BRIDGING THE MODEL-DATA GAP IN PALEOCLIMATOLOGY : DATA ASSIMILATION AND NOISE-REMOVAL FOR MARINE PROXIES

by

Huaran Liu

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

(Atmospheric and Oceanic Sciences)

at the

UNIVERSITY OF WISCONSIN–MADISON

2019

Date of final oral examination: 05/13/19

The dissertation is approved by the following members of the Final Oral Committee:
    Shaun A. Marcott, Assistant Professor, Geoscience
    Stephen R. Meyers, Professor, Geoscience
    Larissa E. Back, Associate professor, Atmospheric and Oceanic Sciences
    Michael C. Morgan, Professor, Atmospheric and Oceanic Sciences
    Matthew H. Hitchman, Professor, Atmospheric and Oceanic Sciences

For my husband Zach,
When we met my life began, and shall continue to our very last breath

Acknowledgments

The journey to accomplish this dissertation is filled with ups and downs. It would not have been possible without all the support from my family and friends, my advisors and my committee, and the Atmospheric and Oceanic Science department.

First, I would like to thank my husband Zach for all the love, kindness, patience and unconditional support you have given me both in my research and life. I feel so blessed that the grad school brought us together and that we are able to accomplish this journey and witness this important moment of life together.

I would like to thank my advisors Shaun and Steve. Thank you both for taking me as your student and providing tremendous support and encouragement to pass through the hardest transition of my PhD career. I appreciate all the guidance and expertise you gave me to become a good geologist. I value all the passion and inspiration you share with me to be a better scientist. I am grateful for all the patience you have shown me in our every scientific discussion, I know my stubborn personality can make things "challenging" sometimes.

I am also grateful for my committee members, Michael Morgan, Larissa Back, and Matthew Hitchman. For all the insightful feedbacks and advice they gave me. They always push me to think harder and keep me on track of doing science when I get lost in the technical aspect of research process.

I would like to thank Feiyu Lu, who serves as an important mentor in the early stage of my PhD careers. Thank you very much for all the patience and expertise

Contents

List of Tables

List of Figures

Abstract

Paleoclimate reconstruction relies on both unbiased proxy observations and competent climate models, as well as appropriate data assimilation methods that utilize both the model and observation. The objective of this dissertation is to help mitigate the model-data gap in paleoclimate and provide better state estimation of the past climate, by addressing the issue of post-depositional mixing on marine proxy observations, and through cross-validation of two popular data assimilation methods for paleoclimate reconstruction.

Chapter 3 seeks to solve for the best data assimilation method for paleoclimate application that has high skill reconstructing both the local temperature and regionally averaged temperature given a very sparse observation network and limited computation power. It is found that the Ensemble Kalman Filter (EAKF) method outperforms the particle filter (PF-SIR) method with only one third of the computation cost. The key to the success of EAKF is a large localization radius and availability of observations in regions that have large-scale co-variability.

Chapter 4 comes from the data perspective: how do we remove the post-depositional mixing effects, namely bioturbation, from the existing marine climate proxy records? What is the actual amplitude and phase of events recorded in climate proxy time series? The bioturbation process is modeled as a linear time-invariant filter. We found that given the modeled mixing layer depths ≥ 10cm, it is impossible to preserve the centennial scale variability even when the sedimentation rate is

above $15cm/kyr$. For ice ages of scale 10^4 - 10^5 years, bioturbation effects are almost negligible considering the event scale. For millennial scale events, the signal attenuation depends on the event scale and specific bioturbation parameters. To remove the bioturbation effect on a given climate proxy series, a deconvolution method is proposed and applied to two benthic foraminifera oxygen isotope records in the LR04 collection for the last interglacial (MIS 5e). We found that the amplitude of the recovered millennial scale variability for the global stack can be approximately 0.5 - 1 ‰larger when the mixing layer depth is 10cm, with implications for reconstructed global average sea level variation and temperature estimates from benthic foramnifera $\delta^{18}O$. The recovered signal at individual site GeoB1117 from the LR04 collection is more volatile compared to that from the global stack, potentially due to higher noise level in a single record. The lower bound of the recovered $\delta^{18}O$ is about 1‰less compared to the value in LR04.

As an independent project from Chapter 3 and 4, Chapter 2 evaluates the quality of ocean state estimation in a coupled data assimilation system when the atmospheric observations are replaced with atmospheric reanalysis. The experiments are conducted in a coupled Lorenz 96 model to mimic a coupled general circulation model (CGCM). Pseudo observations and atmospheric reanalysis are generated from this simple model and four different assimilation schemes are designed. It is found that when the atmospheric observations are replaced with atmospheric reanalysis in setting up a coupled data assimilation system, the ocean analysis quality (RMSE) is degraded by about 16% when there is no model bias and is less than 22% when model bias exists. Different assimilation schemes highlight the

importance of 1) accurate representation of the error covariance of the reanalysis and 2) the temporal coherence along each ensemble member.

Chapter 1

Introduction

1.1 Background

Paleoclimatology provides a perspective for understanding the present-day climate and provides implications and projections for future climate change. Climates of the past either can be 1) inferred from proxy records (Mann and Hughes, 1998; Birdsey and Service, 1998; Adams et al., 2003; Bradley et al., 2005; Mann et al., 2008a,b, 2009; Jones et al., 2009; Shakun et al., 2012; Marcott et al., 2013; Emile-Geay et al., 2013; Dowsett et al., 2016), 2) simulated by state-of-art climate models (Otto-Bliesner et al., 2006; Liu et al., 2009; Otto-Bliesner et al., 2009; Braconnot et al., 2012; Liu et al., 2014; Otto-Bliesner et al., 2017; Zhu et al., 2017b,a), and 3) reconstructed by merging the proxy observations with climate models (Crespin et al., 2009; Goosse et al., 2009, 2010; Widmann et al., 2010; Steiger N., 2013; Hakim et al., 2016; Steiger and Hakim, 2016; Steiger et al., 2017). Each of the three approaches provides useful insights but with its inherent limitations.

Climate proxy records, acting as an objective recorder of past climate, are reconstructed from geological deposits, biological archives and/or chemical compositions. Although often faithful, they are not a direct measurement of the state variables (Shackleton, 1974, 1987; Mix et al., 2000; Ohkouchi et al., 2008; Schneider

et al., 2010; Arbuszewski et al., 2010; Leduc et al., 2010; Laepple and Huybers, 2013), and are generally subject to bias, subjective interpretations and post-depositional distortion. Marine proxy records consist of more than half of the paleoclimate records especially beyond last 10,000 years (Lisiecki and Raymo, 2005; Shakun et al., 2012; Marcott et al., 2013). In addition to the potential bias due to the biological activities of the proxy carriers and mixed signals from multiple climate variables, they are also subject to post-depositional mixing, so-called bioturbation, which acts as a low-pass filter and smooths out the input signal (Goldberg and Koid, 1962; Berger and Heath, 1968; Ruddiman and Glover, 1972; Guinasso and Schink, 1975; Ruddiman et al., 1980a,b; Schiffelbein, 1984a; Bard et al., 1987; Anderson, 2001b; Trauth, 2013). This distortion of proxy time series has long been recognized but seldom addressed when climate inferences or model data comparisons are made from the raw proxy records (Shackleton et al., 2003; Rohling et al., 2008; Turney and Jones, 2010; Dutton et al., 2015a). This can lead to incorrectly dated and significantly underestimated historical events, can exacerbate model-data discrepancy, and can result in misinterpreted physical process.

Model simulations are an "educated guess" of the past climate based on modern knowledge of earth physics and reconstructed historical boundary conditions (Harrison et al., 2016). Although they provide temporally and spatially coherent products that are consistent with model physics, any missing or incorrectly-modeled physical process can lead the simulations away from reality (Braconnot et al., 2012; Perez-Sanz et al., 2014; Otto-Bliesner et al., 2017; Thompson et al., 2019). For example, although the models can reproduce the large-scale feature of the intensified

and expanded Afro-Asian summer monsoon system in the mid-Holocene (6 kyr ago), the simulated monsoon is weaker and has less northward extension than the paleoclimate data suggests (Otto-Bliesner et al., 2006; Harrison et al., 2014). Additional physical processes and mechanisms need to be invoked to simulate the smaller-scale characteristics (Perez-Sanz et al., 2014; Thompson et al., 2019).

The third approach, data assimilation (DA), provides a promising reconstruction dataset that is consistent with both model physics and proxy observations. However, due to the nature of the proxy observations (such as high noise, sparse coverage, and extremely low temporal resolution), DA techniques developed for meteorology are not directly applicable for paleoclimate application. Various methods have been proposed ranging from simple nudging (von Storch and Zorita, 2000), forcing singular vectors (J. Barkmeijer, 2003), to ensemble-based filters (Leeuwen, 2009; Dubinkina et al., 2011; Goosse et al., 2011; Annan and Hargreaves, 2012; Bhend et al., 2012; Steiger N., 2013; Hakim et al., 2016; Steiger and Hakim, 2016; Steiger et al., 2017), but no universal method has been established.

With the objective of better state estimation for paleoclimate, this dissertation addresses two pressing issues facing the paleoclimate community with regard to approaches 1) and 3): The distortion of climate signals caused by post-depositional mixing in marine sediment records, and a suitable data assimilation methodology for paleoclimate state reconstruction.

Chapter 4 presents modeling studies on quantifying bioturbation effects on climate signals recorded in deep sea marine sediments, specifically, addressing the question what is the likely amplitude for $\delta^{18}O$ during the last interglacial period

(125 kyr ago), after the bioturbation effect is removed? And what is the implication for sea level changes and temperatures inferred from $\delta^{18}O$? An inverse modeling method using deconvolution through Fast Fourier Transform (FFT) is proposed to remove bioturbation effects from proxy time series. Forward modeling using convolution is also conducted on artificial time series to evaluate the bioturbation effects across different event scales.

Chapter 3 compares the performance of the two most popular data assimilation methods applied to paleoclimate: particle filter with simple importance sampling (PF-SIR) and Ensemble Kalman filter (EnKF). The comparison is performed for both spatial and temporal construction skills as well as computation efficiency. Both filters are adapted to time-averaged observations and implemented in a Coupled General Circulation Model (CGCM) of intermediate complexity to explore its ability of climate state estimation given the same observation network. We found that EnKF is much more efficient than PF-SIR in terms of both performance and computational cost.

Chapter 2 is an independent project from Chapter 3 and 4. It studies the state estimation for the ocean variables when oceanic observations and atmospheric reanalysis, instead of atmospheric observations, are assimilated into a coupled data assimilation (CDA) system. The question addressed is: what is the error introduced into the ocean analysis when atmospheric observations are replaced with atmospheric reanalysis in a CDA system? The motivation for this study is to provide justification for substituting real atmospheric observations with atmospheric reanalysis when setting up a CDA system to generate ocean analyses. If

the error introduction is acceptable (< 20%), it will avoid the tremendous work of gathering, performing quality control and assimilating all types of real atmospheric observations.

1.2   Data assimilation for state estimation

Data assimilation in meteorology combines instrumental observations and prior model simulations to generate a product that has the full temporal and spatial coverage as the model simulations and at the same time, is constrained by the observations. This product is the "best estimate" of the state, and is used to initialize numerical weather prediction (NWP) models to produce accurate forecasts. The prior model forecasts are usually called *background state* and the data assimilation product *analysis state*. In addition to the initialization of NWP models, the analysis is also the main product for data analysis in the weather and climate science community, for example, NCEP/NCAR reanalysis, which is also the main purpose of introducing data assimilation techniques from meteorology into paleoclimatology: to generate a product that is coherent with both historical proxy records and model simulations, and represents the "best estimate" of the climate state in different geological time periods. In NWP, the main data assimilation techniques fall into two categories: variational methods and Ensemble Kalman Filter (EnKF) with its variants. Variational methods such as 3DVAR and 4DVAR solve for the optimum analysis state by minimizing a cost function. They have been widely implemented in various NWP centers with substantial improvement in forecasts. EnKF is a

Monte Carlo application of the original Kalman filter (Kalman, 1960; Kalman and Bucy, 1961). An ensemble of states are simulated by the models to estimate the *background* error covariance matrix, and the resultant analysis ensemble can be used to initialize a global and regional ensemble prediction system(Houtekamer et al., 2005; Lavaysse et al., 2012). In this dissertation, EnKF with its variants are used and adopted for paleoclimatology state estimation. To achieve the best estimate of the state, data assimilation solves the conditional *probability density function* (PDF) of the state given observations Y:

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)} \tag{1.1}$$

$$\propto P(Y|X)P(X) \tag{1.2}$$

where X represents the current state and $P(X)$ represents the prior knowledge of the state provided by model forecast (background state). Y represents the current observation with PDF $P(Y)$. $P(Y|X)$ is the likelihood of obtaining such observations Y given the prior knowledge X. $P(Y)$ is the normalization constant. The posterior $P(X|Y)$ can be continuously updated as future observations are available.

A complete data assimilation cycle consists of two steps: the forecast cycle and the analysis cycle. In the forecast cycle, the analysis state from the previous the analysis cycle is used as the initial condition and the model is forwarded in time until the next observations are available. In the analysis cycle, the observations are assimilated into the model forecasts to generate the analysis state, which then

initializes the next forecast cycle. In NWP, the forecast cycle is much shorter than the forecastibility of the system: the observations come in frequently enough (6 hours) before the error saturates in the forecast cycle. In paleoclimate, the temporal resolutions of proxy records often range from seasonal to millennial. Annual resolution is considered to be high-resolution in a proxy record. In addition, these proxy observations represent a mean state during a certain period of time instead of instantaneous values of the state variables, for example, coral data represent the annual mean sea water temperature that the coral lived in. This requires a long integration of the forward model during the forecast cycle and demands huge computation power. Furthermore, unlike modern instrumental observations that have a dense network and provide a full coverage of the globe, the paleoclimate observation network is extremely sparse and heavily concentrated in the Northern Hemisphere. This low-frequency, small-coverage of the proxy records pose substantial challenges in constraining the past climate. It is not straightforward to apply data assimilation methods in meteorology to paleoclimate state estimation.

1.2.1 Theory of ensemble-based Kalman Filters

Ensemble Kalman Filter (EnKF) uses an ensemble of model forecasts to represent the PDF of the state. It is a Monte Carlo application of the original Kalman Filter. It assumes that with an *unbiased* model, the PDF of the state and observations follow

a Gaussian distribution:

$$P(X) \;=\; \frac{1}{\sqrt{2\pi}\sigma_b}e^{-\frac{(x^b-x)^2}{2\sigma_b^2}} \;\sim\; N(x,\sigma_b^2) \tag{1.3}$$

$$P(Y|X) \;=\; \frac{1}{\sqrt{2\pi}\sigma_o}e^{-\frac{(y-x)^2}{2\sigma_o^2}} \;\sim\; N(y,\sigma_o^2) \tag{1.4}$$

σ_b^2 and σ_o^2 are the variance of the background/prior and the observation respectively. The multivariate counterpart of σ_b^2 is the background error covariance matrix P^b. In EnKF, P^b is estimated through the sample covariance matrix of the ensemble. In the original Kalman Filter, P^b is estimated by advancing the previous analysis error covariance matrix P^a through a tangent linear model . And due to this requirement, the original Kalman Filter is valid only for linear systems. With EnKF, nonlinear models are allowed since P^b can be approximated from the sample covariance of the ensemble and there is no need to develop the tangent linear models. With the Gaussian assumption, the posterior distribution is:

$$P(X|Y) \;\propto\; P(Y|X)P(X) \tag{1.5}$$

$$\propto\; \frac{1}{\sqrt{2\pi}\sigma_b}e^{-\frac{(x^b-x)^2}{2\sigma_b^2}} \times \frac{1}{\sqrt{2\pi}\sigma_o}e^{-\frac{(y-x)^2}{2\sigma_o^2}} \tag{1.6}$$

$$\propto\; \frac{1}{\sqrt{2\pi}\sigma_b}e^{-\frac{(x^b-x)^2}{2\sigma_b^2}-\frac{(y-x)^2}{2\sigma_b^2}} \tag{1.7}$$

To solve for the posterior state that has the highest probability given the model forecasts and observations (maximize $P(X|Y)$), we have to find the solution that

minimizes the cost function

$$J(x) = \frac{(x^b - x)^2}{2\sigma_b^2} + \frac{(y - x)^2}{2\sigma_b^2} \tag{1.8}$$

For the scaler case, the solution can be readily solved:

$$X_a = \frac{\sigma_o^2}{\sigma_b^2 + \sigma_o^2} X_b + \frac{\sigma_b^2}{\sigma_b^2 + \sigma_o^2} Y \tag{1.9}$$

$$= X_b + \frac{\sigma_b^2}{\sigma_b^2 + \sigma_o^2}(Y - X_b) \tag{1.10}$$

$$= X_b + K(Y - X_b) \tag{1.11}$$

where $K = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_o^2}$ is called *Kalman Gain* and the observation increment $Y - X_b$ is called the *innovation*. The analysis error is given by

$$\sigma_a^2 = (1 - K)\sigma_b^2 \tag{1.12}$$

For the multivariate case, the update equation 1.11-1.12 will become:

$$X_a = X_b + K \times (Y - \mathcal{H}X_b) \tag{1.13}$$

$$K = \frac{P^b \mathcal{H}^T}{\mathcal{H}P^b \mathcal{H}^T + R} \tag{1.14}$$

$$P^a = (1 - K\mathcal{H})P^b \tag{1.15}$$

where $\mathcal{H}$ is an observation operator that projects the model space into the observation space.

1.2.2 Stochastic and deterministic filters

The ensemble-based Kalman Filters use an ensemble to represent the uncertainty of the state. Based on how the ensemble is updated to acquire the analysis ensemble, these filters fall into two categories: stochastic filter and deterministic filters. In the stochastic filters (EnKF), the observation is treated as a random variable with Gaussian uncertainty and is perturbed to form an observation ensemble. The ith member of the background ensemble x_b^i is updated with its corresponding ith member of the perturbed observation. So the equation 1.13 will be written as:

$$x_a^i = x_b^i + K(y^i - \mathcal{H}x_b^i), i = 1, 2, ..., N \tag{1.16}$$

The ensemble-based filter used in Chapter 2 is a stochastic filter. Because of the perturbation of the observation, extra noise is introduced into the system that can lead to degradation of the analysis quality, hence the deterministic filters are proposed, which does not require the perturbation of the observation. In the deterministic filters, the ensemble mean is updated similarly as in equation 1.13

$$\overline{x_a} = \overline{x_b} + K(y - \mathcal{H}\overline{x_b}) \tag{1.17}$$

$$= P_a(P_b^{-1}\overline{x_b} + \mathcal{H}^T R^{-1} y) \tag{1.18}$$

The ensemble x_b^i is updated by fitting the PDF of the ensemble analysis deterministically to match the analysis error predicted by equation 1.15. Different deterministic filters differ in details depending upon how the analysis ensemble is generated. In

Chapter 3 of this dissertation, Ensemble Adjustment Filter (EAKF) is implemented which is a variant of Ensemble Square-Root Filter (EnSRF). In EAKF and EnSRF, the analysis ensemble is obtained by

$$x_a^i = A(x_b^i - \overline{x_b}) + \overline{x_a}, i = 1, 2, ..., N \tag{1.19}$$

The matrix A is calculated so that the sample covariance of the analysis ensemble x_a is identical to equation 1.15. To update the nearby points, a local least square framework is adopted for sequential update (Andersen 2004).

1.2.3   Assimilation of time-averaged observations

As mentioned previously, the climate proxy record type of observations used in paleoclimatology represents time-averaged information instead of instantaneous values. Hence given an observation y that represents τ time-averaged values, the corresponding background ensemble X_b should also be τ time-averaged. To adapt EAKF to paleoclimatology applications, the algorithm in Dirren and Hakim (2005) is implemented. Before the update of equation 1.13-1.15, the background forecasts are time-averaged based on the observation time-scale. For example, if the observation represents the annual mean temperature of year 1800, the model simulations of year 1800 are averaged across all the time steps. The deviations of the forecasts' ensemble at each time step from its time-averaged values are also saved. After the time-averaged quantities are updated, the instantaneous quantities at each time step will be updated by adding the deviations saved previously back to the

updated time-averaged values. The model is then forwarded into the next cycle. The mathematical equivalence of EnKF applied to time-averaged and instantaneous quantities is provided by Huntley and Hakim (2010).

1.3   Bioturbation modeling

Bioturbation refers to biologically-mediated sediment reworking after sediment is deposited onto the sea floor. The re-working process includes both physical mixing process caused by burrowing activities of infauna, and the ingestion-egestion of sediment grains by infauna. It is typically difficult to separate such biological processes from other sea floor mixing mechanisms, such as those due to currents at the sea floor; in this study, as in most deep marine sediment mixing studies (Boudreau, 1986a,b; Boudreau and Imboden, 1987; Teal et al., 2008) where currents are relatively weak, we refer to all observed mixing as ?bioturbation?. The proxy carriers in the sediment are mixed both upward and downward relative the original depth of deposition. The mixing acts as a smoothing filter that damps the amplitude of the input signal series and shifts its phase. The motivation for modeling bioturbation is to quantify its mixing effects on the time series of climate signals preserved in the sediment.

Various approaches have been proposed to model bioturbation. A large group of bioturbation models specialize on the microscopic scale: an individual sediment particle or proxy carrier is simulated, then displaced downcore with a specific sedimentation rate and are shuffled with a stochastic permutations (Choi et al.,

2002; Trauth, 2013) or based on a stochastic probability matrix (Foster, 1983; Shull, 2001) or modeled as a random walk process (Hull et al., 2011; Turner et al., 2017). The other major category treats the bioturbation as a deterministic system that can be characterized by an impulse response function (IRF) (Ruddiman and Glover, 1972; Guinasso and Schink, 1975; Goreau, 1977; Hutson, 1980; Ruddiman et al., 1980a; Schiffelbein, 1984b,a; Bard et al., 1987; Anderson, 2001b). Climate signal deposition into the sediment is equivalent to passing through a filter, and the bioturbation process is hence a convolution between the input signal and the IRF of the system. In this category, IRF can be determined in multiple ways: it can be determined using an impulsive geological deposit, like a volcanic ash layer, where the vertical distribution along the core captures the deep sea sediment mixing process (Ruddiman and Glover, 1972; Ruddiman et al., 1980a; Anderson, 2001b). Alternatively, it can be derived from mathematical equations with certain boundary conditions that are developed to model the system (Guinasso and Schink, 1975; Bard et al., 1987). The majority of the models in this approach simulate the bioturbation as a diffusive process(Guinasso and Schink, 1975; Schiffelbein, 1984b,a). The mixing is then represented as local molecular diffusion and can be parameterized in terms of biodiffusion intensity, mixed layer depth and sedimentation rate.

Compared to the deterministic approach, the models with a stochastic component allow non-local mixing and yields more realistic simulations, which is beneficial in terms of forward modeling. However, the inverse process is not straightforward given the presence of the stochastic component. Therefore the deterministic approach using a diffusive approach to model bioturbation is adopted

in Chapter 4, to develop a bioturbation-removal algorithm. It utilizes FFT deconvolution in the frequency domain. The appendix gives more details about how to apply FFT deconvolution in our paleoclimate application.

1.4   Thesis organization

The thesis consists of three independent projects, each of which aims at providing better state estimation of the climate system. Chapter 2 tests the idea of substituting the atmospheric observations with atmospheric reanalysis when setting up a coupled data assimilation system for an oceanic analysis product, and quantifies the resultant error introduced via various assimilation schemes. Chapter 3 evaluates the performance between EAKF and particle filter approaches for paleoclimate state estimation. Chapter 4 focuses solely on the proxy observation perspective, and presents the modeling studies on bioturbation.

Chapter 2

Assimilating Atmosphere Reanalysis in Coupled Data Assimilation

2.1   Introduction

Coupled data assimilation (CDA) uses a coupled model to extract information from observations that are available in one or more media, and produces continuous time series of the climate states. Compared to single component assimilation, CDA incorporates the full impact of observations across the air-sea interface and allows the covariability of the atmospheric and oceanic states, and thus it can provide consistent state estimation of the coupled system for further study of the climate variability and the initialization of coupled general circulation models (CGCM) (Kitoh and Arakawa, 1999; Arakawa, 2004; Zhang et al., 2007, 2005; Luo et al., 2008; Sugiura et al., 2008; Zhang, 2011; Liu et al., 2013; Tardif et al., 2015). Despite the huge benefits and demand for CDA, the implementation of CDA has both theoretical and technical challenges, for example, the estimation of the coupled error covariance matrix (Han et al., 2013; Lu et al., 2015) and the huge computational costs of CDA

---

[1]Material in this chapter is a slightly modified version of: Liu, H., and F. Lu, Z. Liu, Y. Liu, S. Zhang 2016: Assimilating Atmosphere Reanalysis in Coupled Data Assimilation *JOURNAL OF METEOROLOGICAL RESEARCH*, **42**, Volume 30, Issue 4, pp 572?583

experiments in fully-coupled model Climate Forecast System Reanalysis (CFSR) was completed for the period 1979-2009 (Suranjana Saha and Goldberg, 2010). It is a weak CDA system where the atmospheric and oceanic data assimilation is performed independently and the coupling is only through model dynamics. In this paper, we want to explore the idea of assimilating atmospheric reanalysis data in a CDA system and its resultant consequences. This idea is motivated to find an efficient way to get ocean analysis from a CDA process that incorporates both the atmospheric and oceanic observations. The atmospheric observations include hundreds of types, with different format, coverage, frequency, etc., which makes it nearly impossible for an individual or a small group to collect and assimilate all these observations to set up the CDA system independently. Reanalysis datasets incorporate millions of observations, which include, but are not limited to, radiosonde, satellite, buoy, aircraft, and ship reports, into a stable data assimilation system (e.g. Kalnay et al., 1996; White et al., 2002; Kanamitsu et al., 2002; Uppala et al., 2005; Suranjana Saha and Goldberg, 2010; Dee et al., 2011; HARADA et al., 2015). In addition, these datasets provide global coverage with constant spatial and temporal resolution over three or more decades for hundreds of variables (e.g. Kalnay et al., 1996; White et al., 2002; Kanamitsu et al., 2002; Uppala et al., 2005; Suranjana Saha and Goldberg, 2010; Dee et al., 2011; HARADA et al., 2015), which makes them relatively straightforward to handle from a processing standpoint. If it is feasible to substitute the atmospheric reanalysis datasets for actual observational data, we can set up CDA systems using different models much more easily and expect reasonable output of model analysis, especially oceanic analysis. Zhang

et al. (2007) has assimilated atmospheric reanalysis directly as observations in a fully coupled climate model with a CDA system, without the examination of the possible effects brought out by the substitution. Yet, their results are still promising and their assimilation successfully reconstructs the 20th-century ocean heat content variability and trend in most locations. This indicates that it is feasible to substitute the atmospheric observations with reanalysis in a CDA process. However, the resultant effects on the analysis from the substitution are never carefully studied. In this paper, we will test this idea with an emphasis on the quantification of the resultant effects from the substitution, and investigate the assimilation schemes associated with this substitution.

A coupled Lorenz96 model (Lorenz, 1996) representing the atmosphere and ocean is constructed to test the idea of assimilating atmospheric reanalysis data as observations in a CDA system. The results will be compared to the best-case scenario (benchmark) where both the atmospheric and oceanic observations are assimilated. The paper is organized as follows. The methodology is shown in Section 2, experiments and results are presented in Section 3, tests on different assimilation schemes are shown in Section 4, and Section 5 provides a concluding summary.

2.2 Methodology

2.2.1 Model

A dynamical system is set up by coupling two 40-variable ($nv = 40$) Lorenz96 systems (Lorenz, 1996), one representing the atmosphere (Eq. (1)) and theother representing the ocean (Eq. (2)).

$$\frac{X_j}{dt} = (X_{j+1} - X_{j-2})X_{j-1} - (1 - C_a)X_j + F_a + C_a(Y_j - X_j) \qquad (2.1)$$

$$M\frac{Y_j}{dt} = (Y_{j+1} - Y_{j-2})Y_{j-1} - (1 - C_o)Y_j + F_o + C_a(X_j - Y_j) \qquad (2.2)$$

The atmosphere and ocean are coupled through the flux terms $C_a(Y_j, X_j)$ and $C_o(X_j, Y_j)$, where $C_a = 2.0$ and $C_o = 0.1$ are the coupling coefficients for the atmosphere and ocean, respectively. F_a and F_o represent the external forcing, and in this case, $F_a = 8$, $F_o = 0$, such that the ocean is only forced by the atmosphere. The oceanic timescale is controlled by coefficient M, which is chosen to be 20. Figure2.1 shows the typical time evolution of variables X_1 and Y_1. The climatological standard deviations averaged over 40 variables for X and Y are 3.86 and 0.47, respectively. Therefore, the observational errors are arbitrarily set at 1.0 for the atmosphere and 0.1 for the ocean. In all experiments, the integration time step is 0.005, and 1 time unit is roughly 5 days. We forward the model for 2×10^5 time steps (1000 time units). The first 20000 spin-up steps are discarded when we evaluate the analysis quality.

2.2.2 Assimilation procedure and diagnostics

We employ the ensemble Kalman fil-
ter (EnKF) with perturbed observa-
tion (Evensen, 1994; Burgers and van
Leeuwen, 1998; Houtekamer, 1966).
Eighty ensemble members are used (ens
= 80). Covariance localization used here
is the same as Hamill et al. (2001). It is
performed by applying a Schur product
(an element by element multiplication)
to the forecast error covariance matrix
and a correlation matrix. The correla-

Figure 2.1: Typical time evolution of X_1
(red) and Y_1 (blue).

tion matrix is a fifth order function of Gaspari and Cohn (1999). Covariance in-
flation is also applied with the relaxation method by Zhang et al. (2004). The
root-mean-square error (RMSE) from all analysis steps is calculated to evaluate the
data assimilation performance, which is given by

$$RMSE = \sqrt{\frac{1}{nt}\frac{1}{nv}\sum_t\sum_i(X_{i,t} - X_{i,t}^T)^2} \tag{2.3}$$

where $X_{i,t}$ and $X_{i,t}^T$ are analysis and truth, respectively, at gridpoint i and time step
t; nt is the total time steps and nv is the variable number. In reality, the benchmark
experiment would be the best-case scenario for a CDA system and should produce
the best analysis possible. Therefore, it is used to evaluate the performance of

our proposed schemes. The RMSEs of different experiments that assimilate the reanalysis are then normalized by that of the benchmark experiment as in Eq. (4)

$$Ratio = \frac{R - R_b}{R_b} \times 100\% \qquad (2.4)$$

where R represents the RMSE of the experiment in which the atmospheric reanalysis is assimilated as observation, and R_b represents the RMSE of the benchmark experiment. We repeat 90 simulations for each experiment and the results are displayed via boxplots (Figure2.4).

2.2.3 Model Framework

In this section, we will introduce three models that are used to generate the true state, reanalysis, and conduct the experiments, respectively. The purpose is to allow for model bias to test the robustness of the quantified effects. If these three models are the same, it is a perfect model framework with no model bias.

Model 1: Generate the true state and observation. The true state is a control run of this model and the observations are generated by adding a Gaussian white noise N(0, σ_o) to the true state, where σ_o is the observational error.

Model 2: Generate reanalysis by assimilating observations from model 1. This mimics the fact that different research centers generate the reanalysis through their own GCMs (Kalnay et al., 1996; White et al., 2002; Kanamitsu et al., 2002; Uppala et al., 2005; Suranjana Saha and Goldberg, 2010; Dee et al., 2011; HARADA et al., 2015) and these GCMs are biased with regard to the model used to generate the

true state.

