased temperature error correlations in the CFS-IITM near the sea ice boundaries can be thought as a transition from regime 2 to 3, where only CFS ocean can have large error associated to the

horizontal advection of sea ice.

Although not examined here, it is also possible that we can exploit lagged (Lu et al., 2015b) or nonlocal error correlations between the atmosphere and ocean, which may be represented only in more sophisticated systems. Configuration dependency of such mutual information, and its exploitation should also be investigated in the coupled DA context.

5.7 Summary

We have examined the background error correlations of the atmosphere-ocean coupled system represented by ensemble DA systems. Our comparison of ensemble statistics of two different DA systems revealed features common and uncommon for those systems.

In both data sets, temperatures of the atmosphere and ocean are found to be one of the strongest background error correlations, as expected. This temperature error correlation is strongest at the equatorial upwelling and sea ice boundaries in the CFS-IITM data set. In addition, strong wind-current relationship, especially in the higher latitudes, is found only from the low-resolution FOAM-LETKF. Those background error correlations, if accurately accounted for in strongly coupled data assimilation between the atmosphere and ocean, will provide informative constraints. At the same time, background error correlation between the atmosphere and ocean, seems sparse between most other variables, supporting the use of appropriate variable and spatial localization if only a small ensemble can be used for the background

error covariance estimate.

We have also discussed how the horizontal resolution and internal error growth in the ocean can affect the difference of momentum error correlations between the two experiments. If a high-resolution ocean model is highly chaotic by itself, then the background errors of atmospheric wind and ocean current are close to independent. When this is the case, an atmospheric observation is not a good representation of ocean state and vice versa. Under such condition, the approach by ECCO ocean reanalysis (Stammer et al., 2004) and Sugiura et al. (2008) — to explain ocean state evolution in a year-long window just by the oceanic initial condition and atmospheric forcings — might fail. The need for high ocean resolution especially in the context of ensemble DA, where the error growth represented by the model is responsible for estimating error statistics, should be further explored.

The background error due to model and parameter uncertainties are not accounted for in both experiments. Estimate for those error origins should be conducted via the perturbed-parameter experiments or observation-based statistics.

Further implementation of WCDA systems, not only between the atmosphere and ocean, but also land, sea/land ice, and wave models is also encouraged; with either the NMC method or ensemble method, we will be able to obtain background error statistics. Such studies (e.g., Lin and Pu, 2018) will provide useful guidance for future planning and implementation of SCDA systems.

For variational data assimilation systems that use the static background error covariance, we also need to develop a method to incorporate those sampled background error statistics into a covariance model. Considering the diffusive nature

of thermal coupling (Eqs. 4.3 and 4.4), the diffusion operator (Derber and Rosati, 1989) can be a handy option to model the thermal coupling of background errors in the grid space. Another option is to account for the coupled processes as a form of balanced analysis variables as done by Storto et al. (2018).

Finally, the apparent complexity of coupled background error structures renders knowledge-based optimization of variable and spatial localization difficult. In the next chapter, we explore how we can overcome this apparent complexity of coupled background error correlations by employing data-driven, machine-learning approaches.

Chapter 6: Localization modeling with neural networks and assimilation experiments with FOAM

6.1 Introduction

Previous studies have shown that an appropriate covariance localization is vital for successful strongly coupled data assimilation with ensemble methods, where the background error covariance estimated by the ensemble is not always reliable. However, it is not straightforward to define an appropriate “distance” between variables in the subsystems of the earth which are coupled with various strengths. For example, salinity analysis is found to have little to do with wind observations in a time scale of assimilation windows, even if they are physically collocated. In Chapter 2, we have introduced the correlation-cutoff method, where an observation is assimilated into a part of analysis variables that are expected to have large background error correlations to the observable. The remaining problem is how to encode the average strength of background error correlations into a function, which is fast to evaluate and small on the memory.

For that purpose, we employ neural networks to summarize the error statistics. As we will see, this is a natural extension of the correlation-cutoff method

proposed in Chapter 2. In this way, we can consider analysis/observation variable types and geographical location, for example, as well as physical distance as explanatory variables of the localization weight. We will show an application of strongly coupled assimilation experiments with an atmosphere-ocean coupled general circulation model, in addition to precursor experiments with toy correlation models and theoretical validity as a localization function.

6.2 Methodological concept: correlation-cutoff with neural networks

6.2.1 Overview of the correlation-cutoff method application

In Chapter 2, we have shown that the mean squared background ensemble correlation from an ensemble DA cycle can be used as a “distance” for localization between an observation and an analysis variable.

However, unlike experiments in Chapter 2, it is unrealistic to store the background ensemble correlation for all the pair of observable and analysis variable in real applications. Also, storing the correlation matrix cannot adapt to temporary varying observation networks. Therefore, it is necessary to summarize the climatological strength of background error correlation in a function so that the information can be used for localization. This methodology is illustrated in Figure 6.1, where the localization function is expressed as a composite function of a nonlinear, multivariate regression function $f : (\text{attribute space}) \rightarrow \mathbb{R}$ and an increasing function $g : \mathbb{R} \rightarrow \mathbb{R}$ to be determined. Hereafter we call g as a *cutoff function*. The experiments with the nine-variable coupled model in Chapter 2 can be thought as a

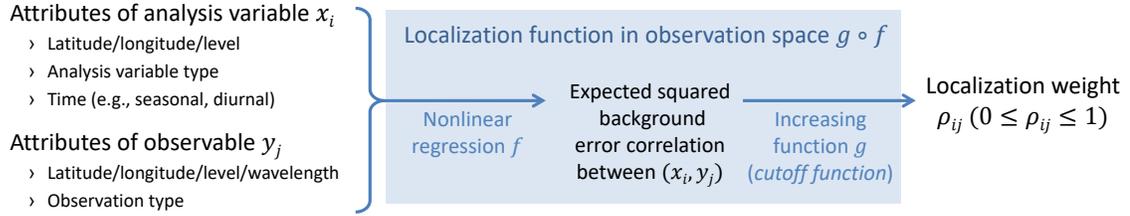


Figure 6.1: Schematic of the correlation-cutoff localization method applied to analysis of coupled geophysical systems, where physical distance alone cannot explain the relevance between an analysis variable and an observable.

special case of this methodology, where f was a function of variable indices, and the cutoff function g was a step function.

With little prior knowledge of the shape of the function f , we choose neural networks as a generic tool for multivariate nonlinear regression. Other possible options for nonlinear regression are compared in Table 6.1. Although a neural network requires relatively expensive training with iterative minimization of the loss function, it is fast to be evaluated once trained and depends on fewer assumptions than other nonlinear regression methods. The actual training and evaluation cost will be examined later. To be more precise, we use one neural network for each pair of observation and analysis variable types. In other words, the regression f is a lookup table of neural networks. In this way, each neural network only uses numerical attributes such as latitude and is trained independently from other neural networks. We choose this approach because it is inefficient, if not impossible, to include those categorical attributes as explanatory variables of a neural network.

Table 6.1: Subjective comparison of regression methodologies. Here, *curse of dimensionality* refers to the exponential growth of nonlinear combinations of input variables, which will affect necessary memory space, training samples needed, and evaluation cost. An *analytical* training means that the regression parameters are obtained by a weighted average of training data without iterative minimization.

Method	Advantages	Disadvantages
Linear regression	Simple to implement Training is analytical	Linear
Lookup table	Nonlinear Simple to implement Training is analytical Fast to evaluate	Discontinuous Assumptions for boundaries Curse of dimensionality
Linear combination of nonlinear basis functions (e.g., polynomial fit)	Nonlinear Training is analytical	Assumptions for basis functions Curse of dimensionality
Neural network	Nonlinear Fewer assumptions Relatively tolerant of input dimensionality	Training requires iteration
Gaussian processes regression	Access to uncertainty Nonlinear Fewer assumptions	More expensive training

6.2.2 Mathematical validity as a localization function

For a function to be a mathematically valid localization function, the localized analysis must exist under general conditions.

Given a positive definite observation error covariance matrix \mathbf{R} and a positive semidefinite (ensemble-based) background error covariance matrix \mathbf{P}^b , the non-localized gain matrix (Eq. 1.1) is well defined because $\mathbf{H}\mathbf{P}^b\mathbf{H}^T + \mathbf{R}$ is positive definite and therefore invertible, where \mathbf{H} is the linearized observation operator.

For model space localization (Eq. 1.2), a well-known sufficient condition for a well-defined gain matrix is that the localization matrix $\rho_m \in \mathbb{R}^{n \times n}$ is positive semidefinite. Then, the localized background error covariance matrix $\rho_m \circ \mathbf{P}^b$ remains positive semidefinite, and the gain matrix always exists. However, this is a restrictive condition when we try to specify a flexible localization function because a localization weight between two model variables cannot be specified independently from the other localization weights. Therefore, the piecewise polynomial function of [Gaspari and Cohn \(1999\)](#) (Eq. 1.5), which produces positive definite correlation matrix in any discretization of \mathbb{R}^3 , has been a popular localization function.

We avoid this restriction by limiting ourselves to two common forms of observation-space localization, namely, the serial ensemble square root filter (EnSRF) and the local ensemble transform Kalman filter (LETKF). For these forms of observation-space localization, a less restrictive sufficient condition for mathematically valid localization functions exists.

For the serial EnSRF, we assimilate each observation sequentially assuming

uncorrelated observation errors. Consider we are assimilating the j th observation ($j = 1, \dots, p$). The update of the i th analysis variable ($i = 1, \dots, n$) reduces to a scalar regression

$$\mathbf{K}_{ij} = \rho_{ij}(K - 1)^{-1} \mathbf{X}_i^b \mathbf{Y}_j^{bT} \left[(K - 1)^{-1} \mathbf{Y}_j^b \mathbf{Y}_j^{bT} + \mathbf{R}_{jj} \right]^{-1}, \quad (6.1)$$

where $\mathbf{K}_{ij} \in \mathbb{R}$ is the localized regression coefficient from the innovation of the j th observation to the increment of the i th analysis variable, ρ_{ij} is the localization weight between the i th analysis variable and the j th observation, $\mathbf{X}_i^b \in \mathbb{R}^{1 \times K}$ is the background ensemble perturbations¹ of the i th variable, $\mathbf{Y}_j^b \in \mathbb{R}^{1 \times K}$ is the background ensemble perturbations of the j th observable, and $\mathbf{R}_{jj} \in \mathbb{R}$ is the (positive) observation error variance. The gain matrix for updating perturbations is the same except that it is multiplied by a scalar function of background and observation error variances of the observable, which is independent of the localization. Now the quantity to be inverted $(K - 1)^{-1} \mathbf{Y}_j^b \mathbf{Y}_j^{bT} + \mathbf{R}_{jj}$ is independent of the localization weight, and the analysis ensemble of the i th variable always exists for any finite ρ_{ij} . Inductively, the global update by a set of observations always exists.

For the LETKF (Algorithm 2), an update of the i th variable requires an inversion of

$$(K - 1) \mathbf{I}_K + \mathbf{Y}^{bT} \mathbf{R}_i^{-1} \mathbf{Y}^b \in \mathbb{R}^{K \times K}, \quad (6.2)$$

where $\mathbf{Y}^b \in \mathbb{R}^{p \times K}$ is background ensemble perturbations of observables and $\mathbf{R}_i \in \mathbb{R}^{p \times p}$ is a localized observation error covariance matrix used for the analysis of the

¹To be precise, for the serial assimilation, the background perturbations for the assimilation of an observation is the analysis ensemble that has assimilated all the previous observations.

i th variable (to be defined). The R-localization of LETKF is done as follows. For an analysis of the i th variable, we discount the inverse observation error covariance as

$$[\mathbf{R}_i^{-1}]_{jk} = \sqrt{\rho_{ij}\rho_{ik}}[\mathbf{R}^{-1}]_{jk}, \quad (6.3)$$

where j and k are observation indices ($j, k = 1, \dots, p$), and \mathbf{R} is the observation error covariance matrix before localization. In other words,

$$\mathbf{R}_i^{-1} = \mathbf{D}_i \mathbf{R}^{-1} \mathbf{D}_i \quad (6.4)$$

for $\mathbf{D}_i = \text{diag}(\sqrt{\rho_{i1}}, \dots, \sqrt{\rho_{ip}})$. We assume that ρ_{ij} is nonnegative for all j . Then because both \mathbf{D}_i and \mathbf{R}^{-1} are positive semidefinite symmetric, \mathbf{R}_i^{-1} will also be positive semidefinite symmetric. Therefore, if only ρ_{ij} is nonnegative for all i and j , then the LETKF analysis exists. In practice, columns and rows of \mathbf{Y}^b and \mathbf{R}_i^{-1} corresponding to observations with zero weights are not evaluated without changing the resulting $\mathbf{Y}^{b\top} \mathbf{R}_i^{-1} \mathbf{Y}^b \in \mathbb{R}^{K \times K}$.

We have shown that for the serial EnSRF and the LETKF, the localized analysis always exists if only each localization weight ρ_{ij} is nonnegative. This result helps us to specify flexible localization functions. We further limit ourselves to $0 \leq \rho_{ij} \leq 1$ considering the nature of localization that reduces the effect of observations.

6.3 Preliminary neural network experiments with toy data

In this section, we will show that a minimal neural network can be used to model a wide range of multidimensional functions.

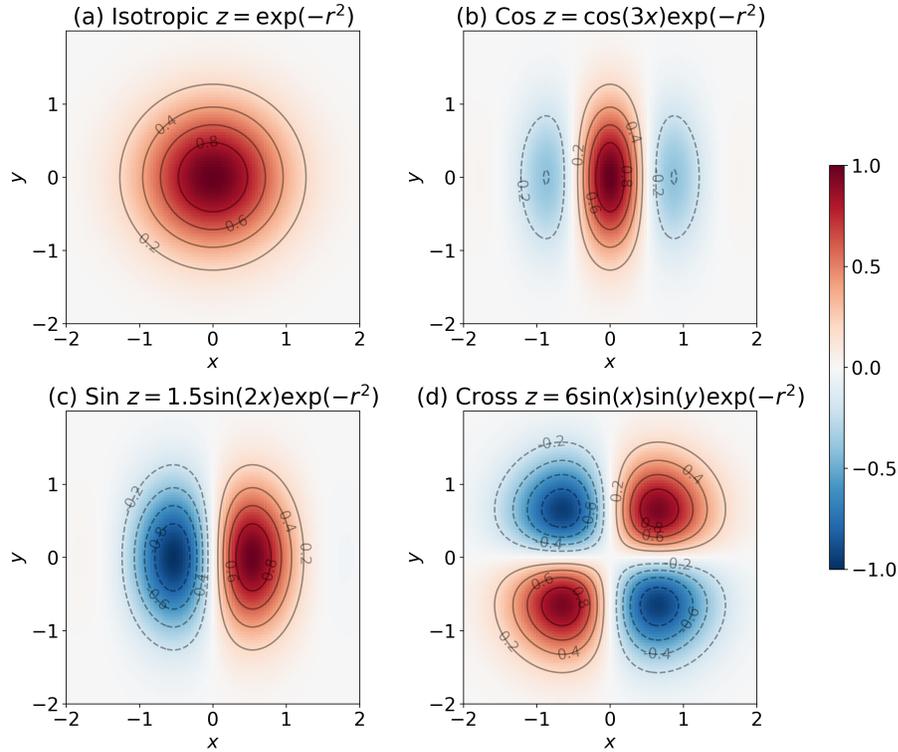


Figure 6.2: Four toy correlation models and their generating functions ($r = \sqrt{x^2 + y^2}$). Shading shows the value z at each horizontal location (x, y) . Contours showing the same quantity are added for visual clarity.

6.3.1 Toy correlation models from geostrophic theory

Figure 6.2 shows the four toy correlation models used in this section. These “correlation functions” mimic the typical multivariate correlation functions found in the mid-latitude atmosphere (Figure 6.3). The amplitude of *Sin* and *Cross* models has been adjusted to have a maximum amplitude close to unity. Similar background error correlations are also observed with FOAM-LETKF and CFS-IITM in the mid-latitude atmosphere when many instances are averaged (not shown). We expect neural networks to reproduce these nonlinear functions and their squared values from erroneous samples.

