Check for updates

JAMES | Journal of Advances in Modeling Earth Systems®

RESEARCH ARTICLE
10.1029/2022MS003094

Special Section:
Data assimilation for Earth system models

Large Ensemble Particle Filter for Spatial Climate Reconstructions Using a Linear Inverse Model

B. Jebri[1] and M. Khodri[1]

[1]LOCEAN-IPSL, Sorbonne Université, CNRS/IRD/UPMC/MNHN, Paris, France

Key Points:

- Linear inverse modeling trained as emulators of CMIP-class models produce realistic climate and skillful ensemble forecasts as compared to the host-based model
- An online particle filter incorporating LIMs (SIR-LIM) overcomes the so-called data assimilation "curse of dimensionality" problem
- SIR-LIM reconstructions for the last millennium allow for uncertainty quantifications due to climate model biases and proxy spatial sampling

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

B. Jebri,
beyrem.jebri@locean.ipsl.fr

Citation:

Jebri, B., & Khodri, M. (2023). Large ensemble particle filter for spatial climate reconstructions using a linear inverse model. *Journal of Advances in Modeling Earth Systems*, *15*, e2022MS003094. https://doi.org/10.1029/2022MS003094

Received 19 MAR 2022
Accepted 14 FEB 2023

© 2023 The Authors. Journal of Advances in Modeling Earth Systems published by Wiley Periodicals LLC on behalf of American Geophysical Union. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

Abstract Proxy records that document the last 2000 years of climate provide evidence for the wide range of the natural climate variability from inter-annual to secular timescales not captured by the short window of recent direct observations. Assessing climate models ability to reproduce such natural variations is crucial to understand climate sensitivity and impacts of future climate change. Paleoclimate data assimilation (PDA) offers a powerful way to extend the short instrumental period by optimally combining the physics described by General Circulation Climate Models (GCMs) with information from available proxy records while taking into account their uncertainties. Here we present a new PDA approach based on a sequential importance resampling (SIR) Particle filter (PF) that uses Linear Inverse Modeling (LIM) as an emulator of several CMIP-class GCMs. We examine in a perfect-model framework the skill of the various LIMs to forecast the dynamics of the surface temperatures and provide spatial field reconstructions over the last millennium in a SIR PF. Our results show that the LIMs allow for skillful ensemble forecasts at 1-year lead-time based on GCMs dynamical knowledge with best prediction in the tropics and the North Atlantic. The PDA further provides a set of physically consistent spatial fields allowing robust uncertainty quantification related to climate models biases and proxy spatial sampling. Our results indicate that the LIM yields dynamical memory improving climate variability reconstructions and support the use of the LIM as a GCM-emulator in real reconstruction to propagate large ensembles of particles at low cost in SIR PF.

Plain Language Summary To detect and attribute anthropogenic climate changes, it is necessary to quantify and understand the wide range of the natural climate variability. The length of the instrumental period is relatively short for investigating slow climate features and their regional impacts. The field of paleoclimate data assimilation offers a way to extend the window of observation and provide physically consistent spatial fields by combining information from both CMIP-class climate models and proxy records documenting the last 2000 years. These approaches are however confronted with limitations due to the use of expensive global climate models over long periods and the very large number of simulations required to correctly describe the space of possible climate states, which increases exponentially with the dimension of the problem. Here we present a new Proxy Data Assimilation strategy that relies on a statistical model called LIM to reproduce the spatiotemporal dynamics of the surface temperatures simulated by costly CMIP-class climate models. This climate model emulator is then used to simulate large sets of particles that are optimally combined with the information provided by proxy records to deduce skillful spatial climate fields reconstructions and their low-frequency variability over the Common Era.

1. Introduction

Much of what is known about the climate system has been deduced from climate models, instrumental observations, climate proxies derived from natural archives (corals, speleothems, marine sediments, etc.) and our physical understanding of robust identified phenomena. While instrumental observations document climate variability over a relatively short period subject to anthropogenic forcing, climate proxies covering the Common Era (CE) are spatially sparser, less precise with a time resolution of about a year and provide partial information on the natural climate variability beyond the instrumental period (PAGES2k Consortium, 2017). Although more complicated to use, the increasing availability of climate proxies has led many studies to innovate and develop new approaches to reconstruct past climate indices or spatial fields (PAGES2k Consortium, 2019). Most of them correspond to statistical approaches which extend the covariance relationships found between proxies and climate features over the instrumental period to reconstruct past climate states (Mann et al., 2008; Neukom et al., 2014; Smerdon & Pollack, 2016). Since these approaches are only based on statistical relationships without strong dynamical constraints, they do not guarantee physical consistency and show important discrepancies depending

on the method used especially when it comes to spatial reconstructions (PAGES2k Consortium, 2019; Wang et al., 2015).

At the same time, significant advances have been made to improve the physics of General Circulation Climate Models (GCMs; Tian & Dong, 2020) and with the increase of computing power, large ensembles (LE) of GCMs simulations are now widely used to understand the contribution of external forcings and explore the range of climate variability (Boucher et al., 2020; Jebri et al., 2020; Stevenson et al., 2019; Zanchettin et al., 2013). However, the conclusions drawn from these studies stay incomplete since the simulations are only constrained with external forcing and a large part of the climate variability is internal to the climate system. Compared to LEs, the field of Paleoclimate data assimilation (PDA) appears as a complementary approach since it optimally combines the climate information embedded in proxy records with the dynamical knowledge included in climate models.

Among the most commonly used data assimilation strategies we find the Kalman filters (KFs), variational approaches and Particle filters (PFs). Variational methods give a unique solution corresponding to the optimum values that maximize the conditional probability given the observations, while ensemble data assimilation approaches such as PF and ensemble KFs (EnKFs) describe the ensemble of plausible states and the evolution of the conditional probability density function (Houtekamer & Zhang, 2016; Van Leeuwen, 2009; Whitaker & Hamill, 2002). While variational strategies are used to assimilate dense observations and produce reanalysis products (e.g., ERA-Interim; Dee et al., 2011), filters such as EnKFs and PFs are preferred in the field of PDA (Liu et al., 2017). Variational approaches are indeed not suitable in the case of PDA due to a much sparser and smaller number of observations, and their large uncertainties (Bhend et al., 2012; Hargreaves & Annan, 2002; Widmann et al., 2010). Despite recent significant progresses, ensemble data assimilation methods are still confronted with stumbling blocks related to the use of computationally expensive GCMs over long periods. Practical solutions had to be developed, giving rise, for example, to so-called "offline" approaches based on simulations already performed before the analysis (Goosse et al., 2006; Hakim et al., 2016; Matsikaris et al., 2015). While the offline approaches use a set of simulations already performed before the analysis to reconstruct the desired field afterward (Goosse et al., 2006; Hakim et al., 2016), the online approaches orient the trajectory of the simulations during the analysis according to the available observations. In theory online approaches are preferable to allow for continuity and physical consistency along the assimilation process (Goosse et al., 2010; Liu et al., 2017; Okazaki et al., 2021). In practice the choice between these two approaches will depend on the system under study (Matsikaris et al., 2015). Online approaches are recommended in the case of predictable systems, whereas offline approaches are simpler to implement in the case of systems with little or no predictability.

A data assimilation approach can be described by the evolution function F of the system states and by the measurement function G giving the relationships between the system states and the available observations. The F and G functions can be formalized as follows:

$$X_{i+1} = F(X_i, f_{i+1}, e) \tag{1}$$

$$Y_{i+1} = G(X_{i+1}, \eta) \tag{2}$$

Where X_{i+1} represents the system states at time $i + 1$ given the system state at the previous time step X_i, f_i a possible term of external forcing and e the model error. The measurement function G specifies the relationships between the complete system state X and the available observations Y with η representing the measurement error. F and G are not necessarily linear functions while η and e can be non-Gaussian errors and do not need to be additive. In the case of offline data assimilation system, since the analysis is produced from an estimated distribution that does not depend on the system states at the previous time steps, no dynamical model is used in Equation 1. For basic KFs, F and G are assumed linear and the errors η and e are taken as Gaussian and additive. For EnKFs the function F is not assumed linear anymore but the measurement operator G stays linear and the measurement noise η is still assumed to be Gaussian. In the case of PFs no assumptions are made on the form of F and G functions and on the associated uncertainties.

Other key differences between EnKFs and PFs, is that for the latter the conditional probability density function is empirically estimated and the measurement function is defined by the dynamical model itself while for the former the prior densities are assumed Gaussian and the measurement function is solved as a linear system of equations for the posterior mean and covariance (Kurosawa & Poterjoy, 2021). Both are confronted with the

large number of simulations needed to overcome the so-called "curse of dimensionality" problem encountered to solve high-dimensional systems and related to the number of particles required to correctly describe the space of possible states that increases exponentially with dimension (Snyder et al., 2008). The theoretical solution being to increase the number of particles exponentially with the number of degrees of freedom, simplifications have been proposed to reduce the size of the problem. Over the past few years, KFs and their ensemble versions have been subject to several adaptations such as the inclusion of covariance inflation and covariance localization to address the high dimension problem specific to spatial reconstructions with a limited amount of GCMs particles (Acevedo et al., 2016; Liu et al., 2017). Another important difference is that for EnKF, each particle is shifted toward the observations during the analysis step while in sequential importance resampling particle filter (SIR-PF) the particles are not moved and no statistical corrections are applied on the particles. Compared to EnKF strategies, very LE are therefore needed for PFs in high dimension system to avoid the particles depletion. On the other hand, especially when considering multi-layer models, no artificial model shocks are potentially caused in PFs, like during the EnKF updates.

Previous works have applied online PFs relying on LOVECLIM intermediate complexity model to successfully reconstruct large-scale or regional past climate features averaged over large domains (Goosse, 2017). However, due to the large number of particles required, no skillful detailed global spatially resolved climate fields have so far been produced with online PFs (Dubinkina & Goosse, 2013; Dubinkina et al., 2011; Goosse, 2017; Goosse et al., 2010; Liu et al., 2017). It is also worth noting that the built-in model-dependency in previous EnKF or PFs PDA products prevents a thorough quantification of the uncertainties related to both the model physics and observational datasets used.

We present here a PF based on a sequential importance resampling particle filter (SIR-PF) that uses LIM as an emulator of GCMs (labeled thereafter SIR-LIM), providing dynamical memory from various CMIP-class model physics and allowing for spatial reconstructions of the climate variability over the CE. This choice has been motivated by the fact that the LIM (Penland & Matrosova, 1994) can approximate the dynamics of GCMs with forecast skill comparable and sometimes even better than what is generally obtained with GCMs (Newman, 2013; Richter et al., 2020) while allowing for the large ensemble of experiments required to sample the high dimension state-space. The LIM is based on a continuous-time Markov chain making the assumption that the predictable part of the represented system is substantively linear and that the non-linear processes can be approached by a Gaussian white noise with a spatial structure but uncorrelated in time. The LIM is used to forward the particles in time to estimate the predictable part of the dynamics and the propagation of the uncertainties associated to the non-linear part between the resampling steps. The PF then allows to select the particles closest to the observations taking in consideration the impact associated to forced unpredictable events and the non-linear dynamics. Based on previous works, the LIM can approximate well many features of the climate system and has been shown to provide satisfactory forecast skill for both oceanic and atmospheric fields at various lead times (Johnson et al., 2000; Newman, 2013; Penland & Matrosova, 1998; Penland & Sardeshmukh, 1995; Winkler et al., 2001). It has also been recently used successfully as a forward model in an Ensemble KF PDA approach (Perkins & Hakim, 2021).

The present study aims to: (a) assess the LIM skills in reproducing the inter annual to decadal dynamics of surface temperatures variability as simulated by expensive CMIP5 and CMIP6-class GCMs, (b) provide a benchmark test-bed in a perfect model framework for a SIR particle filter using a LIM as an integration climate model and (c) quantify limitations and uncertainties of derived climate reconstructions related to the model used and observational spatial sampling. The paper is organized as follows: in Section 2 we introduce the model and data used, and describe the basic equations for the LIM. In Section 3, we describe the PF implementation and introduce the different experiments and metrics used to assess the LIM and SIR-LIM skills. In Section 4, we assess the performances of several LIMs trained on various sets of CMIP5 and CMIP6-class models and discuss their skill sources. In Section 5, we evaluate the SIR-LIMs reconstructions in a perfect model framework and estimate their sensitivity to several parameters. We finally summarize and discuss the results in Section 6.