Model 3: Conduct the benchmark and substitution experiments. Model 3 differs from model 2 in that the model where the CDA system is set up can be different from the model used to generate reanalysis.
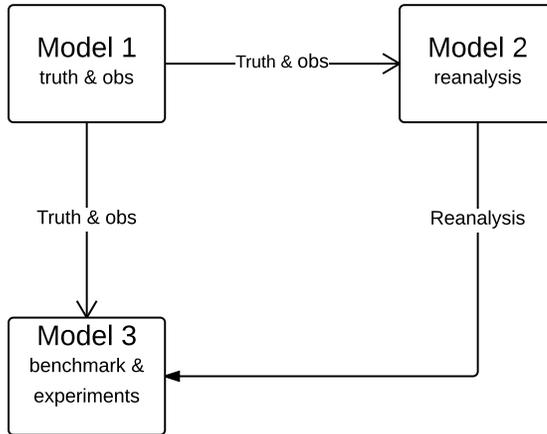
Figure 2.2: Illustration of model setup. Truth and observations (obs) are generated in model 1. Reanalysis is generated in model 2. Experiments are conducted in model 3. In perfect model case, the three models are the same, while in biased model case all three are different.

The relationship among the three models is illustrated in Figure2.2 In the perfect model framework, three models are the same and they use the default parameter values with Runge-Kutta 4 integration scheme (Iseries, 1996). In the biased model framework, model bias is mimicked by different integration scheme (Runge-Kutta 2; Iseries (1996)) and slight variations on model parameters. The detailed setup is summarized in Table. 2.1.

Model	Integration Scheme	Model parameter
Model 1	Runge-Kutta 2	Standard parameters
Model 2	Runge-Kutta 2	5% increase from standard parameters
Model 3	Runge-Kutta 4	Standard parameters

Table 2.1: Model setup in the biased model framework

2.2.4 The benchmark and substitution experiment

In the benchmark experiment conducted in model 3, atmospheric and oceanic observations generated in model 1 are assimilated. Variables X_1, X_3, X_5, ... , X_{39}, are observed every 20 integration time steps and variables Y_1, Y_3, Y_5, ... , Y_{39} are observed every 40 integration time steps, unless specified otherwise. This is the best-case scenario where all observations that are available are assimilated.

The atmospheric reanalysis data used for the substitution experiments are generated in model 2. The observations are the same as that in the benchmark experiment except that no oceanic observations are assimilated since the effects of the ocean data assimilation on the atmosphere are small in this coupled Lorenz96 system. The reanalysis is the ensemble mean output. We also preserve the reanalysis ensemble for the assimilation scheme design later.

In the substitution experiments, the atmospheric reanalysis and oceanic observations are assimilated into model 3. The reanalysis is assimilated with the same frequency as the benchmark experiment, while oceanic observations stay unchanged. Although the atmospheric observations are not available at every grid-point, the reanalysis will provide additional observations at unobserved locations.

2.2.5 Assimilating the reanalysis

The most straightforward way to assimilate the reanalysis is to simply treat the reanalysis as independent observations. First, the error covariance matrix of the reanalysis is calculated, and then the reanalysis is independently perturbed according to the diagonal values of the matrix, namely, the variances of the observations.

The error covariance matrix of the reanalysis can be calculated as:

$$R_t = cov < X - X^T >=$$

$$\begin{pmatrix} cov < X_1 - X_1^T, X_1 - X_1^T > & cov < X_1 - X_1^T, X_2 - X_2^T > & \dots & cov < X_1 - X_1^T, X_{40} - X_{40}^T > \\ cov < X_2 - X_2^T, X_1 - X_1^T > & cov < X_2 - X_2^T, X_2 - X_2^T > & \dots & cov < X_2 - X_2^T, X_{40} - X_{40}^T) > \\ \vdots & \vdots & \ddots & \vdots \\ cov < X_{40} - X_{40}^T, X_1 - X_1^T > & cov < X_{40} - X_{40}^T, X_2 - X_2^T > & \dots & cov < X_{40} - X_{40}^T, X_{40} - X_{40}^T > \end{pmatrix} \quad (2.5)$$

where X and X^T are the time series of the reanalysis and truth, respectively (note that the superscript T means the truth, does not represent the transpose of X), and they are of size $nv \times nt$; X_1 and X_1^T denote the $1 \times nt$ time series of reanalysis and truth of variable 1, and same interpretation for the other elements in Eq. (5). $cov <>$calculates the covariance. R_t is shown in Figure2.3 The reanalysis at any given locations perturbed with Gaussian noise that has the same standard deviation as the square root of the corresponding diagonal element in X_t. Consequently, the error matrix used in calculating the Kalman gain is the $R_t$matrix with its off-diagonal elements set to zero due to the assumption of independence among different observations. In the real world, however, X^T is unknown. We can either use observation to replace X^T in Eq. (5), or use the averaged sample covariance of the original reanalysis ensemble (see R in Eq. (7)), which will be introduced later in Section 4.
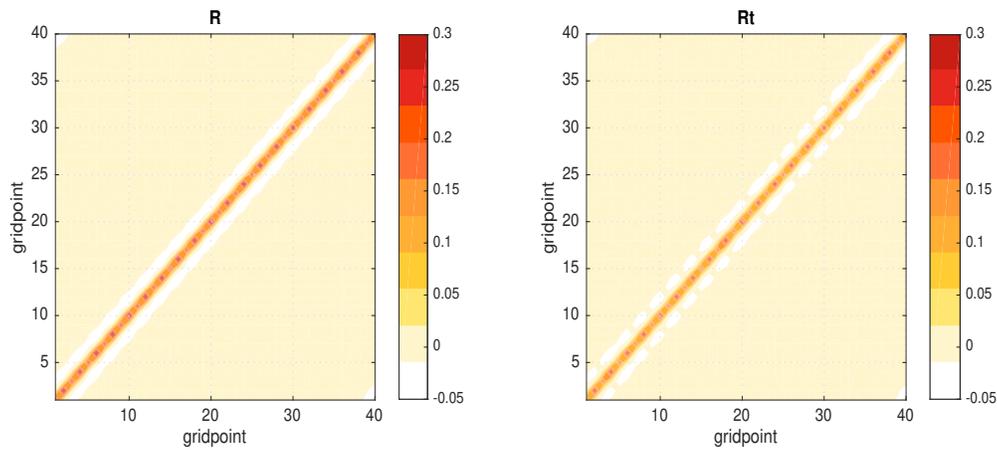
Figure 2.3: (a) The time average of flow-independent ensemble covariance matrix R and (b) the temporal covariance matrix Rt. Point (i, j) indicates the covariance between the ith and j^{th} atmospheric variables.

2.3 Experiments and results

Following the procedure in Section 2, the first type of experiments is denoted as UNCORR in Figure2.4, which stands for "uncorrelated observation ensemble" since the observation ensemble is the reanalysis plus independent Gaussian white noise. Compared to the benchmark, the increase of the average RMSE of the ocean variables over 90 simulations is 11.41% in the perfect model framework and 16.93% in the biased model framework. For the 90 simulations in the perfect model framework (Figure2.4b), the maximum and minimum RMSE increases are 15.46% and 7.42%. In the biased model framework (Figure2.4d), the maximum and minimum non-outlier RMSE increases are 21.79% and 11.81%. If only ocean data assimilation is carried out in this coupled Lorenz96 model, the ocean component

will not be constrained by the oceanic observations alone and the oceanic RMSE can reach the climatological standard deviation. This is because in this simple coupled model, the ocean component is purely driven by the atmosphere and the feedback from the ocean to the atmosphere is small, hence we are not able to get a reasonable oceanic analysis if the atmosphere is not well constrained. In this sense, substituting the atmospheric observations with the atmospheric reanalysis in a CDA process is better than assimilating oceanic observations alone in a coupled system. The performance of assimilating reanalysis is further tested with varied atmospheric observation frequency, atmospheric observation error, atmospheric observation density, and ensemble size. The performance of UNCORR is tested for different atmospheric observation frequencies and the results are shown in Figure2.5. The analysis cycle increases from 20 to 120 time steps as the atmosphere observations become infrequent. The RMSEs of both UNCORR and the benchmark increase due to decreased observational information (figure omitted). The ratios in Figure2.5 show no significant trend when the analysis cycle increases from 20 to 80 steps. However, as the atmospheric observation frequency becomes unrealistically infrequent (every 80 steps and beyond, this frequency is less than 1 observation every 2 days), the ratios tend to decrease. This means that as the analysis quality gets worse for both the benchmark and the substitution experiment due to less available observational information, the difference between the benchmark and the substitution experiment becomes small. That is to say, the resultant analysis deterioration from the substitution is less severe.

The sensitivity of UNCORR to varied atmospheric observation error is shown
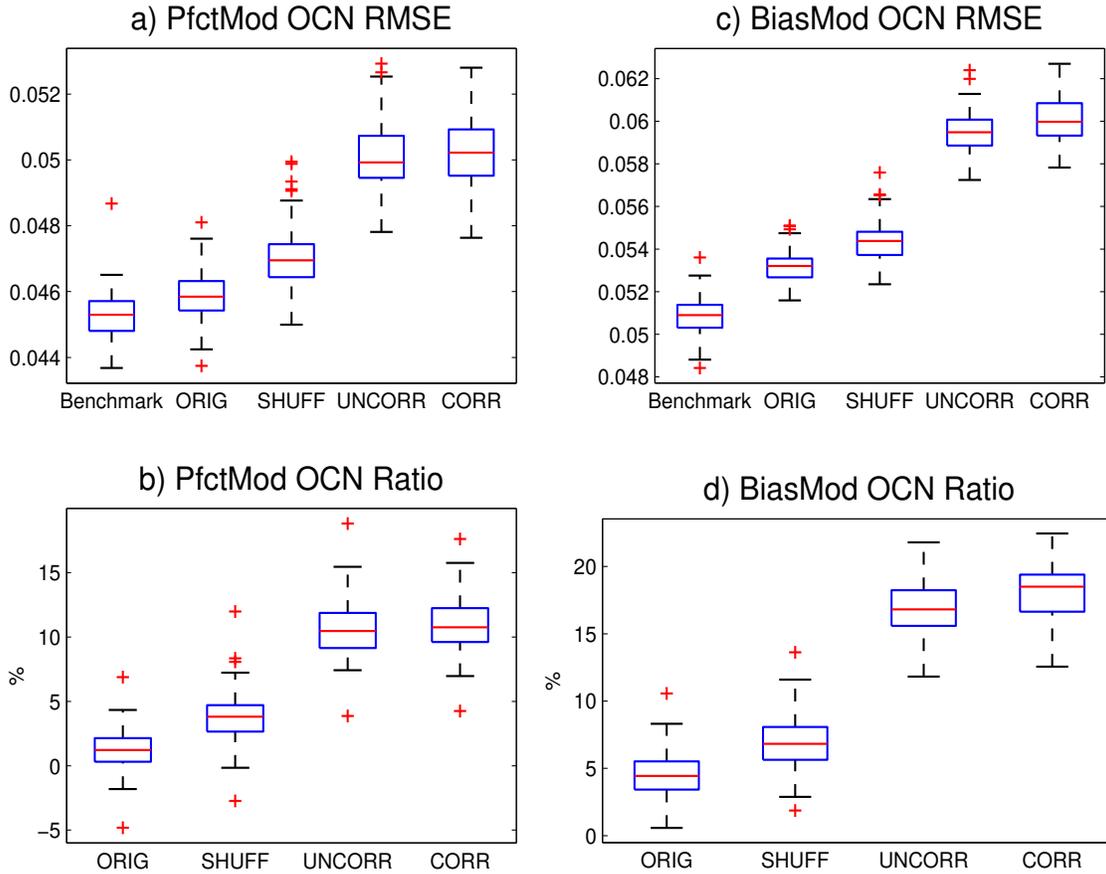
Figure 2.4: Boxplots of 90 simulations for benchmark and the 4 different schemes named ORIG, SHUFF, UNCORR, and CORR. ORIG, SHUFF, and CORR are subsequent sensitivity tests which will be introduced in later section. (a, b) The RMSE and (c, d) the RMSE ratio which is normalized by the benchmark experiment for (a, c) the perfect model case and (b, d) the biased model case. The whiskers below and above the box show minimum and maximum values. The upper and lower bounds of the box are the first and third quartiles. The red line is the median and the red crosses indicate the outliers.

in Figure2.6. The absolute RMSEs of both UNCORR and benchmark increase as the observations become more and more noisier (figure omitted), while the ratios in Figure2.6 show small fluctuations when the atmospheric observation error
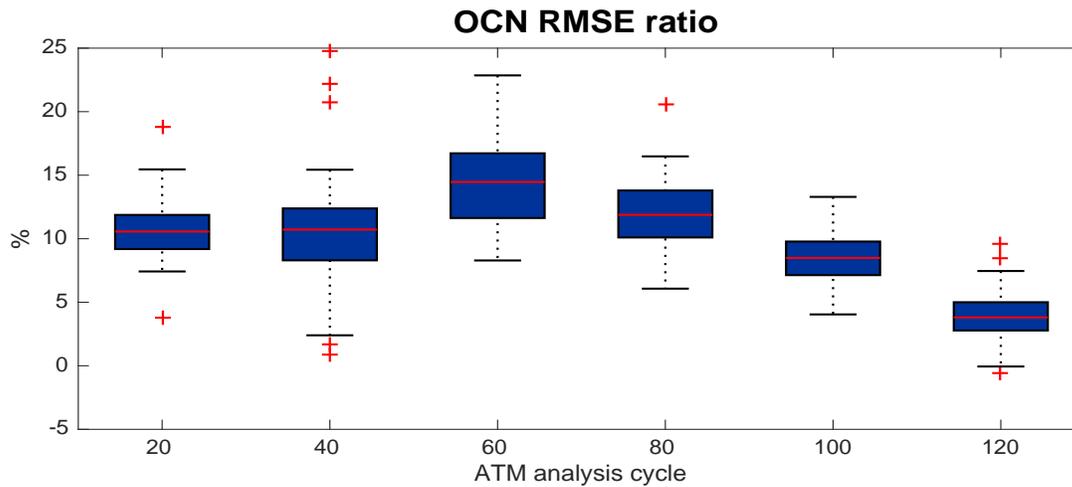
Figure 2.5: Sensitivity of ocean RMSE ratio to atmosphere (ATM) observation frequency for the UNCORR scheme over 90 simulations.

increases from 0.2 to 2.0, and eventually decreases as the error gets unrealistically large (beyond 2.0, more than half of the climatological standard deviation). This indicates that the oceanic analysis deterioration in UNCORR is fairly insensitive to the atmospheric observation error when it is in a reasonable range. In addition, the deterioration is lessened if the oceanic analyses of both the benchmark and the substitution experiment get worse due to the overly noisy observations.

The results of experiments with varied observation density but still evenly distributed observations are shown in Figure2.7. The RMSEs for both experiments are larger with sparser observations and smaller with denser observation (figure omitted). The ratios in Figure2.7 show a consistent decreasing trend as the observations get denser. This suggests that the difference between the substitution and benchmark experiments is less significant if the analyses of both experiments get worse due to less observations, which is similar to the previous sensitivity tests.

OCN RMSE ratio
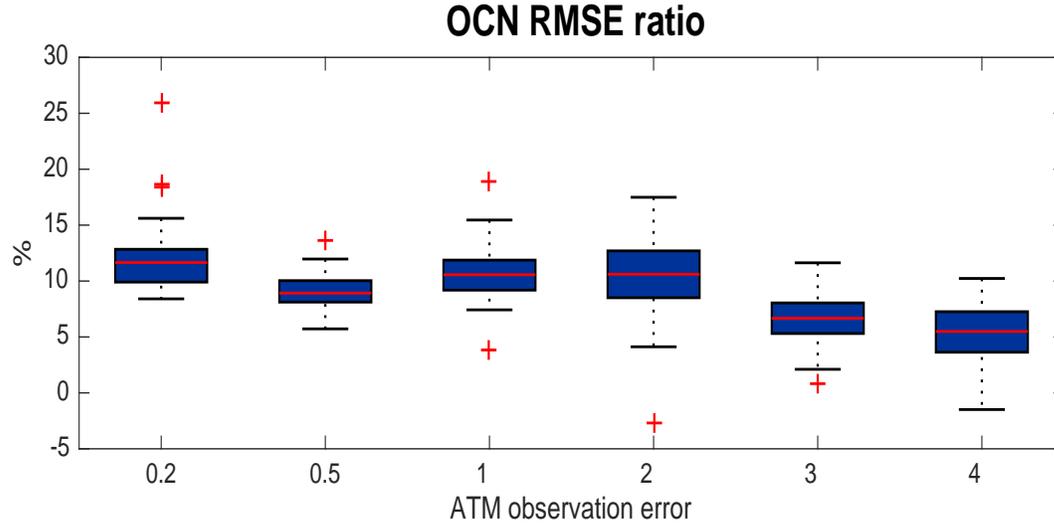
Figure 2.6: As in 2.5, but for different atmopsheric observation errors.

In contrast, if every variable is observed, the increase of RMSE in the substitution experiment relative to the benchmark is the most significant. In reality, the number of atmospheric observations is more or less fixed, and the density change will not be so extreme as in this simple sensitivity test. It is notable that as the observations get intermediately sparse, the data assimilation process gets less stable: the variance of ratios among the 90 simulations is noticeably bigger when there are only 10 or 20 observations. This can be explained by the numerical instabilities developed in a sparse observation network with finite ensemble size (Gottwald, 2014).

The results in Figure2.5, Figure2.6 and Figure2.7 collectively indicate that the decrease in analysis quality due to the substitution is insensitive to the observation quality (frequency, noisiness, and density) within a reasonable range. Meanwhile, when the quality of observation declines too much, the decrease in analysis quality caused by substitution becomes less severe compared to the benchmark because

OCN RMSE ratio
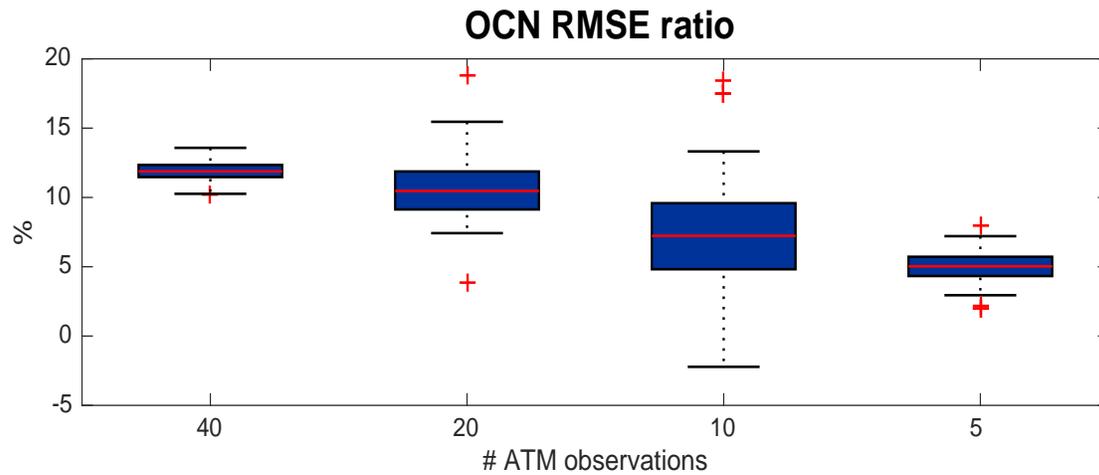
Figure 2.7: As in 2.5, but for different observation densities.

the analysis quality of the benchmark also decreases significantly due to poor observational quality.

Ensemble size is an important factor in the estimation of error covariance and correlation. Additional experiments with ensemble size 20, 40, and 200 are shown in Figure2.8. The RMSEs of both experiments decrease with increasing ensemble size (figure omitted). The ratios in Figure2.8 increase with ensemble size and eventually level off. This suggests that with smaller sample size, the bad analysis quality for both experiments will lead to smaller RMSE contrast between the substitution and the benchmark experiments, thus smaller ratios; and vice versa for sufficiently large sample size.
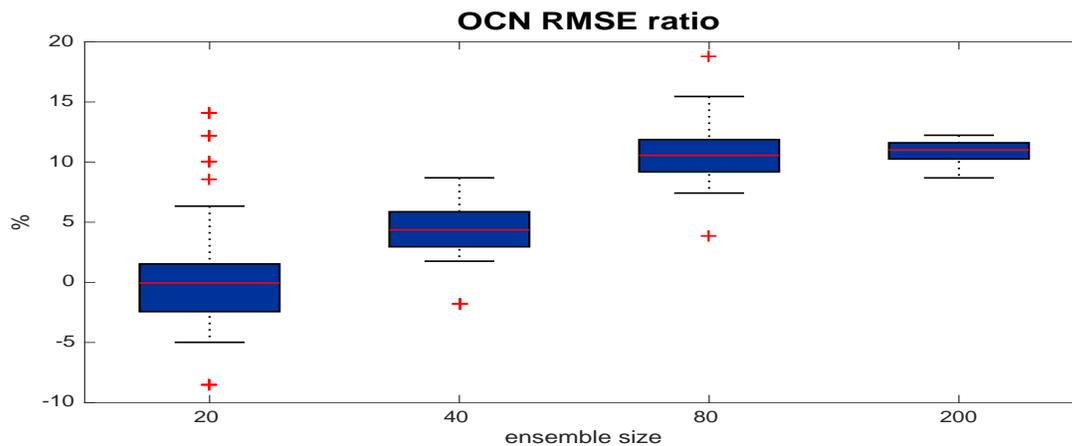
Figure 2.8: As in 2.5, but for different ensemble members.

2.4 Tests on assimilating schemes

In the last section, the UNCORR experiments assume that the analysis has independent errors at different locations, which, however, is not the case in reality. When reanalysis is generated, different model locations are connected through both model dynamics and the use of localization schemes. Thus, the analysis errors will be spatially correlated between nearby or even far-apart locations. In addition, the analysis errors can also persist through time; hence, there is also temporal correlation in the time series of the analysis. The previous UNCORR experiments neglect both the spatial and temporal correlations, which may affect the performance of the CDA scheme. To deal with these correlations and investigate how they affect the CDA, we tested three other schemes for treating the reanalysis error covariance, which are named as CORR, ORIG, and SHUFF.

First, in the CORR scheme, the off-diagonal correlation among different variables of the reanalysis is taken into consideration. Instead of being set to zero as in

UNCORR, the off-diagonal elements in the reanalysis error covariance matrix are retained. In correspondence, a spatially correlated observation ensemble is attained by perturbing the reanalysis with correlated noise. The spatial correlation among different variables can be calculated in a similar way as Rt in Eq. (5):

$$C = corr < X - X^T > \tag{2.6}$$

The average RMSE increases of ocean variables over 90 simulations in CORR are 11.79% for the perfect model framework and 18.13% for the biased model framework. The maximum and minimum non-outlier RMSE increases are 15.76% and 6.96% for 90 simulations in the perfect model framework (2.4a) and 22.45% and 12.55% in the biased model framework (2.4b). Although CORR includes the off-diagonal correlation among different variables, it does not outperform UNCORR and is also less stable. This is mainly caused by the additional sampling errors. Because of the chaotic nature of the model, the correlation among different locations in Lorenz96 decreases below 0.2 within 5 gridpoints. Thus, CORR is subject to significant sampling error in two processes, firstly when the correlation matrix is calculated in Eq. (6) and secondly when the correlated observation ensemble is artificially generated based on the correlation matrix. With a finite sample size, the error in the covariance or correlation estimates increases greatly when the true correlation becomes smaller. Therefore, although including the spatial correlation may improve the performance theoretically, the additional sampling error overwhelms the possible improvement.

Second, the ORIG scheme uses the original reanalysis ensemble as the "per-

turbed" observation ensemble during CDA. The original reanalysis ensemble is the byproduct of the ensemble-based data assimilation filter during the generation of the reanalysis; therefore, they could accurately capture not only the flow-dependent correlation information among different locations, but also the temporal coherence of each ensemble member at every location. The error covariance matrix in this scheme is calculated as the time average of the error covariance matrix of the reanalysis ensemble over each analysis step, as follows

$$R = mean(cov < X_t^{re}(nv, ens) >) \tag{2.7}$$

where $cov <>$ and $mean()$ represent sample covariance and the average of the error covariance matrices at each time step of the covariance over time, respectively; and X_t^{re} represents the original reanalysis ensemble at analysis time step t. R is quantitively similar to R_t and is also diagonally dominant (2.3). The performance of ORIG is noticeably better than UNCORR and CORR (Fig. 4) and is fairly close to the benchmark. The average oceanic RMSE increases over 90 simulations are 2.02% in the perfect model framework and 4.64% in the biased model framework. The ratios range from -1.81% to 4.33% in the perfect model framework over 90 simulations (Fig. 4a), and from 0.58% to 8.31% in the biased model framework.

Third, the SHUFF scheme is used to test the relative importance of accurate spatial correlation and temporal coherence in the improvement from CORR to ORIG. SHUFF is the same as ORIG except that the original reanalysis ensemble is shuffled at each analysis step before it is assimilated. Hence, the temporal coherence carried along each ensemble member is removed in SHUFF while the off-diagonal

spatial correlation is still preserved. The performance of SHUFF is slightly worse than ORIG in both perfect and biased model cases (Fig. 4), which indicates that the temporal coherence of the reanalysis ensemble is less important for the ocean analysis. Meanwhile, SHUFF, similar to ORIG, outperforms UNCORR and CORR significantly: the highest RMSE increase in SHUFF (7.22%) almost approaches the lowest ones in CORR (6.96%) and UNCORR (7.42%). SHUFF and CORR both have the off-diagonal correlation and do not have the temporal coherence, and they primarily differ in generating the perturbations for the reanalysis (Eqs. (5) and (7)), or simply the magnitude of sampling errors for the correlation matrix. The comparisons between SHUFF and CORR and between SHUFF and ORIG therefore suggest that accurate representation of the spatial correlation is relatively more important than the temporal coherence for the ocean analysis. However, for the atmosphere component, the performance of SHUFF is closer to CORR than to ORIG (figure omitted), which suggests a relatively more important role of temporal coherence for the atmospheric analysis. All the assimilation schemes are summarized in Table.2.2.

Scheme	Covariance matrix	Observation ensemble
ORIG	Eq.(7)	Original reanalysis ensemble
SHUFF	Eq.(7)	Shuffled original reanalysis ensemble
UNCORR	Eq.(5), off-diagonal elements set to zero	Perturbed ensemble (uncorrelated)
CORR	Eq.(5)	Perturbed ensemble based on correlation matrix

Table 2.2: Assimilation scheme design

2.5   Summary and conclusion

We substituted the atmospheric observations with reanalysis data to set up a CDA system in coupled Lorenz96 models and quantified the resultant effects on the oceanic analysis. We compared the oceanic RMSE of the substitution experiment where atmospheric reanalysis and oceanic observations are assimilated to a benchmark experiment where both atmospheric and oceanic observations are assimilated. It is found that the substitution results in the deterioration of oceanic analysis quality. The magnitude of this deterioration depends on how the reanalysis is assimilated. When the reanalysis is assimilated directly as independent observations (UNCORR) as in Zhang et al. (2007), the oceanic RMSE increases due to the substitution are about 16% in the perfect model framework and about 22% in the biased model framework compared to the benchmark or best-case scenario. Additional sensitivity tests show that this result is robust with sufficient ensemble size and reasonable atmospheric observation quality (density, frequency, and noisiness). If the ensemble size is smaller, or the observation quality is worse (less frequent, sparser, and noisier), the deterioration will become less severe because the analysis quality of the benchmark also decreases significantly.

In addition to the direct method, three supplementary schemes (CORR, ORIG, and SHUFF) are tested with a focus on the representation of the background error covariance matrix and the generation of the perturbed observations in EnKF. We found that both the spatial correlation among the reanalysis data points and the coherence along each original reanalysis ensemble member are crucial to the analysis quality of the substitution experiments. The oceanic RMSE increase is

significantly reduced when the temporal coherence along each member of the original reanalysis ensemble is preserved (ORIG); the removal of such ensemble member coherence (SHUFF and CORR) and inaccurate capture of the off-diagonal correlation (CORR, UNCORR) will result in the increase of RMSE. However, the relative importance between the off-diagonal correlation and temporal coherence on analysis quality is different for the atmosphere and ocean components. For the ocean component, the RMSE of SHUFF is closer to ORIG than CORR, indicating a relative more important influence from the accurate representation of spatial correlation than temporal coherence, while for the atmosphere, it is the other way round.

This study demonstrates that substituting the atmospheric observations with atmospheric reanalysis is a potentially efficient approach to implementing CDA systems at the cost of moderate degradation of analysis quality. Despite the fact that this degradation cannot be eliminated, the CDA products can still provide state-estimation of the coupled variability in the atmosphere-ocean system, which incorporates both the observational and model information, and the dynamical balance between the atmosphere and ocean components can reduce the initial shock in the initialization of the coupled GCM. There are still remaining issues regarding assimilating atmospheric reanalysis data. First, different schemes, in particular ORIG, UNCORR, and CORR, should be tested on models that have higher spatial correlations, and the impact on the oceanic analysis quality should be evaluated. Second, different ensemble filters such as ensemble adjustment filter (Anderson, 2001b; Zhang et al., 2007), can be employed to assess robustness of assimilating the

reanalysis. Finally, this idea should be further tested in a coupled model of higher complexity.

Chapter 3

A Systematic Comparison of Particle Filter and EnKF in Assimilating Time-Averaged Observations

3.1   Introduction

the particle filter, in theory, is its capability to provide the optimal analysis whether or not the error distribution is Gaussian (Leeuwen, 2009). However, the particle filter in general suffers from the so-called "curse of dimensionality" or particle impoverishment: A phenomenon where the nontrivial weights tend to collapse to only a few particles while majority of the particles degenerate to trivial weights. Resampling after each analysis step is used to alleviate such degeneracy, but PF-SIR still requires a considerable number of particles because the ensemble size required for a successful filtering increases exponentially with the problem size (Snyder et al., 2008). Considering the large degree of freedom of geophysical applications and the limited computation resources, the "curse of dimensionality" has been a severe

problem. However, some recent studies seem to be successful in applying the PF-SIR for PDA using affordable ensemble sizes (N 100) (Dubinkina et al., 2011; Goosse et al., 2011; Goosse, 2012; Mathiot et al., 2013; Mairesse et al., 2013). These studies did not encounter filter degeneracy because they reduced the degree of freedom by performing time averaged, spatial filtering, such as the long-term climatological temperature averaged in a single hemisphere. The particle filter has also been used to investigate the causes and mechanisms of some climate variability in the past (Goosse, 2012; Mathiot et al., 2013).

In contrast to data assimilation (DA) techniques in meteorology, PDA is still in its infancy and no universal methodology has been established. The difficulties posed by spatially and temporally sparse, noisy and mostly indirect proxy records make it challenging to apply most modern DA techniques directly to PDA. Various DA methods have been attempted since the concept was introduced to the paleo field by von Storch and Zorita (2000). Early methods include pattern nudging (von Storch and Zorita, 2000) and forcing singular vectors (J. Barkmeijer, 2003). These methods share the same idea by adding an artificial forcing term to the prognostic equations to nudge the model toward a desired large-scale pattern derived from available observations. The drawbacks of these methods are that they require a relatively dense observation network to capture the large-scale pattern and that they cannot address the uncertainty of the analysis. Recently, ensemble filter methods have been widely used for paleoclimate studies, in which a finite ensemble of model simulations is used to represent the model's statistical behavior and Bayesian estimation theory is used to derive the posterior estimates (Evensen,

1994, 2003; Leeuwen, 2009). These ensemble methods fall into two categories: the particle filter and the ensemble-based Kalman filter.

Past climate can be simulated in state-of-the-art climate models (Braconnot et al., 2012; Otto-Bliesner et al., 2006) or inferred from proxy-based reconstructions (Mann et al., 2008a, 2009; Marcott et al., 2013). Both approaches provide independent yet complementary information about past climate and have been used independently for the reconstruction and understanding of the past climate. Paleoclimate data assimilation (PDA) offers a new approach to further improve the reconstruction of past climate by combining observational information from proxy records and dynamic constraints from climate models. PDA will constrain the model with proxy observations and provide a complete field of the past climate state containing information that cannot be directly derived from the proxy records, equivalent to reanalysis data sets that fill in the spatial and temporal gaps in a sparse observation network (Kalnay, 2003). Furthermore, with careful calibration of the uncertainty in the model and proxy data, PDA can also quantify the uncertainty of the final analysis product.