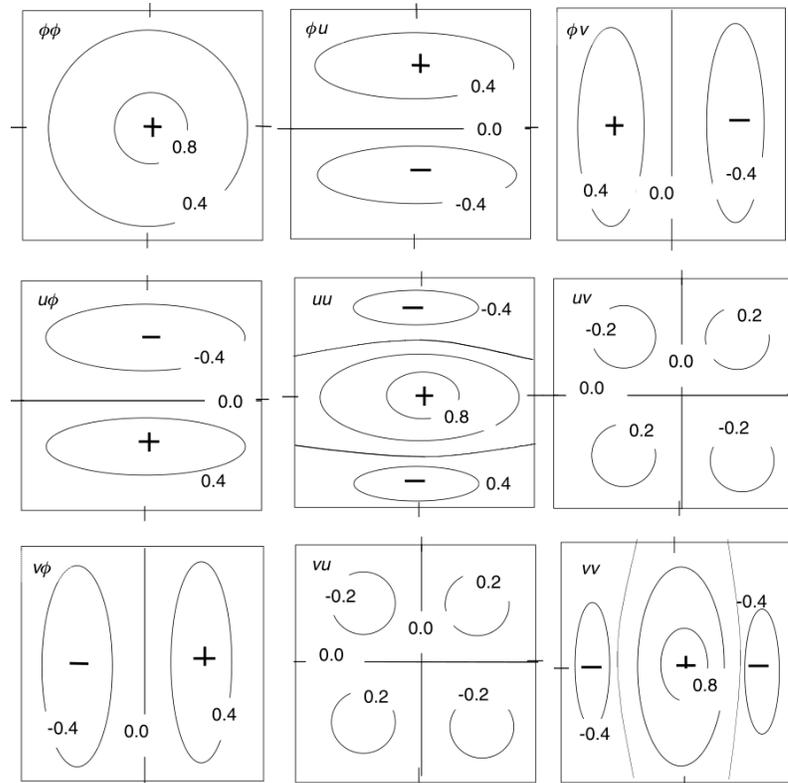


Figure 6.3: Schematic illustration of the multivariate error correlation functions in northern hemisphere mid-latitude atmosphere. These are generated assuming the perfect geostrophic relation between wind (U, V) and geopotential (Φ). Another assumption is the isotropic and Gaussian-shaped spatial correlation of geopotential errors shown in the upper left panel. Figure courtesy [Kalnay \(2003\)](#).

For each toy correlation model, 1000 erroneous training data points are generated as follows. Each of 1000 training data points is generated by first sampling (x, y) from uniformly random distribution $-2 \leq x, y \leq 2$, and then error-free z value is computed from the equation overlaid in Figure 6.2. The z value is squared for some experiments, and finally, Gaussian error with standard deviation 0.2 is added to the (squared) z value. The validation dataset is similarly generated, except no error is added.

6.3.2 Specification of neural networks and training algorithm

Figure 6.4 schematically shows the topology of a neural network, where we try to explain an L -dimensional vector of dependent variables \mathbf{y} with a D -dimensional vector of explanatory variables \mathbf{x} . Formally, the neural network can be written as $\mathbf{y} = \mathbf{W}_2[h(\mathbf{W}_1\mathbf{x} + \mathbf{b}_1)] + \mathbf{b}_2$, where h is an elementwise nonlinear function called an activation function (to be specified), $\mathbf{w} \equiv (\mathbf{W}_1, \mathbf{W}_2, \mathbf{b}_1^T, \mathbf{b}_2^T)^T$ (with matrices flattened) are parameters to be trained². In our example, the output dimension is always $L = 1$, the input dimension is either $D \in \{1, 2, 3\}$ depending on the choice of explanatory variables (to be specified), and the number of hidden units is fixed at $M = 10$. A forward evaluation of the network costs $\mathcal{O}(DM + ML)$ floating point computations.

Two common choices of the activation function h are hyperbolic tangent and a piecewise linear function, $\max(0, \cdot)$, called the rectifier function. We choose hy-

² \mathbf{w} is a vector of $DM + ML + M + L$ real-number parameters.

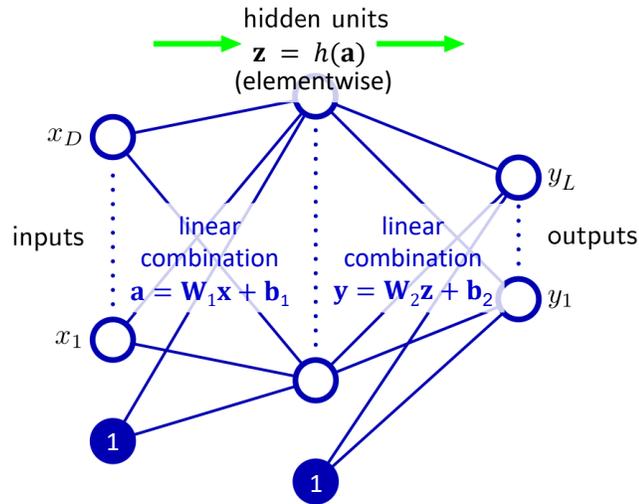


Figure 6.4: Schematic of a two-layer feedforward neural network. The green arrows show how input signals propagate to the output. Modified [Bishop \(2006\)](#).

perbolic tangent as our activation function because it gives smoother fitting³ and generally smaller regression error for our problems than the rectifier function in preceding experiments.

A neural network tries to minimize a loss function, the squared regression error to the training data, which is equivalent to estimating the mean of dependent variable conditioned by the explanatory variables (Chapter 1 of [Bishop, 2006](#)). No regularization term (i.e., weight decay) is added to the loss function because it generally degrades the regression accuracy to the independent validation data set in our problems. In general, regularization is unnecessary if the number of tunable parameters is much smaller than the training sample size, and over fitting to the

³This is expected because affine transformation is C^∞ , and the entire network will have the same smoothness to the activation function. Hyperbolic tangent is C^∞ , and the rectifier function is C^0 .

training data is unlikely.

The training data is normalized so that each independent and dependent variable has zero mean and unit variance; the same linear transformation is applied when we use the trained network for regression (recommendation by [Glorot and Bengio, 2010](#)). The parameters \mathbf{w} of neural networks are uniformly randomly initialized with the range recommended by [Glorot and Bengio \(2010\)](#) and trained with the Adam optimizer ([Kingma and Ba, 2017](#)). Adam is a stochastic gradient⁴-based optimization algorithm with automatic adjustment of the learning rate, which otherwise needs to be tuned manually. Adam also exploits the idea of *momentum*⁵ so that the training does not stop at local minima or saddle points of loss functions. Adam works well for our problems with the recommended hyperparameters; the training converges with reasonable computation time, and the final function is almost insensitive to the random initializations of \mathbf{w} . In contrast, the classic stochastic gradient descent does not usually converge to a single solution when we repeat training from different random initializations of \mathbf{w} . Therefore, we note that these recent advancements of neural network methodology are essential for our applications.

⁴Stochastic gradient descent method evaluates the gradient of the loss function in the parameter (\mathbf{w}) space using one training sample at a time in a randomized order. It is more tolerant of local minima than (non-stochastic) gradient descent, where the gradient is evaluated with all the training samples.

⁵In gradient descent method without momentum, the change of parameters \mathbf{w} in a step is proportional to the gradient of the loss function. With momentum, on the other hand, the change of velocity (i.e., acceleration) of parameter evolution is proportional to the gradient of the loss function as if a ball descending a hill.

The training period is 1000 epochs; that is, each training data point is used 1000 times. Each epoch digests all the training samples sequentially in a randomized order. The 1000 epochs of training (10^6 total training samples digested) is decided subjectively to be long enough for the training to converge.

6.3.3 Explanatory variables for fitting toy data

One purpose of toy correlation model experiments is to enlighten the explanatory variables used to explain the horizontal correlation structure. The four sets of explanatory variables below are tested:

- (x, y) — by the definition of the correlation models (Figure 6.2), these are the most straightforward explanatory variables used in the neural network.
- (r) — an incomplete set of explanatory variables. The azimuthal dependency will be ignored.
- (r, θ) — where $\theta = \arctan(y/x)$ ($-\pi < \theta \leq \pi$). Since these variables have complete information of horizontal coordinate, they may be used as explanatory variables. However, regression with θ is expected to have a discontinuity at $\theta = \pm\pi$.
- $(r, \cos \theta, \sin \theta)$ — it is customary to separate a cyclic variable into its sine and cosine to avoid discontinuities. The same technique can be used for expressing seasonal or diurnal dependencies in regression.

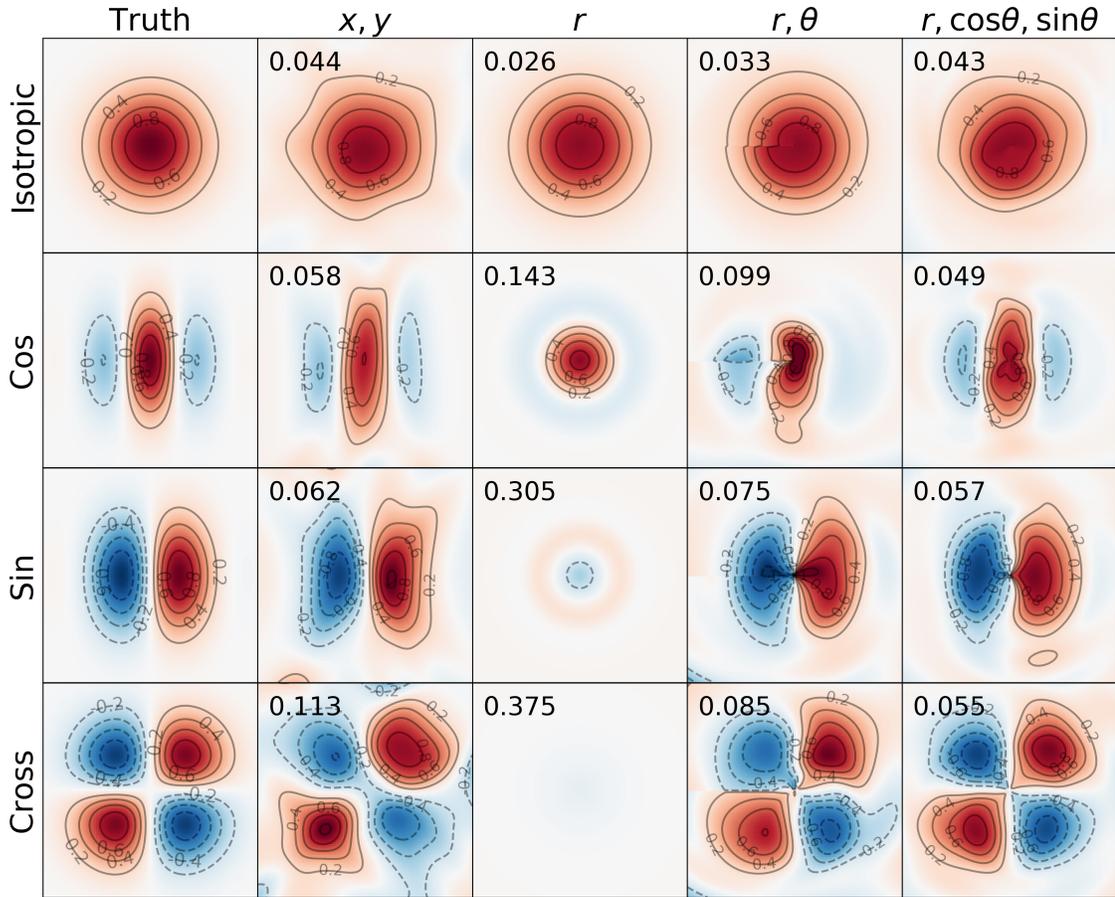


Figure 6.5: Regression by neural networks to toy correlation models. The leftmost column shows the error-free generating function, and the other columns show the regression field with a different set of explanatory variables. Superimposed to each panel are root-mean-squared (RMS) regression errors to the independent validation dataset.

6.3.4 Results and discussion — fitting toy correlation models

Figure 6.5 shows the regression results to the toy correlation models. First, obviously the explanatory variable (r) cannot explain the asymmetric relationship. This incompleteness causes severe misfit for the *Sin* and *Cross* models.

The other three sets of explanatory variables generally produce good regression accuracy to independent validation dataset although (r, θ) has a discontinuity at

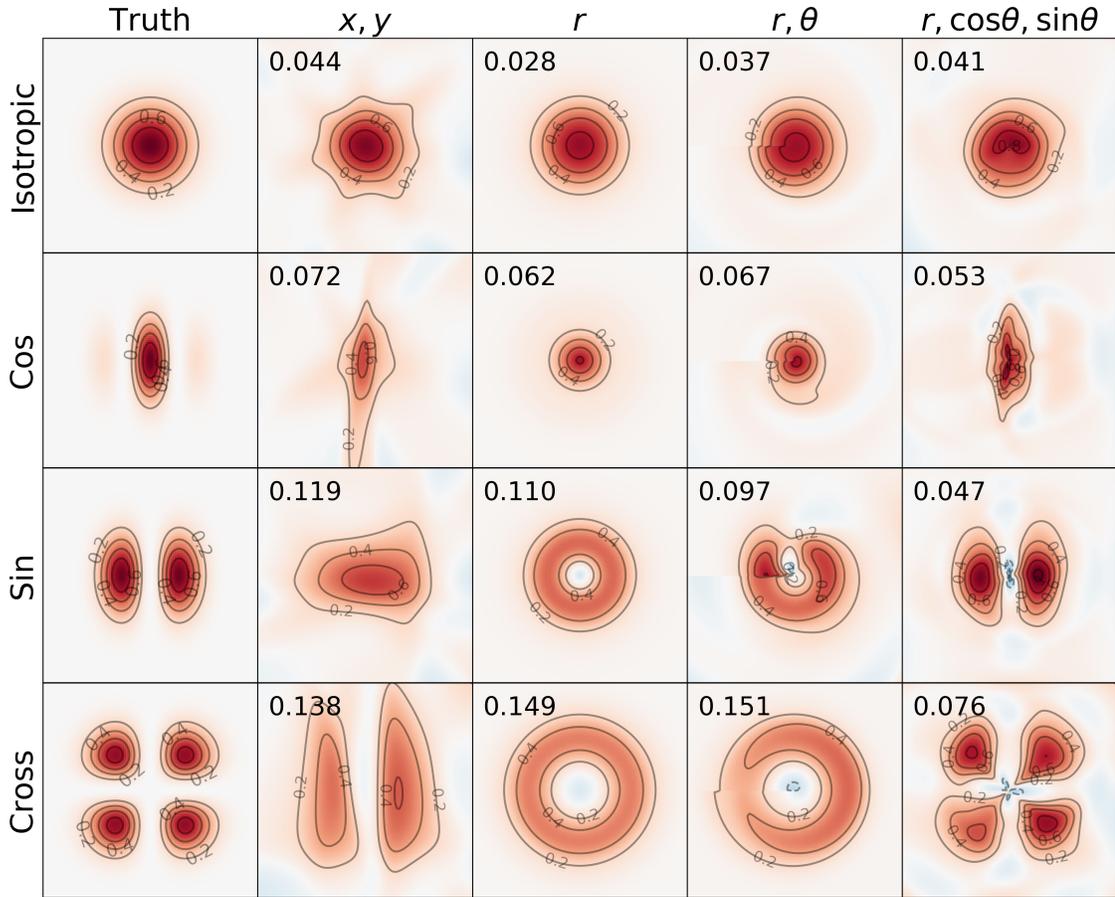


Figure 6.6: Same to Figure 6.5 but for squared generating functions.

$\theta = \pm\pi$. When we compare (x, y) and $(r, \cos\theta, \sin\theta)$, the former tend to have slightly insufficient peak strengths. Considering each training data point has error standard deviation 0.2, the regression is successful.

Figure 6.6 shows regression for the squared correlation models. The regression accuracy to the independent validation dataset is now comparable for (x, y) , (r) , and (r, θ) . Here we find (x, y) and (r, θ) suffer to represent the small-scale features such as *Sin* and *Cross*. Again, (r, θ) has discontinuity at $\theta = \pm\pi$. The regression by $(r, \cos\theta, \sin\theta)$ is generally most successful.

Overall, the neural networks have shown good abilities to reproduce spatially

extended “correlation functions” when an appropriate set of predictors is chosen.

Finally, we discuss an important characteristic of the fitting to squared values by comparing regression to the *Cross* function with (r) in Figures 6.5 and 6.6. In the regression to the raw value (Figure 6.5 bottom center), positive and negative values offset each other, and the resulting function is almost null. However, in the regression to the squared value (Figure 6.6 bottom center), the azimuthal dependency is smoothed out, but the resulting regression function is still nonzero. The regression to the squared value still captures an important geostrophic feature that the “correlation” reaches its maximum at a certain distance. This implies that we do not need to include all the explanatory variables in the regression to get a decent localization function.

6.4 Fitting to FOAM ensemble correlations

In this section we estimate and evaluate the regression function f that will be used for localization of strongly coupled atmosphere-ocean DA.

6.4.1 Generation of training data

Training and validation data sets are generated by sampling the weakly coupled LETKF cycle of FOAM used in the previous chapter. The 64-member LETKF system analyzes the coupled atmosphere-ocean states every 24 hours for a model year (Table 5.1).

In our experiments, there exist 10 types of analysis variables: (T, Q, U, V, P_s)

from the atmosphere and $(T, S, U, V, P_{\text{top}})$ from the ocean. Since only horizontal interpolation operators are considered for observation operators, the set of observation types is the same as the types of analysis variables. Therefore, there are 100 combinations of observation and analysis variable types.

A training data sample for given observation and analysis variable types is generated as follows:

1. Sample time t uniformly randomly from the available period.
2. Sample analysis grid location $(\lambda_{\text{anl}}, \phi_{\text{anl}}, z_{\text{anl}})$ uniformly randomly on the sphere, where λ , ϕ , and z represent longitude, latitude, and vertical coordinate, respectively. Repeat sampling if the analysis grid is topographically masked. For vertical coordinate, the integer model levels in Table 4.3 are used.
3. Sample uniformly random $(r, \theta, z_{\text{obs}})$ ($r \in [0, r_{\text{max}}), \theta \in [0, 2\pi)$), where r is the distance between the analysis grid and the observation, $r_{\text{max}} = 3000$ km, θ is the azimuth angle from the analysis location to the observation location, and z_{obs} is the observation level. Uniformly random sampling in (r, θ) space is designed to more intensively sample observation locations near the analysis grid compared to uniform sampling on the sphere.
4. Get the observation location $(\lambda_{\text{obs}}, \phi_{\text{obs}})$ using $(\lambda_{\text{anl}}, \phi_{\text{anl}}, r, \theta)$. Repeat sampling (step 3-4) if the observable is topographically masked at $(\lambda_{\text{obs}}, \phi_{\text{obs}}, z_{\text{obs}})$.
5. Get background ensemble of the analysis variable and the observable by applying an observation operator to the background ensemble of state vectors.