2. Data Set, Model and Data Assimilation Method

2.1. Data Set

This study relies on a perfect-model framework, which uses as the target to be reconstructed the surface temperatures of the past1000 model experiment covering the 850–1850 CE climate period, performed as part of 4th phase

Table 1
List of CMIP5 and CMIP6-Class Models Considered

	Models	Reference	Institution
CMIP5	IPSL-CM5A-LR	Dufresne et al. (2013)	Institute Pierre Simon Laplace
	GISS-E2-R	Miller et al. (2014)	NASA Goddard Institute for Space Studies
	MIROC-ESM	Watanabe et al. (2011)	Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute, and National Institute for Environmental Studies
	MPI-ESM-P	Jungclaus et al. (2013)	Max Planck Institute for Meteorology
	CCSM4	Gent et al. (2011)	National Center for Atmospheric Research
CMIP6	IPSL-CM6A-LR	This study; Boucher et al. (2020); Lurton et al. (2020)	Institut Pierre Simon Laplace
	MIROC-ES2L	Ohgaito et al. (2021)	Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute, National Institute for Environmental Studies and RIKEN Center for Computational Science
	MPI-ESM1-2-LR	Mauritsen et al. (2019); van Dijk et al. (2022)	Max Planck Institute for Meteorology
	MRI-ESM2-0	Yukimoto et al. (2019)	Meteorological Research Institute

of the paleoclimate modeling intercomparison project (PMIP4; Jungclaus et al., 2017; Kageyama et al., 2017) with the IPSL-CM6A-LR CMIP6 model version of the Institut Pierre-Simon Laplace (Boucher et al., 2020). The IPSL-CM6A-LR model couples the atmospheric component LMDZ (version 6A-LR; Hourdin et al., 2020) to the land surface model ORCHIDEE (version 2.0; d'Orgeval et al., 2008) and to the ocean model NEMO (version 3.6), which includes other models to represent sea-ice interactions (NEMO-LIM3; Rousset et al., 2015) and biogeochemistry processes (NEMO-PISCES; Aumont et al., 2015). The atmospheric and land-surface grid have a resolution of 2.5° in longitude and 1.3° in latitude with 79 vertical layers. The oceanic component has 75 vertical levels with a mean spatial horizontal resolution of about 1° and a refinement of 1/3° near the equator. This model reproduces fairly well the El Niño Southern Oscillation (ENSO) seasonality despite the sea surface temperatures anomalies extending too westward in the central Pacific during El Niño events. The spatial pattern of the Atlantic Multidecadal Variability (AMV) teleconnection in the Pacific is consistent with observations but the Atlantic tropical variability is relatively weaker. Unlike most current state-of-the-art CMIP6 models, the IPSL-CM6A-LR model simulates a predominant secular variability in the Atlantic with AMV peaks separated by about 200 years (Boucher et al., 2020).

We also used a set of independent surface temperature fields from climate model simulations generated as part of 5th and 6th phase of the Coupled Model Inter-comparison Project (CMIP5 and CMIP6) to built several LIMs. We specifically relied on surface temperatures from the pre-industrial control (piControl) and past1000 (850–1849 CE) experiments from models participating in PMIP3 (Schmidt et al., 2012) and PMIP4 (Jungclaus et al., 2017; Kageyama et al., 2017). The complete list of models is given on Table 1.

Finally, we used surface temperatures from NOAA-CIRES 20th Century Reanalysis version 3 (NOAA; Slivinski et al., 2019) over the 1851–2014 period to evaluate the climate models biases and their influences on our reconstructions skills. All surface temperature fields used in this study are interpolated on the same 96 × 96 latitude-longitude grid and anomalies are calculated by removing the mean annual cycle over the entire period before smoothing with a 12-month running mean.

2.2. Linear Inverse Model (LIM)

In this section, we briefly describe the main equations and characteristics of the LIMs that we built to reproduce the dynamics of simulated surface temperatures by CMIP5 and CMIP6-class models. The LIM is a multivariate linear Markov process that represents the evolution of a dynamical system x by:

$$\frac{dx(t)}{dt} = Bx + \varepsilon \tag{3}$$

where x is a vector of dimension P and describes the system state at time t, the linear predictable part of the system is represented by the operator B and where all the nonlinear stochastic processes are represented by ε corresponding to a Gaussian white noise with a spatial structure but uncorrelated in time. The solution to Equation 3 is:

$$x(t+\tau) = G(\tau) * x(t) + \sigma(t+\tau) \tag{4}$$

where $G(\tau)$ is the Green matrix equal to:

$$G(\tau) = \exp(B\tau) \tag{5}$$

and $\sigma(t+\tau)$ is the forecast error expressed as:

$$\sigma(t+\tau) = G(\tau) \int_{t}^{t+} G\left(t'\right)^{-1} \varepsilon\left(t'\right) dt' \tag{6}$$

The most probable prediction $x(t+\tau)$ given the initial condition $x(t)$ for a lead of τ months is then:

$$x(t+\tau) = G(\tau) * x(t) \tag{7}$$

and following Equation 5 the linear operator matrix B can be estimated for any lag τ_0 as:

$$B = \tau_0^{-1} \ln[G(\tau_0)] = \tau_0^{-1} \ln\left[\frac{C(\tau_0)}{C(0)}\right] \tag{8}$$

with $C(\tau_0) = < x(t+\tau_0), x(t)^T >$ and $C(0) = < x(t), x(t)^T >$ the covariance matrix at lag τ_0 and 0. The transition probability density of a Markov process satisfies the Fokker-Planck equation and assuming that the statistic of x is stationary, we get from Equation 6 the Fluctuation-Dissipation relationship (Penland & Matrosova, 1994):

$$Q + BC(0) + C(0)B^T = 0 \tag{9}$$

Given B from Equation 8 and the covariance matrix Q from Equation 9, we can then generate a continuous Markov model and perform a sample of stochastic trajectory by integrating Equation 3. The appropriate method of integration is a two-step process corresponding to a second order Runge-Kutta method and described in Penland & Matrosova, 1994.

$$\begin{cases} z_i(t+\Delta t) = z_i(t) + \sum_{j=1}^{H} b_{ij}\, z_j(t)\Delta t + \sum_{k=1}^{H} q_{ik} \sqrt{n_k \Delta t}\, \mathrm{r}_k \\ x_i\left(t + \frac{\Delta t}{2}\right) = \frac{1}{2}(z_i(t) + z_i(t+\Delta t)) \end{cases} \tag{10}$$

where q_{ik} is the kth eigenvector of the matrix Q and n_k its corresponding eigenvalue. The b_{ij} are the coefficient of the matrix B and r is a P-dimensional vector of independent random numbers drawn at each time step from a Gaussian distribution with unit variance. Since we work with monthly data, in the following section the time step Δt is chosen equal to 1/120 month, that is, 6 hr (Penland & Matrosova, 1994). As it is generally the case, our LIMs are built using the main climate modes of variability retained in the first Empirical Orthogonal Functions (EOFs) (Penland, 1989; Penland & Matrosova, 1994). Before the EOFs analysis, each grid cell is normalized by its root-mean-square amplitude and weighted according to its surface area in order to take into account the grid cell area variations.

2.3. Sequential Importance Resampling Particle Filter

The PDA approach adopted in this study is an online sequential importance resampling particle filter (SIR-PF; Doucet et al., 2001; Van Leeuwen, 2009). The SIR-PF is a Bayesian filter using a set of particles $x_k^0, x_k^1 \ldots, x_k^N$ and weights $w_k^0, w_k^1 \ldots, w_k^N$ to recursively estimate at each time step k, the unknown model state probability density $p_x(x_k|y_{1:k})$:

$$p_x(x_k|y_{1:k}) = \sum_{i=1}^{N} w_k^i * \delta_{x_k}\left(x_k^i\right) \tag{11}$$

with x_k^i the particle i at the time step k, $\delta_{x_k}(\cdot)$ a delta function centered at x_k and w_k^i the weights associated to this particle. The SIR-PF is a recursive approach initialized by drawing a set of particles x_0^N according to the common probability density $p(x_0)$ with the w_0^N all fixed equal to $1/N$. Given the probability density function $p_x(x_k|y_{1:k})$, a dynamical model is used to integrate the set of particles from the time step k to $k+1$ following the transitory probability density function $p_x(x_{k+1}|x_k)$. With the Chapman-Kolmogorov formula we get the probability density function:

$$p_x(x_{k+1}|y_{1:k}) = \int p_x(x_{k+1}|x_k) * p_x(x_k|y_{1:k}) dx_k \tag{12}$$

The Bayes theorem is then used to estimate the model state probability density:

$$p_x(x_{k+1}|y_{1:k+1}) = \frac{p_y(y_{k+1}|x_{k+1}) * p_x(x_{k+1}|y_{1:k})}{p_y(y_{k+1})} \tag{13}$$

where $p_x(x_{k+1}|y_{1:k})$ is called the prior distribution, $p_y(y_k)$ corresponds to the observational probability density function and $p_y(y_k|x_k)$ is generally called the likelihood and corresponds to the conditional probability of the observations given the model state. Since the observational probability density function $p_y(y_k) = \int p_y(y_k|x_k) * p_x(x_k) dx_k$ is just a normalization constant, we do not need to estimate it and we can just consider the following relation of proportionality:

$$p_x(x_{k+1}|y_{1:k+1}) \propto p_y(y_{k+1}|x_{k+1}) * p_x(x_{k+1}|y_{1:k}) \tag{14}$$

given the posterior probability density function:

$$p_x(x_{k+1}|y_{1:k+1}) = \sum_{i=1}^{N} w_{k+1}^i * \delta_{x_{k+1}}\left(x_{k+1}^i\right) \tag{15}$$

with:

$$w_{k+1}^i = \frac{p_y\left(y_{k+1}|x_{k+1}^i\right)}{\sum_{j=1}^{N} p_y\left(y_k|x_k^j\right)} \tag{16}$$

Since the likelihood $p_y\left(y_k|x_k^i\right)$ is often assumed to be Gaussian, we get:

$$w_{k+1}^i = \frac{\exp\left(\frac{-1}{2}\left[y_{k+1} - T\left(x_{k+1}^i\right)\right]^T C^{-1}\left[y_{k+1} - T\left(x_{k+1}^i\right)\right]\right)}{\sum_{j=1}^{N} \exp\left(\frac{-1}{2}\left[y_{k+1} - T\left(x_{k+1}^j\right)\right]^T C^{-1}\left[y_{k+1} - T\left(x_{k+1}^j\right)\right]\right)} \tag{17}$$

where C is the error covariance matrix which describes the discrepancy between the observations and the model variables (we dedicate a detailed discussion in Section 3.1 on the way we estimated this matrix C for our experiments), and T is a simple function that maps the model states into the observation phase space by selecting in the model state x_k^i, the locations where the observations y_k are available. Given this new set of w_{k+1}^i the particles are resampled to generate a new set of equally likely particles x_{k+1}^i in order to drop out the particles far from the observations and duplicate (according to their likelihood) the ones close to the observations (Kuptametee & Aunsri, 2022; Liu & Chen, 1998). The weights w_{k+1}^i of these particles are then all fixed equal to $1/N$ and the procedure is repeated over time taking at each iteration the new set of particles and weights as the new starting set.