In contrast to data assimilation (DA) techniques in meteorology, PDA is still in its infancy and no universal methodology has been established. The difficulties posed by spatially and temporally sparse, noisy and mostly indirect proxy records make it challenging to apply most modern DA techniques directly to PDA. Various DA methods have been attempted since the concept was introduced to the paleo field by von Storch and Zorita (2000). Early methods include pattern nudging (von Storch and Zorita, 2000) and forcing singular vectors (J. Barkmeijer, 2003).

These methods share the same idea by adding an artificial forcing term to the prognostic equations to nudge the model toward a desired large-scale pattern derived from available observations. The drawbacks of these methods are that they require a relatively dense observation network to capture the large-scale pattern and that they cannot address the uncertainty of the analysis. Recently, ensemble filter methods have been widely used for paleoclimate studies, in which a finite ensemble of model simulations is used to represent the model's statistical behavior and Bayesian estimation theory is used to derive the posterior estimates (Evensen, 1994, 2003; Leeuwen, 2009). These ensemble methods fall into two categories: the particle filter and the ensemble-based Kalman filter.

Application of the particle filter to paleoclimate has experienced three stages. In the first stage, a simplified (offline) particle filter was used and a single member that best matches the proxy record is selected among an ensemble of preexisting model simulations as the final analysis (Goosse, 2006b,a); hence, no forwarding model is needed. In the second stage, this method was modified toward an online degenerate particle filter (Crespin et al., 2009; Goosse et al., 2009, 2010; Widmann et al., 2010), in which one member was selected as the best fit of observations at the target period and is then used as the initial conditions for the subsequent simulation. In the third stage, the online degenerate particle filter evolved into the standard particle filter with simple importance resampling (PF-SIR) (Annan and Hargreaves, 2012; Dubinkina et al., 2011; Goosse et al., 2011; Goosse, 2012; Mairesse et al., 2013; Mathiot et al., 2013). In the PF-SIR, analysis is derived as a weighted sum across all ensemble members. The weights are calculated based on the closeness of each member to the

available observations. A resampling step is performed on the ensembles before proceeding to the next cycle where the members with trivial weights are removed and the ones with nontrivial weights are duplicated to maintain the same ensemble size. One advantage of the particle filter, in theory, is its capability to provide the optimal analysis whether or not the error distribution is Gaussian (Leeuwen, 2009). However, the particle filter in general suffers from the so-called "curse of dimensionality" or particle impoverishment: A phenomenon where the nontrivial weights tend to collapse to only a few particles while majority of the particles degenerate to trivial weights. Resampling after each analysis step is used to alleviate such degeneracy, but PF-SIR still requires a considerable amount of particles because the ensemble size required for a successful filtering increases exponentially with the problem size (Snyder et al., 2008). Considering the large degree of freedom of geophysical applications and the limited computation resources, the "curse of dimensionality" has been a severe problem. However, some recent studies seem to be successful in applying the PF-SIR for PDA using affordable ensemble sizes (N 100) (Dubinkina et al., 2011; Goosse et al., 2011; Goosse, 2012; Mathiot et al., 2013; Mairesse et al., 2013). These studies did not encounter filter degeneracy because they reduced the degree of freedom by performing time averaged, spatial filtering, such as the long-term climatological temperature averaged in a single hemisphere. The particle filter has also been used to investigate the causes and mechanisms of some climate variability in the past (Goosse et al., 2011; Mathiot et al., 2013).

The EnKF (Evensen, 1994, 2003) is a modern DA method that can be modified to accommodate time-averaged observations in the paleo context (Dirren and Hakim,

2005; Hakim et al., 2016; Huntley and Hakim, 2010; Steiger N., 2013; Steiger and Hakim, 2016; Bhend et al., 2012). In the EnKF, the final analysis equation is a linear combination of observation information and model estimate with the weight determined according to the uncertainties (Evensen, 1994, 2003; Kalnay, 2003). In the case of Gaussian error, the EnKF can still achieve the optimal analysis. This optimal analysis in theory maximizes the likelihood of achieving such posterior estimates and minimizes the error between itself and the prior estimates and the observations. In general, when the error is non-Gaussian, there is no guarantee that the EnKF can achieve the optimal analysis as the PF-SIR can. Nevertheless, the EnKF can provide robust and consistent analysis with relatively small ensemble sizes, as has been shown in numerous studies (Evensen, 1994, 2003; Zhang et al., 2007).

Our goal is to perform a systematic comparison of the PF-SIR and the EnKF for paleoclimate field reconstruction. We adopted online DA approach for both PF-SIR and EnKF, which provides better temporal consistency (Matsikaris et al., 2015). The two methods will be performed in pseudoproxy experiments (PPEs) that provide a synthetic, controlled test bed (Smerdon, 2012). We found that the EnKF with 16 ensemble members outperforms the PF-SIR using 48 members in both the local reconstruction and hemispheric reconstruction. The conclusion is robust even when the ensemble size of the PF-SIR is doubled to 96. This paper is organized as follows. In section 2, we introduce the model and proxy network, evaluation metrics, and the details of the assimilation methods. Section 3 compares the assimilation of the two filters and discuss the advantages and disadvantages of

each filter. The conclusion and some general discussion are given in section 4.

3.2 Model, Methods, and Experimental Design

3.2.1 Model Description

The Fast Ocean Atmosphere Model (FOAM) is used in this study. FOAM is a fully coupled global atmosphere and ocean model (Jacob, 1997). The atmosphere component is based on National Center for Atmospheric Research CCM2 and CCM3 models with R15 resolution (40 latitudes $\times$ 48 longitudes) and 18 vertical layers. The ocean component is based on Modular Ocean Model created by Geophysical Fluid Dynamics Laboratory. It has a regular 128×128 polar grid and 24 vertical levels. The quality of the simulated climate compares well with higher-resolution models. FOAM has been used for various paleoclimate studies especially in Holocene (Liu et al., 2000, 2004, 2006, 2007). It has also been used to test DA techniques and algorithms successfully (Y. Liu, 2014a,b; Lu et al., 2015, 2017).

3.2.2 EAKF and Its Implementation

Ensemble Adjustment Kalman Filter (EAKF) is used in FOAM. The EAKF is a variant of EnKF and is a deterministic filter that does not require the perturbation of observations as EnKF does (Anderson, 2001b, 2003; Tippett et al., 2003; J. S. Whitaker, T. M. Hamill, 2002). A brief introduction of the EnKF is provided below, which also applies to the EAKF. For full comprehensive description, please refer to Anderson (2001b, 2003); Kalnay (2003).

The EnKF solves for the least squares estimate of the system state given model simulations and observations. Model simulations serve as a prior estimate (background) of the true state, which is then updated with the observational information to solve for the posterior estimate (analysis). The final analysis represents the most likely state of the climate system given model estimates and observations. This process expressed in Bayesian framework is

$$P(X|y) = \frac{P(y|X) * P(X)}{P(y)} \tag{3.1}$$

where $P(X)$ represents the prior distribution of the state X, $P(y|X)$ represents the likelihood of obtaining observation y given model estimate X. $P(y)$ is a normalization term. $P(X|y)$ is the posterior distribution after considering observation information. The observations and the prior estimate of the state are assumed to be unbiased with a Gaussian distribution. The model estimate is characterized by its mean x^b and covariance matrix P^b, where x^b is the prior state vector with the dimension of the state size. The model estimate of the observation y is then extracted from the background $\bar{x^b}$ as $\mathcal{H}(\bar{x^b})$, where H is the observation operator that projects the model space into the observation space. The final analysis $\bar{x^a}$ is derived as a weighted sum of the observation y and the model background $\bar{x^b}$, and the weights are inversely proportional to their relative uncertainty/error. The classic Kalman filter update equations are

$$
\begin{aligned}
\bar{x^a} &= \bar{x^b} + K(y - \mathcal{H}(\bar{x^b})) \\
K &= \frac{P^b \mathcal{H}^T}{\mathcal{H} P^b \mathcal{H}^T + R}
\end{aligned}
$$

From the analysis equation, it can be seen that the correction $K(y - \mathcal{H}(\bar{x^b}))$ of a grid point from a nonlocal observation is determined by the covariance $P^b\mathcal{H}^T$ between these two locations in the gain matrix equation (3). If these two highly covary, the value of K is larger and hence it receives a heavier correction, and vice versa. Therefore, the EnKF and its variants are a covariance-based method: as long as there is covariance, there is correction. In paleoclimatology, all proxy observations contain time-averaged information, such as annual mean and seasonal mean temperature, instead of instantaneous values like those of modern instrumental observations. The algorithm is modified based on Dirren and Hakim (2005) to assimilate the time-averaged observations. For example, if the observation y in equation (2) represents seasonal mean temperature, the corresponding model background x^b is also the seasonally averaged temperature. Note that the covariance matrix Pb is calculated from the time-averaged ensemble; therefore, it captures the low-frequency variability since the high-frequency variability has been smoothed out during the time average.

To remove the long-distance spurious correlations caused by a limited ensemble size, a covariance localization is implemented using the fifth-order function of Gaspari and Cohn (1999). Localization is an important technique used in the EnKF, which determines the influence radius of the observation at a given location (Houtekamer and Mitchell, 2001; Hamill et al., 2001). In a dense observation network, the localization radius is usually chosen to be small since most model grids have a local observation constraint (Y. Liu, 2014b; Lu et al., 2017). In a spatially sparse observation network such as in the paleo context, a broader localization scale

is necessary to extend the observation impact to larger regions. The localization scale in our EAKF experiment is roughly 16 grid points in the ocean, which gives a distance of 5,000 km on the equator, and decreases poleward with the cosine of latitude. We will use "5,000 km" to denote the EAKF experiment with a localization scale of 16 grid points. Covariance inflation is also applied using the relaxation method by Zhang et al. (2004) and the relaxation factor of 0.5. Further sensitivity tests show that the performance is not sensitive to this parameter. For the analysis cycle, we use an "online" approach (with cycling); that is, the analysis of the current step is used as the initial condition of the next cycle. This EAKF method has been used to develop the first ensemble-based coupled data assimilation system in a fully coupled general circulation model (CGCM) (Zhang et al., 2007). It has been implemented in FOAM to complete the first ensemble-based parameter estimation experiment in a CGCM (Y. Liu, 2014a,b).

3.2.3   Particle Filter and Its Implementations

A standard particle filter with simple importance resampling (Doucet, 2001; Gordon et al., 1993; Leeuwen, 2009), denoted as PF-SIR, is adopted in this work. The theoretical framework of PF-SIR is also Bayes theorem as in equation (1). Here the particles refer to ensemble members, and they provide a Monte Carlo approximation of the model distributions. For example, given N particles, the prior distribution P(X) can be constructed as a sum of N delta functions centered at each particle:

$$P(x) = \frac{1}{N} \sum_{i=1}^{N} \delta(x - x_i) \tag{3.2}$$

Given the prior distribution $P(X)$ and the likelihood $P(y|X)$, the posterior density $P(X|y)$ can be approximated as

$$P(x|y) = \sum_{1}^{N} w_i \delta(x - x_i)$$

$$w_i = \frac{P(y|x_i)}{\sum_{i=1}^{N} P(y|x_i)}$$

where wi is called the weight of the particle and it is a normalized likelihood. The likelihood $P(y|x_i)$ is often taken to be Gaussian:

$$P(y|x_i) = e^{-\frac{(\mathcal{H}(x_i)-y)^2}{2C}} \tag{3.3}$$

where C is the error matrix and $\mathcal{H}$ is the observation projector. Note that the likelihood of each particle is a product of Gaussian weights at every observation location. Hence, if the model estimate differs significantly from the observation at only a few locations, the final weight will be reduced to zero even if the majority of locations show a good match between the model and observation. To mitigate the effects of the trivial weights at some locations on the final weight of the particle, we add inflation to the weight calculated at every observation location before taken the product of them. With the posterior distribution and weights, the final analysis will be

$$\overline{f(x)} = \int f(x)P(x|y)dx \approx \sum_{j=1}^{N} w_i f(x_i) \tag{3.4}$$

Equations (4) through (7) construct the procedure of particle filter of sequential importance sampling. The proposal density in this case is chosen to be the prior

probability density, and the weights of the prior particles are equal (Doucet, 2001). This equal prior weight assumption is achieved by resampling the particles at every filtering time step. Therefore, after importance weights are calculated in equations (5) and (6), the particles are resampled according to their importance weights. The particles with weights less than $\frac{1}{N}$ are dropped, and particles with large weights are duplicated. Though there are several resampling methods, we apply a systematic resampling scheme (Hol et al., 2006) in this paper. After resampling, a small perturbation is added to the identical particles. The identical particles share the same ocean state but differ in atmospheric fields. These different atmospheric fields act as the perturbations. The new model states will be used as the initial conditions for the next cycle. In our study, about half of the 48 particles end up having trivial weights and get removed at every analysis step, they are then replaced by the duplications of the remaining particles. Again, note that we are assimilating observations that contain time-averaged information; hence, the background is temporally averaged before calculating the importance weights. This time averaging also helps to reduce the number of degree of freedom (Dubinkina et al., 2011). The assimilation procedure is essentially the same as in the previous studies (Dubinkina et al., 2011; Goosse et al., 2011; Goosse, 2012) except that no spatial filtering is applied. Both our PF-SIR and the PF-SIRs in the previous studies referenced above do not have the localization implemented (the more recently developed techniques in Poterjoy (2015) and Lee and Majda (2016)).

3.2.4 Pseudo Proxy Network and Experiments

We use pseudo seasonal mean sea surface temperature (SST) observations to conduct the PPEs. The pseudoproxy locations are chosen from Marcott et al. (2013), which provides 47 observations globally (Figure 1a) with a higher density in the Northern Hemisphere. The pseudoproxy time series are generated by adding Gaussian noise with a standard deviation of 0.5 °C to a specific model control run representing the actual state or "truth". The Gaussian noise mimics the errors associated with the proxy observation. The performance of each DA method is measured using the root-mean-square error (RMSE) and correlation coefficient between the analysis and the truth.

$$RMSE_i = \sqrt{\frac{1}{nt} \sum_t [X_{i,t} - X_{i,t}^T]^2} \tag{3.5}$$

where i represents the location, t represents the time, $X_{i,t}^T$ is the truth, and $X_{i,t}$ is the analysis. In all experiments, we use 16 members for EAKF and 48 for PF-SIR. We did not use the same ensemble size because EAKF shows decent performance even using a relatively small number of ensemble members Lu et al. (2017). Our results will show that even with only 16 members, EAKF still significantly outperforms PF-SIR with 48 members.

A 200 year-long experiment is carried out for both EAKF and PF-SIR assimilating seasonal SST pseudo observations. Two EAKF sensitivity experiments were also carried out using smaller localization scale, 2,000 km and 500 km, to illustrate the importance of localization radius to the performance of the EAKF. Furthermore, to test the sensitivity of ensemble size, we also carried out a PF-SIR experiment of 96

members and a EAKF-5000 km experiment of 32 members. These two ensemble size sensitivity tests are only 50 year long because of the expensive computational costs.

3.3   Results

The performance of the two filters was mainly assessed by computing the correlation on both local scale and hemispheric scale between the analysis SST and the truth. We also investigate the ability of each filter to capture the modes of variability by performing empirical orthogonal function (EOF) analysis on both the simulation and the truth.

3.3.1   Local Reconstruction Skills

Figure3.1 displays the analysis quality of SST characterized by RMSE and correlation coefficient for the PF-SIR and the EAKF of 5,000 km localization scale. The RMSE is calculated from the differences between seasonally averaged ensemble mean analysis and the "truth" at each grid point and normalized by its corresponding natural variability from Figure3.1a. The local reconstruction skill of the EAKF is significantly better than that of the PF-SIR at both the proxy location and the nearby regions. In the North Atlantic where the observations are dense, the correlation with truth is above 0.6 basin wide for the EAKF (Figure 1e) but below 0.6 for the PF-SIR (Figure3.1c); and the RMSE ranges from 0.4 to 0.6 for the EAKF (Figure3.1d) but above 0.8 for the PF-SIR (Figure3.1b). In regions that have observation clusters such

as the south of Iceland and the east of Newfoundland, the correlation coefficient for the EAKF reaches 0.85. In the tropical eastern Pacific where there is only one observation in the far east, the EAKF still maintains a correlation of around 0.4 and RMSE around 0.9 (Figures3.1e and d), but the PF-SIR's reconstruction skill is fairly weak (Figures3.1b and c).

The spatial reconstruction skill of the EAKF greatly depends on the localization scale. Recall that localization determines how far can the observation at a given location influence its nearby locations. In Figures3.1h and i, we show the results of the EAKF with decreasing localization scales of 2,000 km (Figures3.1f and g) and 500 km (Figures3.1h and i). When the localization scale is reduced, the area of higher correlation and lower RMSE also shrinks toward the observation location. When the localization scale is reduced to 500 km, well-constrained regions shrink to small areas tightly surrounding the proxy locations, as colored patches in Figures3.1d and e diminish to colored dots in Figures3.1h and i.

In PDA, larger localization scale can extend the observational impact of the sparse proxy network, especially in regions where large-scale variability exists and only a few proxies are available. For example, there are basin-wide covarying patterns, such as Pacific Decadal Oscillation (PDO) in the north Pacific and El Nino-Southern Oscillation (ENSO) in the tropical Pacific. One observation along the West Coast of North America (Figure3.1a) is able to partially constrain the evolution of PDO in the entire basin. This is also demonstrated in section 3.3 in terms of modes of variability. When localization radius is small, the reconstruction skill is lost away from the observation sites.
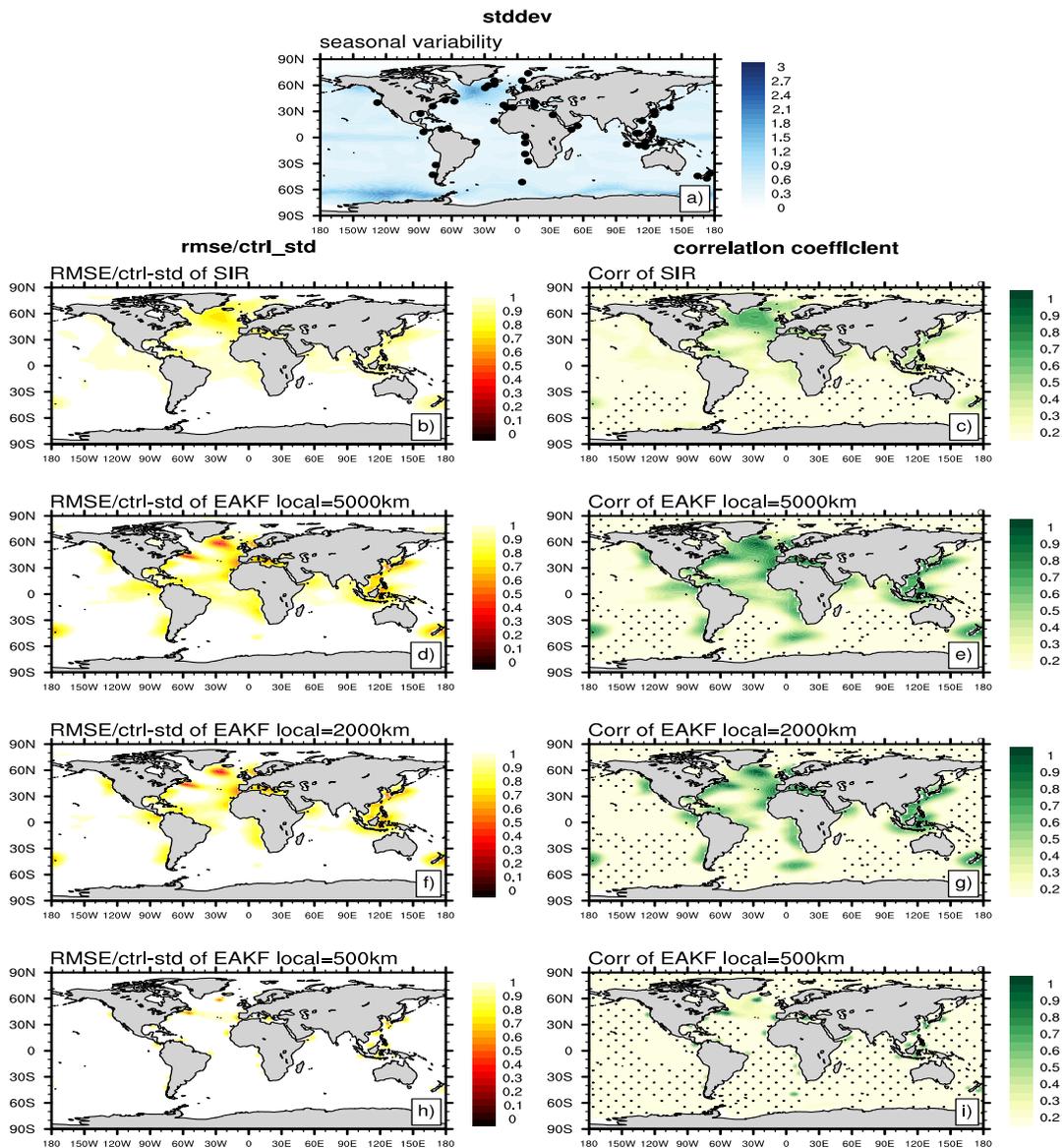
Figure 3.1: Observation locations and spatial reconstruction skill represented by RMSE and correlation coefficient for four experiments. (a) Shows the location of 47 proxies, the shading shows the seasonal natural variability/standard deviation of the SST from the truth run. The four experiments: (b and c) PF-SIR, (d and e) EAKF with localization 5,000 km, (f and g) EAKF with localization 2,000 km, (h and i) EAKF with localization 500 km. RMSE divided by natural variability from Figure 1a (Figures 1b, 1d, 1f, and 1h). The spatial correlation map between each experiment and the truth, and the dotted area is not significant at 0.05 significance level (Figures 1c, 1e, 1g, and 1i).

3.3.2 Hemispheric Averaged SST Time Series

Figure3.3 shows the hemispheric averaged SST time series, derived as area weighted mean of the analysis. For the Northern Hemisphere (NH) where most of the observations are located and where the ocean area is not overwhelmingly dominant (60% coverage) (Figure3.3d), the EAKF with 5,000 km localization scale (red line) and PF-SIR (blue line) have a correlation with the true state (black line) of 0.85 and 0.67, respectively. They both capture the decadal scale variability very well, except that during the year 30 to 80 the PF-SIR starts to suffer from particle impoverishment where it cannot select good members that match the observations out of its 48 members. This mismatch greatly diminishes the final correlation coefficient calculated between PF-SIR and the truth. To take such situations into account, we calculate the correlation coefficient between the truth curve and the other three curves in Figures 3.1a, d, and g at different time segments (Figure3.2), such as 20 years, 40 years, and 100 years. These segments are sampled from the 200 year simulation. Thus, each dot in Figure3.2 represents the correlation coefficient of a sample. We can see that the EAKF-5000 km has higher-correlation coefficient than the PF-SIR at 20 years, 40 years, 100 years, and 200 years sample length (Figure3.2a). The distribution of the dots is more compact for the EAKF-5000 km than the PF-SIR, which suggests that the performance of the EAKF-5000 km is more stable than the PF-SIR. Similar results can also be found in Figures3.2b and c. In addition, despite that PF-SIR captures the general variability in the Northern Hemisphere, there is a small phase shift between the PF-SIR and the truth. This again reflects the nature of the PF-SIR in that the analysis is reconstructed from existing ensemble

members that are closest to the observation, and the analysis will deviate from the truth if all the members are far away from the observations. In contrast, the EAKF corrects each member using the observation increment so that every member in the ensemble will be "dragged" toward the true state. Hence the final analysis has better synchronization with the truth time series compared to the PF-SIR. For the Southern Hemisphere (SH), both methods failed to track the true state due to very limited observations and the large area of ocean (80% coverage) (Figure3.3g). On the global scale (Figure3.3a), the correlation with truth for the global averaged SST is 0.61 for the EAKF and 0.24 for the PF-SIR. The low-correlation coefficient of the PF-SIR is further reduced by the poor performance from the SH. Therefore, even though the PF-SIR is deficient at local reconstruction as shown in Figures3.1b and c, it seems to be able to capture some hemispheric-scale variability if the ensemble size is sufficiently large. However, even with 3 times the computational expense, the PF-SIR with 48 members still underperforms the EAKF with 16 members in the reconstruction of hemispheric-scale SST.

The uncertainty of the final analysis is also shown in Figure3.3, as represented by its ensemble members. The second column is for the EAKF with 5,000 km localization radius (the pink curves as the 16 ensemble members), and the third column is for the PF-SIR (light blue curves as the 48 ensemble members). In general, the PF-SIR tends to underestimate the uncertainty of the model forecasts, an indication of the problem of filter degeneracy. All the 48 ensemble members cluster tightly at the hemispheric and global scales, and the spread is not large enough to encompass the truth time series. This is because the posterior ensemble
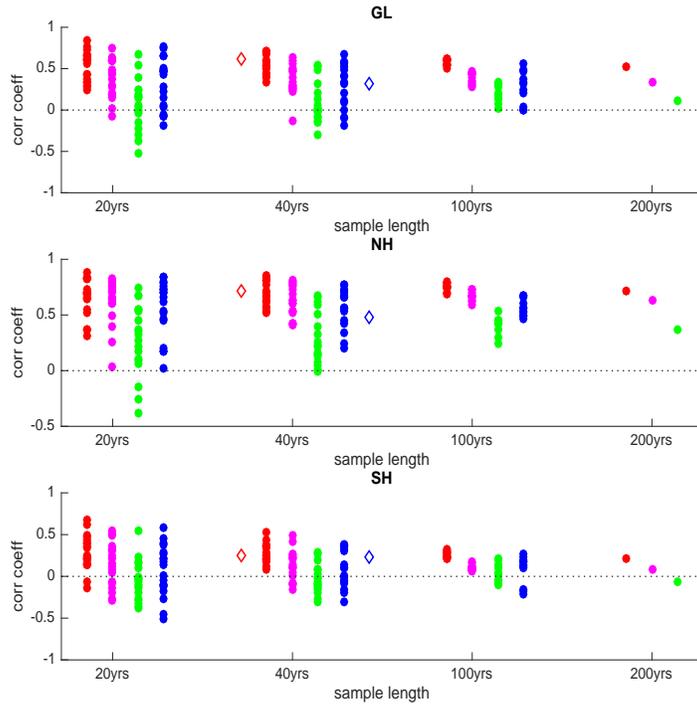
Figure 3.2: Correlation coefficients between hemispheric averaged SST time series of EAKF-5000 (red dot), EAKF-2000 (magenta dot), EAKF-500 (green dot), PF-SIR (blue dot), and that of the truth at (a) global scale, (b) Northern Hemisphere, and (c) Southern Hemisphere. Each dot represents a 20 year, 40 year, or 100 year sample from the 200 yearlong simulation. The red and blue diamonds are the correlation coefficient of the 50 yearlong experiments for EAKF with 32 members and PF-SIR with 96 members.

is heavily bounded by the prior ensemble: it is impossible to generate the posterior ensemble outside the span of the prior ensemble even when the observation is outside the span. For the EAKF, the truth series always lies within the uncertainty of the 16 ensemble members. In the NH (Figure3.3f), abundant observations lead to stronger constraint of the analysis and hence smaller uncertainty, while in the SH (Figure3.3i), weak observational constraint causes large spread among the 16 ensemble members. This suggests that the EAKF provides a better representation of the uncertainty with only 16 members than the PF-SIR with 48 members.
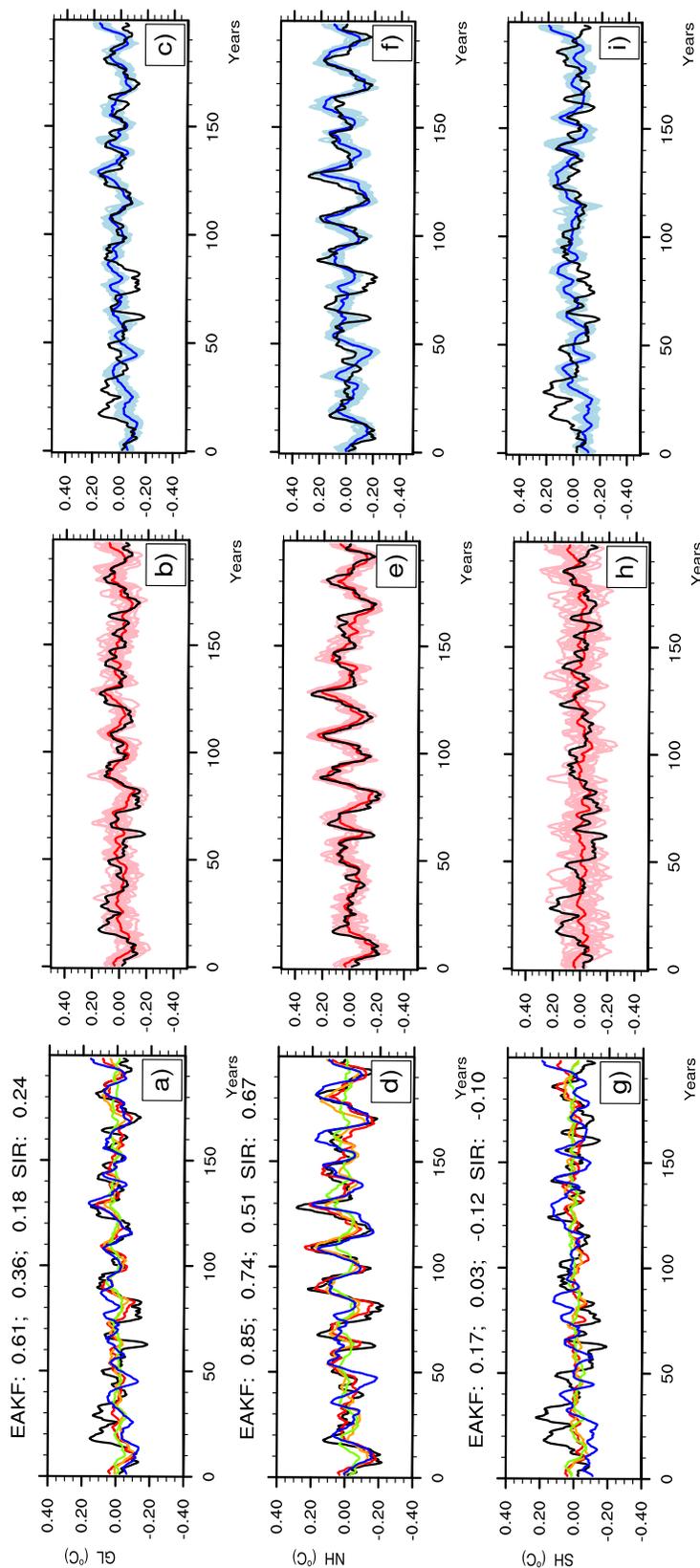
Figure 3.3: Area averaged SST anomalies for (g-i) Southern Hemisphere, (d-f) Northern Hemisphere, and (a-c) global; black curve is for truth; blue curve is for PF-SIR; red curve is for EAKF with localization scale 5,000 km; orange curve is for EAKF with localization scale 2,000 km; and green curve is for EAKF with localization scale 500 km. The correlation coefficient with truth is shown for each experiment in Figures 2a, 2d, and 2g; in Figures 2b, 2e, and 2h, the light blue curves are for 48 ensemble members from PF-SIR; in Figures 2c, 2f, and 2i, the pink curves are the 16 ensemble members from EAKF with localization scale 5,000 km.

3.3.3 Modes of Climate Variability

In this section, we investigate the ability of the two methods in capturing the modes of climate variability. Empirical orthogonal function (EOF) analysis is performed on the four PDA analyses (PF-SIR and three EAKF experiments with different localization radius) as well as the true state on different regions. The EOF patterns and principle components (PCs) of the first mode of variability are shown in Figures 3.4 and Figure3.7. In the observation-dense North Atlantic (Figures 3.4b, d, f, h, and j), both the PF-SIR and the EAKF-5000 km capture the first EOF pattern that is almost identical to the truth in both pattern and strength, and the pattern correlation is close to 1 for both experiments (Table 1). The corresponding PCs of these two methods show similar decadal variability to the truth. The temporal correlation coefficients between the PC time series of each filter and that of the truth are as high as 0.78 for EAKF-5000 km and 0.67 for PF-SIR (Table 1). However, the explained variance of the first mode is much higher in the PF-SIR (25.9%) and EAKF-5000 km (24.7%) than that in truth (18.8%), which suggests that both filters have trouble constraining the precise variability given so limited observational information. As the localization radius in the EAKF is decreased, the spatial pattern gradually becomes weaker (Figures3.4f, h, and j) and the temporal correlation coefficient also decreases (Table 1). Consistent with the grid point analysis and hemispheric analysis, a large localization scale is necessary even in relatively dense observation area.