Here, observation operators are horizontal interpolation.

6. Calculate the background ensemble correlation between the analysis variable and the observable. Save it with the attributes

$$(t, \lambda_{\text{anl}}, \phi_{\text{anl}}, z_{\text{anl}}, \lambda_{\text{obs}}, \phi_{\text{obs}}, z_{\text{obs}}, r, \theta).$$

In practice, we first complete steps 1-4 for all necessary samples to generate a list of attributes $(t, \lambda_{\text{anl}}, \phi_{\text{anl}}, z_{\text{anl}}, \lambda_{\text{obs}}, \phi_{\text{obs}}, z_{\text{obs}}, r, \theta)$ and then steps 5-6 in the t -order (simultaneously for all the variable types) to enhance computational efficiency.

For each pair of variable types, we prepare 8×10^6 training samples and 1×10^6 validation samples. Since there are 10 analysis variable types and 10 observation types in our experiments, these add up to $10 \times 10 \times (8 \times 10^6 + 1 \times 10^6) = 9 \times 10^8$ total samples. Sampling 9×10^8 samples takes several hours with a single processor. The number of necessary training samples depends on the number of explanatory variables and the smoothness of the underlying function, which will not be very sensitive to the model resolution as long as the process-of-interest (e.g., Rossby waves) remains resolved (Ying et al., 2018). The input/output cost is expected to increase for higher-resolution models, but the sampling (steps 5-6) can be done parallelly for each t . Also, the sampling is done only once, and this sampling cost seems acceptable.

6.4.2 Neural network and training configurations

In our experiments, a neural network has $M = 30$ hidden units with hyperbolic tangent nonlinearity. We have increased the number of hidden units from the

toy data experiments ($M = 10$, Section 6.3) because the networks have to consider vertical and latitudinal effects as will be discussed. The training is done with the Adam optimizer with the recommended hyperparameters for 3 epochs. Each network, therefore, digests total of $3 \times 8 \times 10^6$ samples. The other training settings, including loss function, regularization, normalization of training samples, and initialization of the parameters \mathbf{w} , are the same to the experiments with the toy correlation data set.

Training of each neural network takes tens of minutes with a single processor and can be done in parallel for different pair of variable types. Again, this training is done only once before starting the (production) assimilation cycle.

6.4.3 Explanatory variables for regression of FOAM ensemble correlation

We test three sets of explanatory variables for regression of the (squared) background ensemble correlations:

- $(r, z_{\text{anl}}, z_{\text{obs}})$ — this is analogous to (r) in the toy correlation model experiments with additional vertical dependency.
- $(r, z_{\text{anl}}, z_{\text{obs}}, \phi_{\text{anl}})$ — latitudinal dependency is added, considering different atmosphere-ocean dynamics in the lower and higher latitudes.
- $(r, z_{\text{anl}}, z_{\text{obs}}, \phi_{\text{anl}}, \cos \theta, \sin \theta)$ — this is analogous to $(r, \cos \theta, \sin \theta)$ in the toy correlation model experiments. Azimuthal dependency can be explained.

Recall that the analysis variable and observation types are also considered in our regression f because different neural networks are trained and used for different variable types.

Fitting is independently repeated for raw and squared ensemble correlations. The fitting to the squared ensemble correlation serves directly for the correlation-cutoff localization (Figure 6.1). The fitting to raw ensemble correlation mainly serves for visualizing its ability to reproduce known background error statistics; the term “*mean ensemble correlation*” refers to this regression.

6.4.4 Regression results and discussion

Figure 6.7 shows the RMS regression errors of squared background ensemble correlations to the validation data set. First, in the left column, we see that the diagonal blocks tend to have larger regression errors. This is expected because larger background ensemble correlations tend to exist within each subsystem; the regression is trivial if ensemble correlations between two variables are always close to zero. The relative regression errors for different sets of explanatory variables are more apparent when we plot the ratio of regression errors (right column). The negative values in the panel (c) shows that the regression with additional latitudinal dependency ϕ_{anl} achieved smaller regression error. The inclusion of latitudinal dependency is beneficial for explaining atmospheric internal, oceanic internal, and cross-correlations of background errors. On the other hand, the regression error is mostly insensitive to the inclusion of azimuthal dependency $(\cos \theta, \sin \theta)$ as can be

seen in the panel (e). Relative regression accuracies for the raw background ensemble correlations are similar to those for squared ensemble correlations (not shown).

Note that with more available training samples and a larger degree of freedom (i.e., number of hidden nodes), more subtle relationships will be accurately resolved. As a result, more inclusive predictors such as $(r, z_{\text{anl}}, z_{\text{obs}}, \phi_{\text{anl}}, \cos \theta, \sin \theta)$ will increasingly perform better than exclusive predictors such as $(r, z_{\text{anl}}, z_{\text{obs}})$. Therefore, the comparison in the previous paragraph only shows optimality for our specific configurations but does not conclude the general independence of background error correlations to the azimuth angle.

From the above observations, hereafter, we will examine the set of explanatory variables $(r, z_{\text{anl}}, z_{\text{obs}}, \phi_{\text{anl}})$ more carefully as our primary choice. We will show several prominent characteristics of the four-dimensional function of $(r, z_{\text{anl}}, z_{\text{obs}}, \phi_{\text{anl}})$.

Figure 6.8 shows how the neural networks with predictors $(r, z_{\text{anl}}, z_{\text{obs}}, \phi_{\text{anl}})$ can respond to the latitudinal sensitivity of the vertical error correlation length of zonal wind. It is known that the wind error is more vertically correlated at the higher latitudes (panel b). The background error correlation reproduced by a neural network (panel a) captures most of the qualitative features such as tighter vertical wind error correlation in the lower latitudes. This observation shows that the neural networks and training data sampled from the FOAM-LETKF system can incorporate the latitudinal dependency reported in the literature.

Next, we examine multivariate error correlation within the atmosphere. Since we have opted out the azimuth angle (θ) from the predictors by choosing $(r, z_{\text{anl}}, z_{\text{obs}}, \phi_{\text{anl}})$, the geostrophic relationship between the geopotential and wind will only be cap-

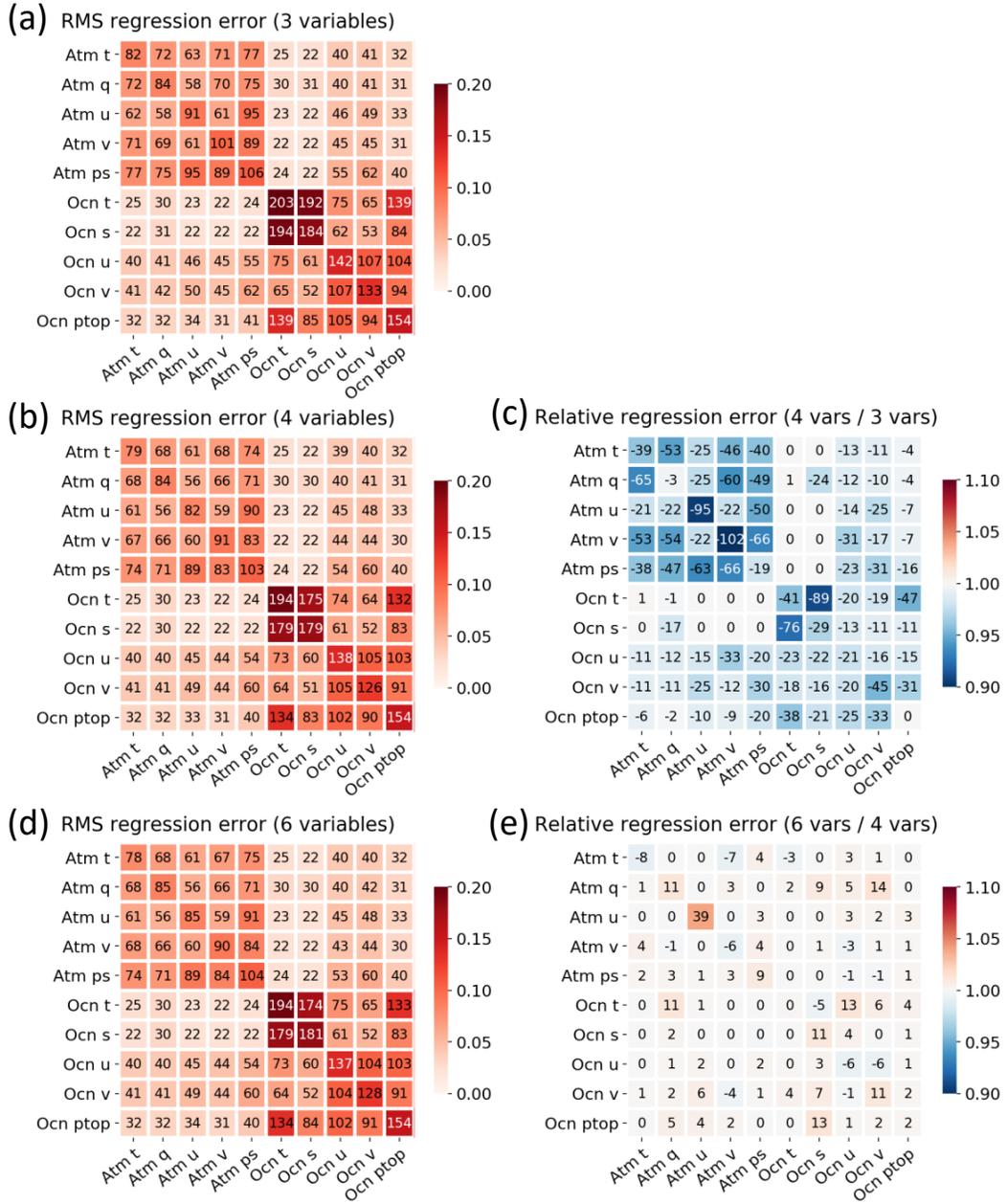


Figure 6.7: Panels (a, b, d): RMS regression error to validation data set of neural networks to squared ensemble correlations with different set of explanatory variables: $(r, z_{\text{anl}}, z_{\text{obs}})$, $(r, z_{\text{anl}}, z_{\text{obs}}, \phi_{\text{anl}})$, and $(r, z_{\text{anl}}, z_{\text{obs}}, \phi_{\text{anl}}, \cos \theta, \sin \theta)$, respectively. Superimposed numbers are RMS regression error multiplied by 1000 (redundant to color). Panels (c, e): relative regression errors, (b) divided by (a) and (d) divided by (b), respectively. Blue color means improved regression by additional explanatory variable(s). Superimposed numbers are the ratio minus one multiplied by 1000 (redundant to color).

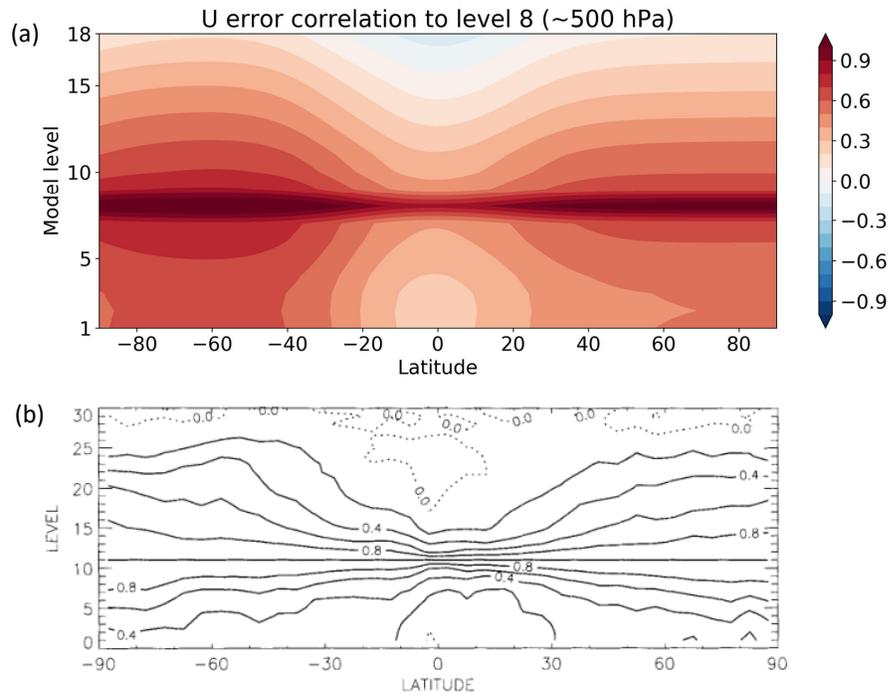


Figure 6.8: Latitudinal sensitivity of vertical error correlation of U -wind. Panels (a) vertical background ensemble auto-correlation of U -wind reproduced by neural networks; correlations are to the 8th level of FOAM (approximately 500 hPa). Panel (b): vertical error auto-correlation of U -wind to the 11th level of another model (approximately 500 hPa), as a function of latitude, estimated by forecast difference method (the NMC method) in January 1999; figure courtesy [Ingleby \(2001\)](#).

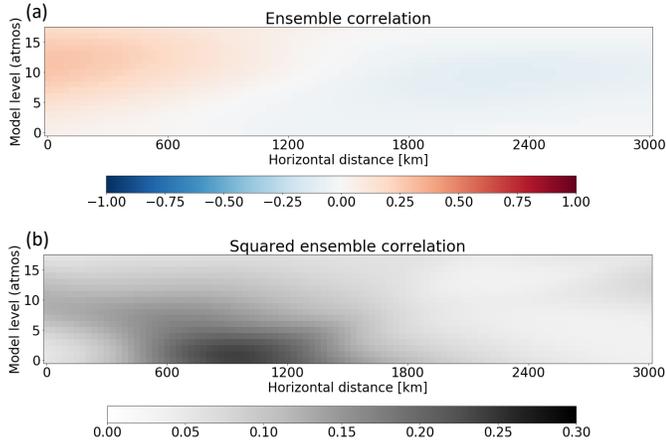


Figure 6.9: Background ensemble correlation reproduced by a neural network, between surface pressure observation and V -wind field at 60°S . Panels (a) and (b) are regressions to the raw and squared background ensemble correlations, respectively.

tured by mean squared correlation (see regression of Sin generation function by r of Figures 6.5 and 6.6). Figure 6.9b shows the squared background ensemble correlation reproduced by a neural network, between surface pressure observation and V -wind field as a function of horizontal distance r and analysis level z_{anl} . The result is consistent to the regression of Sin function by r in Figure 6.6 and the relationship $V\Phi$ in Figure 6.3 that the squared error correlation has its peak near the surface, a thousand kilometers away from the observation location. There also exists weaker sensitivity to upper V -wind just above the surface pressure observation ($z_{\text{anl}} \sim 10$). This sensitivity might be explained by warm/cold air advection, where the southerly wind (positive- V) brings colder, denser air, and it may increase the surface pressure (positive- P) in the southern hemisphere. These errors are positively correlated (panel a), and the inverse relationship is seen in the northern hemisphere (not shown).

Finally, to show its applicability for the SCDA problems, we see how the neural

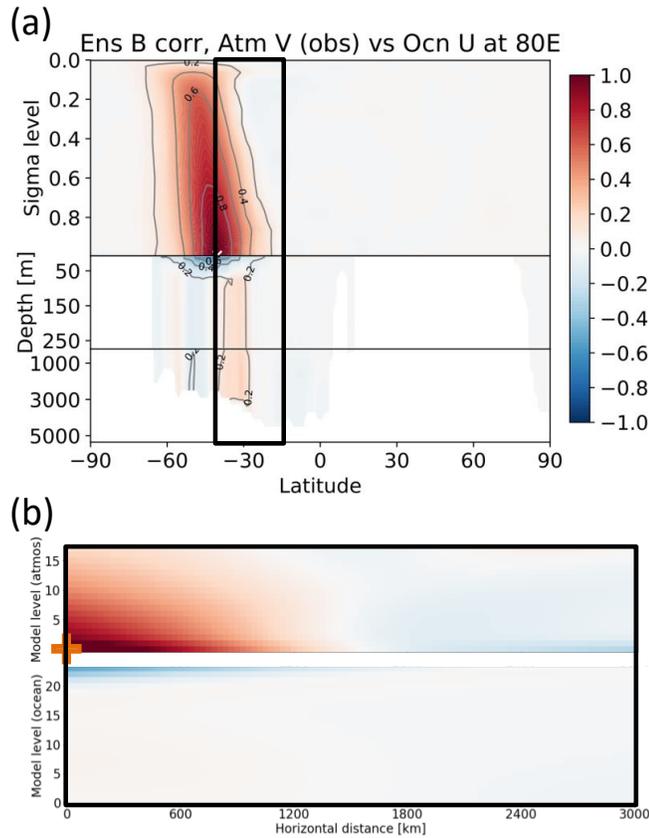


Figure 6.10: Capability of neural networks to reproduce the time-mean background ensemble correlation between the atmosphere and the ocean. Top: time-average of background ensemble correlation, between a hypothetical observation of surface meridional wind at 40°S 80°E (southern Indian ocean) and meridional wind and zonal current meridional cross-sections along 80°E . Bottom: mean background ensemble correlation reproduced by a neural network, between a hypothetical surface meridional wind observation at 40°S and meridional wind and zonal current field, as a function of horizontal distance and the analysis levels. Note that the former is a meridional cross-section, but the latter is a radially symmetric function.

networks reproduce the coupled error correlations. Figure 6.10 compares the simply time-averaged background ensemble correlation (panel a) and the background ensemble correlation reproduced by the neural networks (panel b). Both panels show a proxy to the sensitivity to a V -wind observation at 40°S , and the heavy black rectangles are showing almost comparable domains. Both panels are showing that the surface V -wind observation can constrain the V -wind field above (background error positively correlated with the observable) and the U -current field below (background error negatively correlated with the observable). As discussed in Section 6.2, the quantity shown in the former panel cannot be computed and stored for all the pairs of analysis variable and observable.