3. Experimental Design and Skill Metrics

3.1. SIR-LIM Implementation

The implementation of the SIR-LIM, described in Figure 1, is performed in the following way: (a) we first sample randomly, N particles $x_0^0, x_0^1 \ldots, x_0^N$ according to the initial model density probability $p(x_0)$; which in our case is done by pooling randomly N initial states from a long past1000 simulation; (b) each particle is then propagated for 1 year with the LIM (Equation 4) according to the model distribution $p_x(x_{k+1}|x_k)$; (c) available observations

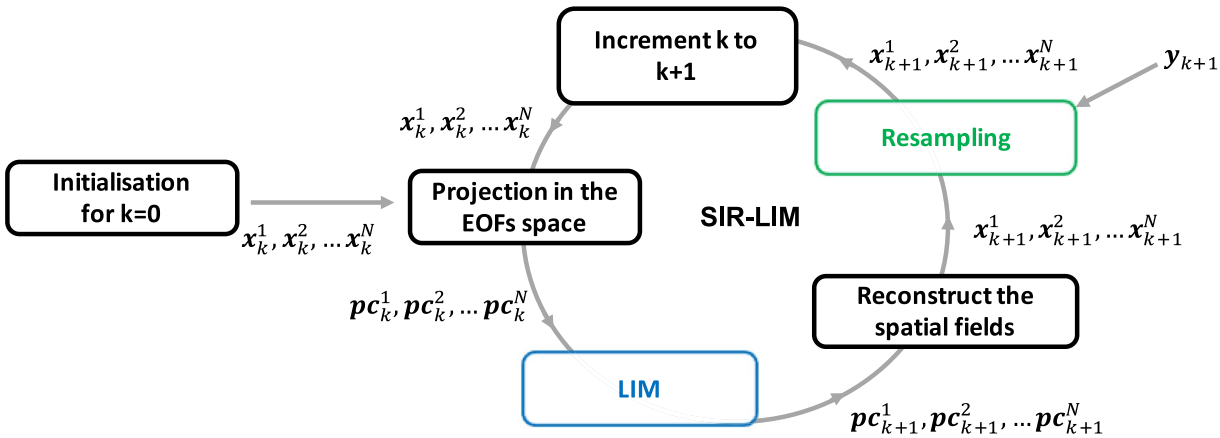

Figure 1. Conceptual representation of the SIR-LIM. The particles integration of an ensemble of N particles by the linear inverse modeling is done in the Empirical Orthogonal Functions space. The notation $\boldsymbol{pc}_k^i$ stands for the vector of PCs considered at the time step k for the particle i. The resampling step is done at the time step $k+1$ according to the ensemble of reconstructed spatial fields $\boldsymbol{x}_{k+1}^i$ and the observations available $\boldsymbol{y}_{k+1}$.

y_{k+1} are then used to estimate w_{k+1}^i following Equation 17. Based on these new weights we resample the set of particles following the residual resampling approach proposed by Liu & Chen, 1998, which ensures a good duplication of the particles close to the observations by first allocating to each particle the integer part of its expected resampling, and then draw randomly the rest of the particles following the distribution given by the decimal part of the expected resampling. Since the LIM includes by construction a stochastic component, no sampling noise needs to be added after the resampling step unlike what is usually done with deterministic models (Dubinkina et al., 2011; Liu et al., 2017). The weights of the particles are then all fixed equal to $1/N$ and the procedure is repeated over time with the new sets of weight and particles.

As the set of particles is propagated in time in the EOFs space, some components of the surface temperature variability are not considered. To take into account this source of uncertainty, we assumed the error matrix C to be, as proposed by Keeler and Ellis (2000) and Dubinkina et al. (2011), the addition of an observational error C_{obs} corresponding to the uncertainty associated to the proxies, plus a representativeness error C_r corresponding to uncertainties associated to the partial representation in the EOFs space of the real un-truncated variability:

$$C = C_r + C_{\text{obs}} \tag{18}$$

As the instrumental errors are supposed to be uncorrelated, the coefficients of the matrix C_{obs} are non-zero on the diagonal and zero elsewhere. For its part, the representativeness error C_r is taken to be equal to the mean square error between the partial representation of the surface temperature in the EOFs space with the real surface temperatures.

3.2. Sensitivity Experiments

The SIR-LIM reconstruction performances are evaluated in a perfect model framework by taking as a target the IPSL-CM6A-LR model surface temperature from the past1000 experiment and by adding a Gaussian white noise to simulate pseudo-proxies. An independent IPSL-CM6A-LR last millennium simulation (not employed for LIM and SIR-LIM calibration) is used to assess the LIM and SIR-LIM scores. In contrast with previous pseudo-proxies experiments (e.g., Dubinkina et al., 2011; Liu et al., 2017) that add a constant noise with a standard deviation of 0.5 regardless the location and variance of the target signal, we consider here a white noise representing 50% of the pseudo-proxy standard deviation in order to avoid giving too much weight to regions with the strongest variability. Several experiments are then setup (see Table 2):

1. To assess the influence of the proxies spatial distribution on the quality of the reconstructions (e.g., Comboul et al., 2015). To do so, we performed two experiments with different spatial sampling (Figure 2). In the first experiment (labeled *Pages2k*) the pseudo-proxies are sampled according to the location of proxies for surface temperatures available at the year 1000 in the multi-proxy PAGES2K-v2 database (PAGES2k Consortium, 2019). This experiment is representative of the real problem with an under-sampling of the

Table 2
Experimental Protocol and Climate Indices

Benchmark experiments	
Pages2k	Experiment relying on pseudo-proxies spatial distribution similar to the actual location of surface temperature proxies available in the PAGES2K-v2 database at the year 1000AD (Figure 2).
Dense	Same as Pages2k but with a regular sampling of 167 pseudo-proxies considered across the globe (Figure 2).
Persistence	Conservative experiment used to evaluate the predictive capabilities of the LIM and consists in predicting for the future the same value as the present.
SIR-0	Particle filter similar to the SIR-LIM but where the LIM is replaced by a Gaussian white noise.
noDA	Case without data assimilation used as a reference score.
Climate indices	
Global	Globally averaged surface temperature
NH	Mean surface temperature between 25°N and 90°N
Tropics	Mean surface temperature between 25°S and 25°N
SH	Mean surface temperature between 25°S and 90°S
ENSO	Defined with the Niño 3.4 index, as the mean sea surface temperature between 5°N–5°S and 120°–170°W
AMV	Mean sea surface temperature between 0°–60°N and 75°–7.5°W
IPO	Defined with the filtered Tripole Pacific Index (describe in Henley et al. (2015)), as the central equatorial Pacific sea surface temperature (10°S–10°N, 170°E−90°W) minus the average of northwest (25°–45°N, 140°E−145°W) and southwest (15°–50°S, 150°E−160°W) Pacific sea surface temperature, smoothed using a 13-yr Chebyshev low-pass filter.

oceans relative to the continents and an under-sampling of the Southern Hemisphere (SH) relative to the Northern Hemisphere (NH). A second experiment (called *Dense*) is also performed with a regular sampling of 167 proxies across the globe in order to assess the limitations in the reconstruction skills induced by the sparse and uneven spatial distribution of *Pages2k* experiment (Figure 2).

2. To evaluate the influence of climate model biases on obtained reconstructions. With that aim, we rely on 4 SIR-LIMs based on LIMs derived from different model physics (see Figure A1): two mono-model LIMs constructed from the CMIP5 and CMIP6 versions of the IPSL model and called thereafter IPSL-CM5-LIM

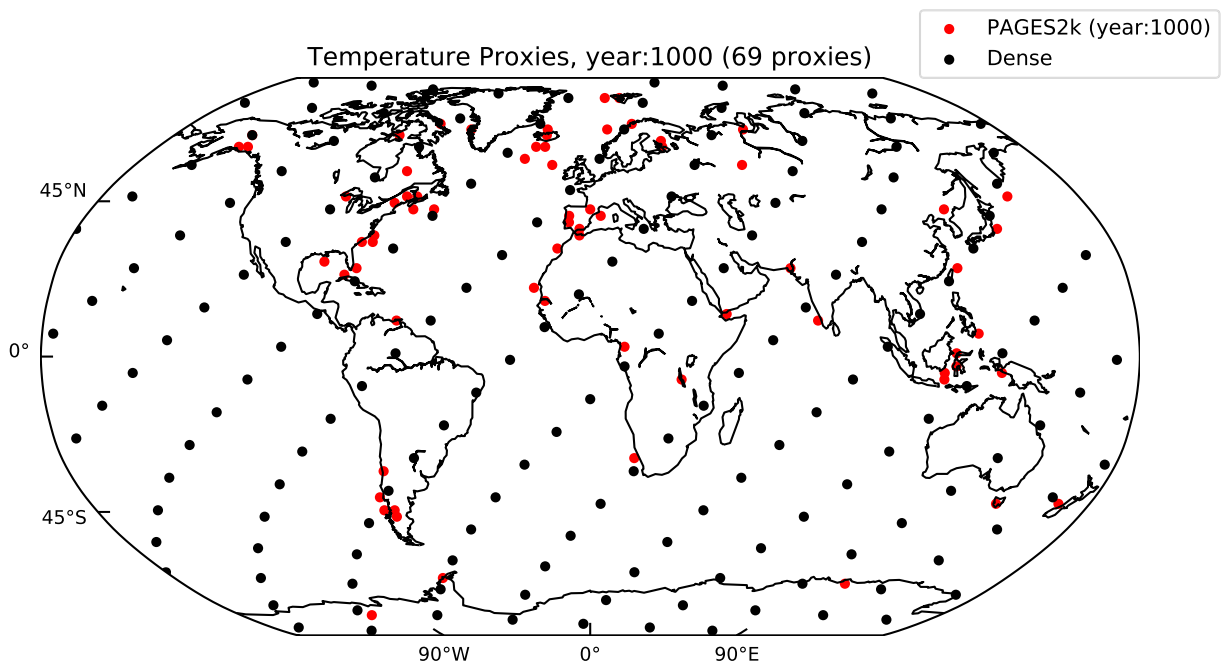

Figure 2. Pseudo-proxies spatial distribution showing the location of the 69 and the 167 surface temperature pseudo-proxies used in the Pages2k (red dots) and Dense experiments (black dots) respectively.

and IPSL-CM6-LIM, and two multi-model LIMs derived from respectively 5 CMIP5 and 4 CMIP6 models and called thereafter MM-CMIP5-LIM and MM-CMIP6-LIM (Table 1).

3. To evaluate the use of the LIM as a dynamical predictive model, we compare the LIM's forecast at various lead-times (considering only the system state at the present time with no additional information) with a very conservative approach called Persistence. This approach simply consists in predicting for the future the same value as the present, giving for any field x at any horizon h and any time step t the following prediction $x_{t+h} = x_t$. The comparison of the predictions of the LIM with Persistence allows evaluating the predictive capabilities of the LIM. The EOFs Persistence is also introduced and corresponds to the Persistence for the components of the surface temperature represented in the EOFs space.
4. To establish a baseline for the internal climate variability and the influences of the resampling on the scores. We therefore introduced two additional experiments called noDA and SIR-0. The noDA experiment represents the case without data assimilation and corresponds to an other past1000 simulation of the IPSL-CM6A-LR model sharing the same external forcing as the target but starting from different ocean initial conditions, and therefore with its own internal variability. The SIR-0 experiment corresponds on the other hand to a SIR-PF where we use a Gaussian white noise in space and time instead of a LIM. It is indeed obvious that in the SIR-LIM, the resampling alone accounts for parts of the scores obtained during the reconstructions regardless of the integration model used. The SIR-0 allows to measure this so-called "re-sampling" score due to the selection of particles close to the observations and which depends on many parameters such as the resampling frequency, the size of the problem and the number of particles.

3.3. Climate Indices and Skill Metrics

In addition to assess the reconstruction skill of the spatial surface temperatures field, we consider 7 main sets of climate indices: the global mean surface temperature, three large scale indices corresponding to the area-averaged surface temperatures in the Tropics, the NH and SH, and three indices for the main modes of climate variability namely the AMV (Ting et al., 2009), the ENSO (McPhaden et al., 2006) and the Interdecadal Pacific Oscillation (IPO, Folland et al., 1999; Power et al., 1999). Note that for ensemble reconstructions, the ensemble-mean index refers to the average of indices previously calculated from outputs across the set of particles. To quantify the uncertainty associated with the reconstructions, the spread of the ensemble is indicated with the 2.5th and 97.5th percentiles. To assess the LIMs and SIR-LIMs skills, we relied on the root-mean-square error (RMSE), the Pearson correlation coefficient (Corr) and coefficient of efficiency (CE). Frequently used, the RMSE gives a measure of the differences between the target and the predicted values. The correlation estimates the similarity between two signals phases while the CE takes into account the phase, amplitude and bias between the two considered signals (Nash & Sutcliffe, 1970). Finally, for ensemble forecasts we also assess the quality of the ensemble spread by plotting the histogram of the target position in the forecast spread.

The detailed and complete list of sensitivity experiments and climate indices described in this section are summarized in Table 2.