In the North Pacific, PF-SIR (Figure3.4c) captured clearly a PDO pattern similar to the truth (Figure3.4a).The correlation coefficient between the corresponding PC

time series and the truth in the Northern Hemisphere is 0.23 for PF-SIR and 0.57 for EAKF-5000 km (Table 1). Note that there are very few observations in the North Pacific: one around the Oregon coast and a few around east China Sea (Figure3.1a). The vast majority of the ocean basin remains unobserved. Nevertheless, both PF-SIR and EAKF-5000 km can capture the PDO spatial pattern and constrain the temporal variability. However, the temporal variability of the PF-SIR is much less accurate than that of the EAKF-5000 km. The ability of the EAKF to constrain the entire basin using very few observations is due to two factors: the existence of a large-scale covarying pattern, namely PDO, and the existence of observations in the critical region, in this case, around the Oregon coast. The control of this basin-wide covariability through the critical observation location is reflected by the one-point correlation map in Figure3.8a. The observation in the crucial region therefore provides information about the phase of the large-scale covariability. With this information, the PF-SIR can select members that have the correct phase and the EAKF can nudge the impact area toward the targeted phases. In this case, EAKF requires a large localization radius such that the entire co-varied area can be corrected. As can be seen from Figures3.4e, g, and i, the spatial pattern gradually fades away as the localization scale is decreased, as does the temporal correlation with truth (Table 1).
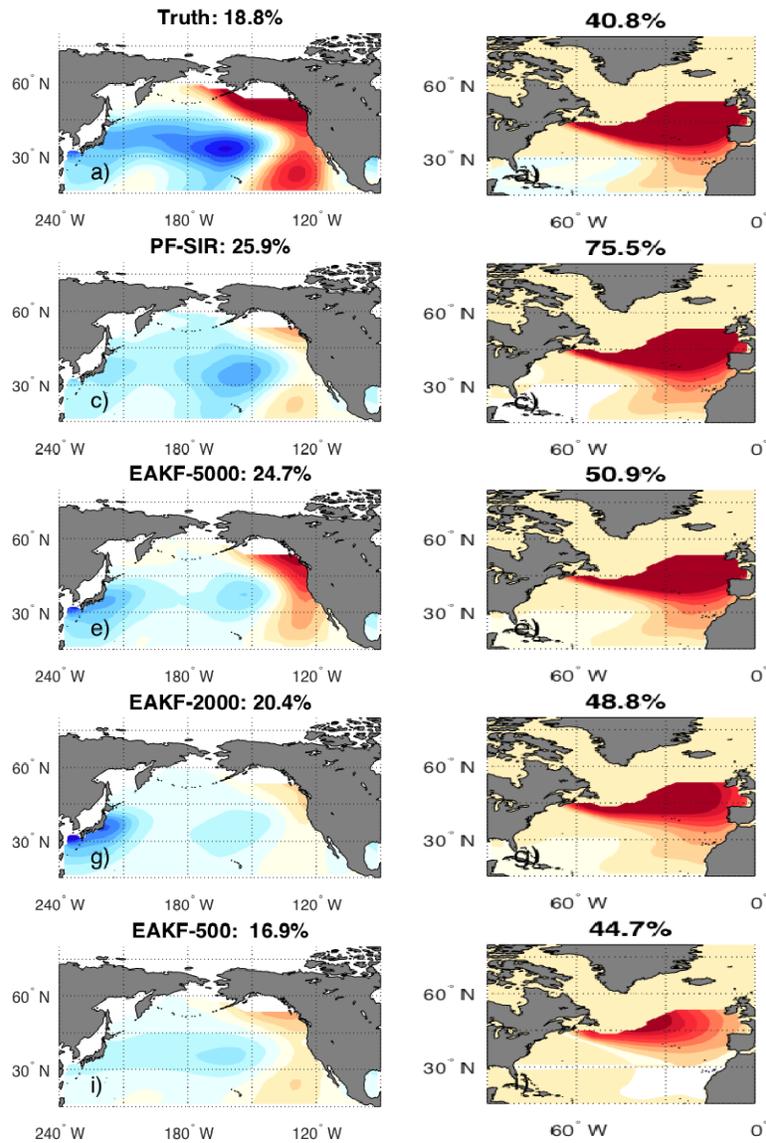
Figure 3.4: First EOF field of (a, c, e, g, and i) North Pacific hemisphere, (b, d, f, h, and j) North Atlantic. Figures 4a and 4b are for the truth state, Figures 4c and 4d are for PF-SIR, Figures 4e and 4f are for EAKF with localization 5,000 km, Figures 4g and 4h are for EAKF with localization 2,000 km, and Figures 4i and 4j are for EAKF with localization 500 km. All figures are drawn in the same color scale and the unit is celsius. On top of each figure, the percentage of explained variance is shown.
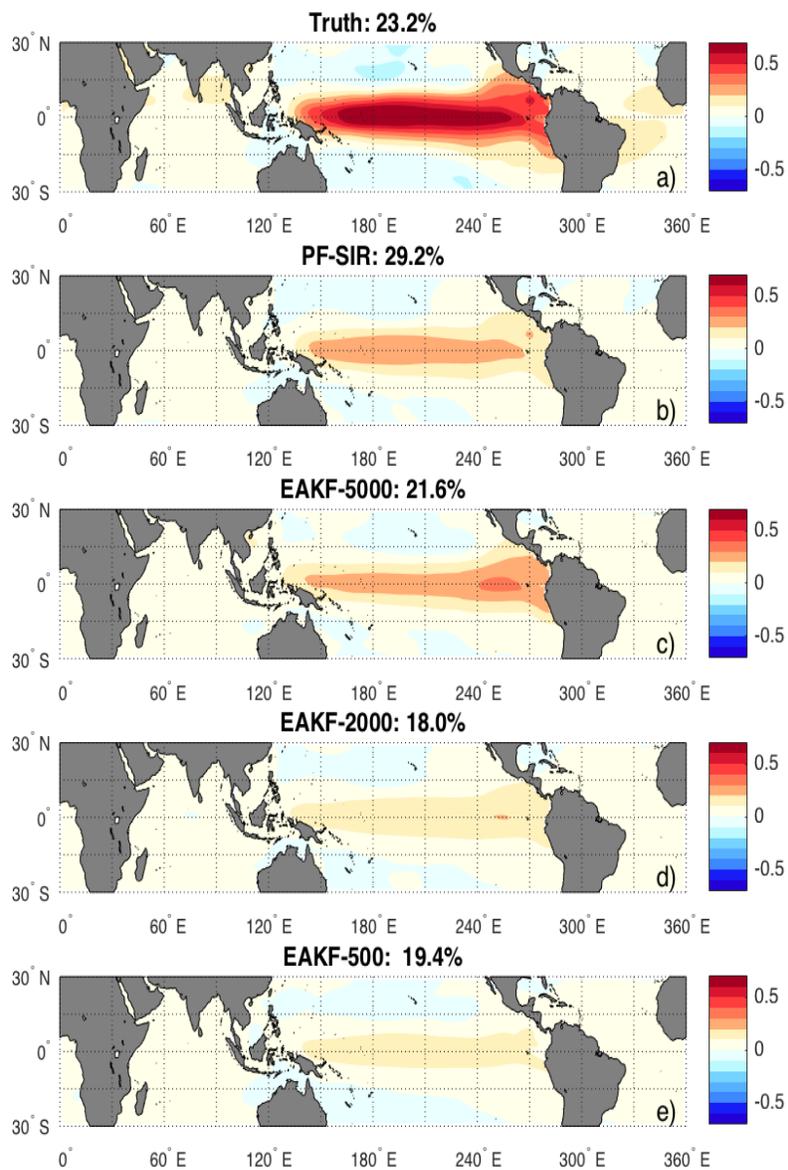
Figure 3.5: . First EOF field of the tropics are for the (a) actual state, (b) PF-SIR, (c) EAKF with localization 5,000 km, (d) EAKF with localization 2,000 km, and (e) EAKF with localization 500 km, respectively. All figures are drawn in the same color scale and the unit is celsius.

As is the case in the North Pacific, the tropical Paific also has large-scale co-variability with the observation location (Figure3.8b) and only one observation located in the far eastern portion of the basin (Figure3.1a). Like North Pacific, both PF-SIR and EAKF-5000 km captured the pattern of ENSO with pattern correlation coefficient close to 1, although the magnitude is weaker than truth. The EAKF-5000 km shows better temporal variability, as the correlation in the first PC is 0.43 compared to 0.25 for the PF-SIR. The weaker spatial pattern but better temporal variability in the tropics and North Pacific for the EAKF compared to the PF-SIR is due to the fundamental differences of the two filters. The analysis of EAKF is ensemble mean output, thus when observational constraint is weak, the ensemble mean tends to average out the internal variability and result in a weaker spatial pattern. Also, recall that in the EAKF the physical field in each member is modified by the observation innovation. Hence, given the phase information of the large-scale pattern from the key observation, the EAKF will directly correct all members and drag them toward the correct phase. The PF-SIR does not have this feature because the analysis of the PF-SIR is a weighted sum across all members (the physical field in each member is not modified), and the weight attached to each member is based on observations from all locations. Therefore, the local constraint of a given observation will be mitigated by observations from other locations. Thus, the temporal constraint of these key observations on their neighboring regions is weaker in the PF-SIR than that in the EAKF. As localization radius is decreased in the two additional experiments of the EAKF, the first EOF pattern gets weaker in the tropics and the PC correlation with truth also decreases. When localization

radius is decreased to 500 km, the EAKF barely has any constraint of the large-scale variability in this area (Figure3.7 and Table 1).
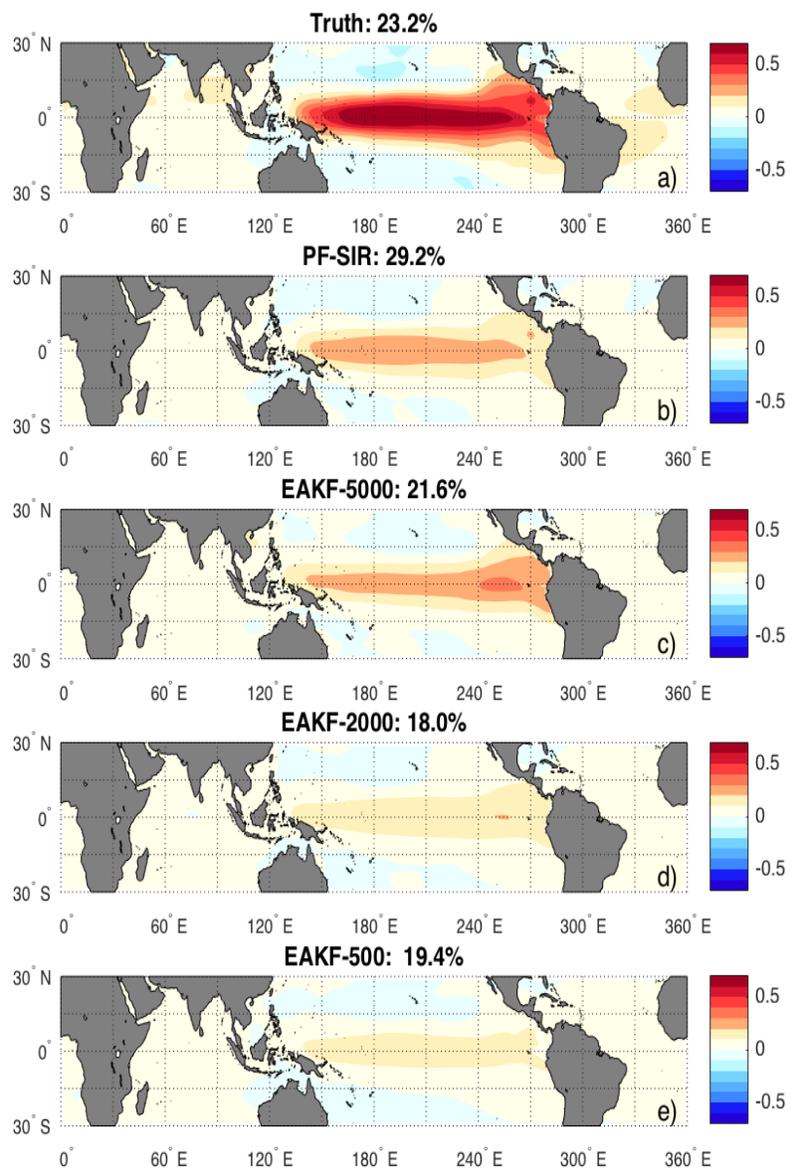
Figure 3.6: . Same as Figure 4 but for the Southern Hemisphere.
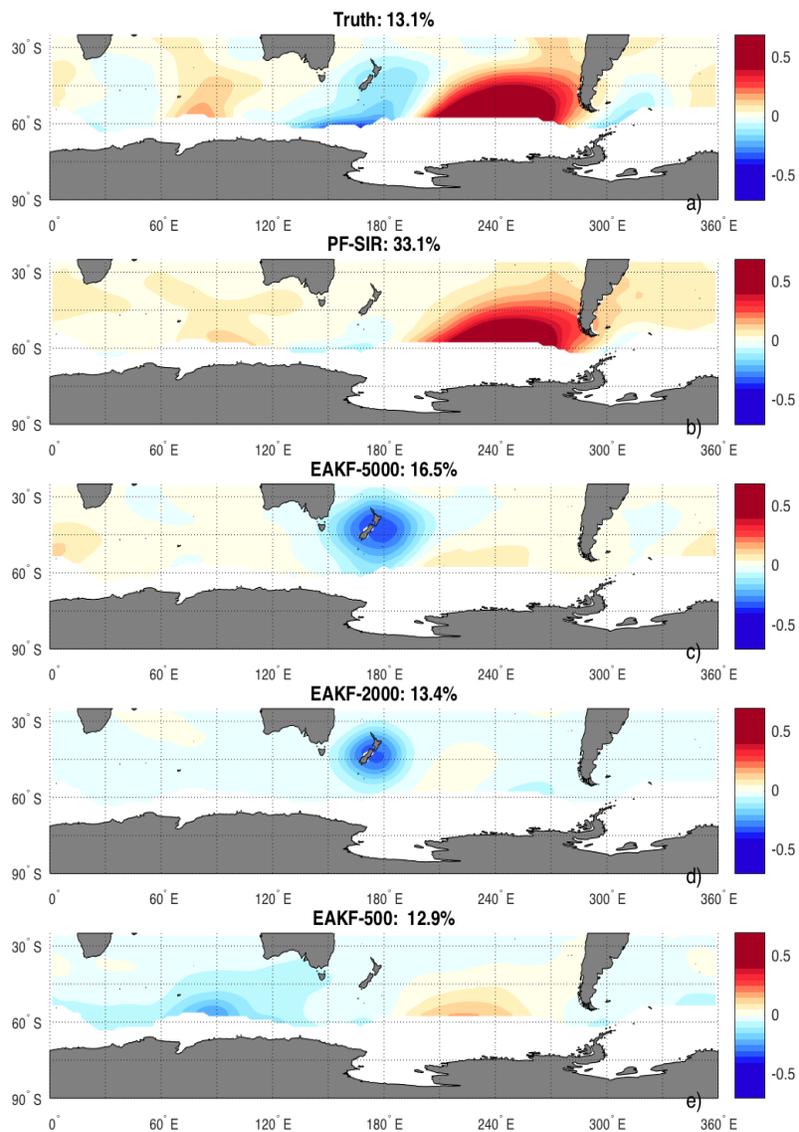
Figure 3.7: . The normalized PCs for the first EOFs from Figures 3, 4, and 8 for North Pacific, North Atlantic, Tropics, and Southern Hemisphere. The black curve is for truth; the blue curve is for PF-SIR; the red curve is for EAKF with localization scale 5,000 km; the magenta curve is for EAKF with localization scale 2,000 km; and the green curve is for EAKF with localization scale 500 km.

In the SH where there are very few observations, both methods failed to reconstruct the leading modes of similar patterns to the truth (Figure3.6). The spatial field of the PF-SIR bears some similarities to the truth in the South Pacific, but the temporal variability of this pattern has no correlation with that of truth at all. The three EAKF experiments capture a pattern that covaries with the truth with a correlation coefficient around 0.3 (Table3.1), but these patterns do not resemble the truth that much. The EAKF-5000 km maintains some performance near observation locations such as around South Atlantic and the west coast of Chile and New Zealand (green patches in Figure3.1e). But these limited observations are far from enough to constrain the entire Southern Hemisphere considering that there is no large-scale covariability in the Southern Hemisphere as in the Northern Hemisphere (Figures3.8c and d). Hence, it is not surprising that both methods perform poorly in the Southern Hemisphere (Table3.1).

	Hemispheric Scale			Modes of variability			
	NH	SH	GL	North Pacific	North Atlantic	Tropics	SH
PF-SIR	0.67	0	0.24	0.23 (0.85)	0.67 (0.99)	0.25 (0.98)	0 (0.95)
EAKF-5000km	0.85	0	0.61	0.57 (0.85)	0.78 (0.98)	0.43 (0.95)	0.25 (0.46)
EAKF-2000km	0.74	0	0.36	0.42 (0.67)	0.70 (0.99)	0.30 (0.97)	0.27 (0.23)
EAKF-500km	0.51	0	0	0 (0.95)	0.44 (0.99)	0 (0.98)	0.30 (0.54)

Table 3.1: Correlation Coefficients Between Different Experiments and Truth Based On Hemispheric-Scale Averaged SST Time Series From Figure3.2; Correlation Coefficients for PCs Series Between Each Experiment and the Truth for the First Mode From Figure3.6. Note. Pattern correlation coefficients for the leading mode EOF between each experiment and the truth (in brackets). The correlation coefficients that are not significant at 0.05 confidence level are marked as zero.
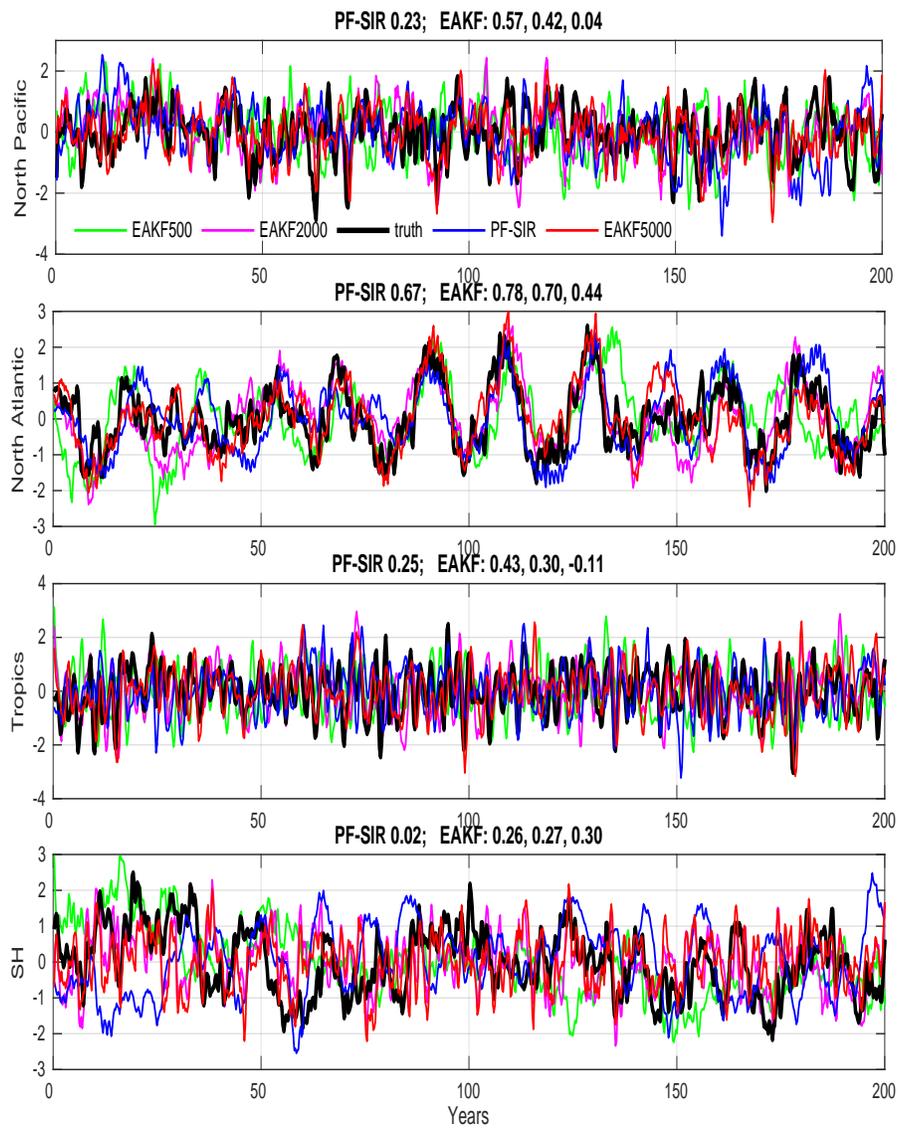
Figure 3.8: . One-point correlation map calculated from seasonal mean model control at selected observation locations. The black dot represents the observation location. The dotted area is not significant at 0.05 significance level.

3.3.4 Reconstruction Skills on Other Variables

From the reconstructed SST analysis, we see that the EAKF-5000 km has better performance than PF-SIR is both the local reconstruction and hemispherically averaged reconstruction. In this section, we will take a look at the spatial reconstruction skills on other variables, namely air surface temperature (AST), 500 hPa geopotential height (Z), and precipitation (PRECP) for both filters (Figure3.9). Since RMSE ratio map shows similar skills to spatial correlation map, we only show correlation map for these variables. The correlation coefficients are calculated over the entire 200 year simulation after smoothing the time series at each grid point to remove the high-frequency noise in the atmosphere. The AST reconstructions (Figures3.9a and b) are similar to the SST reconstruction (Figures3.1c and e) over the ocean area but have a slightly larger impact region away from the observation location (Figures 3.9a and b). Over land, the reconstruction skills mainly exist in coastal regions where there are observations in the nearby ocean, such as eastern North America, Greenland, North Africa, and East Asia. Again, EAKF outperforms the PF-SIR by showing larger area of high-correlation coefficients especially in the Pacific Ocean and the South Atlantic Ocean. For 500 hPa geopotential height, the patterns are similar to AST (Figures3.9c and d). Both PF-SIR and EAKF have poor reconstruction skills for precipitation. This is because that precipitation is not only influenced by temperature but also strongly influenced by the moisture distribution and the wind field. Therefore, it is not surprising to see such poor reconstruction skills on precipitation from both filters.

Correlation Map
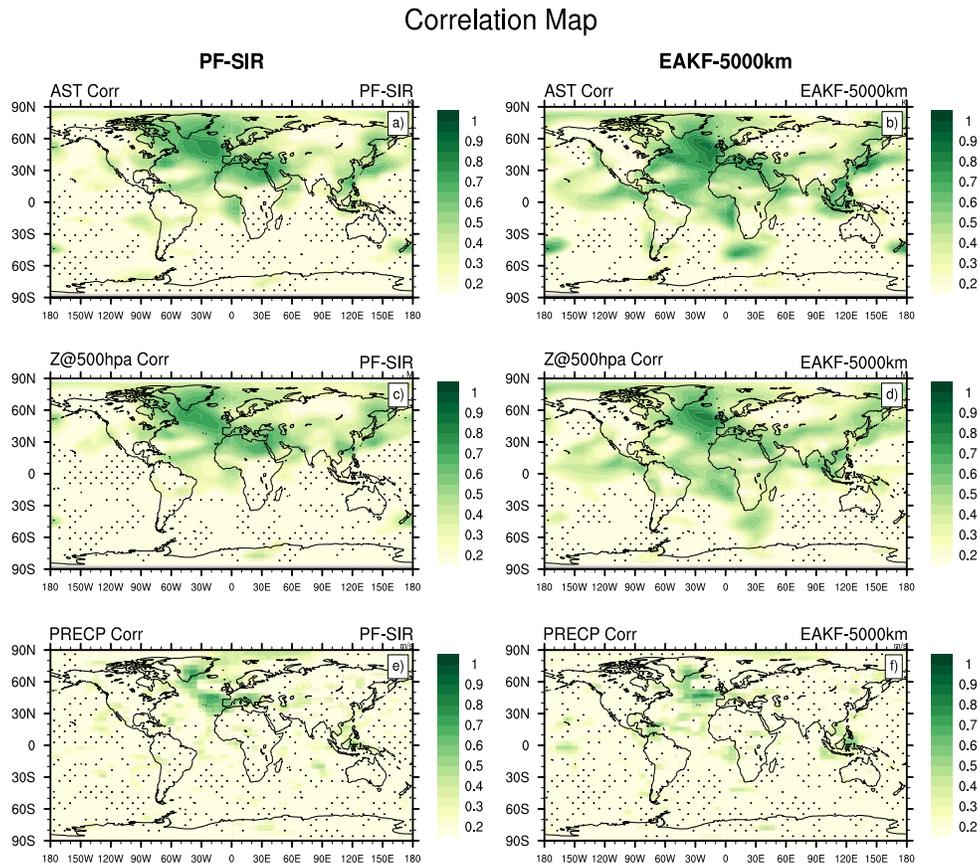
Figure 3.9: . Spatial reconstruction skills for (a, c, and e) PF-SIR and (b, d, and f) EAKF-5000 km on air surface temperature (AST, Figures 9a and 9b), 500 hPa geopotential height (Z, Figures 9c and 9d), and precipitation (PRECP, Figures 9e and 9f).

3.3.5   Sensitivity to Ensemble Size and Localization Scale

Considering a sparse observation network in PDA, the large localization radius is crucial for EAKF to extend the observation impact and hence to constrain large-scale variability as seen from previous sections. This leads to the question on the optimal localization radius for such PDA application. To find such optimal

localization radius, we further increased 5,000 km to 10,000 km in the EAKF. This is almost equivalent to that each observation updates the entire globe. We found that in terms of local reconstruction skill, North Pacific and tropical Pacific have stronger PDO and ENSO pattern compared to the EAKF-5000 km, but except these regions, the patterns in other regions are almost identical (figure not shown). In terms of hemispheric averaged SST time series, the correlation for the EAKF-10000 km is even slightly smaller than the EAKF- 5000 km. This is because in regions that do not have the large-scale covariability, a large localization radius allows sampling error to come into play, which eventually deteriorates the hemispheric-scale performance. Therefore, although large localization scale is necessary for PDA applications, it is not "the larger, the better". This suggests that the optimal localization radius for PDA applications is not a globally uniform value but a spatially varying parameter. Zhen and Zhang (2014) have proposed a variational approach to adaptively determines the optimum radius of influence for ensemble covariance localization. However, the spatially varied localization parameter is beyond the scope of this paper but is worthy of future investigation.

Ensemble size is an important factor for ensemble filters, especially for the PF-SIR that requires an ensemble size that increases exponentially with the dimension of the system (Snyder et al., 2008), and previous studies adopting online PF-SIR uses 96 ensemble members (Dubinkina et al., 2011; Goosse et al., 2011; Goosse, 2012). We double the ensemble size of the EAKF-5000 km and PF-SIR and run the experiment for 50 years to see if this will greatly improve the performance of the two filters and potentially change the relative performance. For the EAKF-5000 km,
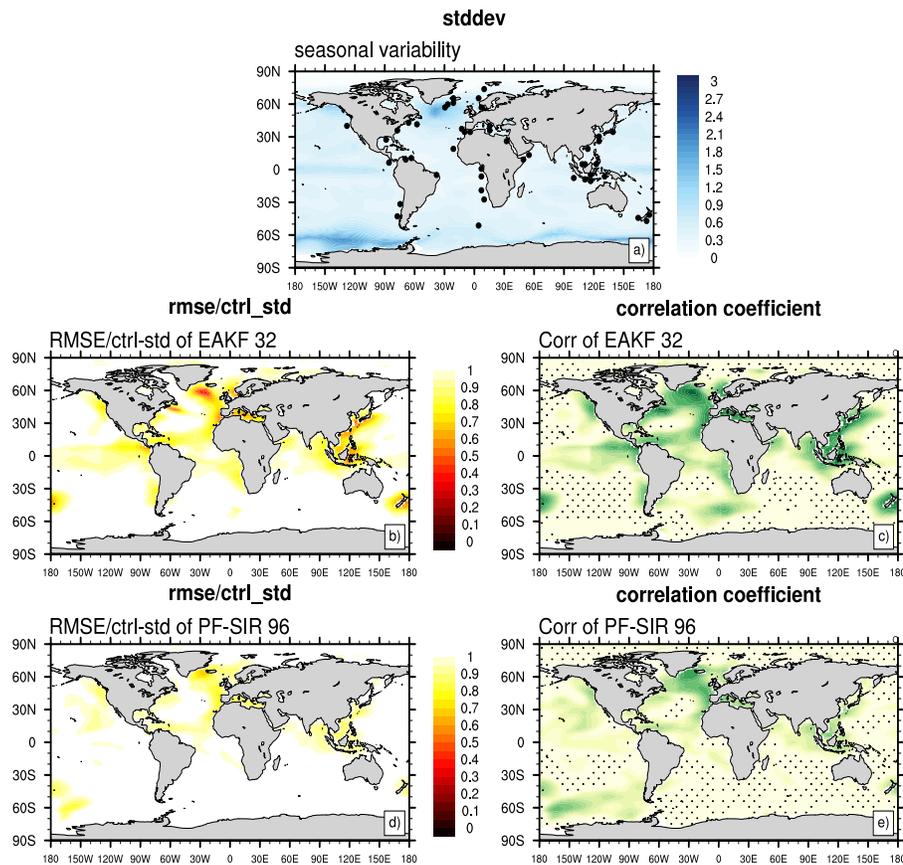
Figure 3.10: . Same as Figure 1 but for (b and c) EAKF-5000 km with 32 ensemble members and (d and e) PF-SIR with 96 ensemble members.

doubling the ensemble size yield slight improvement in local skill (Figures3.10b and c) compared to previous case using only 16 members (Figures3.1d and e). One noticeable improvement is in the tropical Pacific. For the Northern Hemispheric SST time series, the correlation with truth has increased from 0.79 to 0.91. The global mean temperature correlation has increased from 0.54 to 0.72 (Figure3.2). For the PF-SIR, there is noticeable improvement in the local reconstruction scale (Figures3.10d and e), such as in the North Pacific and South Pacific. But for hemispheric averaged

SST, there is no significant improvement. EAKF-5000 km still outperforms the PF-SIR when ensemble size is doubled for both filters. In fact, EAKF-5000 km with only 16 members still significantly outperforms the PF-SIR with 96 members in both local skill and hemispheric SST time series reconstruction (Figure3.11).
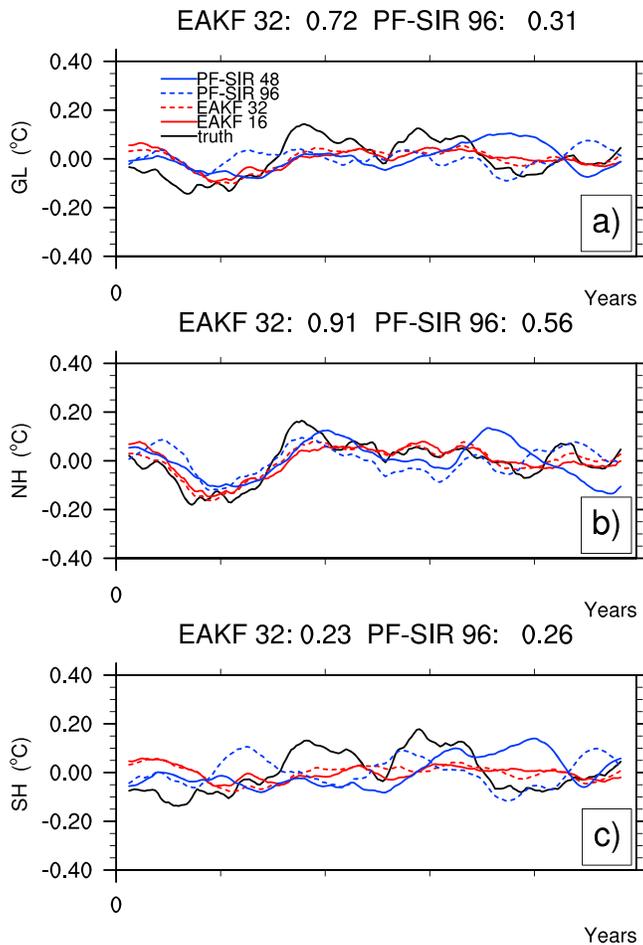
Figure 3.11: Same as Figures 2a, 2d, and 2g but for EAKF-5000 km with 32 ensemble members (red dashed line) and PF-SIR with 96 ensemble members (blue dashed line)

3.4   Summary and Discussion

Particle filter with simple importance resampling and ensemble Kalman filter are emerging as two popular ensemble-based data assimilation methods for paleo-climate studies. Here, using a CGCM (FOAM), we systematically compare the performance of the two filters in assimilating seasonally averaged observations of SST in PPEs and provide some insights into the choice for PDA methodology. It is found that EnKF has overall better performance than the PF-SIR using only one third the number of ensemble members. On local scales, EnKF shows good reconstruction skill in the Northern Hemisphere where observations are relatively abundant, as well as in regions that have few observations but has large-scale variability of coherent pattern, such as the tropical Pacific and the North Pacific. For the PF-SIR, the correlation with truth is much smaller compared to the EnKF and only the north Atlantic region has nontrivial correlation coefficient. In the hemispheric reconstruction, both methods can capture the decadal variability in the data-rich Northern Hemisphere and lose skill in the data-sparse Southern Hemisphere. In general, the EnKF has higher correlation with the actual time series than the PF-SIR. PF-SIR is also subject to particle impoverishment where all members end up far from the observations and hence few particles can be selected out of its 48 ensemble members. When such degeneracy occurs, the reconstruction will deviate from the truth substantially as shown in the year 30-80 in the 200 yearlong experiment (Figure 3.2). Doubling the ensemble size of the PFSIR improves the local reconstruction, but the overall reconstruction skill is still smaller than the EnKF with only 16 members. Hence, we conclude that the EnKF is more consistent and

stable with much cheaper computation cost compared to the PF-SIR. The success of the EAKF over the PF-SIR in this PPE is attributed to the fact that (a) Gaussian assumption is appropriate for this application, and (b) the EAKF is less sensitive to sampling errors than PF-SIR due to different natures of the two filters. Sixteen members is enough to capture the covariance used in equation (3), but 48 (even 96) members still underrepresent the state space of such high-dimensional system. This underrepresentation is manifested as the bounding of the posterior particles of prior ensemble, since it is impossible to generate posterior samples outside the span of the prior ensemble even when the observation is far outside of the prior ensemble. Therefore, a sufficient ensemble size is critical to the PF-SIR, and this "sufficient" ensemble size is much greater than that required by the EAKF. Note that one important practice of applying EAKF to PDA is the choice of a large localization radius since there is large-scale covariability on slow time scales. This practice benefits the covariance-based filter substantially. As the localization radius of the EnKF is decreased in the two addition experiments, the performances on both the local scale and hemispheric scale are diminished. This localization radius effect is most obvious in regions that have large-scale covariability such as tropical Pacific and the North Pacific. One observation in these regions is able to constrain the entire basin given a larger observation impact radius. Three EAKF experiments with different localization scales have clearly illustrated this point, but further study can be conducted to apply different localization radius at different grid points: regions that have large-scale covariability are assigned larger localization radius, while regions that do not have such feature or have relatively dense observations are assigned

smaller localization scale. This localized localization radius network will exploit the sparse observation information and at the same time avoid the long-distance spurious correlation, which will be very suitable to PDA applications. These results raise the question of which filter should be preferred in the future for paleoclimate studies. In our experiments, we show that the EnKF with only 16 members outperforms the PF-SIR with 48 (even 96) members in almost every way. In theory particle filter allows non-Gaussianity, which is an advantage over the EnKF. The challenge for particle filter is, however, the "curse of dimensionality". To better apply particle filter for paleoclimate studies, more techniques such as localized and clustered particle filter (Poterjoy, 2015; Lee and Majda, 2016) should be considered instead of using just the particle filter with simple importance resampling.