Overall, the regression to the (squared) background error correlation sampled from the FOAM-LETKF analysis cycle is successful. However, the regression sometimes overshoots to impossible value. For example, the regression may produce a negative estimate for squared error correlation. This overshoot can be removed *a posteriori* when we apply a cutoff function g (next section). Also, the self-correlation is often less than unity. These problems are alleviated by increasing training samples (the regression is much worse with eight times less training data; not shown). Also, an increased number of hidden units may contribute to better regression to fine structures.

6.5 Assimilation experiments with correlation-cutoff

In this section, we present proof-of-concept experiments of the strongly coupled atmosphere-ocean LETKF with the correlation-cutoff method.

6.5.1 Choice of cutoff function

Before conducting assimilation experiments, we briefly discuss a cutoff function g . There are several desired characteristics for cutoff functions, some physical and some computational.

First of all, a cutoff function should be a non-decreasing function from the concept of the correlation-cutoff method introduced in Chapter 2. An observation should be assimilated only to the analysis variables that have highly correlated background error to the observable quantity. This characteristic may also help the analysis to be balanced if we can consistently correct dynamically intimate variables that should have highly correlated background errors.

Cutoff functions should be in a range $[0, 1]$ from the discussion in Subsection 6.2.2. Combined with the non-decreasing condition, the cutoff function is likely to map 0 to 0 and 1 to 1. Note that by setting $g(x) = 0$ for $x \leq 0$ and $g(x) = 1$ for $x \geq 1$, we can ensure the localization weight to be in an appropriate range even with overshooting regression f .

A cutoff function should be smooth enough to avoid unnecessary disturbance of dynamics. Our choice of hyperbolic tangent activation ensures, for each observation and analysis variable types, the estimated squared correlation is a C^∞ function of

numerical predictors $(r, z_{\text{anl}}, z_{\text{obs}}, \phi_{\text{anl}})$. Therefore, the localization weight ρ is as smooth as the cutoff function g . However, the desirable smoothness is unknown; for example, [Houtekamer and Mitchell \(1998\)](#) and [Ott et al. \(2004\)](#) set localization weight zero outside the cutoff radius (not even continuous), while many localized ensemble DA systems use the piecewise polynomial function of [Gaspari and Cohn \(1999\)](#) (twice continuously differentiable).

From a computational perspective, a cutoff function should provide a sparse localization function. That is, localization weights should be identically zero if the estimated squared error correlation is below some threshold. For both the serial EnSRF and the LETKF, the analysis cost increases with the average number of observations assimilated with nonzero weight into each analysis variable. Therefore, observations with small estimated squared error correlations should not be assimilated rather than assimilated with a small weight. We would probably like to set the threshold larger than $1/(K - 1)$ for a K -member ensemble because the squared sample correlation estimated by K random samples drawn from an uncorrelated distribution converges to $1/(K - 1)$ ([Pitman, 1937](#)), and any value not much larger than $1/(K - 1)$ would not be reliable⁶.

Finally, it may involve an appropriate number of tuning parameter(s). Too many parameters unnecessarily increase manual intervention, while a few tunable

⁶However, some evidence shows that the ensemble perturbations of *deterministic* EnSRFs can represent even smaller underlying error correlations under idealized situations ([Sakov and Oke, 2008](#)). This behavior is also apparent in [Figure 2.1](#), where a 10-member ensemble can tell us mean squared error correlation between some variables is much less than $1/9$.

parameters may be helpful to ensure the localization weight is appropriate for the ensemble size and counteract other imperfections. The theoretically optimal cutoff function is yet to be developed.

Upon these considerations, we tentatively opt a piecewise quadratic function with a cutoff parameter c and vanishing derivative at $x = 1$:

$$\rho = g(x) = \begin{cases} 0 & (x \leq c) \\ 1 - \left(\frac{1-x}{1-c}\right)^2 & (c < x \leq 1) \\ 1 & (x > 1), \end{cases} \quad (6.5)$$

where x is a squared background error correlation reproduced by the regression function f . This cutoff function is continuous, produces sparse localization function, and has heavier weight when $x \sim 1$.

6.5.2 One-year forecast-analysis cycle experiment

Table 6.2 shows the configuration of two experiments, namely, *Control* and *Cutoff*. The cutoff parameter 0.05 is experimentally chosen from 0.05 and 0.1. The localization parameters for the Control experiment are manually tuned from several configurations.

The regionally and temporally averaged analysis and background errors of these experiments are shown in Figure 6.11.

First, we notice that the Cutoff experiment achieves overall smaller analysis and background (i.e., 24-hour forecast) errors than the Control experiment in the atmosphere (panels a-d). This improvement is apparent in all the variables and

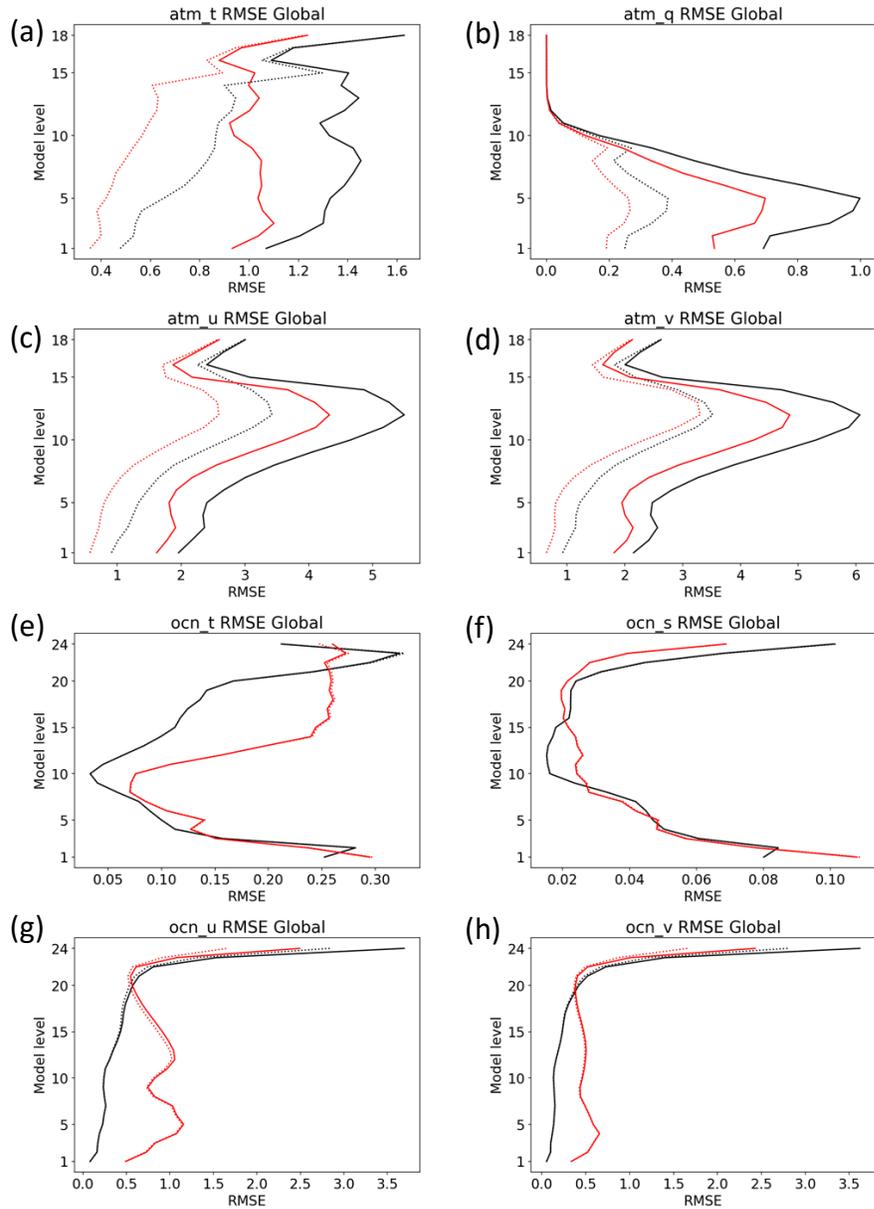


Figure 6.11: Global RMSE of Control (black) and Cutoff (red) experiments as a function of model levels. Solid and dotted lines show the background (i.e., 24-hour forecast) and analysis errors to the known truth. Panels (a-d) show atmospheric temperature (K), specific humidity (g/kg), zonal wind (m/s), and meridional wind (m/s), respectively. Panels (e-h) show oceanic temperature (K), salinity (PSU), zonal current (cm/s), and meridional current (cm/s), respectively. Out of 360 days, the first 30 days are excluded from averaging.

Table 6.2: Settings of strongly coupled assimilation experiments *Control* and *Cutoff*. See text for the differences between *Cutoff*, *Cutoff-nodeep*, and *Cutoff-nodeep-ocnhalf* experiments.

	Control	Cutoff
Experiment period	One model year (from January 1)	
Analysis interval	24 hours	
Observation network	As Table 4.2, at the end of each window	
Analysis method	64-member, strongly coupled LETKF with IAU 0	
Localization	Horizontally 1000 km (atmosphere) and 400 km (ocean), vertically 3 model levels (atmosphere/ocean) Eq. 1.6	Correlation-cutoff with neural networks [predictors $(r, z_{\text{anl}}, z_{\text{obs}}, \phi_{\text{anl}})$ and Eq. 6.5 with parameter 0.05] up to horizontally 3000 km, vertically 16 levels away
Analysis variables	T, Q, U, V, P_s (atmosphere), $T, S, U, V, P_{\text{top}}$ (ocean)	
Covariance inflation	RTPP; 30% prior perturbations (atmosphere) and 90% prior perturbations (ocean)	

levels. We will later examine the geographical features of errors.

The ocean analysis quality of Cutoff is generally worse than Control, especially in the deep, unobserved ocean (panels e-h). With a conjecture that the one-year offline experiment has not provided sufficiently reliable information for the deep ocean, whose timescale reaches multiple years, we conduct two additional experiments, *Cutoff-nodeep* and *Cutoff-nodeep-ocnhalf*.

In the Cutoff-nodeep experiment, the analysis of unobserved deep ocean variables below 2300 meters is turned off. Further, in the Cutoff-nodeep-ocnhalf experiment, the localization weights for the ocean analysis variables are halved to reduce the observation impacts on the ocean. Everything else remains the same as the Cutoff experiment.

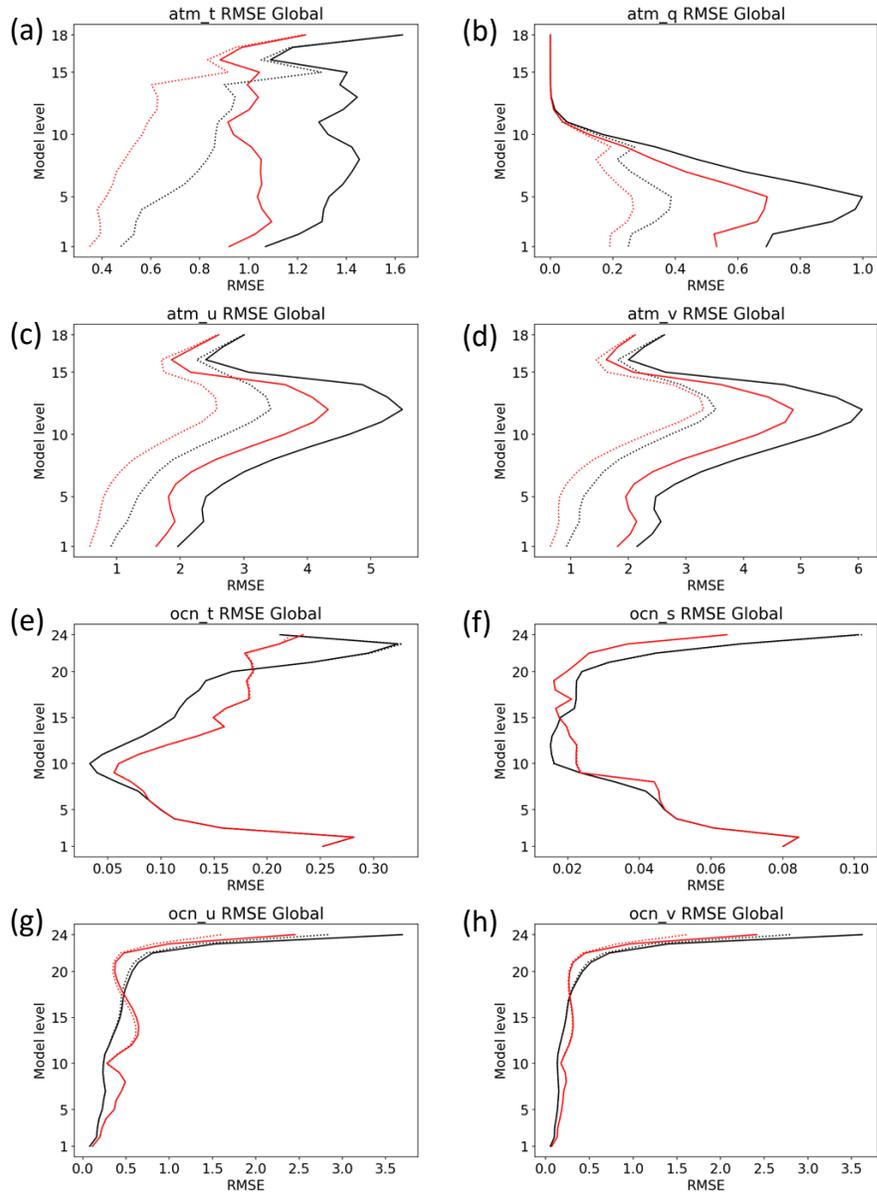


Figure 6.12: Same as Figure 6.11 but for Control (black) and Cutoff-noddeep (red) experiments.

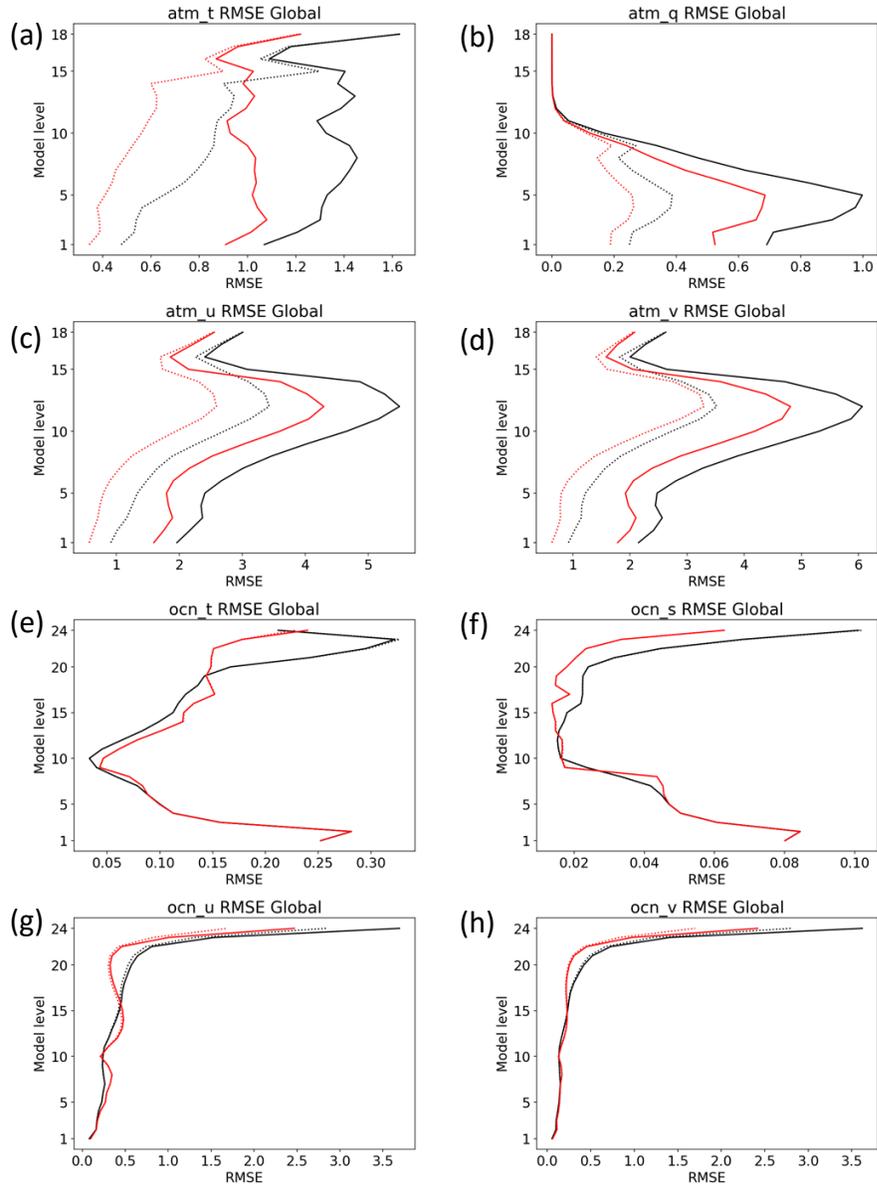


Figure 6.13: Same as Figure 6.11 but for Control (black) and Cutoff-nodeep-ocnhalf (red) experiments.