4. The LIM as a GCM Emulator

4.1. LIM Skills

The LIM projection in the first EOFs allows reducing the dimensionality of the system and is therefore well suited for PFs that tends to degenerate. The SIR-LIMs performances will however depend on the number of EOFs considered, which has to be large enough to preserve as much variance as possible and describe the evolving field pattern, but also small enough to eliminate much of the unpredictable noise. As shown on Figure 3a, the truncation of the first 30 EOFs explains more than 65% of the variances of surface temperature inter-annual variability for the different models. In terms of local variance explained, we find a pretty similar uneven spatial distribution between the different models, with a higher percentage of variance explained in the tropics than in the extra-tropics and in the Pacific than in the Atlantic (Figures 3b–3e). The first and second EOFs mainly describe the external forcing and ENSO variability and account for more than 30% of the total variance while the variability described by the third EOFs will depend on the model considered (Figure A2). Since the first 30 EOFs explain a large part of the surface temperature inter-annual variability, the LIM and SIR-LIM in the following sections are built considering the first 30 EOFs.

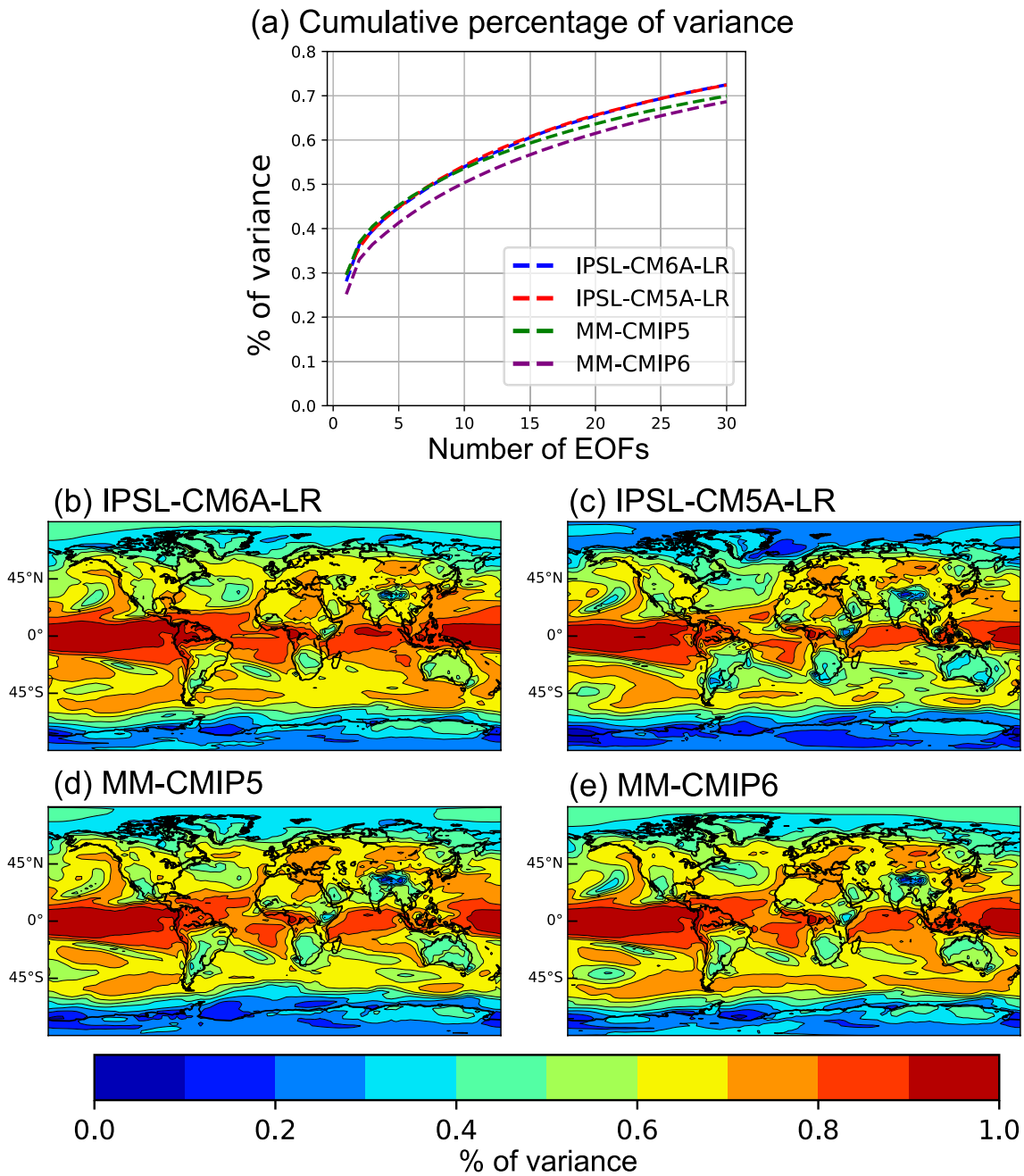

Figure 3. Percentage of variance explained by the first 30 Empirical Orthogonal Functions (EOFs). (a) Cumulative percentage of the total variance explained according to the number of EOFs retained. (b)–(e) Percentage of local variance explained by the first 30 EOFs of the (b) IPSL-CM5A-LR, (c) IPSL-CM6A-LR, (d) MM-CMIP5 and (d) MM-CMIP6 models.

If the dynamics of the surface temperatures can be approximate by Equation 3 then the noise matrix Q, as for any covariance matrix, should be positive definite and its eigenvalues should indicate the amount of variance explained by each "EOFs of the noise" (Penland & Matrosova, 1994). This is the case in our experiment with noise matrix Q positive definite for all the LIMs used in the following sections. Figure 4 shows the different LIMs linear component local skill in predicting the surface temperatures simulated by the IPSL-CM6A-LR model at a lead-time of 1 year. Since the climate models possess a wide range of dynamical interactions and variability patterns (see Figure A1), the forecast obtained by the IPSL-CM6-LIM illustrates the sources of linear predictability specific to the IPSL-CM6A-LR model. On the other hand, the forecast skills of the LIMs derived from other models give useful information about the linear predictable dynamics shared by the GCMs and the confidence

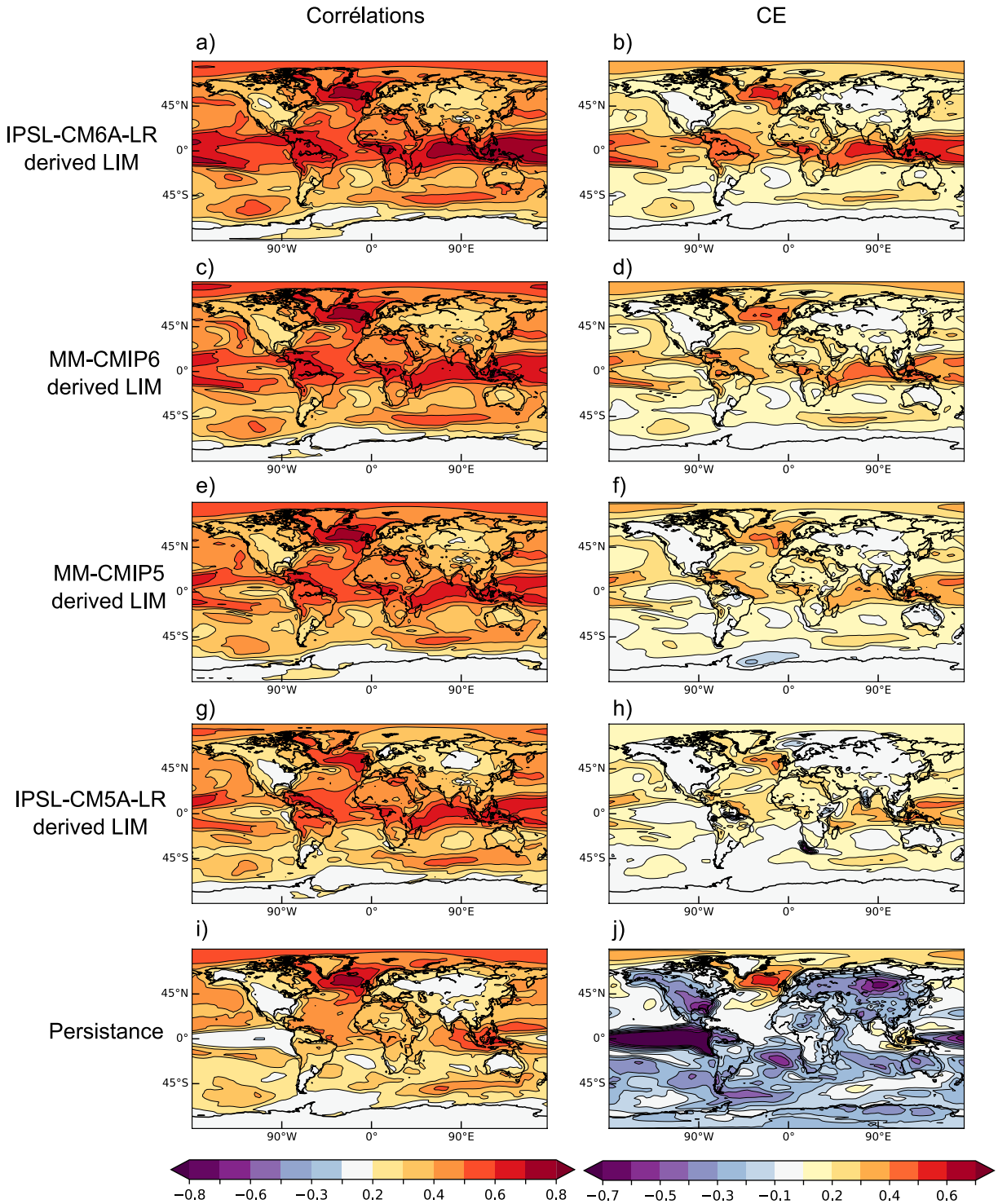

Figure 4. Linear inverse modeling forecast scores at lead 12 months. Correlations (left column) and Coefficient of Efficiency (CE, right column) for (a and b) the IPSL-CM6-LIM, (c and d) MM-CMIP6-LIM, (e and f) MM-CMIP5-LIM, (g and h) IPSL-CM5-LIM and (i and j) Persistence.

we should have in the current CMIP5 and CMIP6-class models as forecast tools. As we could expect, we get the highest scores for the LIM derived from the host model (IPSL-CM6-LIM, Figures 4a and 4b), followed by the MM-CMIP6-LIM (Figures 4c and 4d). All the LIMs show high correlations and CE scores in the North Atlantic and in the tropics specifically in the Indo-Pacific region (Figures 4a–4h). The forecasts are also in general better

over the sea than over continental regions (Perkins & Hakim, 2020). In most regions, all the LIMs outperform the Persistence (Figures 4i and 4j), especially in terms of amplitude (CE). The Persistence shows comparable skill and even sometimes performs better than some LIMs in the extra-tropical North Atlantic which may be due to the strong low frequency variability simulated by the target model in this region. The dynamics in the tropical Pacific being zonally asymmetric, the LIMs get better forecasts in the western tropical Pacific than in the eastern tropical Pacific where the variability is more non-linear (Shin et al., 2020). We also found for the IPSL-CM6-LIM higher forecast skills in the tropical Pacific than in the tropical Atlantic (Figures 4a and 4b) probably related to the higher ENSO predictability (Perkins & Hakim, 2020; Richter et al., 2020). This is however not the case anymore when using the others LIMs, with a larger deterioration of the LIM forecast skill in the Pacific than in the Atlantic (Figures 4c–4h and Figure A3), which suggests that the predictability could be more model dependent in the tropical Pacific than in the tropical Atlantic Ocean.

Such results demonstrate that the CMIP5 and CMIP6 models share some predictable linear dynamics that are reproduced by the LIMs. For ENSO, all the LIMs outperform the Persistence below 3 years with correlation at lead time 2-year greater than 0.4 for the IPSL-CM6-LIM and around 0.2 for the others LIMs (Figure S1 in Supporting Information S1). For the AMV, the Persistence alone shows correlations around 0.7 at 1-year lead and greater than 0.5 even at 4-years lead (Figure S1 in Supporting Information S1). Since the AMV corresponds to a lower frequency variability mode compared to ENSO, it is not surprising to get better forecast skills for the AMV. All in all, these results demonstrate that the LIMs add useful information to predict spatial surface temperatures fields for most of the globe and especially in the tropics including the tropical Atlantic and Pacific oceans. However, in the North Atlantic and for the AMV, a sizable part of the LIMs forecast skill is due to the persistence and ocean dynamical memory effects.