Chapter 4

Un-mixing Deep-sea Sediment Records: a Study on Bioturbation Effects through Convolution and Deconvolution

4.1   Introduction

Post-deposition sediment reworking by benthic macro fauna during their biological activities induces mixing of the climate signals recorded in marine sediment cores. This mixing, which is termed bioturbation, acts to smooth the signal as well as induce phase shifts (Goldberg and Koid, 1962; Berger and Heath, 1968; Guinasso and Schink, 1975; Schiffelbein, 1984a; Bard et al., 1987; Anderson, 2001a; Trauth, 2013). Because of the stochastic and complicated nature of this process, its effects on climate time series, such as $\delta^{18}O$-derived temperature and/or ice volume derived from marine sediment cores, are seldom quantified and evaluated before further climate inferences are made (Shackleton et al., 2003; Rohling et al., 2008; Turney and Jones, 2010; Dutton et al., 2015a). This could be problematic in some environmental settings because the amplitude reduction and phase shift induced by bioturbation

will affect the timing and magnitude of the climate events, which could alter the original climate interpretations that are drawn from the relative phasing of proxy records located in different geographic regions (Waelbroeck et al., 2011; Shakun et al., 2012; Zhang et al., 2017). In addition, bioturbation can add potential inconsistency between proxy observation and model simulations in model-data comparison since the model output typically represents a climate signal without post-deposition reworking. Therefore, modeling the bioturbation and quantifying its impact on preserved climate signals has profound implications for climate studies of the past.

Since 1960, various models have been proposed to simulate bioturbation, which generally fall into two categories based on whether they include a stochastic component or not. The deterministic component of bioturbation effects can be characterized by an impulse response function (IRF) and the subsequent mixing process can be modeled as a time-invariant convolution (Ruddiman and Glover, 1972; Goreau, 1977; Hutson, 1980; Ruddiman et al., 1980b; Schiffelbein, 1984a; Schiffelbein and Dotman, 1986; Bard et al., 1987; Anderson, 2001a). Estimates of the IRFs can be provided from instantaneous geological deposits, such as volcanic layers, that capture the deep-sea sediment mixing process (Ruddiman and Glover, 1972; Ruddiman et al., 1980a) and/or derived from mathematical equations that model bioturbation (Schiffelbein, 1984b; Schiffelbein and Dotman, 1986; Bard et al., 1987; Guinasso and Schink, 1975; Berger and Heath, 1968; Goldberg and Koid, 1962). This deterministic approach facilitates the construction of an inverse model that can be used to remove the bioturbation effects from the preserved climate signal. In contrast, stochastic bioturbation models simulate individual proxy-carriers or sediment particles, and

the physical mixing process is modeled as stochastic permutations (Foster, 1983; Shull, 2001; Choi et al., 2002; Trauth, 2013) or a random walk of these particles (Hull et al., 2011; Turner et al., 2017).While the stochastic bioturbation models generate more realistic mixing output and allow non-local mixing (Boudreau, 1986a; Meysman et al., 2003), inverting the processes is usually not straightforward due to the stochastic component.

In this chapter, we quantify the bioturbation effects under a range of possible conditions across events of different time scales. Specifically, we aim to evaluate the benthic foraminifera $\delta^{18}O$ measurements for the last interglacial period (LIG, 129-116 kyr ago) or the so-called Marine Isotope Stage 5e (MIS 5e). Benthic $\delta^{18}O$ records are used to estimate global ice volume and deep ocean temperature (Ravelo and Hillaire-Marcel, 2007; Kennett et al., 2006). An 0.1‰$\delta^{18}O$ change is related to about 10m of global sea level change. Without evaluate the bioturbation effects imbedded in the proxy record, the subsequent global ice volume and sea level estimates can be greatly underestimated. MIS 5e is the most recent interglacial before the Holocene, when the global mean temperature is estimated to be at least $2°$C warmer (Clark and Huybers, 2009; Otto-bliesner et al., 2009; McKay et al., 2011) and sea level is estimated as 4-9m higher than present day (Veeh, 1966; Kopp et al., 2009). This time interval is frequently used as a comparable analog for future sea level projections under global warming (Rohling et al., 2008; Kopp et al., 2009; Dutton and Lambeck, 2012; Dutton et al., 2015a). To reverse the mixing process, we adopt the deterministic modeling approach and propose an inverse modeling toolkit using deconvolution via Fast Fourier Transform (FFT). Our toolkit will be made

accessible to the community in order to remove potential bioturbation effects on any climate time series from paleoproxy derived data. There is a special pre-processing needed to deal with causality issues of our application, and post-processing is needed to remove amplified noise, which is elaborated upon the methodology section. The entire procedure will be implemented in a R package for open access. In section 2, a comprehensive review of the theoretical foundation of this method is presented. In section 3, artificial proxy records are created and evaluated to test the stability of the algorithm and examine the deconvolution results in cases when the IRF is incorrectly parameterized. In section 4, this method is applied to the globally averaged benthic foraminifera oxygen isotope stack of LR04 (Lisiecki and Raymo, 2005) and to an individual record GeoB 1117 tuned in LR04. Section 5 presents additional discussion about the limitations and caveats when using this method.

4.2   Methodology

4.2.1   Impulse Response Function

For IRFs that characterize bioturbation, we choose the analytical solutions from the Guinasso and Schink (1975) diffusion model. They are appropriate for most benthic mixing environments and span the range of ash-layer profiles commonly found in deep-sea sediments (Schiffelbein, 1984a). In this model, bioturbation is modeled as down-gradient transport analogous to molecular or eddy diffusion. The mixing effects are governed by three parameters, sedimentation rate v (cm/kyr), mixed layer depth L (cm) and mixing intensity D (cm^2/kyr). Figure 4.2a shows the

IRFs under different bioturbation parameters. The shape of the IRF is determined from the dimensionless parameter $G = D/Lv = 1/Pe$, where Pe is the Peclet number and G represents the relative importance of diffusion to advection. Larger Gs indicate stronger diffusion/bioturbation and the IRFs have greater asymmetry (e.g. Figure 4.2). The most important feature of bioturbation is that the mixed event peak always lies below the depth of original deposition because of the subsequent upward mixing. This downward displacement of the mixed event peak makes the onset appear older and the degree of this downward displacement is associated with the skewness of the IRF and ultimately determined by G. As G gets smaller, the IRF becomes thinner and more symmetric, and the mixed peak approaches the original depth of deposition. As G approaches zero, the IRF thins into a vertical line.

This biodiffusive model, although simple, has been shown to be sufficient in simulating major features of bioturbation (Schiffelbein, 1984a). It provides a generic IRF that can systematically model all possible bioturbation scenarios via different combinations of L, v and D. Note that the IRFs shown in Figure 4.2a are unitless. One unit represents one mixed layer depth L (in cm), or one-unit time (L/v in kyr) that it takes to travel through the mixed layer. Hence given the same G, but different D, L and v combinations, the actual depth that the IRF spans also varies (Figure **??**). In summary, G determines the IRF shape, while L and v determines the depth/time span of the IRF.
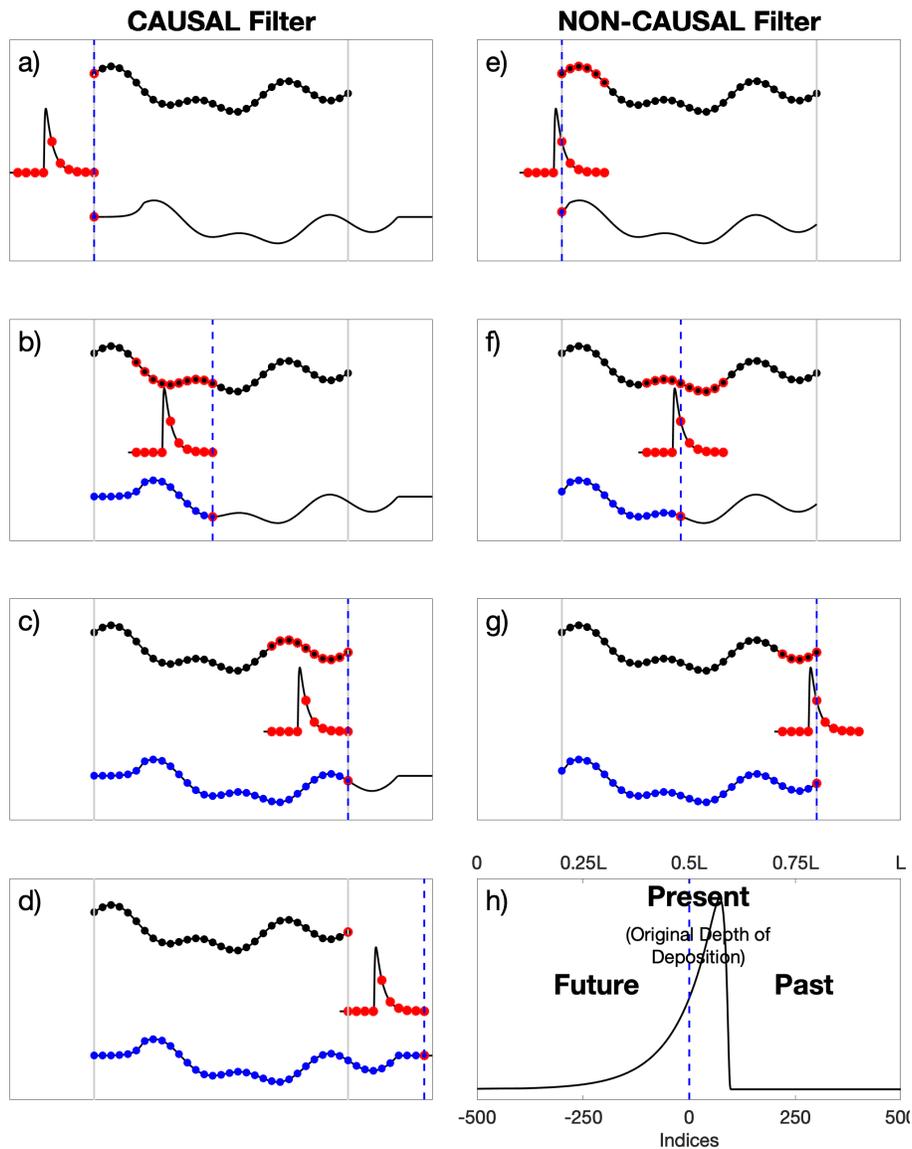
Figure 4.1: Illustration figures for convolution using causal (a)-d)) filter and non-causal (e)-g)) filter. The black dotted curves in a)-g) are the same input time series and the blue dotted curves are output time series. The red dotted curves are the filter. The blue dashed vertical line represents current step t. The current output value y[t] in the blue dotted curve is highlighted by a red edge. The points in the input time series used to calculate the current output y[t] are also highlighted with red edges. Note the phase shift in the final blue curve in d) and g). h) is an illustration of our non-causal IRF from Guinasso and Schink (1975).

4.2.2 Convolution and deconvolution via FFT

After the original climate signal x is deposited at the sediment-water interface, it undergoes bioturbation/mixing, which is modeled using a linear time-invariant filter with an IRF designated h. The output sequence y at time t (or depth z) is a weighted sum of the input series that occurs before (below) and after (above) it (Figure 4.1e-g and Equation 1). Note that for convolution in the time domain (Figure 4.1), the IRF is flipped before the weighted sum is performed, and the IRF and time series in Figure 4.1 are all in depth order (young -> old). m_o in Equation (4.2.2) is the original layer of deposition which corresponds to the red point in the IRF that is in line with the blue dashed curve in Figure 1 e-g. M is the length of the IRFs, which is 9 in Figure 4.1. The points in sequence x that are needed to calculate the current $y[t]$ are highlighted by red edges in Figure 4.1 e-g, and ranges from $t - (M - m_o)$ to $t + m_o - 1$ (Equation (4.2.2)). Hence $y[t]$ contains values mixed from both the past (deeper depth) and the future (shallower depth), and the system is not causal.

$$y[t] = (x * h)[t] = \sum_{n=t-(M-m_o)}^{t+m_o-1} x[n]h[t + (m_o - 1) - n + 1] \tag{4.1}$$

In frequency domain, the convolution is simply the point-wise multiplication of the frequency responses of the input sequence $x[s]$ and IRF $h[s]$ according to the Convolution Theorem.

$$x * h \iff X[\omega]H[\omega] \tag{4.2}$$

Where $X[\omega]$ and $H[\omega]$ are the frequency responses of x and h. The advantage of utilizing the frequency domain analysis is that the inverse process (deconvolution) is simple to compute: when $y[s]$ and $h[s]$ are known, $x[s]$ can be easily solved by an inverse Fourier Transform via the division of $Y[\omega]$ over $H[\omega]$:

$$x[s] = F^{-1}(\frac{Y[\omega]}{H[\omega]}) \tag{4.3}$$

For discrete time series, to obtain a unique solution for deconvolution, the theorem requires the system to be linear, causal and time-invariant. A causal system is illustrated in Figure 4.1a-d and Equation (4.2.2), where $y[t]$ is calculated using only the current and past points in x.

$$y[t] = (x * h)[t] = \sum_{n=t-(M-1)}^{t} x[n]h[t-n+1] \tag{4.4}$$

The final output in Figure 4.1a-d has a phase/time-shift compared to our non-causal system Figure 4.1e-g. Beyond this time shift, the two outputs are identical in terms of shape and amplitude excluding the end portion on both sides. Since the computer algorithm only deals with positive indices, the FFT convolution output for our non-causal filter application can be obtained by adding some shifts to the output of the causal IRF during post-processing. The number of shifts to be added is determined by m_o as can be seen by Figure 4.1d and Figure 4.1g. Conversely, to deconvolve a bioturbated series (inverse modeling), the bioturbated series has to be advanced in time during pre-processing to imitate the output of a causal

filter. The procedure takes the output (blue dotted curve) in Figure 4.1g, and advances it in time (shifting it downcore in depth) to align it in phase with the blue dotted curve in Figure 4.1d. Again, the number of advances is associated with m_o. The short-of-length issue can be circumvented by zero-padding on both ends of the bioturbated sequence, which will introduce some bias at the end portion of the recovered sequence but will guarantee the correct phase. This zero-padding preprocessing is similar to the common zero-padding practice on one end of a time series when performing convolution on two series of unequal length, but with a different purpose. After the time-shifting processing, deconvolution through FFT can be performed on the padded time series using our shifted IRF, to reconstruct the un-bioturbated signals. To avoid the edge effects, a longer time series than the period of interest should be taken. A cosine taper is applied to the time series to remove the discontinuities at the two ends, and more details about the parameters of the the taper are given in the supplemental section.

4.2.3 Algorithm output

Figure 4.2b shows the bioturbated results of an input climate signal (black line) using the IRFs shown in Figure 4.2a. The input time series (black line) is the surface temperature evolution extracted from a model gridpoint in the Northeast Atlantic from TraCE 21ka, a transient simulation of the last deglaciation (He, 2011; Liu et al., 2009). The input climate signal covers several abrupt climate changes such as the the cooling during the Heinrich event 1 (H1, $\sim$ 17 kyr), the Allerød warming (BA, $\sim$ 14.5 kyr), Bølling warming ($\sim$13 kyr), and the younga dryas (YD, $\sim$ 12 kyr). It reflects
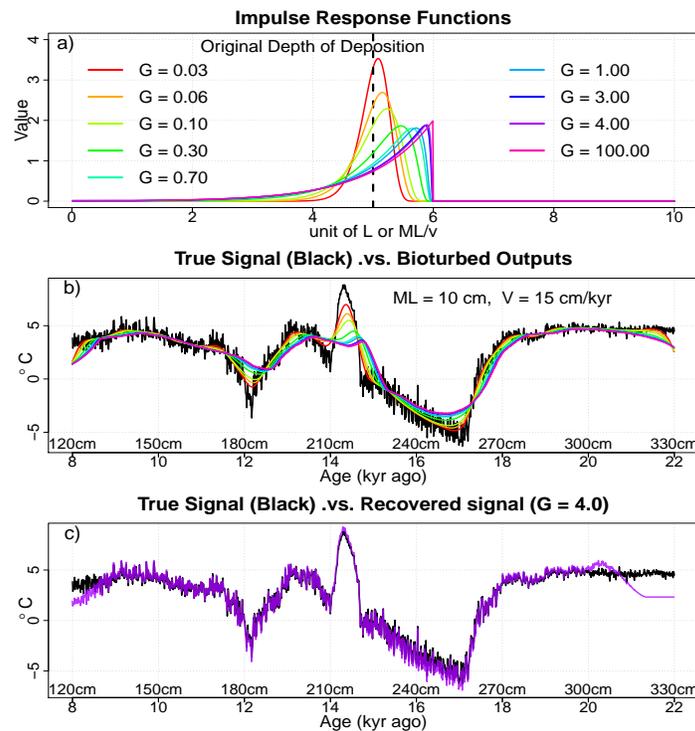
Figure 4.2: Convolution and deconvolution of a time series. a) shows the IRFs of different G values from **?**. The black dashed line is the original depth of deposition. b) shows the convolution output between IRFs from a) and a time series that has both millennial scale variability and high frequency noise. Mixed layer depth is 10 cm and sedimentation rate is 15 cm/kyr in all these cases. axis shows both age (kyr ago) and depth (cm). c) the purple line in c) is the recovered signal using the purple lines (G = 4.0) in b) and a). The back line is the input time series in b).

the high-frequency noise, short climate signals as well as longer-term signals like the transition out of the glacial conditions. Assuming a bioturbation environment that has a mixing layer depth 10 cm and a sedimentation rate of 15 cm/kyr, the IRFs in 4.2 act as a low-pass filter, and all the high-frequency noise in the input signal is smoothed out. As G gets bigger, the input climate signal experiences higher attenuation and larger phase shift. For example, the Bølling warming event at 14.5 kyr (depth 214 cm) has shifted close to 15 kyr (depth 221 cm) after G exceeds 3. This abrupt event is damped more than 5 times as G exceeds 1. Another important feature is that events of different duration are subject to different attenuation given

the same bioturbation environment (IRF). Events of shorter duration experience larger attenuation in amplitude compare to those of longer duration. The Allerød warming event is much less attenuated compared to the Bølling event under all IRFs. When G exceeds 0.7, the Allerød event is even more pronounced than the Bølling event after bioturbation. This different attenuation on events of different time scales under the same bioturbation environment has also been discussed in Anderson (2001a).

Figure 4.2c shows the recovered true signal from the bioturbated series G = 4.0 (purple line, in Figure 4.2b) and its corresponding IRF in Figure 4.2a. This is the perfect case scenario where we know exactly what bioturbation parameters (characterized by L, v, G) the true signal has experienced and there is no random mixing associated with the bioturbation, and no sampling errors are present. Therefore, the recovered signal is almost identical to the true signal except for the edge effects at the two ends of the series. The length of the area that is affected by the edge effects is associated with the parameter used in the cosine taper during the data pre-processing before FFT deconvolution is performed. When performing a FFT on a signal of finite length, we assume the signal actually repeats itself to infinity: the end point connects with the starting point. The discontinuity at each connection location can introduce noise at other frequencies, a phenomenon known as the ?Gibbs effect?. We choose a tapering parameter of 0.2, where the first and last 10% of the samples are modified to gradually decreases to zero using a cosine taper (Bloomfield 2000). Figure A.2 shows some sensitivity tests on this parameter.

4.3   Forward modeling on impacted event scales

Paleo proxy records from marine sediment cores, for example, $\delta^{18}O$ of both planktic and benthic foraminifera, and alkenone-derived sea surface temperature (SST), often exhibits climate oscillations ranging from 10^0 kyr to 10^2 kyr (Seki et al., 2002; Kim et al., 2004; Lisiecki and Raymo, 2005; Kennett et al., 2006; Max et al., 2012), As shown previously in section 4.2, bioturbation has different impact on events of different time scales. By simply looking at the most bioturbated IRF (Figure **??**), we can get a rough estimate of time scale of the bioturbation impact on phase shift: given a large mixing layer depth 15 cm and very small sedimentation rate 1 cm/kyr, there is an approximately 15 kyr phase shift between the recorded event peak and the actual onset of the event. This indicates that centennial to millennial scale events are potentially subject to strong attenuation and for events of scale 10 kyr or longer, bioturbation is very likely to be insignificant. Note that mixing layer depth usually is not explicitly measured, we use the estimates of its upper (15 cm) and lower bound (5 cm), and global average (10 cm) from previous studies (Boudreau, 1994; Teal et al., 2008). To further evaluate the spectrum of impacted events, artificial sinusoid signals of different periodicities are tested under the same set of bioturbation parameters (Figure 4.3): mixing layer depth ML = 10 cm with varied sedimentation rates from 1 cm/kyr to 15 cm/kyr. The artificial sinusoid signals have centennial and millennial scales with periodicity 0.5 - 9 kyr (Figure 4.3a-f) and 50 kyr - 100kyr perodicity (Figure 4.3g-h). Note that the range of the x-axis is very different in each subplot in Figure 4.3. These signals are then convolved with the same set of IRFs. The 0.5 kyr event (Figure 4.3a) is significantly

damped in all cases. The amplitude attenuation is more than 90% and the phase shift is on centennial scales. For the ice-age scale events 50 kyr (Figure 4.3g) and 100 kyr (Figure 4.3h), the bioturbation effect is barely noticeable except in the most extreme case where sedimentation rate is very small (1cm/kyr, red curves). For the millennial scale events (1kyr and 9kyr, Figure 3b-f), there is a progressive reduction of the relative attenuation under the same bioturbation conditions. For example, given ML = 10 cm, v = 3 cm/kyr (orange curves in Figure 4.3), the 1 kyr signal (Figure 4.3b) is completely smoothed out while the 9 kyr signal still has 50% of the original amplitude. In terms of phase shift, the 3 kyr (Figure 4.3c) signal has an approximately 1.5 kyr shift, while the phase shift for the 9 kyr signal is less than 1 kyr (Figure 4.3f).

To further quantify the phase shift and amplitude attenuation of bioturbation on events across different time scales, the experiments in Figure 4.3 are expanded to include more bioturbation conditions (Figure 4.4 and Figure 4.5): in addition to testing various sedimentation rates, G is also varied from 0.03 to 3 to cover all possible IRF shapes; the upper and lower bound of ML are also tested. Instead of presenting the results in the time/depth domain like in Figure 4.3, Figure 4.4 and Figure 4.5 plot the phase shift (amplitude) as a function of G and sedimentation rate v in each subplot, which corresponds to a given ML assumption and event scale. For example, the phase shifts under different sedimentation rates of the 5 kyr event in Figure 3d can be found in Figure 4i by drawing a vertical line at G = 0.3, and their preserved amplitude can be found in the same way in Figure 5i. For centennial scale event, they are completely smoothed out (Figure 4.3a-c) except when the mixing