The analysis and background errors of these experiments as a function of model levels are shown in Figures 6.12 and 6.13, respectively. Both of these tunings substantially reduced the ocean analysis and background errors from the Cutoff experiment. The ocean errors of the Cutoff-nodeep-ocnhalf experiment are now comparable or slightly superior to the Control experiment (Figure 6.13 e-h) with improvements in shallower layers. Atmospheric analysis and background errors are almost unaffected with these changes, showing overall improvement over Control.

Geologically, atmospheric background error is most reduced in the tropics (Figure 6.14). This can be a result of improved variable localization; in the tropics, the geostrophic relationship between mass and wind fields should be weaker than in the higher latitudes. Hence, appropriately localizing this relationship should improve the analysis. On the other hand, the Control experiment assimilates mass and wind observations to the mass and wind fields without distinction.

Ocean surface errors of zonal and meridional currents are smaller than Control almost everywhere, but the temperature, salinity, and altitude fields experience regional degradation near the Antarctic (Figure 6.15). The reason for degradation in these regions remains unclear.

Overall, the proof-of-concept experiments in this section support the neural network usage for localization and the concept of the correlation-cutoff method, with improved atmospheric analysis and comparable ocean analysis to Control. Reliability of the one-year offline experiment in the internal ocean is dubious and it will need further sophistication. Manual tuning such as those applied to our Cutoff-nodeep-ocnhalf experiment or tighter physical cutoff distance in the ocean may be

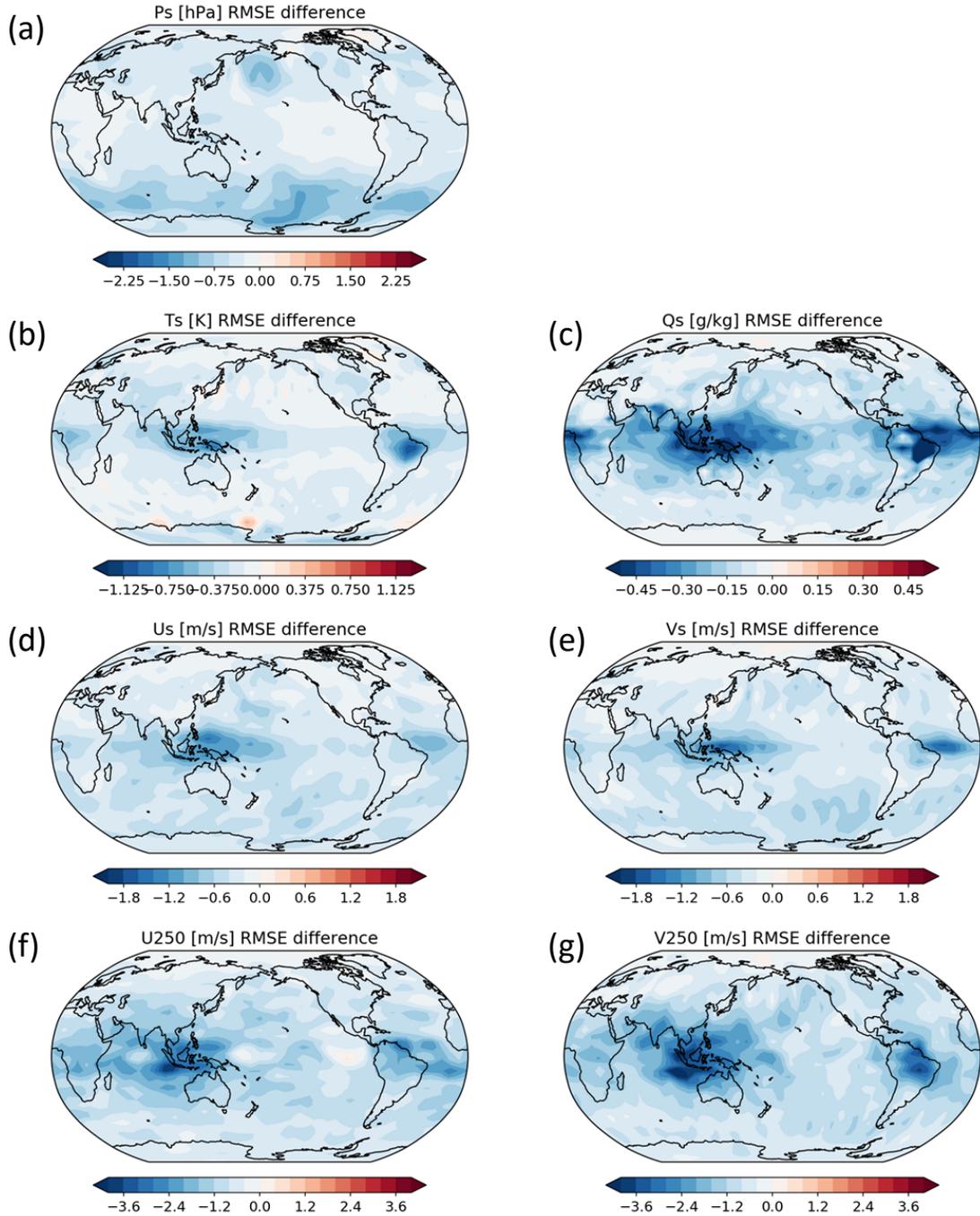


Figure 6.14: Atmospheric background (24-hour forecast) RMSE of the Cutoff-nodeep-ocnhalf experiment relative to the Control experiment ($\text{RMSE}_{\text{Cutoff-nodeep-ocnhalf}} - \text{RMSE}_{\text{Control}}$). Blue (red) colors show improvement (degradation) from the Control experiment. Panels (a-e) show surface pressure, temperature, specific humidity, zonal wind, and meridional wind, respectively. Panels (f) and (g) show zonal and meridional winds at 250 hPa (11th model level). Units are written in each panel. Out of 360 days, the first 30 days are excluded from averaging.

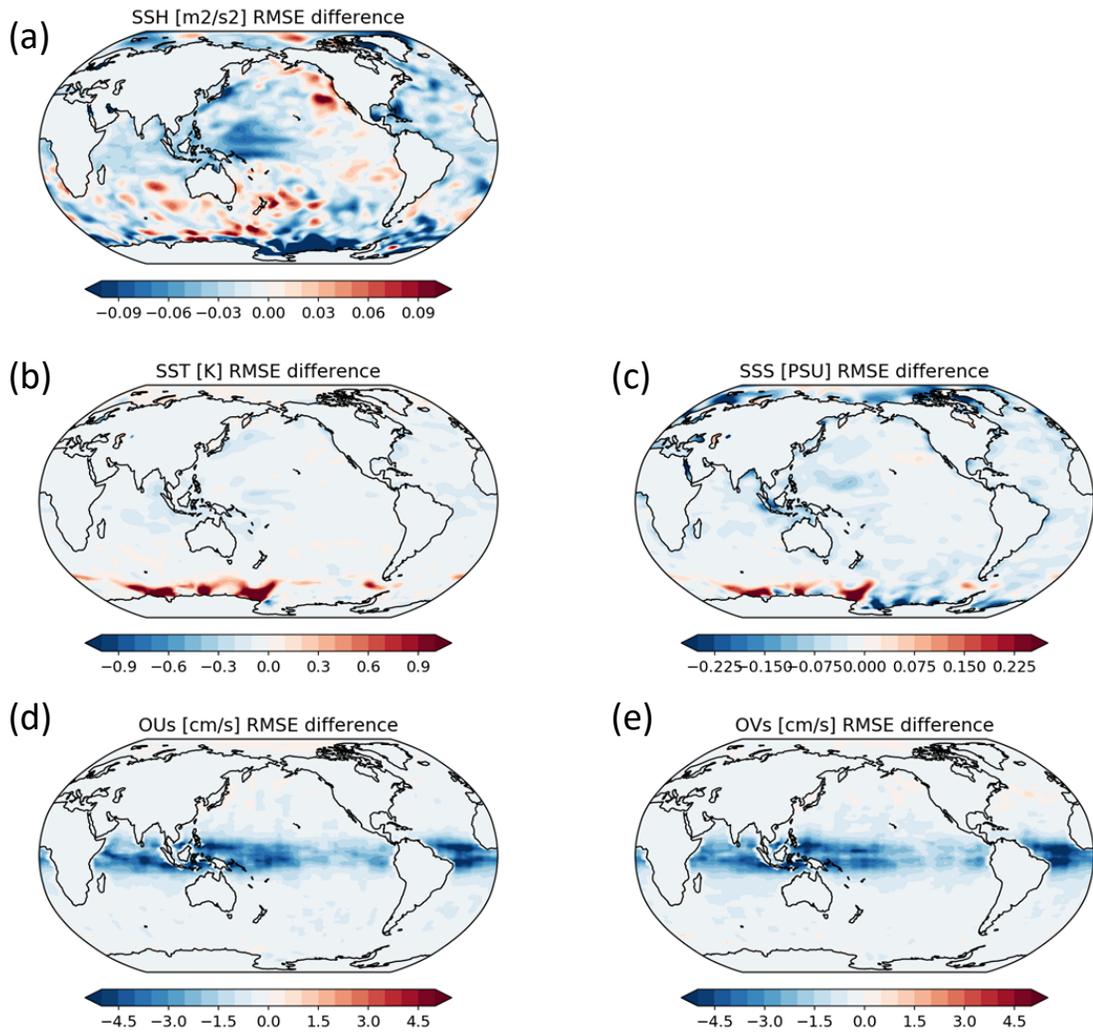


Figure 6.15: Same as Figure 6.14 but for oceanic variables. Panels (a-e) show sea surface height, sea surface temperature, sea surface salinity, surface zonal current, and surface meridional current, respectively.

necessary for domains where offline experiment alone cannot provide reliable climatological information on the error structures.

6.5.3 Computation cost

In this section, we will examine the direct increase of analysis computation time due to the introduction of neural network and then discuss auxiliary matters for readers interested in more efficient implementation.

A forward evaluation of a neural network takes $\mathcal{O}(DM + ML)$ floating point computations; this is a few hundreds in our configuration $(D, M, L) = (4, 30, 1)$. On the other hand, the analysis of a variable with the LETKF takes $\mathcal{O}(K^2 p_{\text{loc}} + K^3)$ floating point computations for a K -member ensemble and p_{loc} local observations assimilated. Therefore, the evaluation of neural networks (which will be repeated for $\mathcal{O}(p_{\text{loc}})$ times) will not dominate the analysis cost if $DM + ML \lesssim K^2$, which is true for our configuration ($K = 64$). We confirm this statement with experiments.

For a direct comparison of computation cost, we conduct two additional experiments, *Control-3000* and *Neural-3000*, both of which allow nonzero localization weights up to 3000 km horizontal distance and ± 16 vertical levels⁷. The Control-3000 experiment uses the Frolov et al. (2016) normalized distance (Eqs. 1.5 and 1.6) so that the localization weights become identically zero beyond 3000 km and 16 vertical levels. The Neural-3000 experiment uses neural networks with predictors $(r, z_{\text{anl}}, z_{\text{obs}}, \phi_{\text{anl}})$ and the cutoff function Eq. (6.5) with a cutoff parameter $c = 0.2$,

⁷The Neural-3000 experiment is found to be using the regression coefficients to *raw* error correlations, which does not significantly affect discussion here.

and any observations beyond 3000 km or 16 vertical levels are given identically zero localization weights. Both experiments assimilate all the observation networks in Table 4.2 for three consecutive 24-hour windows.

With 256 Intel Ivy Bridge processors (2.8 GHz) and Intel Fortran compiler with -O3 optimization, the computation times of Control-3000 and Neural-3000 experiments are 348 and 368 seconds, respectively (the Neural-3000 experiment is 5.7% more expensive). This number includes the time spent on the ensemble forecast, which is not dominant in our settings⁸.

By profiling the Neural-3000 experiment, the evaluation of neural networks turns out to account for about 10% of the LETKF computation time; this number is estimated by a line profiler and imprecise due to compiler optimization. Furthermore, two-thirds of the evaluation cost of neural networks is spent for the hyperbolic tangent function (the rest is for linear combination). Therefore, there is a chance to triple the number of hidden nodes without increasing the evaluation cost if we choose a faster activation function like the rectifier function.

Having confirmed that the direct increase of computation cost is acceptable, below we mention two causes that can indirectly increase the analysis cost and discuss how to alleviate it.

⁸The ocean model has much more analysis variables than the atmosphere model due to their relative resolutions. On the other hand, the 24-hour forecast for the ocean is cheap due to the ocean model's longer time step and the absence of expensive parameterized physics. As a result, the ocean analysis dominates the entire forecast-analysis cost in our experiments. In contrast, for atmosphere-only forecast-analysis systems, the ensemble forecast is usually more expensive than the analysis (Houtekamer and Zhang, 2016).

First, for localization depending only on the physical distance, there are established and efficient algorithms of observation selection (k D-tree, Octree, and other spatial decomposition; also see Appendix A). In these algorithms, many faraway observations can be rejected by a single evaluation of distance between an analysis variable and a subset of observations. On the other hand, we can only evaluate localization weight of an observation at a time with neural networks. To reduce the number of observations whose localization weight is actually evaluated, we can combine physical cutoff distance, beyond which localization weights are set identically zero. Physical cutoff distance is used in the Neural-3000 experiment above (3000 km and 16 levels). Evaluation of localization weight can also be avoided by setting localization weight identically zero for some variable pairs (e.g., between salinity and wind). This decision can also be guided by offline statistics.

Second, for distance-only localization, the LETKF (Algorithm 2) can reuse $\bar{\mathbf{w}}^a$ and \mathbf{W}^a for all the collocated analysis variables because those variables’ analyses are identical in the ensemble space, which can speed up the analysis by a few times (Hunt et al., 2007). This shortcut is not adopted in the Control-3000 experiment above. For any localization that takes analysis variable types into account, including our correlation-cutoff method and the variable localization of Kang et al. (2011), this shortcut is unavailable. To alleviate the analysis cost, we might want to reduce the number of *analysis variable types* that use independent localization weights; for example, we might want to use a category “wind” instead of zonal and meridional winds as two independent categories.

Overall, the correlation-cutoff method with neural networks has shown its

feasibility for strongly coupled global atmosphere-ocean analysis.

6.6 Concluding remarks and future directions

In this chapter, we have demonstrated that the correlation-cutoff method can be realized with neural networks even for a global atmosphere-ocean model, where the computational cost was a primary concern.

With toy correlation models inspired by the geostrophic theory, we have shown that neural networks, even with a small number of hidden units, can be used to explain spatially extended “correlation functions”. The neural networks can be used as an end-to-end measure of regression only with the recent development of machine learning methodologies.

The neural networks are used to fit the (squared) background error of an atmosphere-ocean coupled model, where we have shown that it can reproduce some known features of background errors. For example, the dependency of the vertical correlation length of wind error on latitude is reproduced. Another example is the geostrophic error structure, where the peak sensitivity of the wind analysis to a pressure observation resides a thousand kilometers away from the observation. Also, mathematical validity of these localization functions for serial EnSRF and the LETKF is shown.

The regression function that maps $(r, z_{\text{anl}}, z_{\text{obs}}, \phi_{\text{anl}})$ to the squared background error correlation is used in a localization function of a strongly coupled atmosphere-ocean LETKF system. The experiments showed computational feasibility for global

strongly coupled analysis. We have also discussed how to alleviate the indirect increase of computation cost accompanying the correlation-cutoff method.

For the analysis accuracy, our proof-of-concept experiments have shown substantially improved atmospheric analysis with the largest improvements in the tropics, and comparable ocean analysis to the Control experiment after some tuning. The improved atmospheric analysis in the tropics may be attributed to the appropriate variable localization between mass and wind variables. For the internal ocean, whose timescale is longer than the offline experiment, tuning based on the first principle seems necessary.

Here are future directions suggested:

- The use of neural networks in a localization function — one of the most elemental parts of ensemble data assimilation algorithm that shall be evaluated billions of times in an analysis — has been attempted for the first time. Therefore, we have prioritized implementing the method with the global atmosphere-ocean model to show its computational feasibility. However, we may want to step back to simpler models (but more complex than the 9-variable model tested in Chapter 2) and thoroughly examine the balance and accuracy of the analysis compared to other localization methods. To show its ability to naturally combine spatial and variable localizations, the simple model should be multivariate and spatially extended like shallow water equations or the coupled model of [Lorenz \(2005\)](#) used by [Bishop et al. \(2017\)](#).
- Further tuning of the cutoff function g , including exploration of its theoretical

optimal.