4.2. Ensemble Forecast and Non-Linear Dynamics

It is well known that the predictability of the climate varies over time and will depend on the initial conditions, especially when considering non-linear systems (Borchert et al., 2019; Brune et al., 2018; Planton et al., 2018). The deterministic LIM component is based on linear dynamics and the confidence in these forecasts also depends on the climate initial state due to the presence of some non-linear processes at work (see Figures S2 and S3 in Supporting Information S1). The LIM gives an estimation of the forecasts uncertainties caused by the influence of these non-linear dynamics and also the influence of the external forcings $\boldsymbol{\varepsilon}$. This non-linear part of the LIM allows for ensembles forecasts following the two step stochastic integration described in Equation 10. In Monte-Carlo approaches, like PFs, the ensemble spread is essential to get the range of possible future system states and quantify the uncertainties associated to these predictions. To assess the ability of the LIM to produce appropriate forecasts range, we performed 100-members LIMs ensembles for several global indices at lead 12-month over the 850–1850 period. In Figure 5, we illustrate this with ENSO ensemble forecasts skill over the 1200–1300 period chunk, for neutral ENSO conditions (case 1), strong ENSO transition from a positive to a negative phase (case 2), a neutral to a negative phase (case 3), a negative to a positive phase (case 4) and a negative to a neutral phase (case 5) as simulated by the target model. The IPSL-CM6-LIM shows the best performances (Figure 5b) when the other LIMs tend to underestimate the target variability and struggle to predict the amplitude of large ENSO events (Figures 5c–5e). This is especially true for the strong La Niña visible in the case 3 with a target outside the forecast ensemble spread for the IPSL-CM5-LIM and MM-CMIP5-LIM (Figures 5d and 5e).

The skill of these ensemble forecast is a key factor for the success of PDA and especially when considering PFs. However, since it may change according to the model used, we show on Figure 6, the histograms of the target positions within the ensemble forecasts spread at lead 12-month over the whole 850–1850 CE period, for the global mean surface temperatures, the AMV and ENSO indices for the different LIMs. When using the IPSL-CM6A-LR and MM-CMIP6 derived-LIMs, the histograms display overpopulated middle ranks for global mean temperature (Figures 6a and 6b) indicating an over-dispersive spread, with the target falling more often in the middle deciles of the ensemble spread. The other LIMs also perform very well at lead 12-month with a target more than 95.7% of the time in the ensemble spread (Figures 6c and 6d). All the LIMs tend however to miss some rare extreme temperature anomalies but with asymmetrical probabilities. For example, with the IPSL-CM6-LIM, the target falls 2.3% of the time outside the lower edge and 0.6% of the time above the upper edge of the spread (Figure 6a red bars). These results indicate that the forecasts tend to underestimate more often rare extreme cooling events associated with, for example, unpredictable strong volcanic eruptions.

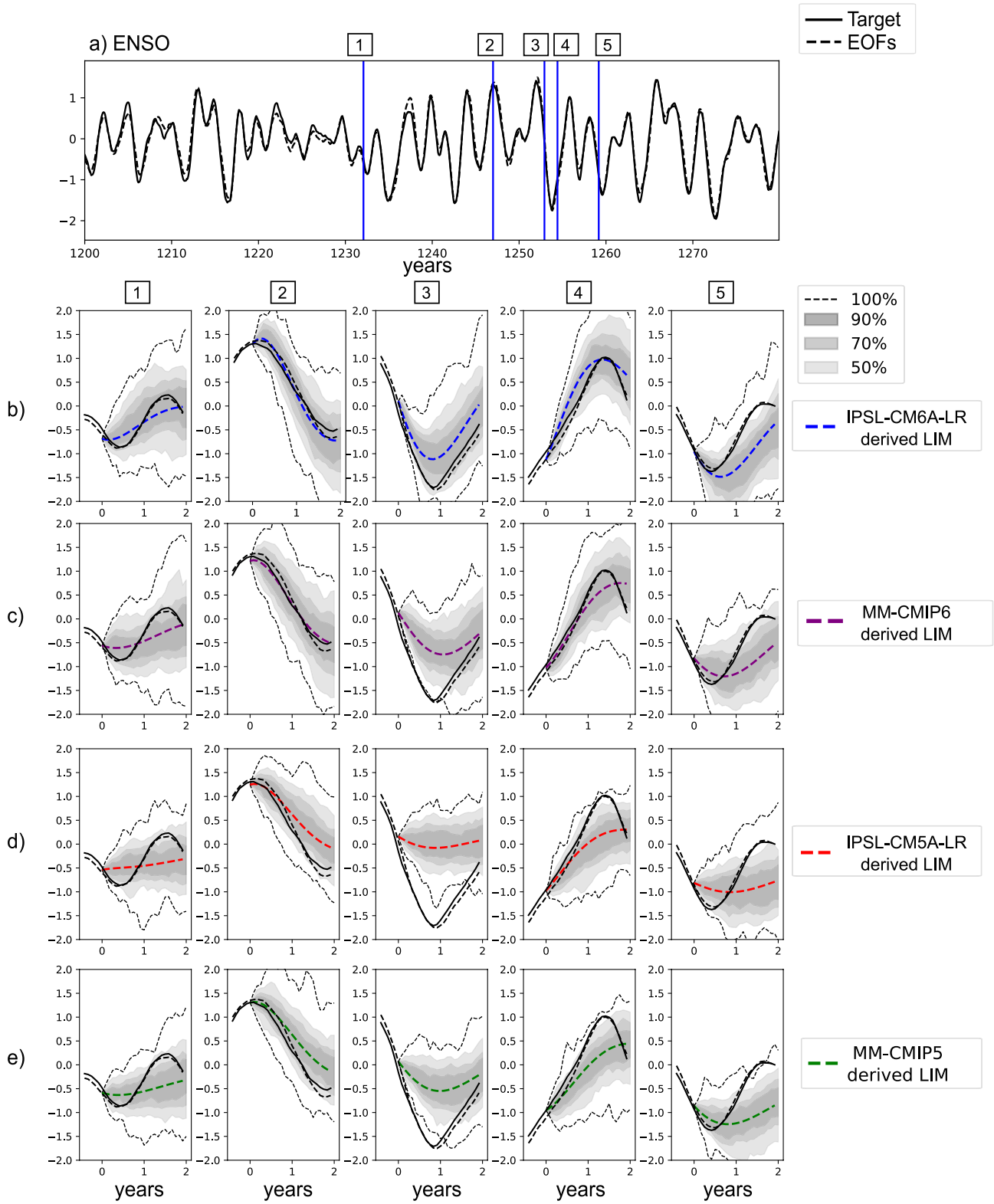

Figure 5. El Niño Southern Oscillation (ENSO) ensemble forecast over the 1200–1300 simulated period by the IPSL-CM6A-LR model. (a) ENSO index of the target simulation (black curve) and its projection in the base of the first 30 Empirical Orthogonal Functions (black dashed curve). The blue vertical lines show the starting date of different forecast experiments noted from 1 to 5. For the 5 starting dates the (b) IPSL-CM6A-LR, (c) MM-CMIP6, (d) IPSL-CM5A-LR and (e) MM-CMIP5 derived LIMs are used to produce forecast ensembles with the integration of 100 particles over a period of 2 years. The ensemble mean is shown by the color dashed lines and the envelope by the fine dotted black lines. The gray shadings indicate the quantiles.

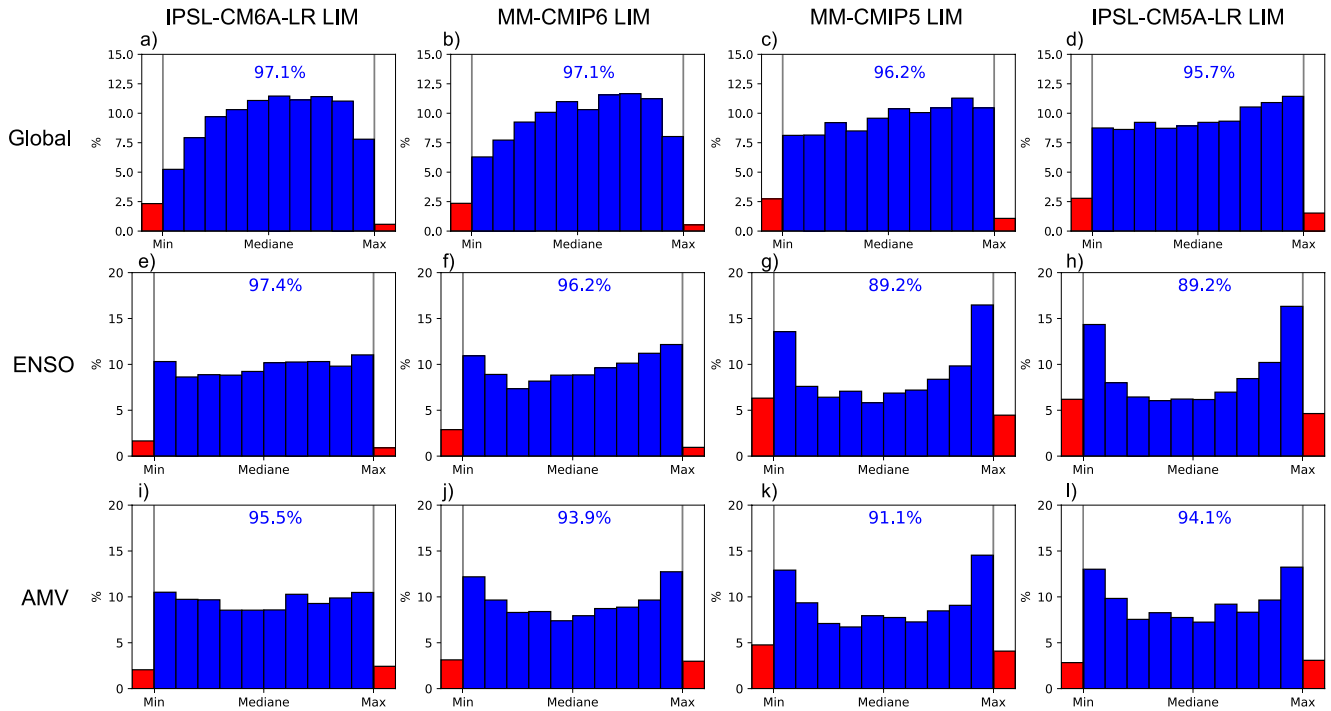

Figure 6. Probability histograms of the target position within 100-members ensemble linear inverse modeling (LIM) forecast spread at 12-month lead. Each blue bar indicates the percentage of the times the target is in the spread by bins of one decile. The red bars show the percentage of time the target is below (for the left red bar) and above (for the right red bar) the ensemble spread edges. The distributions are shown for the (a–d) the global mean surface temperature, (e–h) El Niño Southern Oscillation, and (i–l) Atlantic Multidecadal Variability indices using the (first row) IPSL-CM6-LIM, (second row) MM-CMIP6-LIM, (third row) MM-CMP5-LIM and (fourth row) IPSL-CM5-LIM. The cumulated percentage of times the target is within the spread envelope is indicated at the top of each panel. In the case of a perfectly well calibrated spread we should get a flat blue histogram with a target 98% of the time within the ensemble spread.

For ENSO, we get a well calibrated spread for the IPSL-CM6-LIM and MM-CMIP6-LIM with a flat histogram and a target respectively 97.4% and 96.2% of the time within the ensemble envelop (Figures 6e and 6f). The U-shape histograms visible with the IPSL-CM5-LIM and MM-CMIP5-LIM are on the other hand characteristic of under-dispersive spreads, the target being more often in the border deciles than the middle ones (Figures 6g and 6h). This is not the case anymore when looking at the training error obtained with LIMs forecasting their own GCM-parents (Figures A4d–A4f) indicating that the U-shape distributions are due to dynamical differences across GCMs rather than to LIM errors. Similar results are found for the AMV (Figures 6i–6l), with the IPSL-CM5-LIM and MM-CMIP5-LIM that tend to underestimate the forecast uncertainties when predicting the IPSL-CM6A-LR AMV but describe well the forecast uncertainty associated to their own GCM-parents (Figures A4g–A4i). Such features emphasize the importance of relying on various sets of GCMs to better assess the uncertainties that may be related to model-dependent physics in real PDA derived reconstructions.