layer depth is the minimum of 5 cm and the sedimentation rate exceeds 10 cm/kyr, when 20% of the original amplitude can be preserved. For ice-age scale events, the amplitude almost remains intact in most conditions (Figure 4.5s-x) and the phase shift is very small (*le* 2 kyr) compared to its event scales (Figure 4.4s-x), except in the most bioturbated case where the mixing layer is 15 cm and sedimentation rate is smaller than 2 cm/kyr, which can also be seen in Figure 4.3g-h. For the 3 kyr scale event, the phase shift can be limited to hundreds of years when the sedimentation rate is above 4 cm/kyr given the ML = 5 cm condition (Figure 4.4g), and more than 50% of the amplitude can be preserved (Figure 4.5g). As mixing layer depth increases from 5 cm to 10 cm and 15 cm, the sedimentation rate has to exceed approximately 8 cm/kyr and 12 cm/kyr respectively to preserve the same signal as in the ML = 5 cm case. The intensification of bioturbation effects as ML increases can be observed in other millennial scale events. One noticeable feature of Figure 4.4 and Figure 4.5 is that the sedimentation rate and the mixing layer depth variations have a more noticeable impact than the variation of G. As mentioned in section 4.1, G controls the shape of the IRF, aka the skewness, while ML and v controls the time/depth span of the IRF. Hence if ML/v is small, the IRF cannot span a long time/depth, and the bioturbation will not have strong impact.
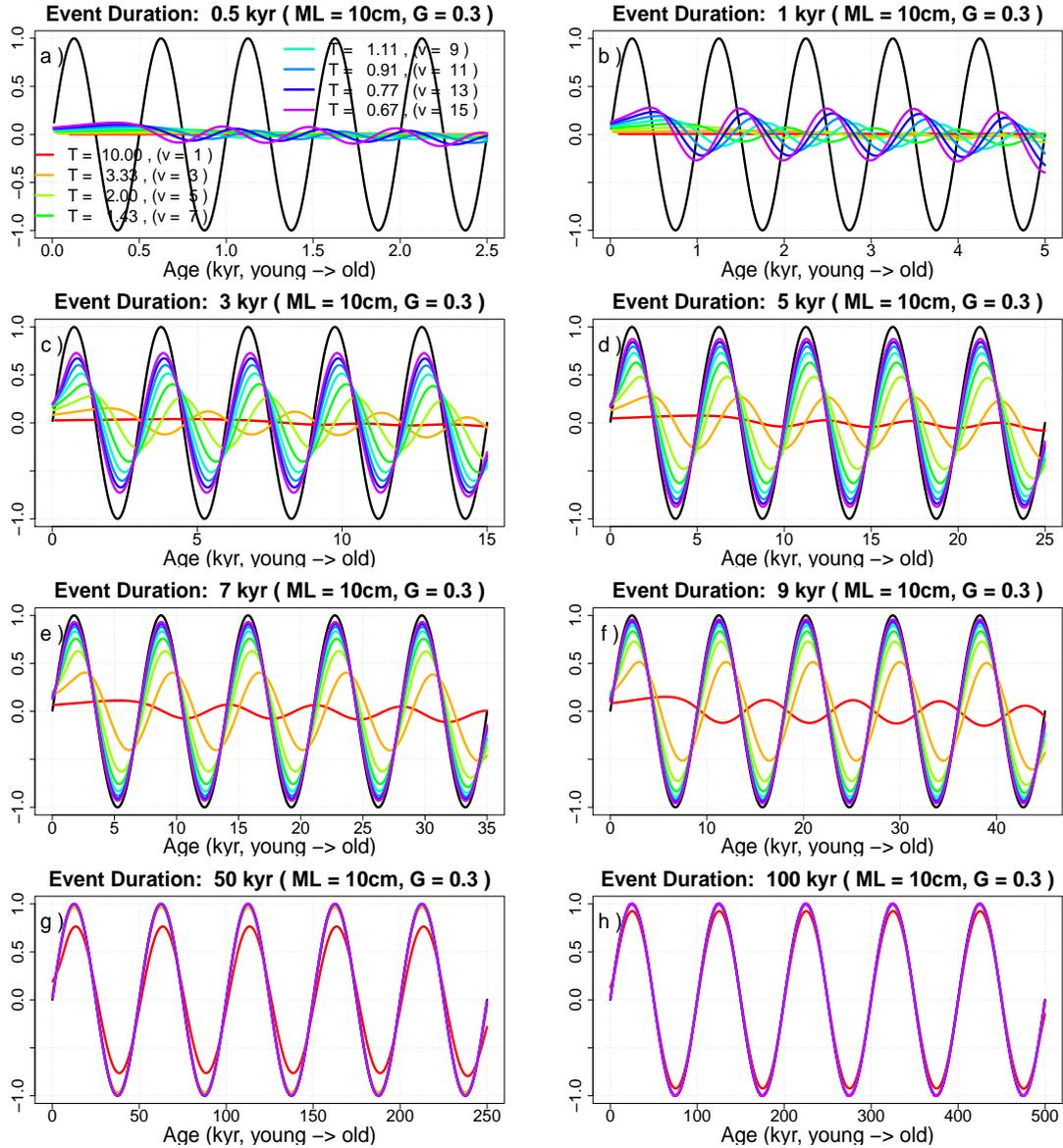
Figure 4.3: Different attenuation of events of various time scales under the same bioturbation environment. black curves in a)-h) shows sinusoid events of 0.5 kyr, 1 kyr, 3 kyr, 5 kyr, 7 kyr, 9 kyr, 50 kyr and 100 kyr period respectively. They are bioturbated with the same set of bioturbation parameters ML = 10 cm, G = 0.3 and v = 1 cm/kyr (red curves), 3 cm/kyr (orange curves), 5 cm/kyr (light green curves), 7 cm/kyr (green curves), 9 cm/kyr (cyan curves), 11 cm/kyr (blue curves), 13 cm/kyr (royal blue curves) and 15 cm/kyr (magenta curves). Note that the range of x-axis is very different for each subplot
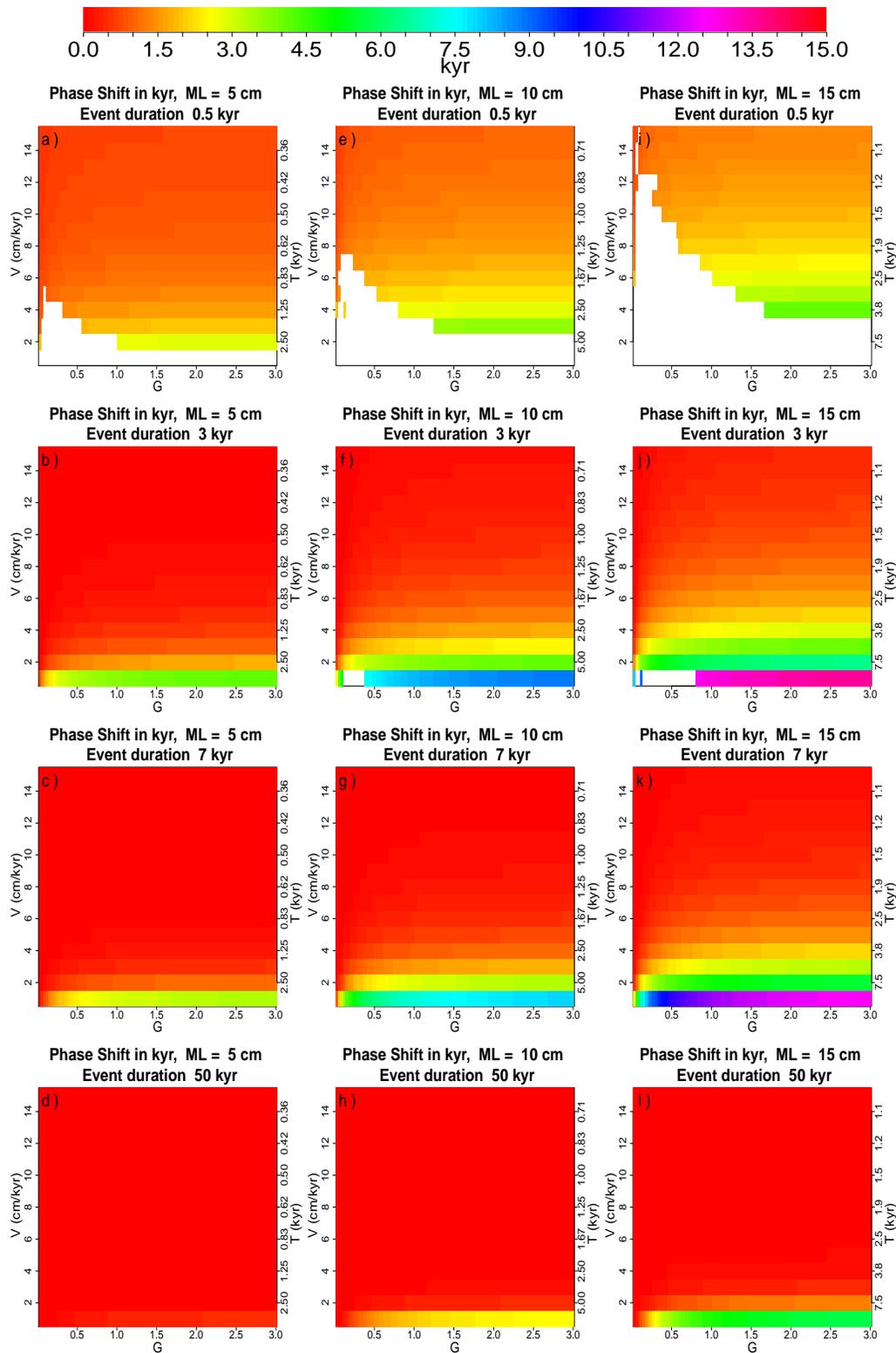
Figure 4.4: Phase shift of events of various time scales under various bioturbation environment. x-axis is change of G parameters and y-axis is the change of sedimentation rates. The first, second, and third column represents ML = 5cm, 10cm and 15cm scenarios respectively. The first through eighth row represents events of 0.5 kyr, 3 kyr, 7 kyr, and 50 kyr respectively.
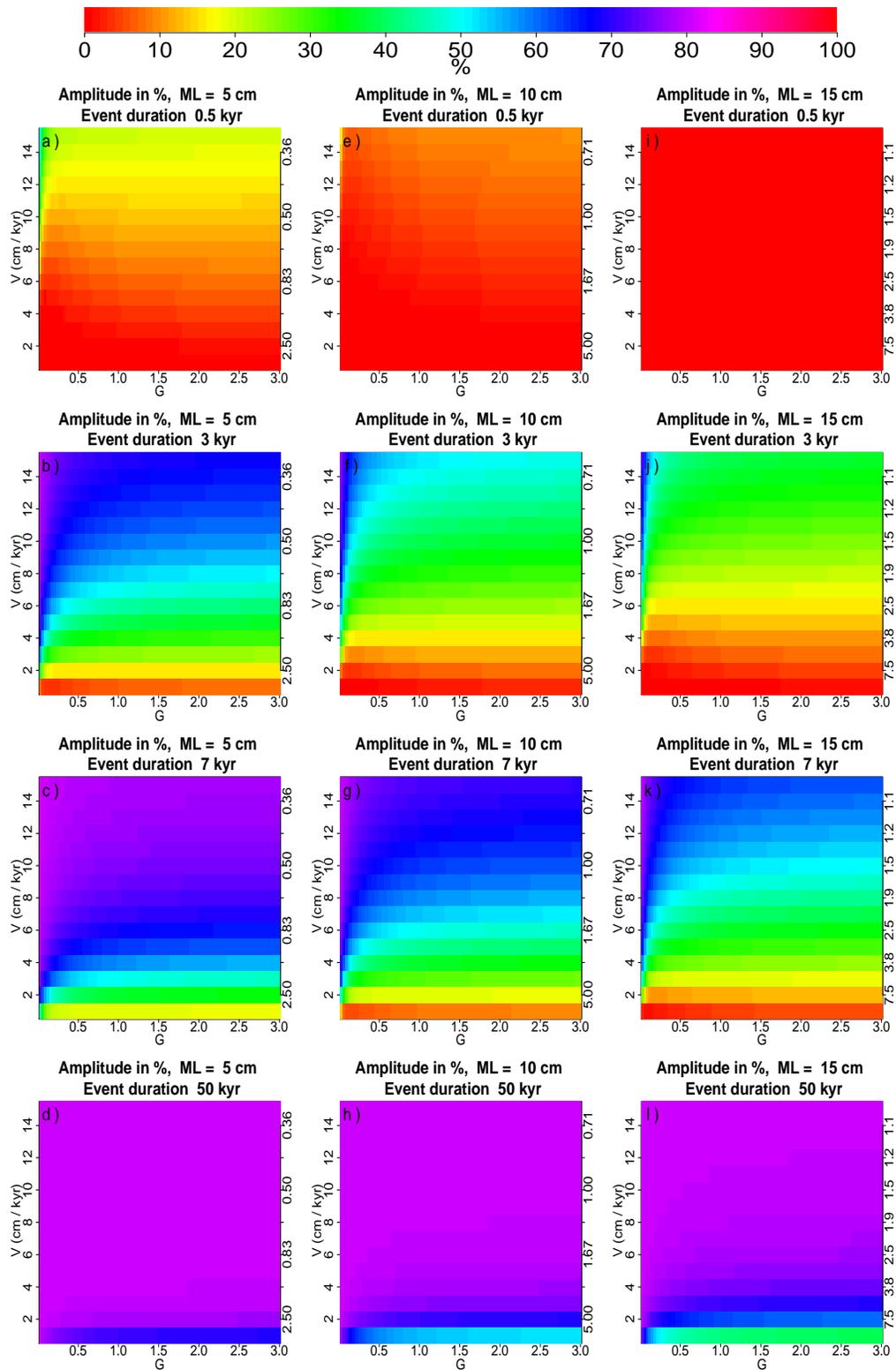
Figure 4.5: Same as Figure 4.4, but for amplitude in % (relative to input amplitude).

4.4 Inverse modeling

Section 4.2.3 introduced the algorithm in an idealized situation where there is no noise and bias caused from interpolation and parameter estimation (Figure 4.2). In real-world scenario, 1) the proxy records have irregular sampling, which requires temporal interpolation before deconvolution un-mixing is performed. 2) there is no explicit estimate on the bioturbation environment for most proxy records, which relies on empirical estimation, and 3) there is measurement error. Each of these issues can affect the recovered signal, and are investigated here.

4.4.1 Pseudo Proxy Test

In this section, we test the algorithm on pseudo proxy records. These records are generated using a similar forward modeling procedure as in section 4.3 except that the final simulated proxy records are generated by sampling unevenly from the bioturbated series, to mimic the uneven temporal resolution of marine proxies. Two individual experiments are performed to evaluate the utility of the un-mixing approach:

   A. Test on interpolation effect: Each pseudo proxy is un-bioturbated by applying identical bioturbation parameters as used to generate them.

   B. Test on incorrect parameter estimation: Each modeled pseudo proxy record is un-bioturbated using incorrect deconvolution parameters (v, M and/or L).

   C. Test on measurement error: The pseudo proxy record is added with Gaussian noise, and then interpolated and deconvolved.

Experiment A is designed to assess artifacts that can be introduced associated with interpolation. Experiment B is designed to evaluate the recovered signal when incorrect parameters are used to deconvolve the record, where we also need to estimate the true bioturbation parameters in addition to the application of temporal interpolation. Experiment C tests the model sensitivity to measurement error, and two different noise levels are evaluated.
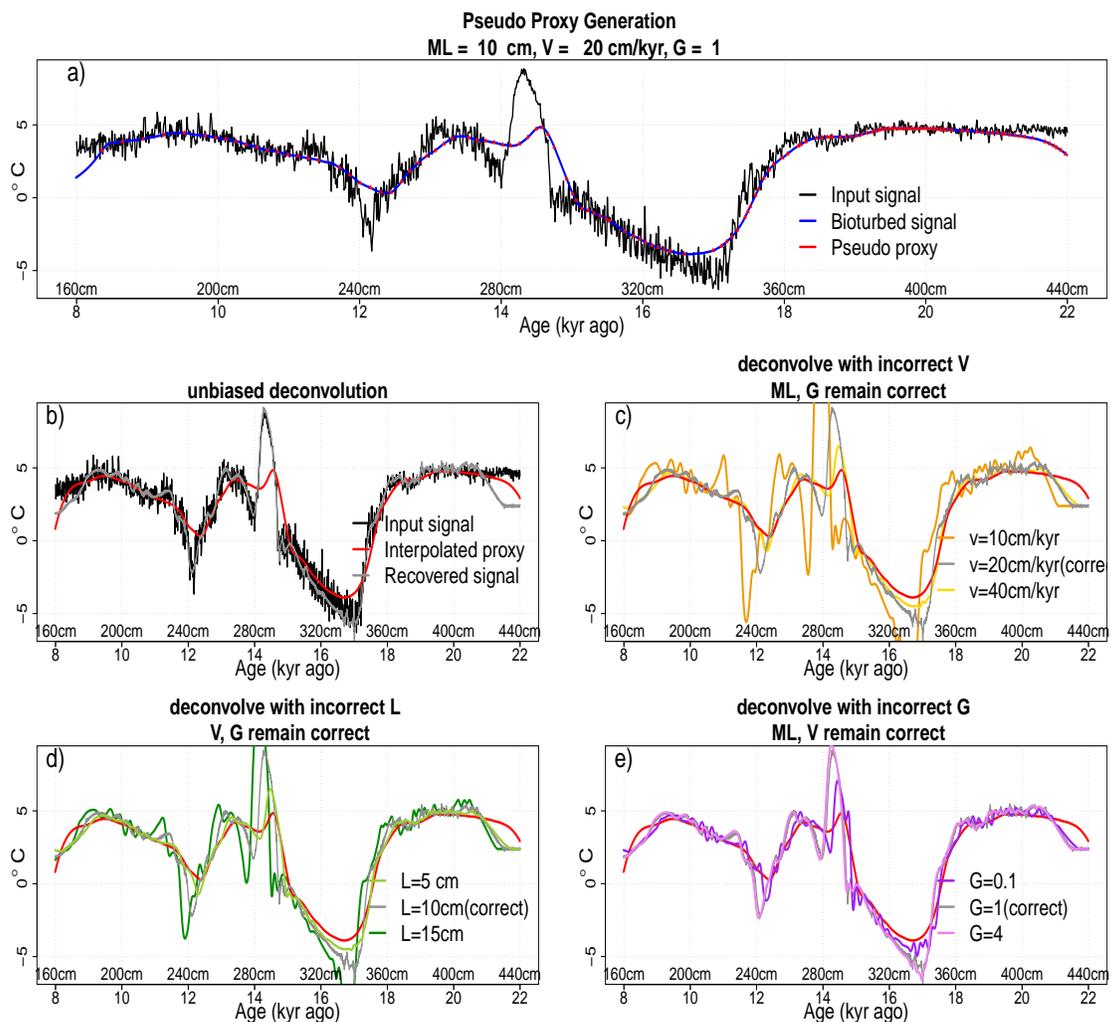
Figure 4.6: Artificial proxy tests. The input time series (truth) is shown as black curve in a). It is bioturbated (with parameters L = 10cm, v = 20cm/kyr and G = 1) to obtain blue curve, which is the climate signal recorded in the sediments. The pseudo proxy record (red dots) is generated by sampling irregularly from the blue curve. For b)-e), the proxy record is linearly interpolated (red curves) and the correct deconvolution output is shown as a grey curve. In b), the interpolated proxy record is deconvolved with an IRF that is parameterized with correct L, v and G, the output is the grey curve. In c), the interpolated proxy record (red curve) is deconvolved with an IRF that has correct L and G, but overly estimated v = 40cm/kyr (yellow curve) and underestimated v = 10 cm/kyr (orange curve); In d), the interpolated proxy record (red curve) is deconvolved with an IRF that has correct v and G, but overly estimated L = 15cm (dark green curve) and underestimated L = 5 cm (light green curve). In e), the interpolated proxy record (red curve) is deconvolved with an IRF that has correct v and L, but overly estimated G = 4 (dark pink curve) and underestimated L = 0.1 (light purple curve).

Figure 4.6a illustrates the simulated input climate signal (black line), the bio-turbated signal (blue line) with $G = 1.0, L = 10cm, v = 20cm/kyr$, and our pseudo proxy time series (red dots) that is sampled from the bioturbated input climate signal. The pseudo proxy time series is sampled with irregular temporal resolution that has a median of 86 years, the 1st and 3rd and quantiles being 63 and 126 years respectively (Its temporal resolution is taken from core MD01-2461 in the Northeast Atlantic). This pseudo proxy time series is then linearly interpolated (red lines in Figure 4.6b-e) for later experiments; piecewise linear interpolation to a regular sampling grid is a common approach that is used in paleoclimate analysis, and is required for the Fourier deconvolution. Figure 4.6b shows the recovered series (grey curve) in experiment A using the correct parameters. As expected, it accurately recovers the slow variability on the millennial scale, while it fails to recover the short-term random noise compared to Figure 4.2c. This constitutes information

that is lost associated with sampling and subsequent interpolation. This curve is also shown in in Figure 4.6c-e for comparison purposes.

In experiment B, an incorrect set of bioturbation parameters are used for the deconvolution "unmixing". Figure 4.6c shows the un-mixed, interpolated pseudo proxy using incorrect sedimentation rates of 10 cm/kyr (orange line) and 40 cm/kyr (yellow line) compared to the correct value of 20 cm/kyr (grey line). Considering the Bølling warming event ($\sim$ 14.5 kyr, $\sim$ 215 cm) as a benchmark, the recovered un-mixed signals are all shifted up-core after deconvolution with the different IRFs. When a 10 cm/kyr sedimentation rate is used for deconvolution, the recovered signal has the largest phase shift and largest amplitude increase relative to the pseudo proxy, because of the underlying assumption that the observed signal has gone through much stronger bioturbation due to a smaller sedimentation rate. Conversely, when a 40 cm/kyr sedimentation rate is used for deconvolution, the observed signal is considered to have gone through relatively weaker bioturbation due to a higher sedimentation rate. Hence the recovered signal has the smallest phase shift and largest amplitude relative to the pseudo proxy among the three cases. The phase of the recovered signal using the correct parameter v = 20 cm/kyr lies between the overestimated v case and underestimate v case. Therefore, by considering the uncertainty in our knowledge of the bioturbation parameters in real-world scenarios, and using them for deconvolution, we can provide uncertainty bounds for the true phase and amplitude of a given paleoclimate event. Figure 4.6d-e shows the same experiment as Figure 4.6c but for variable L and G parameters; the same interpretation applies. It can be problematic to interpret the recovered

signal using only one set of parameters since bioturbation is seldom measured explicitly and large bias in parameter estimation is likely to exist, for example an overestimation of the mixing intensity leads to an exaggerated "Melt water pulse" in Berger et al. (1977)(Jones and Ruddiman, 1982; Schiffelbein, 1984b). Deconvolution of the record using a plausible range of IRFs that considers all possible bioturbation conditions will yield confidence bounds for the true signal.

Note that in experiment B, errors introduced by biased parameters will be greatly amplified when the denominator in Equation (4.2.2) is close to zero, which can completely obscure the true signal; post-processing is required to suppress this computational noise. In experiment B, a Wiener filter is applied to $H[\omega]$ during division in the frequency domain where values close to zero are replaced with a small non-zero constant. This is a common practice in signal processing and seismology (Neelamani, 2008).

To test the sensitivity of the algorithm to measurement error, Gaussian white noise is added to the climate pseudo proxy (red dotted lines in Figure **??**, Figure 4.7a-b) series before interpolation. Two noise level are tested: 2% of the pseudo proxy amplitude (yellow dotted lines in Figure 4.7 left column) and 20% (orange dotted lines in Figure 4.7 right column). When the deconvolution un-mixing procedure is performed without a Wiener filter, the raw output is a series of amplified noise (grey lines in Figure 4.7c-d). When a Wiener filter is turned on, the noise is greatly reduced (grey lines in Figure 4.7e-f). But for the high noise level case, the millennial scale variability is still obscured in the high-frequency noise (Figure 4.7f). Hence when substantial measurement error / random noise is present, the Wiener filter

alone is not sufficient. A Taner band-pass filter is needed for further noise reduction (taner function in the Astrochron package in R; (Taner, 1992; Meyers, 2014). It specifies a frequency cut-off parameter fh above which the components of the time series are removed. After a Taner filter is applied, the millennial scale variability in low noise level case is accurately recovered (Figure 4.7g), but in the high noise level case, to preserve a clear millennial scale variability, a smaller cut-off frequency is needed. However, because only low frequency harmonic components are retained, the recovered series looks more periodic/sinusoidal (Figure.4.7 h). This can lead to the inference of non-existing variability. Therefore, extra caution should be exercised when interpreting a recovered signal using a relatively small cut-off frequency. More discussion on the choice of the frequency cut-off parameter fh used in Taner filter is shown in the appendix. But importantly, the oxygen isotope records evaluated in this dissertation are characterized by a low measurement noise level, more analogous to panels a-g of Figure 4.7.
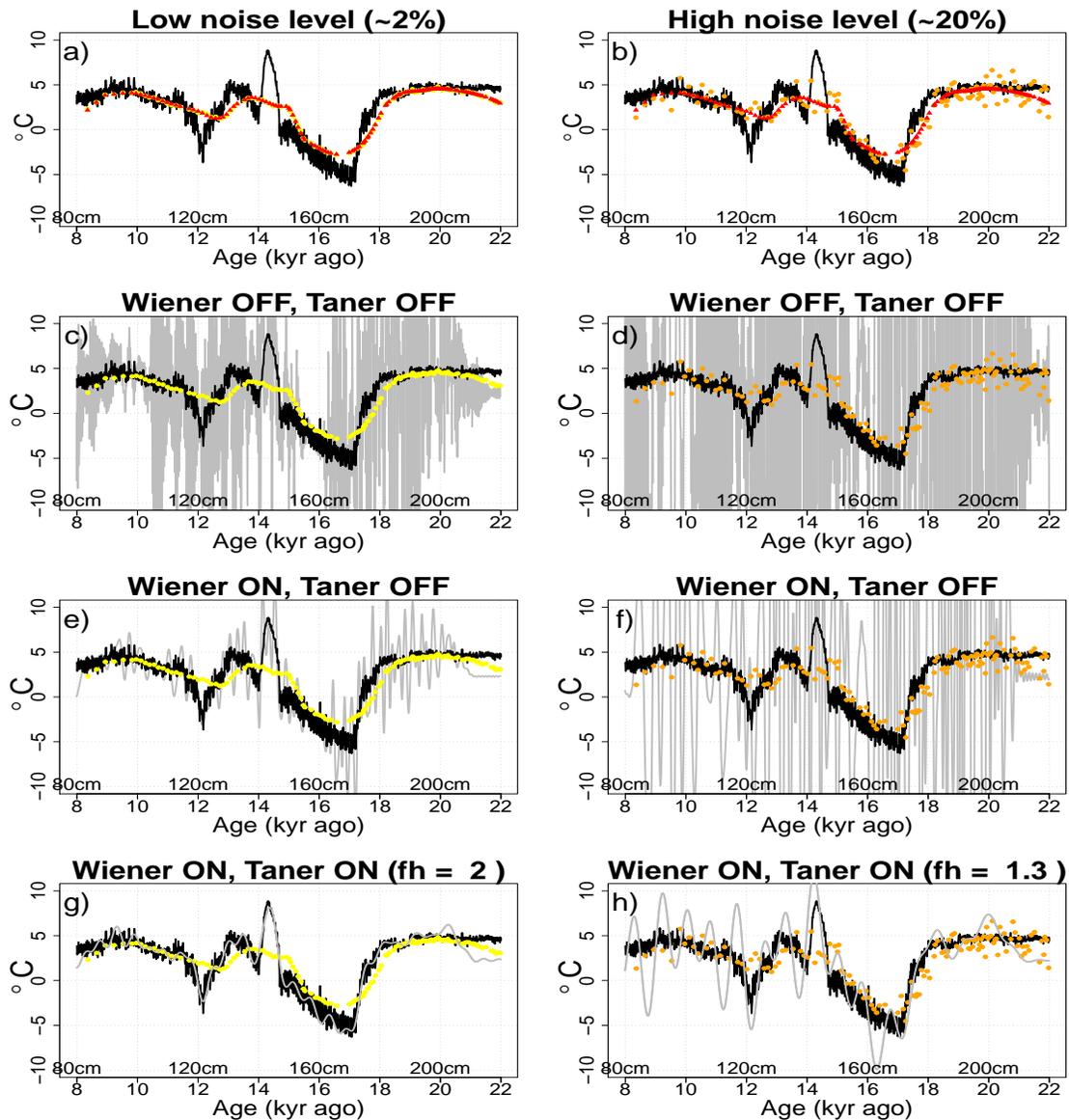
Figure 4.7: Noise amplification.Gaussian noise of $0.1°C$ (left column) and $1°C$ (right column) are added to the pseudo proxy record (red triangle dotted lines in a)-b)) from Figure 4.6a). The noisy records are plotted as yellow dotted line in the left column and orange dotted line in the right column. The noisy records are then interpolated and deconvolved to generate the grey lines in c)-h). The black curves are the true climate signal that is deposited into the sediment. In c)-d), no Wiener filter or Taner filter are applied, and the deconvolved series have values that go to infinity; in e)-f), a Wiener filter is applied with no Taner filter, and the noise is greatly suppressed in the deconvolved series; In g)-h), both Wiener and Taner filter are applied, and the recovered series are much clearer

4.4.2   Application to LR04

In this section, $\delta^{18}O$ records from the benthic foraminifera isotope stack LR04 (Lisiecki and Raymo, 2005) will be used. It is a composite of 57 globally distributed oxygen isotope records that span the last 5.3 million years. The 57 records were placed on the same age model using the graphic correlation technique developed by Lisiecki and Lisiecki (2002). They were orbitally tuned under the constraints of the mean sedimentation rate of the 57 records. The final global stack is an average across the 57 records. This is one of the most widely used and well-tuned datasets in the paleoclimatology community. The benthic $\delta^{18}O$ measures the global ice volume and deep ocean temperature (Ravelo and Hillaire-Marcel, 2007; Kennett et al., 2006), and has been widely used to reconstruct sea level changes in different geological intervals(Dutton et al., 2015a; Rohling et al., 2008). The period of interest is the last interglacial period (LIG, 129-116 kyr ago), or the so-called the Marine Isotope Stage 5e (MIS 5e, Shackleton et al. (2002)).The reason for choosing this period is that the global mean temperature is estimated to be at least $2°C$ warmer (Clark and Huybers, 2009; Otto-bliesner et al., 2009; McKay et al., 2011) and sea level 4-9 meters higher than present day (Kopp et al., 2009; Dutton et al., 2015a; Rohling et al., 2008; Dutton et al., 2015b), making it analogous to future climate change.
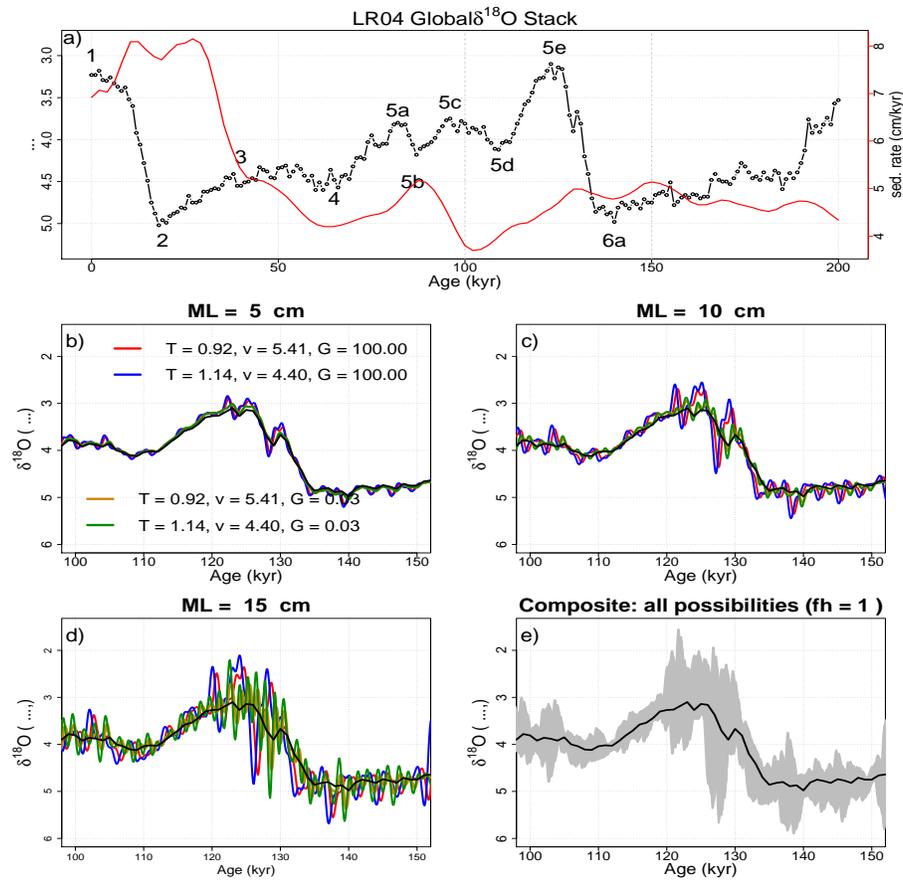
Figure 4.8: $\delta^{18}O$ record from the globally stacked LR04 (averaged time series from the 57 tuned records). a) time series of sedimentation rate (blue dots, y-axis on the left) and resolution (red dots, y-axis on the right). b) $\delta^{18}O$ time series from 200 kyr ago. The period of interest (last interglacial) is highlighted by grey dotted line. c)-e) shows the deconvolved records under $ML = 5cm$, $10cm$ and $15cm$ scenario respectively. In each of c)-e), different combinations of G and v are tested. G ranges from 0.03 to 4 with an extreme value, and v are sampled as the every 20th percentile from the sedimentation rates from this period. Only the combinations of extreme parameters are chosen to be shown in c)-e). All the combinations from ML, G and v are shown as grey curves in f)

First, the LR04 globally stacked $\delta^{18}O$ record will be examined. The temporal resolution of this record is 1 kyr. The sedimentation rate and $\delta^{18}O$ time series are shown in Figure 4.8a. Note that during the last interglacial (LIG), the sedimentation rate is below 5 cm/kyr. To choose the IRF for deconvolution, proper estimation of sedimentation rate, mixing layer depth and biodiffusivity (or G) are needed. Sedimentation rates can be readily calculated from depth and age series from the published record or in this case, provided by tuning (Lisiecki and Raymo, 2005). The sedimentation rate is usually not constant through time. Mixing layer depth and biodiffusivity are not explicitly measured in most proxy records, which requires empirical estimation. According to previous research on modern marine depositional systems, there are well-accepted upper and lower bounds for mixing layer depth L, which falls into a range from 5cm to 15cm (Boudreau, 1994; Teal et al., 2008) with a global average of 10cm (Boudreau, 1994). For bioturbation intensity, there is no well-established universal range. Boudreau (1994) established an empirical relationship between D_B and v but it can explain only 20-30% of the variance. Based on this relationship, the possible D_B range given a fixed v and global average $ML = 10cm$ yields G values from 0.1 to 40 (Figure. 1 in Boudreau (1994)), which covers all the possible shapes of the IRF. Therefore, it is reasonable to sample G values that can cover all possible shapes of the IRF (Figure 4.2a). Note that we are using a time-invariant IRF with the underlying assumption that the bioturbation environment has stayed the same through time, hence the parameters remain constant through the entire depth. This time-invariant assumption may result in an inaccurately recovered signal based on one set of parameters. However,

given the uncertainty of each bioturbation parameter, an ensemble of different parameter combinations can be used to deconvolve the proxy record. This will generate uncertainty bounds for the recovered signal.