- More accurate offline estimation of squared error correlation, especially for the internal ocean. Longer offline experiments with appropriate covariance inflation methods will provide more accurate information for localization of the internal ocean. For this purpose, the relaxation to prior spread method (RTPS; [Whitaker and Hamill, 2012](#)) seems more appropriate than the relaxation to prior perturbation method (RTPP; [Zhang et al., 2004](#)) because the former does not change the analysis ensemble correlation structure in the model space. Also, the initial ensemble must be carefully prepared because an oceanic ensemble is slow to reach its steady state.
- The squared background error correlation estimated with training data from an offline EnKF cycle (weakly coupled with spatial localization in our case) may misfit that of the production EnKF cycle (strongly coupled with correlation-cutoff). Also, we may want to incorporate the gradually changing observation network, which is known to affect the background error correlation length (e.g., [Ying et al., 2018](#)). Therefore, it is worth exploring the iterative estimation of squared error correlation using training data from correlation-cutoff EnKF cycles ([Brian Hunt, 2018](#), personal communication) or online training of squared error correlation ([Kayo Ide, 2019](#), personal communication).
- Usage of graphics processing units (GPUs). Although current computation cost with central processing units (CPUs) is acceptable, training and evaluation of neural networks are most efficiently done with GPUs. This will help

to improve the regression accuracy without increasing training and evaluation time. Training can also speed up by mini-batch training.

- Ultimately, our methodology should be tested with more complex and realistic systems, including assimilation of real observations or extension to Earth system models beyond the atmosphere and the ocean.

Chapter 7: Concluding summary

7.1 Summary of this thesis

Strongly coupled data assimilation of the Earth system is thought to be an ultimate methodology for its state estimate. Its concept is simple because we can deal with the coupled system as a single, autonomous dynamical system. However, several practical problems remain before we realize the operational strongly coupled data assimilation systems, caused by the different temporal and spatial scales of the dynamics. Another problem is our ignorance to the three-dimensional coupled error structures, an essential element of strongly coupled DA.

Localization is one of the largest problems to realize the strongly coupled DA with ensemble methods, where the background error covariance is generally less reliable than in the weakly coupled DA. Therefore, we have explored a localization method that applies to the analysis of the coupled system, where the physical distance is not the only descriptor of the coupling strength between two quantities.

In Chapter 1, we have reviewed the current understanding and approaches for coupled data assimilation and localization of ensemble-based analysis.

In Chapter 2, we have shown that the time-mean squared background ensemble correlation can serve as a “distance” between an analysis variable and an

observation. Based on this observation, we proposed the correlation-cutoff method, where we only couple the analysis between pairs of subsystems with strong error correlations. The method is tested with a nine-variable coupled model of the chaotic atmosphere and ocean, showing smaller analysis errors with a limited ensemble size.

As a related topic, the effect of coupling on the error growth and the characteristics of the attractor is examined with a simple chaotic coupled model with fast and slow timescales (Chapter 3). Here, discontinuous sensitivity of the attractor's characteristics to the coupling strength is reported. The results call for more attention to the forward dynamics of the coupled systems, which will affect the uncertainty estimate, including those in the ensemble predictions.

Chapter 4 served for our system description, which is commonly used for the investigation in Chapters 5 and 6.

In Chapter 5, we have investigated the structure of the atmosphere-ocean coupled background errors and discussed where the strongly coupled DA will be most beneficial. Certain background error correlations between some surface variables exist, but at the same time, the correlation matrix is sparse enough to support the importance of appropriate variable and spatial localization. We compared the background ensemble statistics of two global atmosphere-ocean EnKF systems with different level of sophistications. With more realistic CFS-IITM, more detailed geological features of local error correlations have revealed. Results of this chapter will help future design of strongly coupled DA, either with the ensemble or variational methods.

We have further extended the correlation-cutoff method for a non-static ob-

servations network and a global atmosphere-ocean coupled model in Chapter 6 by leveraging neural networks. With neural networks, we can build a localization function that considers the analysis variable and observation types as well as their geographical location and distance, with realistic storage and computation costs. Our proof-of-concept experiments showed improved analysis in the atmosphere, but some tuning was necessary for internal ocean, where the error statistics sampled from an offline DA cycle could be inaccurate.

Throughout the thesis, we have examined a spectrum between uncoupled/localized and fully coupled systems with a stress on the coupled atmosphere-ocean system. However, active research on coupled data assimilation is likely to continue for the coming decade with the emergence of increasingly sophisticated coupled models and improved understanding of coupled systems.

7.2 Future direction

In addition to the future directions of the correlation-cutoff method discussed in Section 6.6, the effect of the coupled forward model in the forecast-analysis system (i.e., the difference between uncoupled and weakly coupled DAs) should also be investigated more thoroughly. Our collaborative work with a coupled quasi-geostrophic model (Penny et al., 2019) partially serves for this purpose, which has examined the broad range of methodologies including a spectrum from uncoupled to strongly coupled DA paradigms and a few assimilation algorithms including variational, ensemble, and hybrid methods on the accuracy and stability of coupled

state estimate. Local causality or predictability¹ analysis (Bach et al., 2019) will also be a promising approach for investigating coupled forward dynamics in multiple timescales, including those of DA windows.

Finally, coupled DA will also provide a basis on which we can estimate and correct the model deficiencies related to coupled processes (Bhargava et al., 2018; Carton et al., 2018; Liu et al., 2014), accurately account for uncertainties in coupled observation operators (Geer et al., 2018), and estimate observation impact on coupled forecasts (Kalnay et al., 2012; Chen and Kalnay, 2019). These by-products of coupled DA will be as beneficial for future improvements of coupled numerical predictions as the improved coupled state estimates, the main product of coupled data assimilation.

¹Global predictability has been examined by, for example, Atmospheric Model Intercomparison Project (AMIP)-type experiments, where prescribed SST drives an atmospheric model. Local predictability will be similarly important to understand more elemental coupling processes.

Appendix A: Speedup observation lookup with Octree

The standard algorithm of the LETKF was shown in Algorithm 2 (Chapter 4). An essential part of this algorithm is the selection of observations within the localization (query) radius (line 18 of Algorithm 2). When the simple linear search algorithm (i.e., brute force) is employed, the computational cost of observation lookup for an analysis is $\mathcal{O}(np)$ for the number of model variables n and the number of observations p . This process often dominates the computational cost of the analysis when the localization radius is small enough (i.e., $|\mathcal{J}| \ll p$, where \mathcal{J} is the set of observations assimilated into each model variable with nonzero weight) and no other process costs $\mathcal{O}(np)$ or larger.

To speed up the observation lookup of the LETKF, an algorithm called k D-tree is often used (e.g., Szunyogh et al., 2008). The k D-tree algorithm recursively divides the set of observations into two disjoint sets. The division is done alternatively in x , y , and z directions until a leaf (i.e., a box without children) of the tree satisfies some stopping criteria (e.g., less than ten observations in the box). A drawback of k D-tree is its complexity of implementation, especially when combined with addition and removal of observations, which are sometimes unavoidable for DA applications because of non-synoptic observations and quality control.

In our implementation of LETKF for FOAM, a similar algorithm called *Octree*¹ is adopted. The Octree algorithm divides each set of observations into eight disjoint boxes so that the *volumes of the boxes* are almost equal (on the other hand, *kD-tree* divides the set so that the *numbers of observations* are almost equal). The advantage of Octree over *kD-tree* is, therefore, the static structure of the tree, which enables easy addition or removal of observations. An Octree is often used for multi-body simulations including video games, which require the heavy calculation of collision detection in three-dimensional space.

Algorithms 3 and 4 are pseudocode to add and get observations using Octree. Both *octree_add* and *octree_query* procedures are called recursively to walk through the tree in the depth-first order.

To apply Octree for spherical geometry, we employ a technique called *ball-tree* in the detection of the horizontal intersection of the query radius and the box. Each box is given its *center* location at construction. When an observation is added to a box, the *radius* of the box is updated so that all the observations in the box is contained in the radius of the box. (Line 17 of Algorithm 3). The radius of the box is later used to detect the possibility of having observation within the query radius (Line 8 of Algorithm 4). The beauty of the ball-tree technique is that it allows horizontal intersection detection between the query radius and the box by just a single evaluation of horizontal distance. This simplicity can also speed up the query compared to complex intersection evaluation between the rectangle box and the query radius on the curved coordinate. This technique applies even to the

¹The name Octree is a combination of oct (8) and tree (a recursive data structure).

curved coordinate including poles because the triangle inequality holds everywhere on the sphere [i.e., $dist(A, C) \leq dist(A, B) + dist(B, C)$ for any three points A , B , and C].

Figure A.1 is the horizontal visualization of some boxes of Octree. Note that the ball-tree technique may cause false detection of intersection between the box and the query radius (see the blue circle around the Florida peninsula, which intersects with the green circle but not with the green box). This false detection is acceptable because any observation outside the query radius is finally filtered out (Line 13 of Algorithm 4) and cannot be added to \mathcal{J} .

To see the utility of Octree, computation time spent for an analysis is measured with and without Octree. Our implementation divides the global three-dimensional domain (latitude, longitude, and level) four times, which results in $8^4 = 4096$ smallest boxes. In this experiment, the 64-member LETKF assimilates total $p = 34,520$ observations into $n = 1,822,380$ model variables. The assimilation is parallelly conducted using 128 MPI processes on the Deepthought2 supercomputer at the University of Maryland (Ivy Bridge, 2.8GHz). Horizontal localization allows assimilation of observations up to 2000 km distance, and vertical localization up to 3 model-level difference. Cross-assimilation between the atmosphere and the ocean is also allowed (i.e., SCDA). With this localization, the average number of observations assimilated into each model variable with nonzero weights $|\mathcal{J}|$ is 155.9. Table A.1 compares the time spent for observation lookup with and without Octree. The results show that Octree speeds-up the observation lookup by more than 40 times, with the final cost much less than the LETKF core calculations.

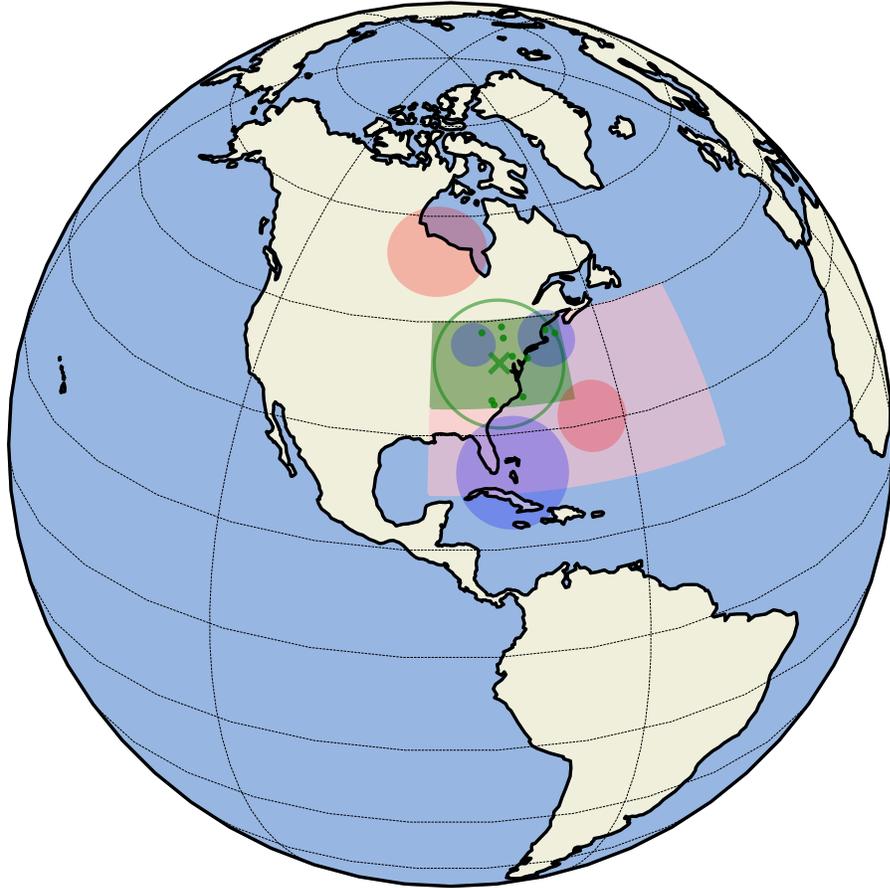


Figure A.1: Horizontal schematic of Octree application on the spherical coordinate. The pink box is one of the three-time-divided boxes, which is 45 degrees in longitude and 22.5 degrees in latitude. The green box is one of the four-time-divided boxes. The green cross, dots, and circle show the *center* of the green box (at 39.375°N 78.75°W), hypothetical observations in the green box, and the *radius* of the box after adding the observations (approximately 943 km), respectively. Note that the observations (green dots) exist only in the intersection of the green box and circle. Queries like the blue (red) circles intersect (does not intersect) with the green circle and may (may not) find observations in the green box.

Algorithm 3: Add an observation to Octree. Note that in Fortran nomenclature, *structures* are called as *derived types*.

```
1 structure BoxStruct contains
2   is_leaf : True if the box is leaf
3   children : Pointers to 8 child boxes unless the box is leaf
4   observations : Pointers to observations if the box is leaf
5   center : Horizontal center location
6   radius : Maximum horizontal distance from the center to observations,
   initialized to zero
7   top : Top level
8   bottom : Bottom level
9 end structure
10
11 box : An instance of BoxStruct
12 j : Index of the observation added to the tree
13 obsloc : 3-dimensional location of the observation
14 dist( $l_1, l_2$ ) : Horizontal distance of two locations ( $l_1, l_2$ )
15
16 procedure octree_add(box, j, obsloc)
17   box.radius  $\leftarrow$  max(box.radius, dist(obsloc, box.center))
18   if box.is_leaf then
19     Add j to box.observations
20   else
21     l  $\leftarrow$  index of the child contains the observation
        $\triangleright l \in [1, 8]$ , obtained using box.center, box.top, box.bottom,
       and obsloc
22     octree_add(box.children[l], j, obsloc)
23   end
24 end procedure
```

Algorithm 4: Get a list of observations from Octree

```
1 box, dist : As defined in Algorithm 3
2  $\mathcal{J}$  : List of observation indices, initialized to be empty
3 gridloc : Location of analysis grid (i.e., query center)
4 qr : Query radius
5 zrange : Vertical range of query
6
7 procedure octree_query(box,  $\mathcal{J}$ , gridloc, qr, zrange)
8   if  $\text{dist}(\text{box.center}, \text{gridloc}) > \text{qr} + \text{box.radius}$  or
    $\text{zrange} \cap [\text{box.bottom}, \text{box.top}] = \emptyset$  then
9     return  $\triangleright$  No observations within the query radius
10  end
11  if box.is_leaf then
12    for  $j \in \text{box.observations}$  do
13      if  $\text{dist}(\text{location of } j\text{th observation}, \text{gridloc}) \leq \text{qr}$  and
      level of  $j$ th observation  $\in \text{zrange}$  then
14        Add  $j$  to  $\mathcal{J}$ 
15      end
16    end
17  else
18    for  $l \in [1, 8]$  do
19      octree_query(box.children[ $l$ ],  $\mathcal{J}$ , gridloc, qr)
20    end
21  end
22 end procedure
```

Table A.1: Time spent on observation lookup with and without Octree. The computation time for the slowest process and the mean of all 128 processes are shown. Each number is the average of three experiments. For comparison, the computation time for LETKF core calculation (roughly equivalent to lines 19 to 32 of Algorithm 2) is also shown.

	Linear search	Octree	LETKF core
Slowest process	78.88 sec	1.91 sec	21.56 sec
Mean	43.00 sec	0.97 sec	11.76 sec

To conclude this appendix, the simplicity of Octree with the ball-tree technique can be a favored choice over the commonly adopted k D-tree algorithm for localized data assimilation applications. Although the balanced structure of k D-tree may have an advantage for other problems like nearest neighbor search, it is expected that these algorithms have similar performance for range search like finding observations in a localization radius. Therefore, for our applications, we can choose one of these algorithms by the ease of implementation if no ready-made libraries are available.