All in all, despite some limitations concerning rare unpredictable extreme cold events that may be underestimated, we demonstrate that the LIM produces skillful ensemble forecasts. The LIM is able to capture useful information on the dynamics of the system at leads up to 1 year and these results motivate the use of the LIM as a GCM-emulator to propagate at low cost a large ensemble of particles with a resampling every year in data assimilation analysis such as our PF.

5. Pseudoproxy Climate Reconstruction

5.1. SIR-LIM Scores

We now assess the SIR-LIM performances as part of the *Pages2k* and *Dense* experiments relying on an ensemble of 900 particles over 850–1850 CE. The scores for several reconstructed climate indices with the 4 SIR-LIMs and the two benchmark experiments are listed in Table 3 and local scores for surface temperatures are shown in Figure 7. In both *Dense* and *Pages2k* experiments, the IPSL-CM6-SIR-LIM reconstructs very well the global mean surface temperature

Table 3
Correlation and Coefficient of Efficiency (Within Brackets) Between the Reconstructions Obtained by the Different Experiments Over the 850AD-1850AD and the Target

	Indices					
Experiments	Global	NH	SH	ENSO	IPO	AMV
IPSL-CM6-SIR-LIM Pages2k	**0.89 (0.79)**	**0.89 (0.79)**	**0.52 (0.25)**	**0.61 (0.37)**	**0.61 (0.36)**	**0.87 (0.76)**
IPSL-CM6-SIR-LIM Dense	0.91 (0.82)	0.90 (0.81)	0.75 (0.56)	0.81 (0.66)	0.83 (0.68)	0.80 (0.64)
IPSL-CM5-SIR-LIM Pages2k	0.88 (0.76)	0.89 (0.71)	0.35 (0.08)	0.56 (0.31)	0.60 (0.35)	0.86 (0.74)
MM-CMIP5-SIR-LIM Pages2k	0.89 (0.79)	0.90 (0.81)	0.50(−0.04)	0.57 (0.33)	0.67 (0.45)	0.86 (0.74)
MM-CMIP6-SIR-LIM Pages2k	0.89 (0.79)	0.90 (0.80)	0.48 (0.21)	0.57 (0.32)	0.60 (0.31)	0.87 (0.74)
SIR-0 Pages2k	0.13 (−0.17)	0.12 (−0.19)	0.07 (−0.18)	0.06 (−0.21)	0.20 (0.04)	0.12 (−0.17)
noDA	0.57 (0.1)	0.49 (−0.05)	0.36 (−0.22)	0.03 (−0.97)	0.04 (−0.66)	0.31 (−0.37)

Note: The bold values correspond to the scores obtained in the Pages2k experiment with the SIR-LIM derived from the target model.

variability at inter-annual to decadal timescale. It tends however to underestimate large cooling anomalies caused by super eruptions such as Samalas (1257CE) or the Kuwae (1452–1453 CE) as illustrated on Figure 7a. Even though the LIMs are not directly driven by external forcing, such unpredictable events are still in some way captured by the ϵ parametrization which underestimates the amplitude only for rare extreme cases (see Section 4.2).

When looking at spatial scale, the IPSL-CM6-SIR-LIM provides positive scores in *Pages2k* for both the phase and amplitude in most of the tropical regions and in the NH (Figures 7b and 7c). We get the highest scores in the sub-polar regions bordering the Atlantic and Arctic Oceans while the lowest scores are obtained in the SH due to the scarce availability of pseudo-proxy in this hemisphere. Considering the ~~quite~~ limited observational sampling in the Pacific Ocean, we found however quite satisfactory scores for ENSO and IPO reconstructions (Table 3), illustrating the role of dynamical teleconnections with pseudo-proxies available in remote regions. With a uniform and denser pseudo-proxy distribution, we get globally better scores over most of the globe and especially in the tropical Pacific, Indian Ocean and in the SH (Figures 7d and 7e). We can however notice higher scores in some regions for *Pages2k* such as in Southern Alaska and in the North Atlantic near Iceland where we have a higher density of pseudo-proxies as compared to *Dense* (Figure 2 and Figures 7b–7e). The noDA baseline experiment (which corresponds to another IPSL-CM6A-LR model past1000 simulation) shows correlation larger than 0.2 near the maritime continent and Indian Ocean, illustrating a relatively larger contribution of external forcings in this area (Figures 7f and 7g). We also get pretty high correlation for the global and hemispheric indices but very bad CE scores for all the indices and most of the regions (Table 3; Figure 7g). In SIR-0 PF experiment, which uses an integration model that does not propagate climate information over time, the scores are very low indicating the relatively low importance of the "resampling score" alone in the SIR-LIMs performances.

For the other SIR-LIMs, we found relatively similar results in the *Pages2k* experiment with again the best scores in the NH and for the AMV (Figures 8a–8f and Table 3). We get globally better performances for the multi-model SIR-LIMs than the IPSL-CM5-SIR-LIM. The difference of skill between the SIR-LIM derived from the IPSL-CM6A-LR model and the other ones illustrates the impact that GCMs biases may induce (ΔCE in Figures 8g–8i) and show the confidence we should have in PDA reconstructions based on CMIP5 and CMIP6-class models. For example, in Figure 8h the differences visible in the Western Pacific between the IPSL-CM6-SIR-LIM and the MM-CMIP6-SIR-LIM near the maritime continent is related to different ENSO teleconnection spatial fingerprints among models with anomalies that extend too far into the Western Pacific along the equator in the IPSL-CM6A-LR model as compared to the MM-CMIP6 models (Figure A1). We could therefore expect better ENSO reconstructions with the MM-CMIP6-SIR-LIM than the IPSL-CM6-SIR-LIM in the case of a real PDA since the MM-CMIP6-LIM exhibits weaker biases in the tropical Pacific with a more faithful ENSO spatial pattern with the observations (Figure A1). Another example is the negative ΔCE visible in the Southeast Tropical Atlantic region and the Angola-Benguela front (Figures 8g–8i) that may be due to significant bias and discrepancies in the current CMIP5 and CMIP6-class model in these regions (Richter & Tokinaga, 2020; Xu et al., 2014).

Figure 9 shows that the SIR-LIM can also capture well the frequency content with peaks between 2 and 4-years for ENSO, 8 to 32-years for IPO and 120 to 200-years for AMV. It may however underestimate the frequency

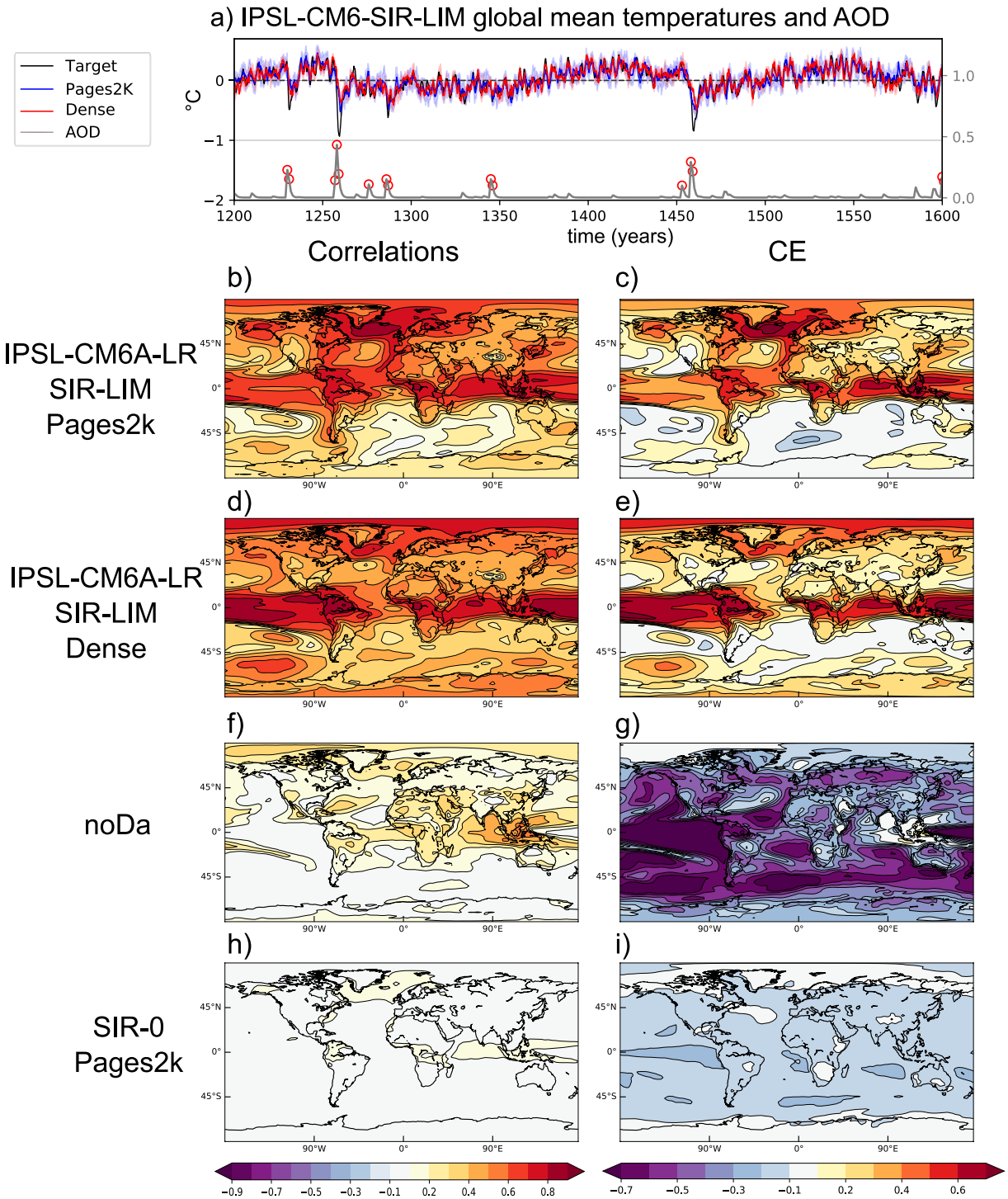

Figure 7. Reconstructions skill for the different experiments. (a) IPSL-CM6-SIR-LIM global mean reconstructed surface temperature over 1200–1600 in the Pages2k (in blue) and Dense (in red) experiments. The target is shown by the black curve and the blue and red shading indicate the 2.5th and 97.5th percentiles of the Pages2k and Dense reconstructions ensemble spreads respectively. The global mean Aerosol Optical Depth (AOD, in gray) is also shown to indicate the timing of volcanic eruption with red circles when the global mean AOD is higher than 0.1. (b)–(i) Correlations (left column) and CE (right column) local scores for the period 850–1850 for the IPSL-CM6-SIR-LIM in (b and c) the Pages2k and (d and e) Dense experiments. The same scores are displayed for (f and g) the noDA and (h and i) SIR-0 benchmark experiments.