For deconvolution, three scenarios of mixing layer depth estimation are conducted: $ML = 5cm$, $10cm$ and $15cm$, corresponding to an extreme shallow mixing layer (Figure 4.8b), average mixing layer (Figure 4.8c) and extreme deep mixing layer (Figure 4.8d) respectively. In each scenario, different G values ranges from 0.03 to 3, and an extreme value 100 is tested. Five different sedimentation rates sampled as the 20%, 40%, 60%, 80% and 100% percentile of the observed sedimentation rate distribution through 0-200kyr ago are chosen. Figure 4.8b-d shows the extreme combination in each scenario: a small G value (0.03) with sedimentation rates at 20% (4.40 cm/kyr) and 80% (5.41 cm/kyr) percentiles, then a large G value (100) with the same two sedimentation rates. Figure 4.8a shows that when ML is very small and sedimentation rates are at intermediate values (4.40 cm/kyr-5.41 cm/kyr) with small variability, the change of G (from 0.03 to 100) does not induce a noticeable difference. The recovered signal is similar to the input signals under all combinations. As ML increases from 5 cm to 10 cm (Figure 4.8b) and 15 cm (Figure 4.8c), the amplitude of the recovered signals gets progressively larger, and the phase shifts also become more noticeable. When $ML = 10cm$, the impact of G is noticeable as the phase difference between $G = 0.03$ and $G = 100$ is easily distinguished. The amplitude of the millennial scale variability is increased by about 0.5 ‰. Figure 4.8d shows the ensemble of recovered signals under all possible parameter combinations. Note that the variability in this interglacial is more

pronounced after the un-mixing deconvolution and there is also a time shift in this recovered paleoclimate signal. This suggests previously unrecognized changes in temperature and/or ice volume, as discussed in section 4.5.
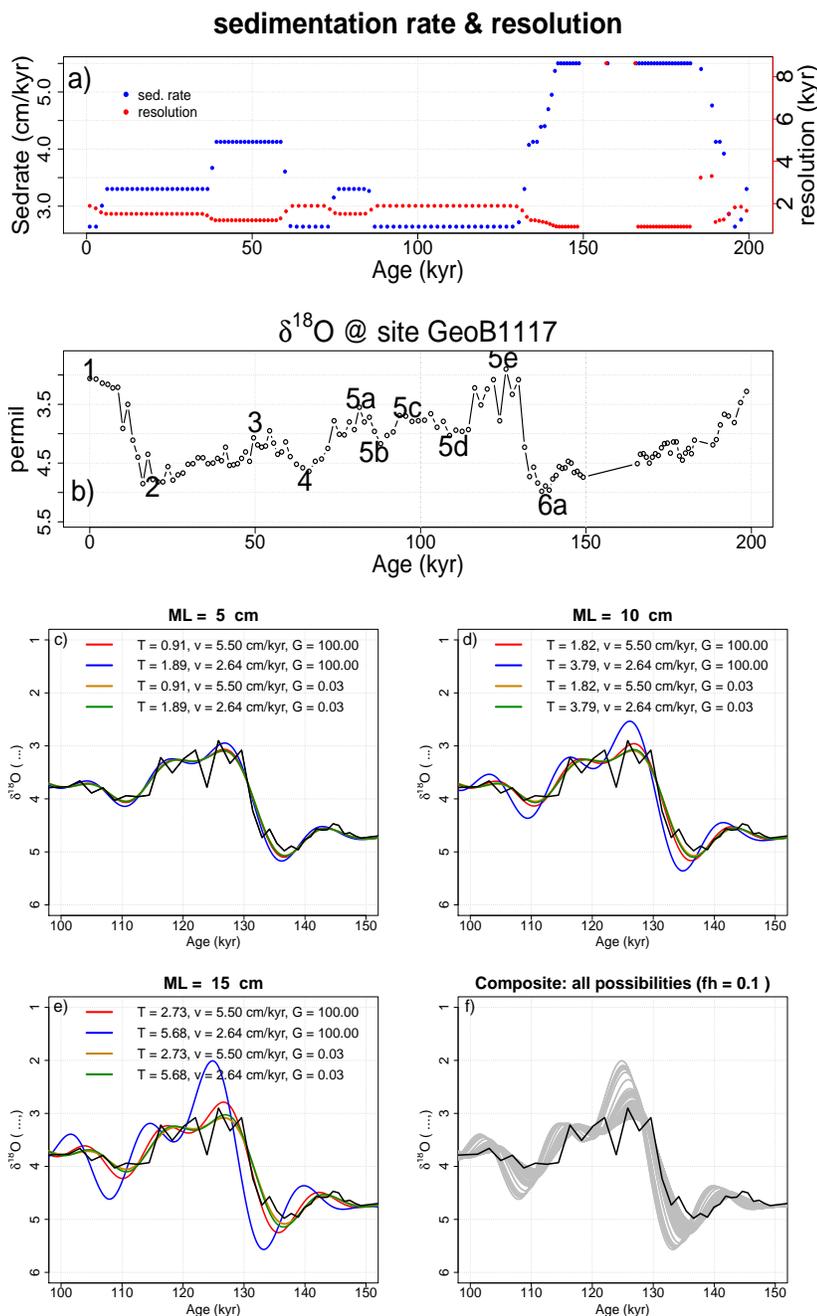
Figure 4.9: Benthic foraminifera δ^{18}O record from GeoB site 1117 during last interglacial. a) time series of sedimentation rates (blue dots, y-axis on the left) and temporal resolution (red dots, y-axis on the right). b) $\delta^{18}O$ time series from 200 kyr ago to present. The interval of interest (last interglacial) is labeled as 5e. c)-e) shows the deconvolved records under the $ML = 5cm$, $10cm$ and $15cm$ scenarios respectively. In each of c)-e), different combinations of G and v are tested. G ranges from 0.03 to 4 with an extreme value, and v are sampled as the every 20th percentile of the sedimentation rates from this interval. Only the combinations of extreme parameters are chosen to be shown in c)-e). All the combinations from ML, G and v are shown as grey curves in f)

One individual $\delta^{18}O$ record from site GeoB 1117 (Bickert and Wefer, 2012) in Lisiecki and Raymo (2005) is shown. This site is chosen because it has a relatively high sampling resolution ($\leqslant$ 2kyr) during MIS 5e together with a small and stable sedimentation rate (below 3 cm/kyr to 5.5 cm/kyr during the LIG, Figure 4.9a) . Figure 4.9 c-e shows the deconvolution using the extreme parameter combinations in different ML scenarios. Similarly, there is a progression of amplified variability as mixing layer depth increases from 5 cm to 15 cm. Compared to the LR04 globally stacked data, the amplitude of the recovered signal increases by about 1‰, which suggests previously unrecognized changes in temperature and/or ice volume, as discussed in section 4.5. The recovered signal is much more volatile compared to that of the globally stacked record (Figure 4.8), this can be due to a higher noise level in a single record compared to a globally averaged stack (the noise level test in section 4.4.1 Figure 4.7). The noise in the global stack is greatly reduced when averaging across the 57 sites.

4.5   Discussion

In this chapter, a systematic study of the bioturbation effects on paleoclimate records is conducted. We adopt a deterministic approach where the mixing procedure is treated as a time-invariant linear filter. The bioturbated signal is the convolution output between the impulse response function of the bioturbation system and the input paleoclimate signal. Following this forward modeling approach, the bioturbation effects on events across different time scales (10^2-10^5 years) are investigated.

We conclude that

1) Centennial scale signals will be completely smoothed out by bioturbation, it is possible to preserve 20% of the signal if the mixing layer is extremely shallow ($5cm$) and the sedimentation rate is above 13 cm/kyr.

2) For events of 10^4-10^5 years scale, bioturbation has vey little impact even when sedimentation rate is below 2 cm/kyr.

3) For millennial scale events, bioturbation impact is strongly dependent on the event scale as well as the bioturbation parameters ML, G and v. We provide a quantitive chart for phase shift and amplitude preservation for all the possible combinations in Figures 4.4-4.5.

To remove the bioturbation effects on a proxy record, a deconvolution method is proposed to recover the original climate signal. It can be applied directly to proxy records to study the uncertainty of signal attenuation induced by bioturbation. The theoretical base of this method is the Convolution Theorem on discrete time series. The pre-requisites for the theorem to be valid are that the input series is linear, causal, and time-invariant. While linearity and time-invariance are reasonably satisfied in our application, our IRFs are non-causal due to the fact that sediment is mixed in both directions. Hence a time shift in proxy records is introduced, achieved by zero-padding the proxy records on both sides, to the proxy records are performed to mimic the output from a causal filter (Figure 4.1). FFT deconvolution is conducted after the pre-processing. Bluntly performing FFT deconvolution on a non-causal discrete filter and a discrete time series without the time-shifting pre-processing will result in incorrect phase of the recovered proxy series. Thus,

the zero-padding procedure in our application has two purposes, one is the time-shifting, and one is to make input time series of equal lengths so that they will have the same number of FFT pairs for point-wise division in the frequency domain.

Characterizing bioturbation in the mixed layer as a time-invariant IRF is a deterministic process. It models the systematic change as any climate signal passes through a given bioturbation environment. That being said, any random mixing, such as lumpy-mixing (e.g. non-local mixing in Boudreau (1986a); Meysman et al. (2003)) caused by large burrowers is not included in this framework. Including these stochastic components in bioturbation models will generate more realistic bioturbated time series in the forward modeling approach. However, it is not feasible to un-mix these records due to the nature of randomness. Hence our framework focuses on the deterministic impact of bioturbation, and all the variability in the records to be deconvolved is treated as real signals. These could include sampling errors and interpolation errors, which will introduce noise amplification during the inverse process and will need to be suppressed in the post-processing step to recover the true signal.

Compared to forward modeling, the major limitation of this inverse modeling is that it has to use a time-invariant IRF, which assumes that the sedimentation rate v, biodiffusion D_B (or G) and the mix layer depth L do not change through time. This can be circumvented by applying a range of possible parameters to provide the upper and lower bounds of recovered signals. Based on the time-varied sedimentation rates, it is also practical to slice the time series into different segments according to the variation of its sedimentation rates and apply different IRF on each

segment.

The unbioturbated marine records can yield new implications for our interpretation of the past climate. In the application to LR04 records, we deconvolve the globally stacked benthic foraminifera $\delta^{18}O$ record to recover the possible amplitude of the input signal at the last interglacial. We found that for the globally stacked $\delta^{18}O$ and its sedimentation rate from LR04, the amplitude of the millennial scale variability of $\delta^{18}O$ during MIS 5e can be increased by more than half permil with an lower bound 1 ‰less. For an individual record at Site GeoB 1117, the recovered signal is much more volatile with a 1 ‰smaller $\delta^{18}O$ value in the lower bound. For example, this 1 ‰variability in $\delta^{18}O$ would correspond to approximately 100m global mean seal level change, if it was entirely attributable to ice volume. In the LR04 collection, $\sim$80% of the records have sedimentation rates that are well below $5cm/kyr$, which implies the millennial scale variation at MIS 5e are very likely subject to non-trivial bioturbation attenuation and the subsequent sea level estimation based on $\delta^{18}O$ values is probably underestimated.

Chapter 5

Summary and conclusion

This dissertation consists of three projects on the state estimation of the climate system. Chapter 2 evaluates the state estimation quality of the ocean in a coupled data assimilation system where the atmospheric observations are replaced with atmospheric reanalysis. Chapter 3 and chapter 4 focus on climate reconstruction for the past. Chapter 3 proposes a suitable data assimilation method for merging model outputs and observation data together to provide the optimum estimate for past climate. Chapter 4 proposes a bioturbation-removal algorithm to recover the input climate signal in marine proxy records.

In chapter 2, we study the degradation of ocean state estimation in a coupled data assimilation (CDA) system when the atmospheric observations are substituted with atmospheric reanalysis. The error quantification experiments are carried out in a coupled Lorenz 96 model that represents the coupled atmosphere-ocean system. Pseudo observations and atmospheric reanalysis are generated in this model, and then assimilated into the system. We compare the analysis quality of the substitution CDA experiment to a benchmark experiment where both atmospheric and oceanic observations are assimilated. Four different schemes of assimilating the reanalysis are tested, based on how the "observation" ensemble are generated from the reanalysis and how the error covariance matrix is represented. We found that

when the reanalysis is assimilated directly as if they are independent observations, the RMSE of the ocean analysis increases by approximately 16% compared to that of the benchmark when there is no model bias. RMSE increases by less than 22% when model bias exists. The findings are consistent when the the ensemble size is sufficient and the quality of the atmospheric observations is decent in terms of frequency, noise level and density. If the ensemble size and atmospheric observation quality are not adequate, the state estimation of the ocean deteriorates significantly in both the benchmark and the substitution experiments, which renders the difference between the two insignificant in percentage. The results from different assimilation schemes highlights two factors that can mitigate the additional error introduction in our substitution experiments: 1) the accurate representation of the error covariance of the reanalysis and 2) the temporal coherence along each ensemble member. This study provides justification for substituting the atmospheric observation with the reanalysis when setting up a CDA system, and suggests this substitution is a reasonable approach for generating future CDA analysis.

In Chapter 3, we study the state estimation performance of the two most popularly used data assimilation methods in paleoclimate: particle filter with simple importance resampling (PF-SIR) and the ensemble Kalman filter (EAKF). They are examined in pseudo proxy experiments in terms of local and hemispherically average performance, as well as their ability to capture the mode of large-scale variability. We found that EAKF outperforms PF-SIR in almost every measure, with only one third of the computation cost. The key to the better state estimation of EAKF in paleoclimate application is the large localization radius needed and the

availability of proxy observations in key regions where large-scale co-variability, such as PDO and ENSO, exists. For PF-SIR, 96 members is still inefficient to overcome EAKF despite its theoretical superiority in allowing non-Gaussian dynamics. Hence we conclude that EAKF is a better state estimation method for paleoclimate data assimilation and more complex techniques need to be implemented on PF-SIR in order to both mitigate the "curse of dimensionality" and improve its performance. This rigorous cross-validation of PF-SIR and EAKF demonstrates the promising reconstruction skill of EAKF for paleoclimatology, and also sheds lights on the preferred geographic locations for gathering future proxy observations that will add the strongest constraint to the unobserved regions in state reconstruction.

In Chapter 4, we present modeling studies on post-depositional mixing of marine climate proxy records. Both forward modeling and inverse modeling are studied. In the forward modeling study, we create a comprehensive quantitative evaluation of how events of different time scales are altered under different bioturbation environments. We suggest that it is almost impossible to reliably preserve centennial events in most deep-sea marine proxy records due to the severe attenuation. On the temporal scale of ice ages scale (10^4-10^5 years), bioturbation effects are almost negligible. For millennial scale events, the relative signal attenuation is extremely parameter- and time scale-specific. For a given event and possible bioturbation environment, the signal attenuation can be estimated from Figures 4.4- 4.5.

For inverse modeling, we propose a deconvolution through the FFT method to un-mix the proxy record, in which bioturbation is treated as a linear time-invariant filter. The method is applied to the LR04 globally stacked bethic foraminfera $\delta^{18}O$

record for the last interglacial and an individual record GeoB1117 in the LR04 collection. We found that the amplitude of millennial scale variability during MIS 5e can be 1 ‰larger for both the global stack and site GeoB1117. Considering that a 0.1‰change in $\delta^{18}O$ is related to 10m global sea level change, our study implies that the subsequent sea level estimation for the last interglacial based on the LR04 record are subject to change; future work will need to evaluate the relative contribution of $\delta^{18}O$ variability between temperature and ice volume. Since this period is considered a comparable analogue for modern sea level projections under anthropogenic forcing, the projected values might be greatly underestimated when they are derived from bioturbated signals. This project provides a quantitative measure of signal alteration for the community to refer to when dealing with marine proxies. Furthermore, it provides a toolkit to recover the potential true climate signal using reasonable bioturbation parameters for the scientific community.

Paleoclimate reconstruction relies on both unbiased proxy observations and competent climate models, as well as appropriate data assimilation methods. While this dissertation addresses the bioturbation effects in the proxy observation and proposes EAKF for paleoclimate reconstruction, much more future work is needed. For un-mixing the proxy time series, the next step will be developing an algorithm that allows a time-variant filter. This will greatly reduce the uncertainty of the recovered signal because sedimentation rate can change dramatically through time. The performance of PF-SIR is mainly limited by degeneracy in current work. New techniques are under development to alleviate this issue (Poterjoy, 2015; Kuensch and Fearnhead, 2018) and the development of a super computation system will

also facilitate the implementation of large ensemble simulations in the future.

Appendix A

Supplemental information

A.1 Time/depth coverage of IRF

Figure.A.1 shows that the depth/time coverage of the function is determined by $\frac{ML}{v}$. Given the same IRF with $G = 100$, it covers more depth/time when $\frac{ML}{v}$ is bigger ($ML = 15$ cm, $v = 1$ cm/kyr) and covers less depth/time on the x-axis when $\frac{ML}{v}$ is smaller ($ML = 5$ cm, $v = 10$ cm/kyr). This also indicates that the bioturbation impact not only depends on the shape of IRF, but also $\frac{ML}{v}$. So given an IRF with $G = 100$ and $\frac{ML}{v} = 1$kyr ($ML = 10$ cm, $v = 10$ cm/kyr), it might be less impactful compared to an IRF with a more symmetric shape with $G = 0.03$ and $\frac{ML}{v} = 10$ kyr ($ML = 10$ cm, $v = 1$ cm/kyr).
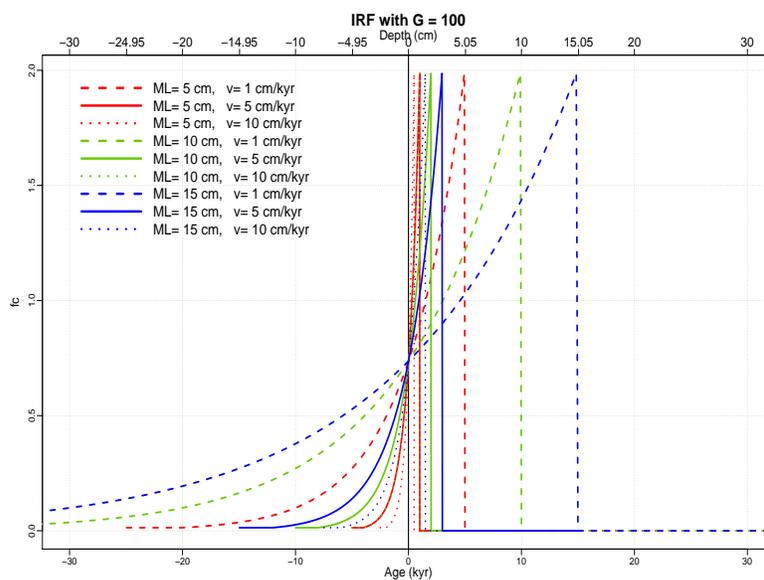
Figure A.1: Impulse response function for G = 100 under different ML and v combinations. Note the depth/time coverage of the function. The depth in cm is shown in the upper x-axis and the time in kyr is shown in the lower x-axis. The axis is labeled relative to the time/depth of original deposition.

A.2 Cosine taper ratio

The mathematical requirement for performing a Fourier transform is that the time series has to be infinite, and in practice, a finite time series is often tapered to minimize the discontinuity at the beginning and end of the series (Oppeheim et al., 1996). A cosine taper with tapering ratio $p = 0.2$ is used in our application. The tapering ratio p determines the percent of the data series tapered (choose 0-1). A tapering ratio that can minimize the edge effects during FFT while having a minimum modification on the input time series is desired. Figure.A.2 provides a justification on our choice of p. We first generate our **finite** pseudo proxy series (grey dotted curves in Figure.A.2b), c), e), g) and i)) by truncating an **infinite** bioturned series (magenta curves in Figure.A.2a)-b)). This bioturbated series is generated from an infinite input signal (black curves in a)). The finite pseudo proxy is first tapered with $p = 0.3$ (blue curve in Figure.A.2c)), notice how the values approach zeros at the two ends compared to the un-tapered series (grey dotted curve). The tapered series is then deconvolved to generate the recovered series (blue curve in Figure.A.2d)). The recovered signal follow the input signal very well except at the edges, where the finite pseudo proxy is modified by the taper. There are still edge effects remaining in the recovered series that makes it look fuzzy. As p decreases (Figure.A.2e)-j)), the discontinuities at the edges are minimized less and less, and the recovered signal becomes noisier and noisier.
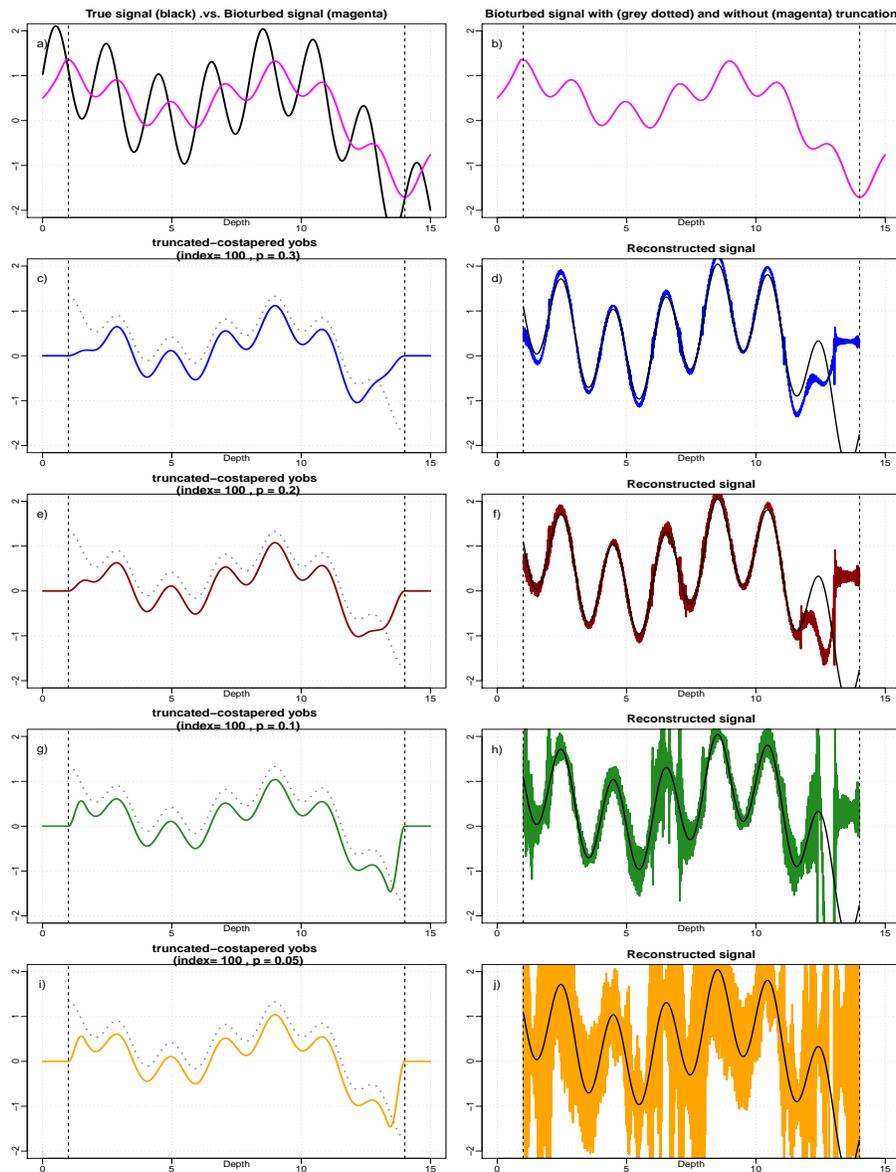
Figure A.2: The impact of the cosine taper ratio on the recovered signal. a)-b) shows how the bioturbated series is generated, which is later used for deconvolution in the rest of the plots. The black curve in a) is the infinite input climate signal, and it is bioturbated (convolved with a certain IRF) to generate an **infinite** pseudo proxy record (magenta) in a)-b). This pseudo proxy record is then truncated at the location of the two vertical dotted black curves to represent our **finite** proxy time series (grey dotted curve in b), c), e), g) i)). The **finite** input signal is also show as black curve in d) f), h), j). The finite pseudo proxy record (grey dotted curve) is tapered with different ratio $p = 0.3$ in c), $p = 0.2$ in e), $p = 0.1$ in g), and $p = 0.05$ in i), and then deconvolved. The recovered signal is show in d), f) h) and j) respectively.

A.3 The choice of frequency cut-off parameter

Given a proxy record with irregular temporal resolution, the highest possible frequency that it can capture is the Nyquist frequency determined by its highest sampling rate (highest temporal resolution). Any variability above this frequency is a result of amplified noise. If a time series has a highest sampling resolution δt kyr, the upper bound of fh will be given by $\frac{1}{2\delta t}$, hence:

$$fh \leq \frac{1}{2\delta t} \tag{A.1}$$

To retain the variability on a certain time scale, the cut-off frequency should be no less than the frequency of the event of interest. Let T denotes the period (time scale) of the interested variability of interest, then fh should be no less than $\frac{1}{T}$, hence:

$$fh \geq \frac{1}{T} \tag{A.2}$$

(A.3) and (A.3) given a weak constraint of fh because $\frac{1}{T} \sim \frac{1}{2\delta t}$ is usually a wide range. In practice, a careful examination of the recovered series using various fh values is needed. Figure**??** shows the recovered signal when different fh are used for the Taner filter. The highest temporal resolution in this time period of the pseudo proxy record is 0.032 kyr, which gives a upper bound for fh: $\frac{1}{2*0.032} \approx 15.6$. The spike at 14 kyr we are hoping to recover is of the time scale 1 ky, which gives a lower bound of fh about 1. As shown in FigureA.3, when $fh \geq 1$, the 14 kyr peak is recovered but there is also artificial peaks imbedded in the recovered series. But when the observation error (noise level) is smaller ($0.5°$C, left column in Figure 4.7), the millennial scale variability can be accurately recovered. This demonstrates that for an accurate reconstruction of the un-mixed events, not only do we need high resolution records, but also proxy observations with small observation error.
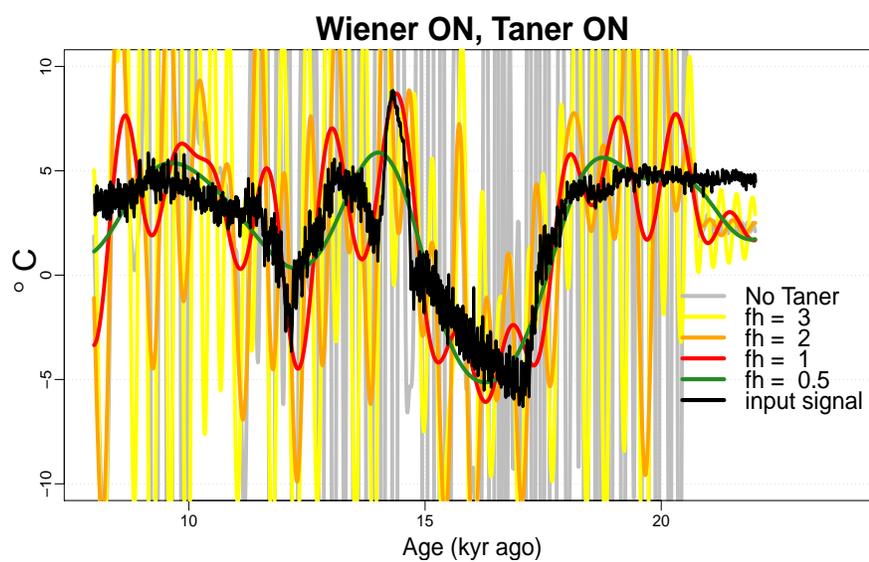
Figure A.3: The impact of the cut-off frequency. Different cut-off frequencies $fh = 3$ (yellow line), $fh = 2$ (orange line), $fh = 1$ (red line), $fh = 0.5$ (green line) are applied to the grey line in FigureA.3f. The black line is the input signal

References

Adams, J. B., M. E. Mann, and C. M. Ammann, 2003: Proxy evidence for an El Niño-like response to volcanic forcing. *Nature*, **426 (6964)**, 274–278, doi:10.1038/nature02101.

Anderson, D. M., 2001a: Attenuation of millenial-scale events by bioturbation in marine sediments. *Paleoceanography*, **16 (0)**.

Anderson, J. L., 2001b: An Ensemble Adjustment Kalman Filter for Data Assimilation. *Monthly Weather Review*, **129 (12)**, 2884–2903, doi:10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2.

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.

Annan, J. D. and J. C. Hargreaves, 2012: Identification of climatic state with limited proxy data. *Climate of the Past*, **8 (4)**, 1141–1151, doi:10.5194/cp-8-1141-2012.

Arakawa, O., 2004: Comparison of local precipitation?SST relationship between the observation and a reanalysis dataset. *Geophysical Research Letters*, **31 (12)**, L12 206, doi:10.1029/2004GL020283, URL http://doi.wiley.com/10.1029/2004GL020283.

Arbuszewski, J., P. DeMenocal, A. Kaplan, and E. C. Farmer, 2010: On the fidelity of shell-derived δ18Oseawater estimates. *Earth and Planetary Science Letters*, **300 (3-4)**, 185–196, doi:10.1016/j.epsl.2010.10.035.

Bard, E., M. Arnold, J. Duprat, J. Moyes, and J.-c. Duplessy, 1987: Reconstruction of the last deglaciation: deconvolved records of delta18O profiles, micropaleontological variations and accelerator mass spectrometric 14C dating E. *Climate Dynamics*, **(1987)**, 101–112.

Berger, W. H. and G. R. Heath, 1968: Vertical mixing in pelagic sediments. *Journal of Marine Research*, **26 (2)**, 134–143.

Berger, W. H., R. F. Johnson, and J. S. Killingley, 1977: © 1977 Nature Publishing Group. *Nature*, **266 (April 21)**, 730–732, doi:10.1038/266309a0, `NIHMS150003`.

Bhend, J., J. Franke, D. Folini, M. Wild, and S. Br, 2012: of the Past An ensemble-based approach to climate reconstructions. 963–976, doi:10.5194/cp-8-963-2012.

Bickert, T. and G. Wefer, 2012: Late Quaternary Deep Water Circulation in the South Atlantic: Reconstruction from Carbonate Dissolution and Benthic Stable Isotopes. *The South Atlantic*, 599–620, doi:10.1007/978-3-642-80353-6_30.

Birdsey, R. A. and U. F. Service, 1998: North Atlantic Oscillation Dynamics Recorded in. *Science*, **282 (October)**, 446–449.

Boudreau, B. P., 1986a: Mathematics of tracer mixing in sediments: 1. spatially-dependent, diffusive mixing. *American Journal of Science*, **286**, 161–198.

Boudreau, B. P., 1986b: Mathematics of tracer mixing in sediments: 2. nonlocal mixing and biological conveynor-belt phenomena. *American Journal of Science*, **286**, 199–238.

Boudreau, B. P., 1994: Is burial velocity a master parameter for bioturbation? *Geochimica et Cosmochimica Acta*, **58 (4)**, 1243–1249, doi:10.1016/0016-7037(94)90378-6.

Boudreau, B. P. and D. M. Imboden, 1987: Mathematics of Tracer Mixing in Sediments : III . The Theory of Nonlocal Mixing within Sediments. *American Journal of Science*, **287**, 693–719, doi:10.2475/ajs.287.7.693.

Braconnot, P., S. P. Harrison, M. Kageyama, P. J. Bartlein, V. Masson-Delmotte, A. Abe-Ouchi, B. Otto-Bliesner, and Y. Zhao, 2012: Evaluation of climate models using palaeoclimatic data. *Nature Climate Change*, **2 (6)**, 417–424, doi:10.1038/nclimate1456, URL `http://dx.doi.org/10.1038/nclimate1456`.

Bradley, R. S., M. Vuille, M. Werner, R. Y. Chan, T. J. Osborne, K. R. Briffa, M. K. Hughes, and P. D. Jones, 2005: Proxy-Based Northern Hemisphere Surface Temperature Reconstructions: Sensitivity to Method, Predictor Network, Target Season, and Target Domain. *Journal of Climate*, **18 (13)**, 2308–2329, URL `http://works.bepress.com/raymond{_}bradley/7{%}5Cnhttp://works.bepress.com/cgi/viewcontent.cgi?article=1013{&}context=raymond{_}bradley{%}5Cnhttp://works.bepress.com/raymond{_}bradley/7/`.

Burgers, G. and P. J. van Leeuwen, 1998: Analysis scheme in the ensemble Kalman filter. *Monthly Weather Review*, 1719–1724, doi:10.1175/1520-0493(1998) 126<1719:ASITEK>2.0.CO;2, URL `http://journals.ametsoc.org/doi/pdf/10.1175/1520-0493(1998)126{%}3C1719:ASITEK{%}3E2.0.CO;2`.

Choi, J., F. Francois-Carcaillet, and B. P. Boudreau, 2002: Lattice-automaton bioturbation simulator (LABS): Implementation for small deposit feeders. *Computers and Geosciences*, **28 (2)**, 213–222, doi:10.1016/S0098-3004(01)00064-4.