Bibliography

- Anderson, J., and L. Lei, 2013: Empirical Localization of Observation Impact in Ensemble Kalman Filters. *Monthly Weather Review*, **141** (11), 4140–4153, doi:10.1175/MWR-D-12-00330.1.
- Anderson, J. L., 2003: A Local Least Squares Framework for Ensemble Filtering. *Monthly Weather Review*, **131** (4), 634–642, doi:10.1175/1520-0493(2003)131<0634:ALLSFF>2.0.CO;2.
- Anderson, J. L., 2007: Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter. *Physica D: Nonlinear Phenomena*, **230** (1-2), 99–111, doi:10.1016/j.physd.2006.02.011.
- Anderson, J. L., 2012: Localization and Sampling Error Correction in Ensemble Kalman Filter Data Assimilation. *Monthly Weather Review*, **140** (7), 2359–2371, doi:10.1175/MWR-D-11-00013.1.
- Anderson, J. L., and S. L. Anderson, 1999: A Monte Carlo Implementation of the Nonlinear Filtering Problem to Produce Ensemble Assimilations and Forecasts. *Monthly Weather Review*, **127** (12), 2741–2758, doi:10.1175/1520-0493(1999)127<2741:AMCIOT>2.0.CO;2.
- Bach, E., S. Motesharrei, E. Kalnay, and A. Ruiz-Barradas, 2019: Local atmosphere-ocean predictability: dynamical origins, lead times, and seasonality. *Journal of Climate*, Accepted.
- Bannister, R. N., 2008a: A review of forecast error covariance statistics in atmospheric variational data assimilation. I: Characteristics and measurements of forecast error covariances. *Quarterly Journal of the Royal Meteorological Society*, **134** (637), 1951–1970, doi:10.1002/qj.339.
- Bannister, R. N., 2008b: A review of forecast error covariance statistics in atmospheric variational data assimilation. II: Modelling the forecast error covariance statistics. *Quarterly Journal of the Royal Meteorological Society*, **134** (637), 1971–1996, doi:10.1002/qj.340.

- Bettge, T. W., J. W. Weatherly, W. M. Washington, D. Pollard, B. P. Briegleb, and W. G. J. Stand, 1996: The NCAR CSM Sea Ice Model. URL <http://www.cgd.ucar.edu/ccr/bettge/ice/>.
- Bhargava, K., E. Kalnay, J. A. Carton, and F. Yang, 2018: Estimation of Systematic Errors in the GFS Using Analysis Increments. *Journal of Geophysical Research: Atmospheres*, **123** (3), 1626–1637, doi:10.1002/2017JD027423.
- Bishop, C. H., S. Frolov, D. R. Allen, D. D. Kuhl, and K. Hoppel, 2017: The Local Ensemble Tangent Linear Model: an enabler for coupled model 4D-Var. *Quarterly Journal of the Royal Meteorological Society*, **143** (703), 1009–1020, doi:10.1002/qj.2986.
- Bishop, C. H., and D. Hodyss, 2009: Ensemble covariances adaptively localized with ECO-RAP. Part 1: tests on simple error models. *Tellus A*, **61** A (1), 97–111, doi:10.1111/j.1600-0870.2008.00372.x.
- Bishop, C. M., 2006: *Pattern Recognition and Machine Learning*. Springer.
- Bloom, S. C., L. L. Takacs, A. M. Da Silva, and D. Ledvina, 1996: Data Assimilation Using Incremental Analysis Updates. *Monthly Weather Review*, **124**, 1256–1271, doi:10.1175/1520-0493(1996)124<1256:DAUIAU>2.0.CO;2.
- Bochi, J., 2008: The Multiplicative Ergodic Theorem of Oseledets. URL <http://www.mat.uc.cl/~jairo.bochi/docs/oseledets.pdf>, 1–12 pp.
- Carton, J. A., G. A. Chepurin, L. Chen, and S. A. Grodsky, 2018: Improved Global Net Surface Heat Flux. *Journal of Geophysical Research: Oceans*, **2**, 3144–3163, doi:10.1002/2017JC013137.
- Charney, J. G., R. Fjörtoft, and J. Neumann, 1950: Numerical Integration of the Barotropic Vorticity Equation. *Tellus*, **2** (4), 237–254, doi:10.1111/j.2153-3490.1950.tb00336.x.
- Chen, T.-C., and E. Kalnay, 2019: Proactive Quality Control: Observing System Simulation Experiments with the Lorenz '96 Model. *Monthly Weather Review*, **147** (1), 53–67, doi:10.1175/MWR-D-18-0138.1.
- Daley, R., 1993: *Atmospheric Data Analysis*. Paperback ed., Cambridge University Press, 457 pp.
- De Cruz, L., S. Schubert, J. Demaeyer, V. Lucarini, and S. Vannitsem, 2018: Exploring the Lyapunov instability properties of high-dimensional atmospheric and climate models. *Nonlinear Processes in Geophysics*, **25** (2), 387–412, doi:10.5194/npg-25-387-2018.
- Derber, J., and A. Rosati, 1989: A Global Oceanic Data Assimilation System. *Journal of Physical Oceanography*, **19** (9), 1333–1347, doi:10.1175/1520-0485(1989)019<1333:AGODAS>2.0.CO;2.

- Deser, C., and M. L. Blackmon, 1995: On the relationship between tropical and North Pacific sea surface temperature variations. 1677–1680 pp., doi:10.1175/1520-0442(1995)008<1677:OTRBTA>2.0.CO;2.
- Evensen, 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), 10 143–10 162, doi:10.1029/94JC00572.
- Feng, X., K. Haines, C. Liu, E. de Boissésou, and I. Polo, 2018: Improved SST-precipitation intra-seasonal relationships in the ECMWF coupled climate reanalysis. *Geophysical Research Letters*, 1–9, doi:10.1029/2018GL077138.
- Fisher, M., 2003: Background error covariance modelling. *Seminar on Recent developments in data assimilation for atmosphere and ocean, 8-12 September 2003*, ECMWF, URL <https://www.ecmwf.int/node/9404>.
- Frolov, S., C. H. Bishop, T. Holt, J. Cummings, and D. Kuhl, 2016: Facilitating Strongly Coupled Ocean-Atmosphere Data Assimilation with an Interface Solver. *Monthly Weather Review*, **144**, 3–20, doi:10.1175/MWR-D-15-0041.1.
- Fujii, Y., and M. Kamachi, 2003: Three-dimensional analysis of temperature and salinity in the equatorial Pacific using a variational method with vertical coupled temperature-salinity empirical orthogonal function modes. *Journal of Geophysical Research*, **108** (C9), 3297, doi:10.1029/2002JC001745.
- Fukumori, I., 2002: A Partitioned Kalman Filter and Smoother. *Monthly Weather Review*, **130** (5), 1370–1383, doi:10.1175/1520-0493(2002)130<1370:APKFAS>2.0.CO;2.
- Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, **125** (554), 723–757, doi:10.1002/qj.49712555417.
- Gasperoni, N. A., and X. Wang, 2015: Adaptive localization for the ensemble-based observation impact estimate using regression confidence factors. *Monthly Weather Review*, **143** (6), 1981–2000, doi:10.1175/MWR-D-14-00272.1.
- Geer, A., P. D. Rosnay, Y. Hirahara, K. Lonitz, P. Browne, S. Tietsche, and S. English, 2018: Coupling through the observation operator. URL <https://www.ecmwf.int/en/learning/workshops/annual-seminar-2018>.
- Gelb, A., J. F. Kasper, Jr., A. N. Raymond, Jr., C. F. Price, and A. A. Sutherland, Jr., 1974: *Applied Optimal Estimation*. M.I.T. Press, 374 pp.
- Ginelli, F., H. Chaté, R. Livi, and A. Politi, 2013: Covariant Lyapunov vectors. *Journal of Physics A: Mathematical and Theoretical*, **46** (25), doi:10.1088/1751-8113/46/25/254005.

- Glorot, X., and Y. Bengio, 2010: Understanding the difficulty of training deep feed-forward neural networks. *AISTATS, volume 9 of JMLR Proceedings*, JMLR.org, 249–256, URL <http://proceedings.mlr.press/v9/glorot10a/glorot10a.pdf>.
- Greybush, S. J., E. Kalnay, T. Miyoshi, K. Ide, and B. R. Hunt, 2011: Balance and Ensemble Kalman Filter Localization Techniques. *Monthly Weather Review*, **139** (2), 511–522, doi:10.1175/2010MWR3328.1.
- Hallberg, R., 2013: Using a resolution function to regulate parameterizations of oceanic mesoscale eddy effects. *Ocean Modelling*, **72**, 92–103, doi:10.1016/j.ocemod.2013.08.007.
- Hamill, T. M., J. S. Whitaker, and C. Snyder, 2001: Distance-Dependent Filtering of Background Error Covariance Estimates in an Ensemble Kalman Filter. *Monthly Weather Review*, **129**, 2776–2790, doi:10.1175/1520-0493(2001)129<2776:DDFOBE>2.0.CO;2.
- Han, G., X. Wu, S. Zhang, Z. Liu, and W. Li, 2013: Error Covariance Estimation for Coupled Data Assimilation Using a Lorenz Atmosphere and a Simple Pycnocline Ocean Model. *Journal of Climate*, **26** (24), 10 218–10 231, doi:10.1175/JCLI-D-13-00236.1.
- Hartmann, D. L., 2015: Pacific sea surface temperature and the winter of 2014. *Geophysical Research Letters*, 1894–1902, doi:10.1002/2015GL063083.
- Hartmann, D. L., 2016: *Global Physical Climatology*. 2nd ed., Elsevier.
- Hoffman, R. N., and E. Kalnay, 1983: Lagged average forecasting, an alternative to Monte Carlo forecasting. *Tellus A*, **35A**, 100–118, doi:10.3402/tellusa.v35i2.11425.
- Hollingsworth, A., and P. Lönnberg, 1986: The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field. *Tellus A*, **38A** (2), 111–136, doi:10.1111/j.1600-0870.1986.tb00460.x.
- Houtekamer, P. L., and H. L. Mitchell, 1998: Data Assimilation Using an Ensemble Kalman Filter Technique. *Monthly Weather Review*, **126** (3), 796–811, doi:10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2.
- Houtekamer, P. L., and H. L. Mitchell, 2001: A Sequential Ensemble Kalman Filter for Atmospheric Data Assimilation. *Monthly Weather Review*, **129** (1), 123–137, doi:10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2.
- Houtekamer, P. L., and F. Zhang, 2016: Review of the Ensemble Kalman Filter for Atmospheric Data Assimilation. *Monthly Weather Review*, **144** (12), 4489–4532, doi:10.1175/MWR-D-15-0440.1.

- Hunt, B. R., E. J. Kostelich, and I. Szunyogh, 2007: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D: Nonlinear Phenomena*, **230** (1-2), 112–126, doi:10.1016/j.physd.2006.11.008.
- Huntley, H. S., and G. J. Hakim, 2010: Assimilation of time-averaged observations in a quasi-geostrophic atmospheric jet model. *Climate Dynamics*, **35** (6), 995–1009, doi:10.1007/s00382-009-0714-5.
- Ide, K., P. Courtier, M. Ghil, and A. C. Lorenc, 1997: Unified Notation for Data Assimilation: Operational, Sequential and Variational. *Journal of the Meteorological Society of Japan. Ser. II*, **75** (1B), 181–189, doi:10.2151/jmsj1965.75.1B.181.
- Ingleby, N. B., 2001: The statistical structure of forecast errors and its representation in The Met. Office Global 3-D Variational Data Assimilation Scheme. *Quarterly Journal of the Royal Meteorological Society*, **127** (571), 209–231, doi:10.1002/qj.49712757112.
- IPCC WG1, 2014: *Climate Change 2013 - The Physical Science Basis*. Cambridge University Press, Cambridge, doi:10.1017/CBO9781107415324.
- Jacob, R. L., 1997: Low Frequency Variability in a Simulated Atmosphere Ocean System. Ph.D. thesis, University of Wisconsin-Madison, 172 pp.
- Janjić, T., and Coauthors, 2018: On the representation error in data assimilation. *Quarterly Journal of the Royal Meteorological Society*, **144** (713), 1257–1278, doi:10.1002/qj.3130.
- Kalman, R. E., 1960: A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, **82** (1), 35–45, doi:10.1115/1.3662552.
- Kalnay, E., 2003: *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, 341 pp., doi:10.1017/CBO9780511802270.
- Kalnay, E., Y. Ota, T. Miyoshi, and J. Liu, 2012: A simpler formulation of forecast sensitivity to observations: application to ensemble Kalman filters. *Tellus A*, **64** (0), 1–9, doi:10.3402/tellusa.v64i0.18462.
- Kang, J.-S., E. Kalnay, J. Liu, I. Fung, T. Miyoshi, and K. Ide, 2011: Variable localization in an ensemble Kalman filter: Application to the carbon cycle data assimilation. *Journal of Geophysical Research*, **116** (D9), D09 110, doi:10.1029/2010JD014673.
- Karspeck, A. R., and Coauthors, 2018: A global coupled ensemble data assimilation system using the Community Earth System Model and the Data Assimilation Research Testbed. *Quarterly Journal of the Royal Meteorological Society*, doi:10.1002/qj.3308.

- Kiehl, J., J. Hack, G. Bonan, B. Boville, B. Briegleb, D. Williamson, and P. Rasch, 1996: Description of the NCAR Community Climate Model (CCM3). *NCAR Technical Note NCAR/TN-420+STR*, 159, doi:10.5065/D6FF3Q99.
- Kingma, D. P., and J. Ba, 2017: Adam: A Method for Stochastic Optimization. URL <http://arxiv.org/abs/1412.6980v9>, 1412.6980v9.
- Kumar, A., L. Zhang, and W. Wang, 2013: Sea Surface Temperature-Precipitation Relationship in Different Reanalyses. *Monthly Weather Review*, **141** (3), 1118–1123, doi:10.1175/MWR-D-12-00214.1.
- Laloyaux, P., M. Balmaseda, D. Dee, K. Mogensen, and P. Janssen, 2016: A coupled data assimilation system for climate reanalysis. *Quarterly Journal of the Royal Meteorological Society*, **142**, 65–78, doi:10.1002/qj.2629.
- Laloyaux, P., S. Frolov, B. Ménétrier, and M. Bonavita, 2018a: Implicit and explicit cross-correlations in coupled data assimilation. *Quarterly Journal of the Royal Meteorological Society*, doi:10.1002/qj.3373.
- Laloyaux, P., and Coauthors, 2018b: CERA-20C: A Coupled Reanalysis of the Twentieth Century. *Journal of Advances in Modeling Earth Systems*, **10** (5), 1172–1195, doi:10.1029/2018MS001273.
- Lea, D. J., I. Mirouze, M. J. Martin, R. R. King, A. Hines, D. Walters, and M. Thurlow, 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, doi:10.1175/MWR-D-15-0174.1.
- Legras, B., and R. Vautard, 1996: A guide to liapunov vectors. *Predictability*, Vol. 1, ECMWF, 135–146, URL <http://gershwin.ens.fr/legras/publis/liapunov.pdf>.
- Lei, L., and J. S. Whitaker, 2015: Model Space Localization Is Not Always Better Than Observation Space Localization for Assimilation of Satellite Radiances. *Monthly Weather Review*, **143** (10), 3948–3955, doi:10.1175/MWR-D-14-00413.1.
- Li, H., E. Kalnay, and T. Miyoshi, 2009: Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter. *Quarterly Journal of the Royal Meteorological Society*, **135**, 523–533, doi:10.1002/qj.371.
- Lin, L.-F., and Z. Pu, 2018: Characteristics of Background Error Covariance of Soil Moisture and Atmospheric States in Strongly Coupled Land-Atmosphere Data Assimilation. *Journal of Applied Meteorology and Climatology*, **57** (11), 2507–2529, doi:10.1175/JAMC-D-18-0050.1.
- Liu, C., Q. Xiao, and B. Wang, 2009: An Ensemble-Based Four-Dimensional Variational Data Assimilation Scheme. Part II: Observing System Simulation Experiments with Advanced Research WRF (ARW). *Monthly Weather Review*, **137** (5), 1687–1704, doi:10.1175/2008MWR2699.1.

- Liu, J., and E. Kalnay, 2008: Estimating observation impact without adjoint model in an ensemble Kalman filter. *Quarterly Journal of the Royal Meteorological Society*, **134** (634), 1327–1335, doi:10.1002/qj.280.
- Liu, Y., Z. Liu, S. Zhang, R. Jacob, F. Lu, X. Rong, and S. Wu, 2014: Ensemble-based parameter estimation in a coupled general circulation model. *Journal of Climate*, **27** (18), 7151–7162, doi:10.1175/JCLI-D-13-00406.1.
- Liu, Y., Z. Liu, S. Zhang, R. Jacob, X. Rong, and F. Lu, 2017: A "online" EnKF-CGCM System. *PSU-UMD DA Workshop*.
- Liu, Z., Y. C. Lai, and M. A. Matías, 2003: Universal scaling of Lyapunov exponents in coupled chaotic oscillators. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, **67** (4), 4, doi:10.1103/PhysRevE.67.045203.
- Lönnberg, P., and A. Hollingsworth, 1986: The statistical structure of short-range forecast errors as determined from radiosonde data Part II: The covariance of height and wind errors. *Tellus A*, **38A** (2), 137–161, doi:10.1111/j.1600-0870.1986.tb00461.x.
- Lorenz, E. N., 1963: Deterministic Nonperiodic Flow. *Journal of the Atmospheric Sciences*, **20**, 130–141, doi:10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2.
- Lorenz, E. N., 2005: Designing chaotic models. *Journal of the Atmospheric Sciences*, **62** (5), 1574–1587, doi:10.1175/JAS3430.1.
- Lu, F., Z. Liu, S. Zhang, and Y. Liu, 2015a: Strongly Coupled Data Assimilation Using Leading Averaged Coupled Covariance (LACC). Part I: Simple Model Study. *Monthly Weather Review*, **143** (9), 3823–3837, doi:10.1175/MWR-D-14-00322.1.
- Lu, F., Z. Liu, S. Zhang, Y. Liu, and R. Jacob, 2015b: Strongly Coupled Data Assimilation Using Leading Averaged Coupled Covariance (LACC). Part II: CGCM Experiments. *Monthly Weather Review*, **143** (11), 4645–4659, doi:10.1175/MWR-D-15-0088.1.
- Ménétrier, B., T. Montmerle, Y. Michel, and L. Berre, 2015a: Linear Filtering of Sample Covariances for Ensemble-Based Data Assimilation. Part I: Optimality Criteria and Application to Variance Filtering and Covariance Localization. *Monthly Weather Review*, **143** (5), 1622–1643, doi:10.1175/mwr-d-14-00157.1.
- Ménétrier, B., T. Montmerle, Y. Michel, and L. Berre, 2015b: Linear Filtering of Sample Covariances for Ensemble-Based Data Assimilation. Part II: Application to a Convective-Scale NWP Model. *Monthly Weather Review*, **143** (5), 1644–1664, doi:10.1175/mwr-d-14-00156.1.
- Miller, A. J., and Coauthors, 2017: Coupled oceanatmosphere modeling and predictions. *Journal of Marine Research*, **75** (3), 361–402, doi:10.1357/002224017821836770.