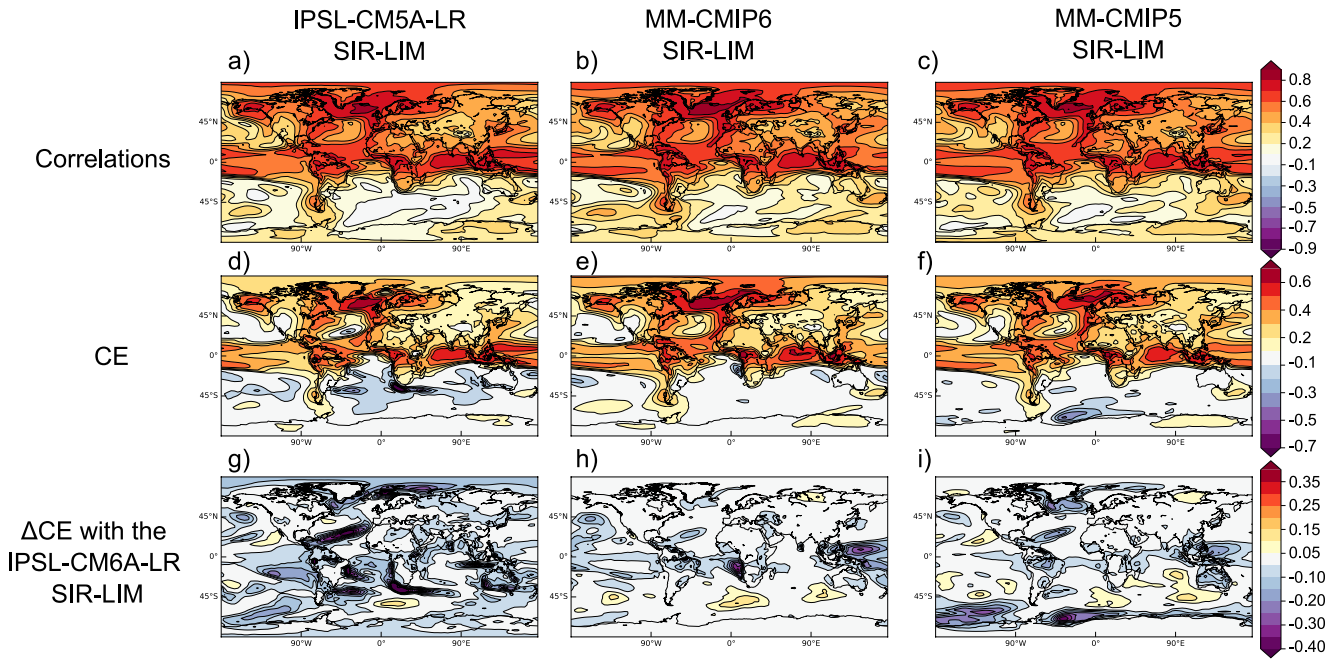

Figure 8. Spatial reconstructions skill for the different SIR-LIMs in the *Pages2k* experiment. (a–c) Correlations and (d–f) coefficient of efficiency (CE) for the IPSL-CM5-SIR-LIM (left column), MM-CMIP6-SIR-LIM (middle column) and MM-CMIP5-SIR-LIM (right column). (g–i) Difference between the CE obtain by each SIR-LIM and the IPSL-CM6-SIR-LIM.

content in the real case especially for ENSO and IPO with an underestimated amplitude in all cases, including the experiments with IPSL-CM6-SIR-LIM (Figures 9a, 9b, 9d, and 9e). However, this problem can be overcome by providing a stronger constrain with more available pseudo-proxy in ENSO and IPO-sensitive regions as shown by the smaller underestimated amplitude in ENSO and IPO frequency peaks in *Dense* as compared to *Pages2k* experiments (Figures 2, 9a, 9b, 9d, and 9e). The larger number of pseudo-proxies in the regions surrounding the North Atlantic Ocean especially in the *Pages2k* experiment (Figure 2) explains also the better representation of the characteristic 200-years peak specific to the IPSL-CM6-LR model (Boucher et al., 2020) in all SIR-LIMs reconstructed AMV (Figures 9c and 9f). It is worth pointing out that even if none of the CMIP5 and CMIP6 GCMs exhibit the same bicentennial low frequency variability in the Atlantic as the target model, all the SIR-LIMs reproduce this low frequency variability. These results demonstrate that when a model simulates pretty well the spatial fingerprint and amplitude of the climate variability, a regular resampling of the particles allows reproducing a range of frequencies present in the observations but not freely simulated by the model. This feature is particularly interesting when considering past climate reconstructions guided by real proxy records with GCMs that might not spontaneously reproduce the observed ranges of the natural climate variability.

5.2. Sensitivity to the Number of Particles and EOFs Truncation

To complete our analysis, we also evaluated the SIR-LIM sensitivity to the number of EOFs and particles used. Figure 10 shows the sensitivity of the IPSL-CM6-SIR-LIM to these parameters for the *Pages2k* experiment. As we could expect we see that for a small number of particles (<100), it is preferable to drastically reduce the dimension of the problem by taking in consideration a few EOFs (<10, Figure 10a). When considering 30 EOFs and 900 particles we get among the best results with no improvements when increasing the number of particles given the uncertainties associated to the instrumental error, lack of proxies and LIM approximations (Figure 10a). By increasing the number of EOFs we reduce the representativeness error especially in the NH but increase the dimensionality of the system with a larger number of particles required to avoid the PFs degeneracy. Since the first EOFs explain a large part of the variability in the tropics we find good results in the tropics with the small number of 5 EOFs (Figure 10b) while in the NH hemisphere a number of at least 10 EOFs is required to get skillful reconstructions (Figure 10c). Our results show that using more than 30 EOFs does not improve the reconstruction skills since the SIR-LIMs reconstruct mainly large-scale dynamics and that a set of 900 particles is sufficient to prevent the filter degeneracy (Figure 10 and Figure S4 in Supporting Information S1).

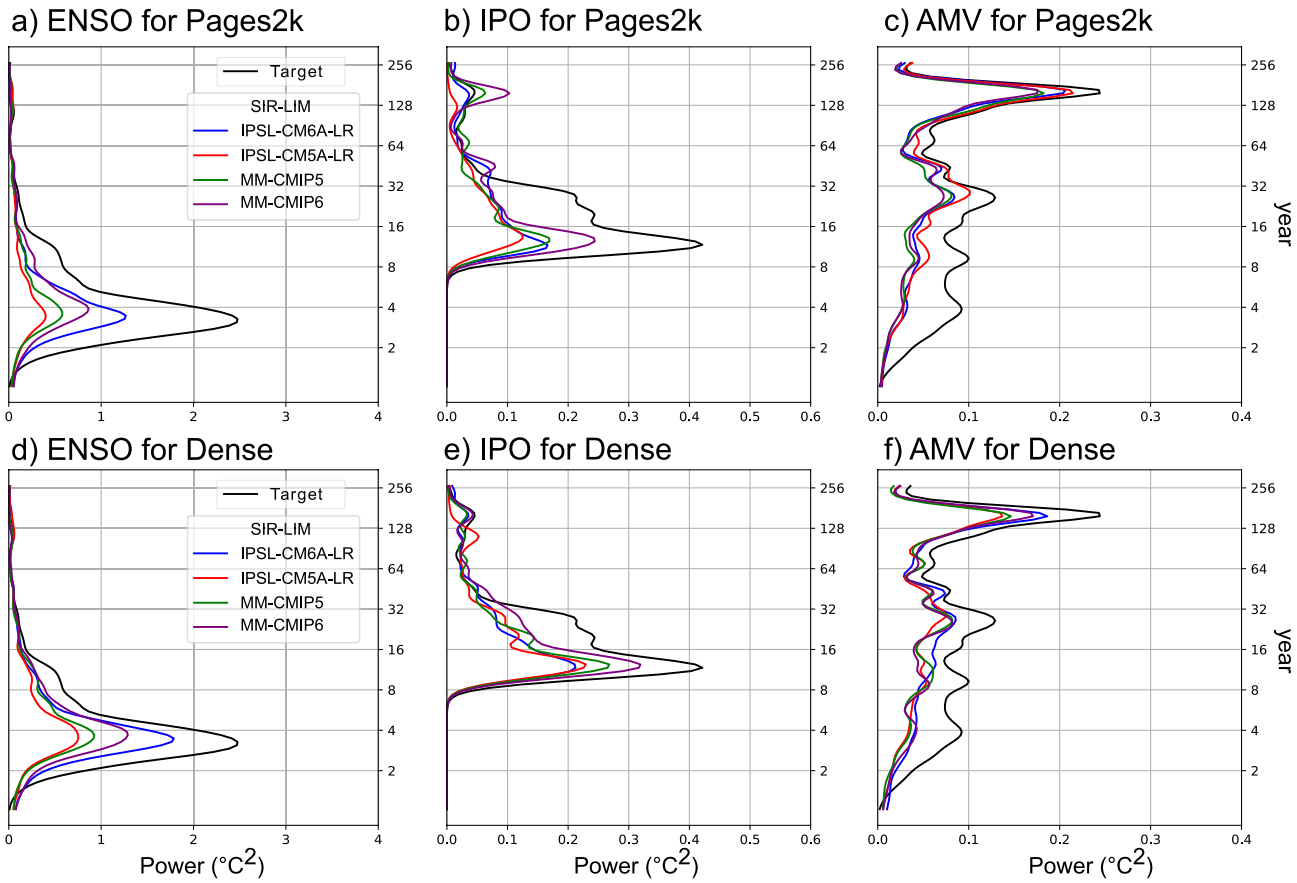

Figure 9. Power spectrum for (a) El Niño Southern Oscillation, (b) Interdecadal Pacific Oscillation and (c) Atlantic Multidecadal Variability indices in the IPSL-CM6A-LR target simulation in black and in reconstructions derived from the Pages2k experiments with the IPSL-CM6-SIR-LIM (blue curve), IPSL-CM5-SIR-LIM (red curve), MM-CMIP5-SIR-LIM (green curve) and MM-CMIP6-SIR-LIM (purple curve). (d)–(f) Same as (a)–(c) but deduced from the Dense experiments.

6. Conclusions

The PDA field offers a way to extend back in time the relatively short period of instrumental observations by assimilating available proxies in state-of-the-art climate models. However, the use of efficient online ensemble-based data assimilation techniques with costly models to reconstruct spatial fields over long period stays until today prohibitively expensive and difficult. Because of the high-dimensional system state space, PDA techniques such

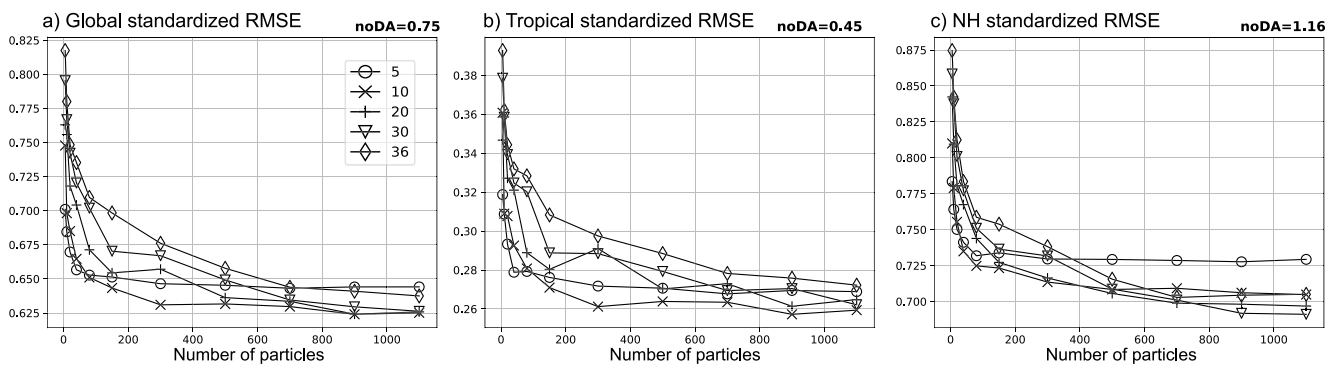

Figure 10. SIR-LIM sensitivity to the number of Empirical Orthogonal Functions (EOFs) and particles considered in the Pages2k experiment. The scores are for the SIR-LIM derived from the IPSL-CM6-LR model for a period of 100 years. The curves correspond to the root-mean-square error (RMSE) divided by the RMSE of noDA for (a) the Global, (b) Tropical and (c) Northern Hemisphere RMSE. Each curve corresponds to a fixed number of EOFs and is expressed as a function of the number of particles used. The RMSE of noDa is shown in the top right corner and allows to retrieve the original RMSE value by multiplying the standardized RMSE by the RMSE of noDA.

as KFs and their ensemble versions are nowadays preferred since PFs encounter problems of degeneracy when the number of particles is limited (Liu et al., 2017). Here we promote the use of a low cost online PF able to give skillful spatial reconstructions by relying on LIMs derived from different GCMs-physics allowing to overcome the "curse of dimensionality" problem.