Clark, P. U. and P. Huybers, 2009: Increasing rates of ice mass loss from the Greenland and Antarctic ice sheets revealed by GRACE. *Geophysical Research Letters*, **36 (19)**, 856–857, doi:10.1029/2009GL040222.

Crespin, E., H. Goosse, T. Fichefet, and M. E. Mann, 2009: The 15th century Arctic warming in coupled model simulations with data assimilation. *Climate of the Past*, **5 (3)**, 389–401, doi:10.5194/cp-5-389-2009.

Dee, D. P., et al., 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, **137 (656)**, 553–597, doi:10.1002/qj.828.

Dirren, S. and G. J. Hakim, 2005: Toward the assimilation of time-averaged observations. *Geophysical Research Letters*, **32 (4)**, 1–5, doi:10.1029/2004GL021444.

Doucet, d. F. N. G. N., A., 2001: *Sequential Monte Carlo methods in practice* . Springer.

Dowsett, H., et al., 2016: The PRISM4 (mid-Piacenzian) paleoenvironmental reconstruction. *Climate of the Past*, **12 (7)**, 1519–1538, doi:10.5194/cp-12-1519-2016.

Dubinkina, S., H. Goosse, Y. Sallaz-Damaz, E. Crespin, and M. Crucifix, 2011: TESTING A PARTICLE FILTER TO RECONSTRUCT CLIMATE CHANGES OVER THE PAST CENTURIES. **21 (12)**, 3611–3618, doi:10.1142/S0218127411030763.

Dutton, A. and K. Lambeck, 2012: Ice Volume and Sea Level During the Last Interglacial. **337 (July)**, 216–220.

Dutton, A., et al., 2015a: Sea-level rise due to polar ice-sheet mass loss during past warm periods. *Science*, **349 (6244)**, aaa4019–aaa4019, doi:10.1126/science.aaa4019.

Dutton, A., et al., 2015b: Sea-level rise due to polar ice-sheet mass loss during past warm periods. *Science*, **349 (6244)**, aaa4019–aaa4019, doi:10.1126/science.aaa4019.

Emile-Geay, J., K. M. Cobb, M. E. Mann, and A. T. Wittenberg, 2013: Estimating Central Equatorial Pacific SST Variability over the Past Millennium. Part II: Reconstructions and Implications. *Journal of Climate*, **26 (7)**, 2329–2352, doi:10.1175/JCLI-D-11-00511.1, URL http://journals.ametsoc.org/doi/abs/10.1175/JCLI-D-11-00511.1.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. 10143 pp., doi:10.1029/94JC00572.

Evensen, G., 2003: The Ensemble Kalman Filter : theoretical formulation and practical implementation. 343–367, doi:10.1007/s10236-003-0036-9.

Foster, D. W., 1983: BIOTURB: A Fortran Program to Simulate the Effects of Bioturbation on the Vertical Distribution of Sediment. **1 (1)**, 39–54.

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:10.1256/smsqj.55416.

Goldberg, E. D. and M. Koid, 1962: Geochronological studies of deep sea sediments by the ionium/thorium method. *Geochimica et Cosmochimica Acta*, **26**, 417?450.

Goosse, A. O. L. J. M. M. E. R. H. R. N. T. A. X. E., H, 2006a: of the Past The origin of the European ? Medieval Warm Period ? 99–113.

Goosse, H., 2012: The Medieval Climate Anomaly in simulations with data assimilation. *Quaternary International*, **279-280**, 173–174, doi:10.1016/j.quaint.2012.08.246.

Goosse, H., E. Crespin, A. D. Montety, M. E. Mann, and H. Renssen, 2010: Reconstructing surface temperature changes over the past 600 years using climate model simulations with data assimilation. **115**, 1–17, doi:10.1029/2009JD012737.

Goosse, H., J. Guiot, M. E. Mann, A. Dubinkina, and Y. Sallaz-Damaz, 2011: The medieval climate anomaly in Europe: Comparison of the summer and annual mean signals in two reconstructions and in simulations with data assimilation. *Global and Planetary Change*, **84-85**, 35–47, doi:10.1016/j.gloplacha.2011.07.002, URL `http://dx.doi.org/10.1016/j.gloplacha.2011.07.002`.

Goosse, H., Æ. W. Lefebvre, A. D. Montety, Æ. E. Crespin, and A. H. Orsi, 2009: Consistent past half-century trends in the atmosphere , the sea ice and the ocean at high southern latitudes. 999–1016, doi:10.1007/s00382-008-0500-9.

Goosse, R. H. A. T. B. R. M. M. E., H, 2006b: Using paleoclimate proxy-data to select optimal realisations in an ensemble of simulations of the climate of the past millennium. *Climate Dynamics*, 165–184, doi:10.1007/s00382-006-0128-6.

Gordon, N., D. J. Salmond, and A. F. M. Smith, 1993: Novel approach to nonlinear and linear Bayesian state estimation. *IEE Proceedings*, **140 (2)**, 107–113.

Goreau, T. J., 1977: Quantitavtive effects of sediment mixing on stratigraphy and biogeochemistry: a signal theory approach. *Nature*, **266 (April 21)**, 730–732, doi: 10.1038/266309a0, `NIHMS150003`.

Gottwald, G. A., 2014: Controlling balance in an ensemble Kalman filter. *Nonlinear Processes in Geophysics*, **21 (2)**, 417–426, doi:10.5194/npg-21-417-2014.

Guinasso, N. L. and D. R. Schink, 1975: Quantitative estimates of biological mixing rates in abyssal sediments. *Journal of Geophysical Research*, **80 (21)**, 3032–3043, doi: 10.1029/jc080i021p03032.

Hakim, G. J., J. Emile-geay, E. J. Steig, D. Noone, and M. David, 2016: The Last Millennium Climate Reanalysis Project : Framework and First Results. *Journal of Geophysical Research: Atmospheres*, **98195**, 1–56.

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 (11)**, 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.

HARADA, Y., et al., 2015: The JRA-55 Reanalysis: General Specifications and Basic Characteristics. *Journal of the Meteorological Society of Japan. Ser. II*, **93 (1)**, 5–48, doi: 10.2151/jmsj.2015-001, `arXiv:1011.1669v3`.

Harrison, S. P., P. J. Bartlein, and I. C. Prentice, 2016: What have we learnt from palaeoclimate simulations? *Journal of Quaternary Science*, **31 (4)**, 363–385, doi:10. 1002/jqs.2842.

Harrison, S. P., et al., 2014: Climate model benchmarking with glacial and mid-Holocene climates. *Climate Dynamics*, **43 (3-4)**, 671–688, doi:10.1007/ s00382-013-1922-6.

He, F., 2011: Simulating transient climate evolution of the last deglaciation with CCSM3. Ph.D. thesis.

Hol, J. D., T. B. Schon, and F. Gustafsson, 2006: 2006 ieee nonlinear statistical signal processing workshop. Curran Associates, Inc., Linkoping, Sweden, Institute of Electrical and Electronics Engineers (IEEE).

Houtekamer, P. L., 1966: Data Assimilation Using an Ensemble Kalman Filter Technique. *Semiconductors and Semimetals*, **1 (C)**, 289–313, doi:10.1016/S0080-8784(08) 62382-0.

Houtekamer, P. L. and H. L. Mitchell, 2001: A sequential ensemble {Kalman} filter for atmospheric data assimilation. *Monthly Weather Rev.*, **129 (1)**, 123–137, doi:10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2.

Houtekamer, P. L., H. L. Mitchell, G. Pellerin, B. Hansen, M. Charron, L. Spacek, and M. Buehner, 2005: Atmospheric Data Assimilation with an Ensemble Kalman Filter: Results with Real Observations. *Monthly Weather Review*, **133 (3)**, 604–620, doi:10.1175/mwr-2864.1.

Hull, P. M., P. J. S. Franks, and R. D. Norris, 2011: Mechanisms and models of iridium anomaly shape across the Cretaceous ? Paleogene boundary. *Earth and Planetary Science Letters*, **301 (1-2)**, 98–106, doi:10.1016/j.epsl.2010.10.031, URL `http://dx.doi.org/10.1016/j.epsl.2010.10.031`.

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.

Hutson, W. H., 1980: uncertainty and stratigraphie uncertainty. doi:10.1130/ 0091-7613(1980)8<127.

Iseries, A., 1996: *A First Course in the Numerical Analysis of Differential Equations*. Cambridge University Press, 447p.

J. Barkmeijer, T. N. P., T. Iversen, 2003: Forcing singular vectors and other sensitive model structures 1. 2401–2423, doi:10.1256/qj.02.126.

J. S. Whitaker, T. M. Hamill, 2002: Ensemble Data Assimilation without Perturbed Observations. 1913–1924.

Jacob, R., 1997: LOW FREQUENCY VARIABILITY IN A SIMULATED ATMOSPHERE OCEAN SYSTEM. Ph.D. thesis, University of Wisconsin-Madison.

Jones, G. A. and W. F. Ruddiman, 1982: Assessing the Global Meltwater Spike. *Quanternary research*, **172**, 148–172.

Jones, P. D., et al., 2009: The Holocene High-resolution palaeoclimatology of the last millennium : a review of current. *Library*, **1**, 3–49, doi:10.1177/0959683608098952.

Kalman, R. E., 1960: A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, **82**, 35–45, doi:10.1115/1.3662552.

Kalman, R. E. and R. S. Bucy, 1961: New Results in Linear Filtering and Prediction Theory. *Journal of Basic Engineering*, **83 (1)**, 95, doi:10.1115/1.3658902.

Kalnay, E., 2003: *Atmospheric modeling, data assimilation and predictability*. Cambridge University Press.

Kalnay, E., et al., 1996: The NCEP/NCAR 40-year reanalysis project. 437–471 pp., doi:http://dx.doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.

Kanamitsu, M., W. Ebisuzaki, J. Woollen, S.-K. Yang, J. J. Hnilo, M. Fiorino, and G. L. Potter, 2002: NCEP?DOE AMIP-II Reanalysis (R-2). *Bulletin of the American Meteorological Society*, **83 (11)**, 1631–1644, doi:10.1175/BAMS-83-11-1631, URL `http://journals.ametsoc.org/doi/10.1175/BAMS-83-11-1631`, 0601429.

Kennett, J., E. Roark, K. Cannariato, B. Ingram, and R. Tada, 2006: Latest Quaternary paleoclimatic and radiocarbon chronology, Hole 1017E, southern California margin. *Proceedings of the Ocean Drilling Program, 167 Scientific Results*, **167**, 249–254, doi:10.2973/odp.proc.sr.167.242.2000.

Kim, J. H., N. Rimbu, S. J. Lorenz, G. Lohmann, S. I. Nam, S. Schouten, C. Rühlemann, and R. R. Schneider, 2004: North Pacific and North Atlantic sea-surface temperature variability during the Holocene. *Quaternary Science Reviews*, **23 (20-22 SPEC. ISS.)**, 2141–2154, doi:10.1016/j.quascirev.2004.08.010.

Kitoh, A. and O. Arakawa, 1999: On overestimation of tropical precipitation by an atmospheric {GCM} with prescribed {SST}. *Geophysical Research Letters*, **26 (19)**, 2965–2968, doi:10.1029/1999GL900616.

Kopp, R. E., F. J. Simons, J. X. Mitrovica, A. C. Maloof, and M. Oppenheimer, 2009: Probabilistic assessment of sea level during the last interglacial stage. *Nature*, **462 (7275)**, 863–867, doi:10.1038/nature08686.

Kuensch, H. R. and P. Fearnhead, 2018: Particle Filters and Data Assimilation. *Ssrn*, doi:10.1146/annurev-statistics-031017-100232.

Laepple, T. and P. Huybers, 2013: Reconciling discrepancies between Uk37 and Mg/Ca reconstructions of Holocene marine temperature variability. *Earth and Planetary Science Letters*, **375**, 418–429, doi:10.1016/j.epsl.2013.06.006, URL `http://dx.doi.org/10.1016/j.epsl.2013.06.006`.

Lavaysse, C., M. Carrera, S. BELAIR, N. Gagnon, R. Frenette, M. Charron, and M. Yau, 2012: Impact of Surface Parameter Uncertainties within the Canadian Regional Ensemble Prediction System. *Monthly Weather Review*, **141 (5)**, 1506–1526, doi:10.1175/mwr-d-11-00354.1.

Leduc, G., R. Schneider, J. H. Kim, and G. Lohmann, 2010: Holocene and Eemian sea surface temperature trends as revealed by alkenone and Mg/Ca paleothermometry. *Quaternary Science Reviews*, **29 (7-8)**, 989–1004, doi:10.1016/j.quascirev.2010.01.004, URL `http://dx.doi.org/10.1016/j.quascirev.2010.01.004`.

Lee, Y. and A. J. Majda, 2016: State estimation and prediction using clustered particle filters. *Proceedings of the National Academy of Sciences*, **113 (51)**, 201617 398, doi:10.1073/pnas.1617398113, URL `http://www.pnas.org/lookup/doi/10.1073/pnas.1617398113`.

Leeuwen, P., 2009: Particle Filtering in Geophysical Systems. 4089–4114, doi:10.1175/2009MWR2835.1.

Lisiecki, L. E. and M. E. Raymo, 2005: A Pliocene-Pleistocene stack of 57 globally distributed benthic D 18 O records. **20**, 1–17, doi:10.1029/2004PA001071.

Liu, Z., S. Wu, S. Zhang, Y. Liu, and X. Rong, 2013: Ensemble data assimilation in a simple coupled climate model: The role of ocean-atmosphere interaction. *Advances in Atmospheric Sciences*, **30 (5)**, 1235–1248, doi:10.1007/s00376-013-2268-z.

Liu, Z., et al., 2009: Transient simulation of last deglaciation with a new mechanism for bolling-allerod warming. *Science*, **325 (5938)**, 310–314, doi:10.1126/science.1171041.

Liu, Z., et al., 2014: Coherent changes of southeastern equatorial and northern African rainfall during the last deglaciation. *Science*, **346 (6214)**, 1223–1228, doi:10.1126/science.1259531.

Lorenz, E., 1996: Predicability: a Problem partially solved. *Proceedings of Seminar on Predictability*, **1**.

Lu, F., Z. Liu, Y. Liu, S. Zhang, and J. Robert, 2017: Understanding the control of extratropical atmospheric variability on ENSO using a coupled data assimilation approach. *Climate Dynamics*, **48 (9-10)**, 3139–3160, doi:10.1007/s00382-016-3256-7.

Lu, F., Z. Liu, S. Zhang, and Y. Liu, 2015: Coupled Data Assimilation Using Leading Averaged Coupled Covariance Part II: CGCM Applications. *Monthly Weather Review*, **In prepara**, 4645–4659, doi:10.1175/MWR-D-14-00322.1.

Luo, J.-J., S. Masson, S. K. Behera, and T. Yamagata, 2008: Extended ENSO Predictions Using a Fully Coupled Ocean?Atmosphere Model. *Journal of Climate*, **21 (1)**, 84–93, doi:10.1175/2007JCLI1412.1, URL http://journals.ametsoc.org/doi/abs/10.1175/2007JCLI1412.1.

Mairesse, A., H. Goosse, P. Mathiot, H. Wanner, and S. Dubinkina, 2013: Investigating the consistency between proxy-based reconstructions and climate models using data assimilation : a mid-Holocene case study. 5194, doi:10.5194/cp-9-2741-2013.

Mann, M. E., J. Park, and R. S. Bradley, 2008a: Global interdecadal and century-scale climate oscillations during the past five centuries. **41 (3)**.

Mann, M. E., Z. Zhang, M. K. Hughes, R. S. Bradley, S. K. Miller, S. Rutherford, and F. Ni, 2008b: Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia. *Proceedings of the National Academy of Sciences of the United States of America*, **105 (36)**, 13 252–7, doi:10.1073/ pnas.0805721105, URL `http://www.pnas.org/content/105/36/13252.full`.

Mann, M. E., et al., 2009: Medieval Climate Anomaly. *Science*, **326 (November)**, 1256–1260, doi:10.1126/science.1166349, URL `http://www.pages-igbp.org/ products/2011-03-28-16-23-06/249-pages-news-vol-19-no-1`.

Mann, R., Michael E.and Bradley and M. Hughes, 1998: Global scale temperature patterns and climate forcing over the past six centuries. *Nature*, **392**, 779–87., URL `https://search-proquest-com.ezaccess.libraries.psu. edu/docview/204482251?pq-origsite=summon`.

Marcott, S. a., J. D. Shakun, P. U. Clark, and A. C. Mix, 2013: A Reconstruction of Regional. *Science (New York, N.Y.)*, **339 (6124)**, 1198–1201, doi:10.1126/science. 1228026, URL `http://www.ncbi.nlm.nih.gov/pubmed/23471405`.

Mathiot, P., H. Goosse, X. Crosta, B. Stenni, M. Braida, H. Renssen, and C. J. V. Meerbeeck, 2013: Using data assimilation to investigate the causes of Southern BP. 887–901, doi:10.5194/cp-9-887-2013.

Matsikaris, A., M. Widmann, and J. Jungclaus, 2015: On-line and off-line data assimilation in palaeoclimatology: A case study. *Climate of the Past*, **11 (1)**, 81–93, doi:10.5194/cp-11-81-2015.

Max, L., et al., 2012: Sea surface temperature variability and sea-ice extent in the subarctic northwest Pacific during the past 15,000 years. *Paleoceanography*, **27 (3)**, 1–20, doi:10.1029/2012PA002292.

McKay, N. P., J. T. Overpeck, and B. L. Otto-Bliesner, 2011: The role of ocean thermal expansion in Last Interglacial sea level rise. *Geophysical Research Letters*, **38 (14)**, 4–9, doi:10.1029/2011GL048280.

Meyers, S. R., 2014: *Astrochron: An R Package for Astrochronology*. URL `https://cran.r-project.org/package=astrochron`.

Meysman, F. J. R., B. P. Boudreau, and J. J. Middelburg, 2003: Relations between local , nonlocal , discrete and continuous models of bioturbation. 391–410.

Mix, A. C., E. Bard, G. Eglinton, L. D. Keigwin, A. C. Ravelo, and Y. Rosenthal, 2000: Alkenones and multiproxy strategies in paleoceanographic studies. *Geochemistry, Geophysics, Geosystems*, **1 (11)**, doi:10.1029/2000GC000056.

Neelamani, R., 2008: *Deconvolution and Optimal Filtering in Seismology*, 1571–1584. Springer New York, New York, NY, doi:10.1007/978-0-387-30441-0_87, URL `https://doi.org/10.1007/978-0-387-30441-0_87`.

Ohkouchi, N., T. I. Eglinton, and L. D. Keigwin, 2008: Spatial and Temporal Offsets Between Proxy Records in a Sediment Drift Spatial and Temporal Offsets Between Proxy Records in a Sediment Drift. *Science*, **1224 (2002)**, 1224–1228, doi:10.1126/science.1075287.

Oppeheim, A. V., R. W. Schafer, and J. R. Buck, 1996: *Discrete-time signal processing*. 1–4 pp.

Otto-Bliesner, B. L., E. C. Brady, G. Clauzet, R. Tomas, S. Levis, and Z. Kothavala, 2006: Last glacial maximum and Holocene climate in CCSM3. *Journal of Climate*, **19 (11)**, 2526–2544, doi:10.1175/JCLI3748.1.

Otto-Bliesner, B. L., A. Jahn, R. Feng, E. C. Brady, A. Hu, and M. Löfverström, 2017: Amplified North Atlantic warming in the late Pliocene by changes in Arctic gateways. *Geophysical Research Letters*, **44 (2)**, 957–964, doi:10.1002/2016GL071805.

Otto-Bliesner, B. L., S. J. Marshall, J. T. Overpeck, and G. H. Miller, 2009: Simulating Arctic Climate Warmth and Ice field Retreat in the Last Interglaciation Bette. *Science*, **311 (2006)**, 1751–1755, doi:10.1126/science.1120808.

Otto-bliesner, B. L., S. J. Marshall, J. T. Overpeck, and G. H. Miller, 2009: Simulating Arctic Climate Warmth and Icefield Retreat in the Last Interglaciation. *Science*, **311 (2006)**, 1751–1755, doi:10.1126/science.1120808.

Perez-Sanz, A., G. Li, P. González-Sampériz, and S. P. Harrison, 2014: Evaluation of modern and mid-Holocene seasonal precipitation of the Mediterranean and northern Africa in the CMIP5 simulations. *Climate of the Past*, **10 (2)**, 551–568, doi:10.5194/cp-10-551-2014.

Poterjoy, J., 2015: A localized particle filter for high-dimensional nonlinear systems. *Monthly Weather Review*, **(2001)**, 151020092806 005, doi:10.1175/MWR-D-15-0163. 1.

Ravelo, A. C. and C. Hillaire-Marcel, 2007: *The Use of Oxygen and Carbon Isotopes of Foraminifera in Paleoceanography*, Vol. 1. 735–764 pp., doi:10.1016/S1572-5480(07) 01023-8.

Rohling, E. J., K. Grant, C. Hemleben, M. Siddall, B. A. A. Hoogakker, M. Bolshaw, and M. Kucera, 2008: High rates of sea-level rise during the last interglacial period. *Nature Geoscience*, **1 (1)**, 38–42, doi:10.1038/ngeo.2007.28.

Ruddiman, W. F. and L. K. Glover, 1972: Vertical mixing of ice-rafted volcanic ash in North Atlantic sediments. *Geological Society of America Bulletin*, **(September)**, 1–251, doi:10.4324/9781315744292.

Ruddiman, W. F., G. A. Jones, T. H. Peng, L. K. Flover, B. P. Glass, and P. J. Liebertz, 1980a: Tests for size and shape dependence in deep-sea mixing. *Sediment Geology*, **25 (2957)**.

Ruddiman, W. F., B. Molfino, A. Esmay, and E. Pokras, 1980b: Evidence bearing on the mechanism of rapid deglaciation. **3 (1980)**, 65–87.

Schiffelbein, P., 1984a: Effects of benthic mixing on the information contend of deep-sea stratigraphical signals. *Nature*, **311**.

Schiffelbein, P., 1984b: Time series analysis program for stratigraphy data. Ph.D. thesis.

Schiffelbein, P. and L. Dotman, 1986: measured in 6180 from planktonic foraminifer good agreement with l? C-based estimates from the sediment) being characterized by lighter 6180 ? 180 ' and % CaCO. *Journal of Geophysical Research*, **91**, 3821–3835.

Schneider, B., G. Leduc, and W. Park, 2010: Disentangling seasonal signals in Holocene climate trends by satellite-model-proxy integration. *Paleoceanography*, **25 (4)**, 1–13, doi:10.1029/2009PA001893.

Seki, O., R. Ishiwatari, and K. Matsumoto, 2002: Millennial climate oscillations in NE Pacific surface waters over the last 82 kyr: New evidence from alkenones. *Geophysical Research Letters*, **29 (23)**, 59–1–59–4, doi:10.1029/2002gl015200.

Shackleton, N. J., 1974: Attainment of isotopic equilibrium between ocean water and the benthonic foraminifera genus Uvigerina: Isotopic changes in the ocean during the last glacial. *Les Méthodes quantitatives d'étude des variations du climat au cours du pléistocène, Colloque international du CRNS. Gif-sur-Yvette.*, **219**, 203–210, doi:10.1021/la3005486.

Shackleton, N. J., 1987: Oxygen isotopes, ice volume and sea level. *Quaternary Science Reviews*, **6 (3-4)**, 183–190, doi:10.1016/0277-3791(87)90003-5.

Shackleton, N. J., M. Chapman, M. F. Sánchez-Goni, D. Pailler, and Y. Lancelot, 2002: The classic marine isotope substage 5e. *Quaternary Research*, **58 (1)**, 14–16, doi:10.1006/qres.2001.2312.

Shackleton, N. J., M. F. Sánchez-Goñi, D. Pailler, and Y. Lancelot, 2003: Marine isotope substage 5e and the Eemian interglacial. *Global and Planetary Change*, **36 (3)**, 151–155, doi:10.1016/S0921-8181(02)00181-9.

Shakun, J. D., et al., 2012: Global warming preceded by increasing carbon dioxide concentrations during the last deglaciation. *Nature*, **484 (7392)**, 49–54, doi:10.1038/nature10915.

Shull, D. H., 2001: Transition-matrix model of bioturbation and radionuclide diagenesis. *Limnology and Oceanography*, **46 (4)**, 905–916.

Smerdon, J. E., 2012: Climate models as a test bed for climate reconstruction methods: Pseudoproxy experiments. *Wiley Interdisciplinary Reviews: Climate Change*, **3 (1)**, 63–77, doi:10.1002/wcc.149.

Snyder, C., T. Bengtsson, P. Bickel, and J. Anderson, 2008: Obstacles to High-Dimensional Particle Filtering. *Monthly Weather Review*, **136 (12)**, 4629–4640, doi:10.1175/2008MWR2529.1.

Steiger, N. and G. Hakim, 2016: Multi-timescale data assimilation for atmosphere-ocean state estimates. *Climate of the Past*, **12 (6)**, 1375–1388, doi:10.5194/cp-12-1375-2016.

Steiger, N. J., E. J. Steig, S. G. Dee, G. H. Roe, and G. J. Hakim, 2017: Climate reconstruction using data assimilation of water isotope ratios from ice cores. *Journal of Geophysical Research*, **122 (3)**, 1545–1568, doi:10.1002/2016JD026011.

Steiger N., S. E. J. B. D. S. R. G. H., Hakim G., 2013: Assimilation of Time-Averaged Pseudoproxies for Climate Reconstruction. **(2012)**, 426–441, doi:10.1175/JCLI-D-12-00693.1.

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 interannual climate variations. *Journal of Geophysical Research: Oceans*, **113**, 1–21, doi:10.1029/2008JC004741.

Suranjana Saha, H.-L. P. X. W. J. W. S. N. P. T. R. K. J. W. D. B. H. L. D. S. R. G. G. G. J. W. Y.-T. H. H.-y. C. H.-M. H. J. J. S. M. I. R. T. D. K. P. V. D. D. K. J. D. M. E. J. M. H. W. R. Y. S. L. H. v. d. D. A. K. W. W. C. L. M. C. Y. X. B. H. J.-K. S. W. E. R. L. P. X. M. C. S. Z. W. H. C.-Z. Z. Q. L. Y. C. Y. H. L. C. R. W. R. G. R., Shrinivas Moorthi and M. Goldberg, 2010: The NCEP Climate Forecast System Reanalysis. *Bulletin of the American Meteorological Society*, **91 (8)**, 1015–1058, doi:10.1175/2010bams3001.1, 9809069v1.

Taner, M. T., 1992: Technical report. Rock Solid Images, Inc, `http://www.rocksolidimages.com/attributes-revisited/#_Toc328470897`.

Tardif, R., G. J. Hakim, and C. Snyder, 2015: Coupled atmosphere?ocean 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.

Teal, L. R., M. T. Bulling, E. R. Parker, and M. Solan, 2008: Global patterns of bioturbation intensity and mixed depth of marine soft sediments. **2 (June)**, 207–218, doi:10.3354/ab00052.

Thompson, A. J., C. B. Skinner, C. J. Poulsen, and J. Zhu, 2019: Modulation of mid?Holocene African rainfall by dust aerosol direct and indirect effects. *Geophysical Research Letters*, 2018GL081225, doi:10.1029/2018GL081225, URL `https://onlinelibrary.wiley.com/doi/abs/10.1029/2018GL081225`.

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 (7)**, 1485–1490, doi:10.1175/1520-0493(2003)131<1485:ESRF>2.0.CO;2.

Trauth, M. H., 2013: TURBO2 : A MATLAB simulation to study the effects of bioturbation on paleoceanographic time series. *Computers and Geosciences*, **61**, 1–10, doi:10.1016/j.cageo.2013.05.003, URL `http://dx.doi.org/10.1016/j.cageo.2013.05.003`.

Turner, S. K., P. M. Hull, L. R. Kump, and A. Ridgwell, 2017: A probabilistic assessment of the rapidity of PETM onset. *nature communications*, doi:10.1038/s41467-017-00292-2.

Turney, C. S. M. and R. T. Jones, 2010: Does the Agulhas Current amplify global temperatures during super-interglacials? *Journal of Quaternary Science*, **25 (6)**, 839–843, doi:10.1002/jqs.1423.

Uppala, S. M., et al., 2005: The ERA-40 re-analysis. *Quarterly Journal of the Royal Meteorological Society*, **131 (612)**, 2961–3012, doi:10.1256/qj.04.176, URL `http://doi.wiley.com/10.1256/qj.04.176`.

Veeh, H. H., 1966: ages of Pleistocene high sea level stand. *Journal of Geophysical Research*, **71 (14)**, 3379, doi:10.1029/JZ071i014p03379.

von Storch, C. U.-G.-R. F. J. J. M. V. R. W. M., H. and E. Zorita, 2000: Combining paleoclimatic evidence and GCMs by means of data assimilation through upscaling and nudging (DATUN). *Proc. 11th Symposium on Global Change StudiesA, merican Meteorological Society, Long Beach, CA.*

Waelbroeck, C., L. C. Skinner, L. Labeyrie, J.-C. Duplessy, E. Michel, N. Vazquez Riveiros, J.-M. Gherardi, and F. Dewilde, 2011: The timing of deglacial circulation changes in the Atlantic. *Paleoceanography*, **26 (3)**, n/a–n/a, doi:10.1029/2010pa002007.

White, G., et al., 2002: The NCEP?NCAR 50?Year Reanalysis: Monthly Means CD?ROM and Documentation. *Bulletin of the American Meteorological Society*, **82 (2)**, 247–267, doi:10.1175/1520-0477(2001)082<0247:tnnyrm>2.3.co;2.

Widmann, M., H. Goosse, G. V. D. Schrier, R. Schnur, and J. Barkmeijer, 2010: of the Past Using data assimilation to study extratropical Northern Hemisphere climate over the last millennium. 627–644, doi:10.5194/cp-6-627-2010.

Y. Liu, S. Z.-R. J. F. L. X. R. S. W., Z. Liu, 2014a: Ensemble-Based Parameter Estimation in a Coupled General Circulation Model. *Journal of Climate*, **27**, 7151–7162, doi:10.1175/JCLI-D-13-00406.1.

Y. Liu, S. Z.-X. R. R. J. S. W. F. L., Z. Liu, 2014b: Ensemble-Based Parameter Estimation in a Coupled GCM Using the Adaptive Spatial Average Method. *JOURNAL OF CL IMATE*, **27**, 4002–4014, doi:10.1175/JCLI-D-13-00091.1.

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, J., Z. Liu, E. C. Brady, D. W. Oppo, P. U. Clark, A. Jahn, S. A. Marcott, and K. Lindsay, 2017: Asynchronous warming and δ 18 O evolution of deep Atlantic water masses during the last deglaciation. *Proceedings of the National Academy of Sciences*, **114 (42)**, 11 075–11 080, doi:10.1073/pnas.1704512114.

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

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 (1)**, 3541–3564, doi:10.1175/MWR3466.1.

Zhang, S.-Q., M. J. Harrison, A. T. Wittenberg, A. Rosati, J. L. Anderson, and V. Balaji, 2005: Initialization of an ENSO Forecast System Using a Parallelized Ensemble Filter. *Monthly Weather Review*, **133**, 3176–3201, doi:10.1175/MWR3024.1.

Zhen, Y. and F. Zhang, 2014: A Probabilistic Approach to Adaptive Covariance Localization for Serial Ensemble Square Root Filters. *Monthly Weather Review*, **142 (12)**, 4499–4518, doi:10.1175/mwr-d-13-00390.1.

Zhu, J., et al., 2017a: Investigating the Direct Meltwater Effect in Terrestrial Oxygen-Isotope Paleoclimate Records Using an Isotope-Enabled Earth System Model. *Geophysical Research Letters*, **44 (24)**, 12,501–12,510, doi:10.1002/2017GL076253.

Zhu, J., et al., 2017b: Reduced ENSO variability at the LGM revealed by an isotope-enabled Earth system model. *Geophysical Research Letters*, **44 (13)**, 6984–6992, doi: 10.1002/2017GL073406.