- Minobe, S., A. Kuwano-Yoshida, N. Komori, S.-p. Xie, and R. J. Small, 2008: Influence of the Gulf Stream on the troposphere. *Nature*, **452** (7184), 206–209, doi:10.1038/nature06690.
- Mitchell, H. L., and P. L. Houtekamer, 2000: An Adaptive Ensemble Kalman Filter. *Monthly Weather Review*, doi:10.1175/1520-0493(2000)128<0416:aaekf>2.0.co;2.
- Miyakoda, K., 1986: Assessment of Results from Different Analysis Schemes. *GARP Publications Series*, **26**, 217–253.
- Miyoshi, T., and K. Kondo, 2013: A Multi-Scale Localization Approach to an Ensemble Kalman filter. *Sola*, **9** (2012), 170–173, doi:10.2151/sola.2013-038.
- Miyoshi, T., K. Kondo, and T. Imamura, 2014: The 10,240-member ensemble Kalman filtering with an intermediate AGCM. *Geophysical Research Letters*, **41** (14), 5264–5271, doi:10.1002/2014GL060863.
- Motesharrei, S., and Coauthors, 2016: Modeling Sustainability: Population, Inequality, Consumption, and Bidirectional Coupling of the Earth and Human Systems. *National Science Review*, **3** (4), nww081, doi:10.1093/nsr/nww081.
- Mulholland, D. P., P. Laloyaux, K. Haines, and M. A. Balmaseda, 2015: Origin and Impact of Initialization Shocks in Coupled Atmosphere-Ocean Forecasts. *Monthly Weather Review*, **143** (11), 4631–4644, doi:10.1175/MWR-D-15-0076.1.
- Nerger, L., 2015: On Serial Observation Processing in Localized Ensemble Kalman Filters. *Monthly Weather Review*, **143** (5), 1554–1567, doi:10.1175/MWR-D-14-00182.1.
- Ng, G.-H. C., D. McLaughlin, D. Entekhabi, and A. Ahanin, 2011: The role of model dynamics in ensemble Kalman filter performance for chaotic systems. *Tellus A*, **63** (5), 958–977, doi:10.1111/j.1600-0870.2011.00539.x.
- Norwood, A., E. Kalnay, K. Ide, S.-C. Yang, and C. Wolfe, 2013: Lyapunov, singular and bred vectors in a multi-scale system: an empirical exploration of vectors related to instabilities. *Journal of Physics A*, **46**, 254021, doi:10.1088/1751-8113/46/25/254021.
- O’Kane, T. J., and Coauthors, 2019: Coupled Data Assimilation and Ensemble Initialization with Application to Multiyear ENSO Prediction. *Journal of Climate*, **32** (4), 997–1024, doi:10.1175/JCLI-D-18-0189.1.
- Oseledets, V. I., 1968: A multiplicative ergodic theorem. Characteristic Ljapunov exponents of dynamical systems. *Transactions of the Moscow Mathematical Society*, **19**, 179–210, URL <http://www.mathnet.ru/eng/mmo214>.
- Ott, E., and Coauthors, 2004: A local ensemble Kalman filter for atmospheric data assimilation. *Tellus A: Dynamic Meteorology and Oceanography*, **56** (5), 415–428, doi:10.3402/tellusa.v56i5.14462.

- Parrish, D. F., and J. C. Derber, 1992: The National Meteorological Center’s Spectral Statistical-Interpolation Analysis System. *Monthly Weather Review*, **120** (8), 1747–1763, doi:10.1175/1520-0493(1992)120<1747:TNMCSS>2.0.CO;2.
- Pecora, L. M., and T. L. Carroll, 1990: Synchronization in chaotic systems. *Physical Review Letters*, **64** (8), 821–824, doi:10.1103/PhysRevLett.64.821.
- Pecora, L. M., T. L. Carroll, G. A. Johnson, D. J. Mar, and J. F. Heagy, 1997: Fundamentals of synchronization in chaotic systems, concepts, and applications. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, **7** (4), 520–543, doi:10.1063/1.166278.
- Peña, M., and E. Kalnay, 2004: Separating fast and slow modes in coupled chaotic systems. *Nonlinear Processes in Geophysics*, **11** (3), 319–327, doi:10.5194/npg-11-319-2004.
- Penny, S., E. Bach, K. Bhargava, C.-C. Chang, C. Da, L. Sun, and T. Yoshida, 2019: Strongly coupled data assimilation in multiscale media: experiments using a quasi-geostrophic coupled model. *Journal of Advances in Modeling Earth Systems*, 2019MS001652, doi:10.1029/2019MS001652.
- Penny, S. G., 2017: Mathematical foundations of hybrid data assimilation from a synchronization perspective. *Chaos*, **27** (12), doi:10.1063/1.5001819.
- Penny, S. G., and T. M. Hamill, 2017: Coupled Data Assimilation for Integrated Earth System Analysis and Prediction. *Bulletin of the American Meteorological Society*, **98** (7), ES169–ES172, doi:10.1175/BAMS-D-17-0036.1.
- Penny, S. G., and T. Miyoshi, 2015: A local particle filter for high dimensional geophysical systems. *Nonlinear Processes in Geophysics Discussions*, **2** (6), 1631–1658, doi:10.5194/npgd-2-1631-2015.
- Penny, S. G., and Coauthors, 2017: Coupled Data Assimilation for Integrated Earth System Analysis and Prediction: Goals, Challenges and Recommendations. *World Meteorological Organization*, **WWRP 2017-3**, 50, URL https://www.wmo.int/pages/prog/arep/wwrp/new/documents/Final_WWRP_2017_3_27_July.pdf.
- Petrie, R. E., and S. L. Dance, 2010: Ensemble-based data assimilation and the localisation problem. *Weather*, **65** (3), 65–69, doi:10.1002/wea.505.
- Pitman, E. J. G., 1937: Significance Tests Which May be Applied to Samples from any Populations. II. The Correlation Coefficient Test. *Journal of the Royal Statistical Society*, **4** (2), 225–232.
- Poterjoy, J., 2016: A Localized Particle Filter for High-Dimensional Nonlinear Systems. *Monthly Weather Review*, **144** (1), 59–76, doi:10.1175/MWR-D-15-0163.1.

- Purser, R. J., W.-S. Wu, D. F. Parrish, and N. M. Roberts, 2003: Numerical Aspects of the Application of Recursive Filters to Variational Statistical Analysis. Part I: Spatially Homogeneous and Isotropic Gaussian Covariances. *Monthly Weather Review*, **131**, 1524–1535, doi:10.1175//1520-0493(2003)131<1524:NAOTAO>2.0.CO;2.
- Rosenblum, M. G., A. S. Pikovsky, and J. Kurths, 1996: Phase Synchronization of Chaotic Oscillators. *Physical Review Letters*, **76** (11), 1804–1807, doi:10.1103/PhysRevLett.76.1804.
- Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bulletin of the American Meteorological Society*, **91** (8), 1015–1057, doi:10.1175/2010BAMS3001.1.
- Saha, S., and Coauthors, 2014: The NCEP Climate Forecast System Version 2. *Journal of Climate*, **27** (6), 2185–2208, doi:10.1175/JCLI-D-12-00823.1.
- Sakov, P., and P. R. Oke, 2008: Implications of the Form of the Ensemble Transformation in the Ensemble Square Root Filters. *Monthly Weather Review*, **136**, 1042–1053, doi:10.1175/2007MWR2021.1.
- Schepers, D., E. de Boissésou, R. Eresmaa, C. Lupu, and P. de Rosnay, 2018: CERA-SAT: A coupled satellite-era reanalysis. URL <https://www.ecmwf.int/en/newsletter/155/meteorology/cera-sat-coupled-satellite-era-reanalysis>.
- Shen, Z., Y. Tang, X. Li, Y. Gao, and J. Li, 2018: On the localization in strongly coupled ensemble data assimilation using a two-scale Lorenz model. *Nonlinear Processes in Geophysics Discussions*, 1–24, doi:10.5194/npg-2018-50.
- Shlyayeva, A., J. Whitaker, and C. Snyder, 2018: Model space localization in serial ensemble filters. URL http://web.meteo.mcgill.ca/enkf/pr_files/EnSRFmodloc_Shlyayeva.pdf.
- Shukla, J., 1998: Predictability in the Midst of Chaos: A Scientific Basis for Climate Forecasting. *Science*, **282** (5389), 728–731, doi:10.1126/science.282.5389.728.
- Singleton, T., 2011: Data Assimilation Experiments with a Simple Coupled Ocean-Atmosphere Model. Ph.D. thesis, University of Maryland, College Park, 116 pp.
- Sluka, T. C., 2018: Strongly Coupled Ocean-Atmosphere Data Assimilation with the Local Ensemble Transform Kalman Filter. Ph.D. thesis, University of Maryland, College Park.
- Sluka, T. C., S. G. Penny, E. Kalnay, and T. Miyoshi, 2016: Assimilating atmospheric observations into the ocean using strongly coupled ensemble data assimilation. *Geophysical Research Letters*, **43** (2), 752–759, doi:10.1002/2015GL067238.

- Smith, P. J., A. S. Lawless, and N. K. Nichols, 2017: Estimating Forecast Error Covariances for Strongly Coupled Atmosphere-Ocean 4D-Var Data Assimilation. *Monthly Weather Review*, **145** (10), 4011–4035, doi:10.1175/MWR-D-16-0284.1.
- Smith, P. J., A. S. Lawless, and N. K. Nichols, 2018: Treating Sample Covariances for Use in Strongly Coupled Atmosphere-Ocean Data Assimilation. *Geophysical Research Letters*, **45** (1), 445–454, doi:10.1002/2017GL075534.
- Sprott, J. C., 1997: Lyapunov Exponent and Dimension of the Lorenz Attractor. URL <http://sprott.physics.wisc.edu/chaos/lorenzle.htm>.
- Stammer, D., K. Ueyoshi, A. Kohl, W. G. Large, S. A. Josey, and C. Wunsch, 2004: Estimating air-sea fluxes of heat, freshwater, and momentum through global ocean data assimilation. *Journal of Geophysical Research*, **109** (C5), C05 023, doi:10.1029/2003JC002082.
- Storto, A., M. J. Martin, B. Deremble, and S. Masina, 2018: Strongly coupled data assimilation experiments with linearized ocean-atmosphere balance relationships. *Monthly Weather Review*, MWR–D–17–0222.1, doi:10.1175/MWR-D-17-0222.1.
- Strogatz, S. H., 2015: *Nonlinear Dynamics and Chaos*. 2nd ed., Westview Press.
- Sugiura, N., T. Awaji, S. Masuda, T. Mochizuki, T. Toyoda, T. Miyama, H. Igarashi, and Y. Ishikawa, 2008: Development of a four-dimensional variational coupled data assimilation system for enhanced analysis and prediction of seasonal to inter-annual climate variations. *Journal of Geophysical Research*, **113** (C10), C10 017, doi:10.1029/2008JC004741.
- Szunyogh, I., E. J. Kostelich, G. Gyarmati, E. Kalnay, B. R. Hunt, E. Ott, E. Satterfield, and J. A. Yorke, 2008: A local ensemble transform Kalman filter data assimilation system for the NCEP global model. *Tellus, Series A: Dynamic Meteorology and Oceanography*, **60 A** (1), 113–130, doi:10.1111/j.1600-0870.2007.00274.x.
- Tardif, R., G. J. Hakim, and C. Snyder, 2014: Coupled atmosphereocean data assimilation experiments with a low-order climate model. *Climate Dynamics*, **43**, 1631–1643, doi:10.1007/s00382-013-1989-0.
- Tardif, R., G. J. Hakim, and C. Snyder, 2015: Coupled atmosphereocean data assimilation experiments with a low-order model and CMIP5 model data. *Climate Dynamics*, **45** (5-6), 1415–1427, doi:10.1007/s00382-014-2390-3.
- Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003: Ensemble Square Root Filters. *Monthly Weather Review*, **131**, 1485–1490, doi:10.1175/1520-0493(2003)131<1485:ESRF>2.0.CO;2.
- Trevisan, A., and L. Palatella, 2011: On the Kalman Filter error covariance collapse into the unstable subspace. *Nonlinear Process Geophys.*, **18**, 243–250, doi:10.5194/npg-18-243-2011.

- Vannitsem, S., and V. Lucarini, 2016: Statistical and dynamical properties of covariant Lyapunov vectors in a coupled atmosphere-ocean model multiscale effects, geometric degeneracy, and error dynamics. *Journal of Physics A: Mathematical and Theoretical*, **49** (22), 224001, doi:10.1088/1751-8113/49/22/224001.
- Wallace, J. M., and P. V. Hobbs, 2006: *Atmospheric Science An Introductory Survey*. 2nd ed., Academic Press, Elsevier, 483 pp.
- Wang, X., and C. H. Bishop, 2003: A Comparison of Breeding and Ensemble Transform Kalman Filter Ensemble Forecast Schemes. *Journal of the Atmospheric Sciences*, **60**, 1140–1158, doi:10.1175/1520-0469(2003)060<1140:ACOBAE>2.0.CO;2.
- Waters, J., D. J. Lea, M. J. Martin, I. Mirouze, A. Weaver, and J. While, 2015: Implementing a variational data assimilation system in an operational 1/4 degree global ocean model. *Quarterly Journal of the Royal Meteorological Society*, **141** (687), 333–349, doi:10.1002/qj.2388.
- Weaver, A., M. Chrust, B. Ménétrier, A. Piacentini, and Y. Yang, 2017: Ensemble-variational assimilation with NEMOVAR: Part 1: design aspects and illustrations. URL https://cerfacs.fr/wp-content/uploads/2017/05/Weaver_AVENUE_0617.pdf.
- Whitaker, J. S., and T. M. Hamill, 2002: Ensemble Data Assimilation without Perturbed Observations. *Monthly Weather Review*, **130** (7), 1913–1924, doi:10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2.
- Whitaker, J. S., and T. M. Hamill, 2012: Evaluating Methods to Account for System Errors in Ensemble Data Assimilation. *Monthly Weather Review*, **140** (9), 3078–3089, doi:10.1175/MWR-D-11-00276.1.
- Yan, Y., A. Barth, and J. M. Beckers, 2014: Comparison of different assimilation schemes in a sequential Kalman filter assimilation system. *Ocean Modelling*, **73**, 123–137, doi:10.1016/j.ocemod.2013.11.002.
- Yang, S.-C., and Coauthors, 2006: Data Assimilation as Synchronization of Truth and Model: Experiments with the Three-Variable Lorenz System. *Journal of the Atmospheric Sciences*, **63**, 2340–2354, doi:10.1175/JAS3739.1.
- Ying, Y., F. Zhang, and J. L. Anderson, 2018: On the Selection of Localization Radius in Ensemble Filtering for Multiscale Quasigeostrophic Dynamics. *Monthly Weather Review*, **146** (2), 543–560, doi:10.1175/MWR-D-17-0336.1.
- Yokota, S., M. Kunii, K. Aonashi, and S. Origuchi, 2016: Comparison between Four-Dimensional LETKF and Ensemble-Based Variational Data Assimilation with Observation Localization. *SOLA*, **12** (3), 80–85, doi:10.2151/sola.2016-019.
- Yoshida, T., and E. Kalnay, 2018: Correlation-Cutoff Method for Covariance Localization in Strongly Coupled Data Assimilation. *Monthly Weather Review*, **146** (9), 2881–2889, doi:10.1175/MWR-D-17-0365.1.

- Zhang, F., C. Snyder, and J. Sun, 2004: Impacts of Initial Estimate and Observation Availability on Convective-Scale Data Assimilation with an Ensemble Kalman Filter. *Monthly Weather Review*, **132** (5), 1238–1253, doi:10.1175/1520-0493(2004)132<1238:IOIEAO>2.0.CO;2.
- Zhang, S., M. J. Harrison, A. Rosati, and A. Wittenberg, 2007: System Design and Evaluation of Coupled Ensemble Data Assimilation for Global Oceanic Climate Studies. *Monthly Weather Review*, **135** (10), 3541–3564, doi:10.1175/MWR3466.1.
- Zhang, S., Z. Liu, A. Rosati, and T. Delworth, 2012: A study of enhance parameter correction with coupled data assimilation for climate estimation and prediction using a simple coupled model. *Tellus A*, **64** (1), 1–20, doi:10.3402/tellusa.v64i0.10963.