To assess the influence of the proxies spatial distribution and get an estimation of the influence of the dynamical model used on obtained reconstructions, we conduct different experiments with different spatial pseudo-proxies sampling and build 4 LIMs derived from different models physics. We found that the dynamics associated to the surface temperatures of the first 30 EOFs is substantively linear, and is relatively well approximated by a LIM. The LIM is able to capture predictable information from its GCM-parents and gives skillful forecasts at lead 1-year for different large-scale indices and spatial fields. For most of the globe, the LIM shows good regional forecasts with higher skills in the tropics, North Atlantic and Pacific Ocean and highlight the common source of linearly predictable dynamic shared by different GCMs. The LIMs give also a representation of the uncertainties associated to their forecast, the non-linear dynamics being approximated by a stochastic process which provide useful information on the system nonlinearities needed to generate faithful large ensemble of possible future climates at 1-year lead time. In all cases, the SIR-LIMs clearly outperform the benchmark experiments only driven by external forcing (noDA experiment), even though the amplitude of extreme short-lived cooling events caused by strong volcanic eruptions are underestimated. However, the use of a LIM as a dynamical model still limits the full potential of PFs and future work with more sophisticated models with non-Gaussian and even additive noise (Bianucci & Mannella, 2021; Martinez-Villalobos et al., 2019) and/or with other non-linear forecast models (Nadiga, 2021) could help improve the reconstructions of extreme anomalies.

Our pseudo-proxies experiments provide a convincing validation of the SIR-LIM approach as a PDA technique. For spatial reconstructions, we found better skill in regions with high pseudo-proxy sampling, although the SIR-LIMs exploit spatial teleconnection to reconstruct areas with quite limited observational constrains. When assimilating a more regular and denser pseudo-proxies set as compared to the real spatial sampling at the beginning of the millennium, we get globally better reconstructions in most of the globe and especially in the tropical Pacific, Indian Ocean and in the SH indicating that more proxies in these regions would improve the reconstructions quality. The scores show overall comparable skills for the global mean surface temperature across the various GCMs derived-LIMs, but with important differences at regional scale. Our results are less model-dependent for climate modes and regions that are well constrained by observations availability in sensitive areas such as the AMV and in the NH as compared to ENSO, IPO and in the SH especially in term of amplitude (CE).

These results increase the confidence we may have in the AMV and NH real reconstructions that could be derived with our PDA approach from the Pages2k database, while the SH, IPO, and ENSO reconstructions may be considered cautiously. We found high correlations and CE scores for the global mean surface temperature in the *Pages2k* experiments despite the pseudo-proxies scarceness. These results increase the confidence we may have in the current reconstructions for the global mean surface temperature variability at decadal timescale and suggest, in agreement with (Parsons & Hakim, 2019), that the variability of the global mean temperature at decadal to inter decadal time scale is mainly driven by the North Hemisphere low frequency variability. Thanks to the particles regular resampling, the SIR-LIMs also allow reproducing a range of frequencies present in the "observations" but not necessarily simulated spontaneously by the GCMs considered. These results are encouraging and show that as long as a model simulates a fairly similar variability in terms of spatial fingerprint and amplitude, the SIR-LIM can reconstruct a range of low frequency variability not produced by the host climate model while still ensuring dynamical consistency.

The quality of the reconstructions achieved here are nevertheless model dependent mostly because of CMIP-class model classical biases such as the cold tongue bias in the equatorial Pacific or the imperfect spatial fingerprints of global climate modes. The translation of these perfect model skills for real reconstructions have to be discussed cautiously considering the biases of the considered model used to build the LIM and the spatial variability of available proxies throughout the last 2000 years. For example, the skill observed in the North Atlantic and for the AMV may be model dependent and due to the strong secular variability present in the IPSL-CM6A-LR model. The question about the real existence of this centennial variability in the Atlantic is all the more important that in our pseudo-proxies experiment, it corresponds to a significant part of the SIR-LIMs scores in the Atlantic. Unfortunately, the instrumental period appears to be too short to easily dissociate this range of low-frequency variability from the external forcing trends during the instrumental period but stay potentially real since the

proxies analysis tends to suggest the existence of this kind of Atlantic centennial variability in the past (Gray et al., 2004; Thirumalai et al., 2018). Finally, our results also suggest that the SIR-LIM PDA can be considered a tool to allow the production of low-cost reconstructions following dynamical constraints deduce from different GCMs that will help shed some light on the discrepancies found among currently available reconstructions for the last 2000 years (PAGES2k Consortium, 2019). Such differences could be attributed to the network of proxy assimilated, the statistical methodology employed and, when it comes to PDA, the dynamical model considered (e.g., PAGES2k Consortium, 2019; Perkins & Hakim, 2021; Wang et al., 2015).

Appendix A

See Figures A1, A2, A3, and A4.

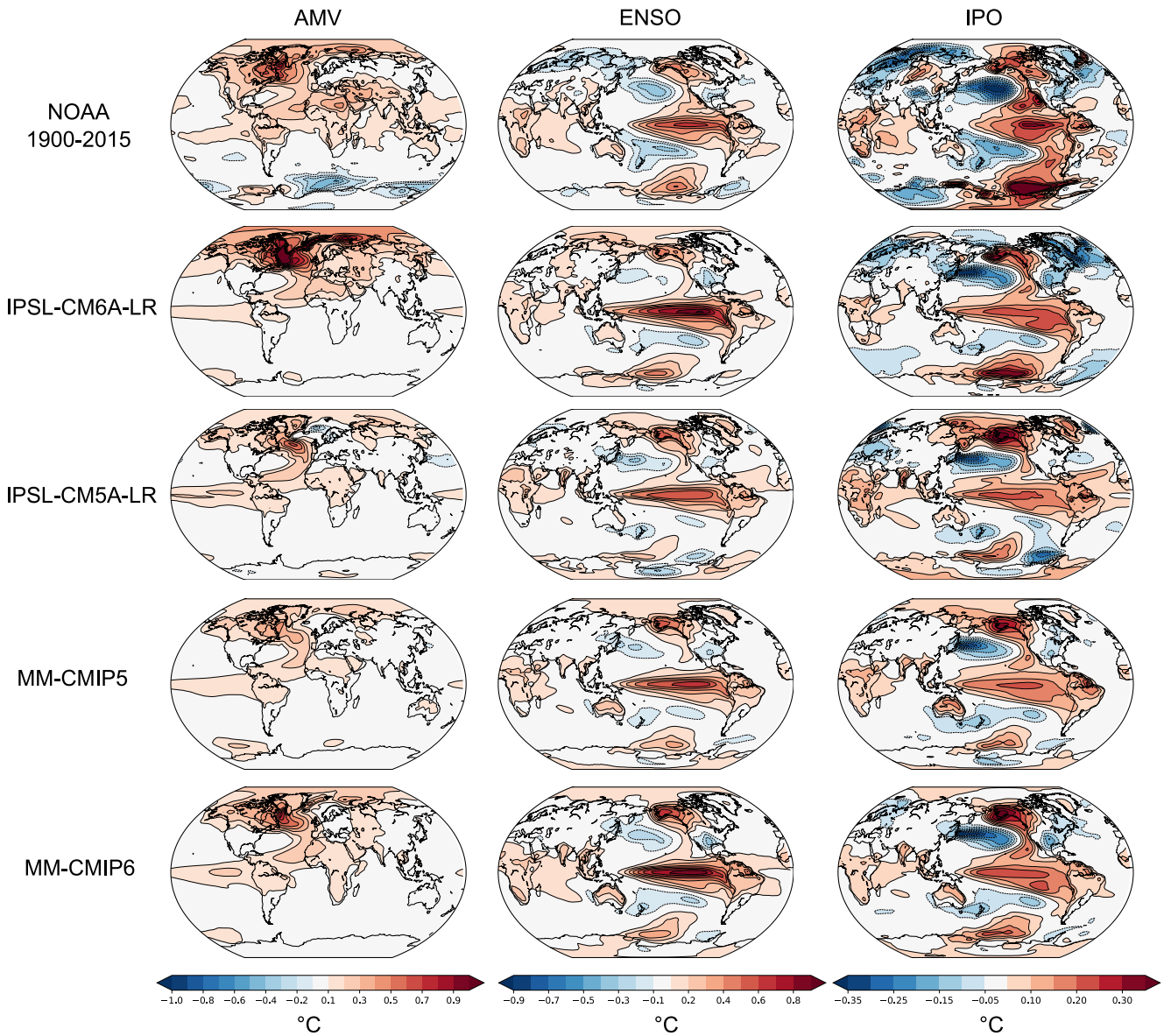

Figure A1. Spatial pattern of the main modes of surface temperatures variability. Regression of the standardized AMV (left column), ENSO (middle column) and IPO (right column) indices on the surface temperature anomalies. These patterns (in °C/std) have been estimated from NOAA-CIRES 20th Century Reanalysis version 3 (NOAA; Slivinski et al., 2019) surface temperature over the 1900–2015 period (upper panels) and on 500 years piControl simulations for the IPSL-CM6A-LR, IPSL-CM5A-LR, CMIP5, and CMIP6 models respectively. The CMIP-class models and methods used to compute the AMV, ENSO, and IPO indices are summarized in Tables 1 and 2.

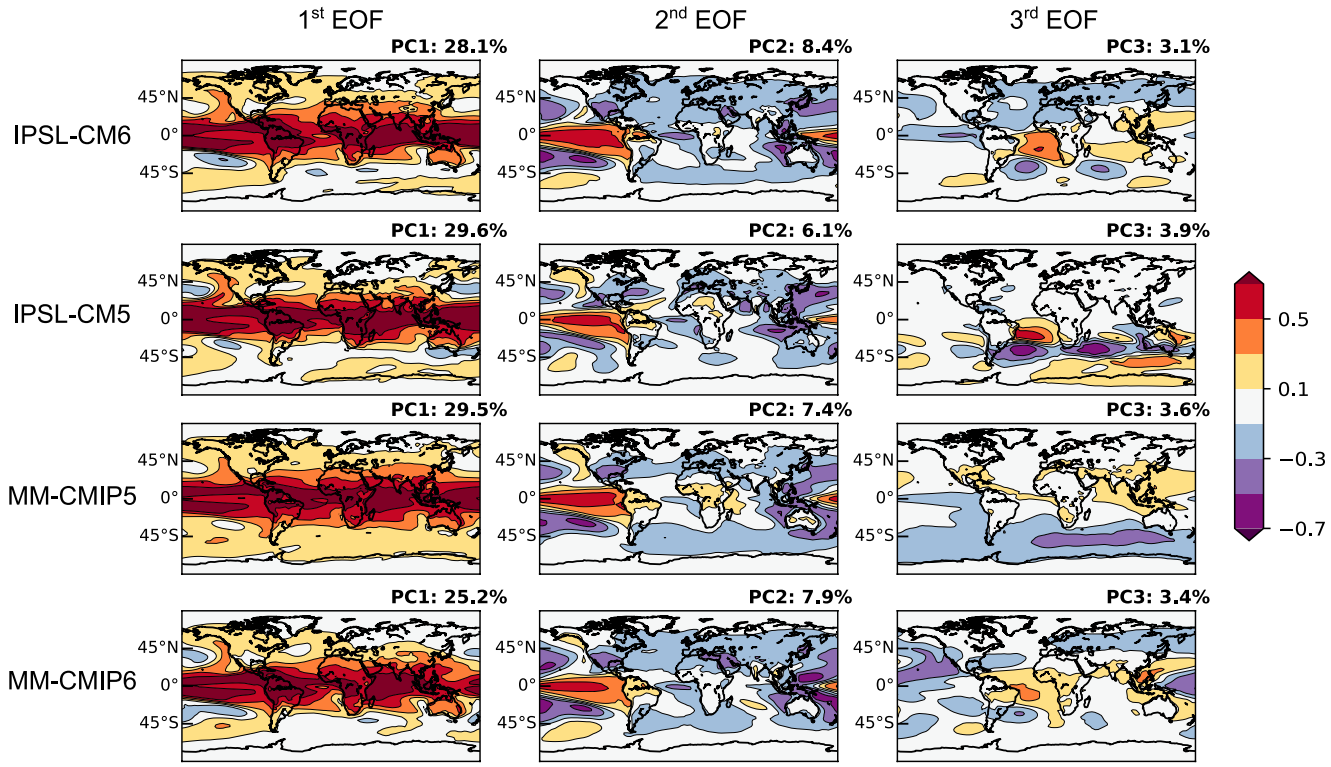

Figure A2. Percentage of variance explained by the first three leading Empirical orthogonal functions (EOFs) for the different set of model simulations. EOFs are multiplied by the square-root of their eigenvalues and the total fraction of the variance explained is given on top of each panel.

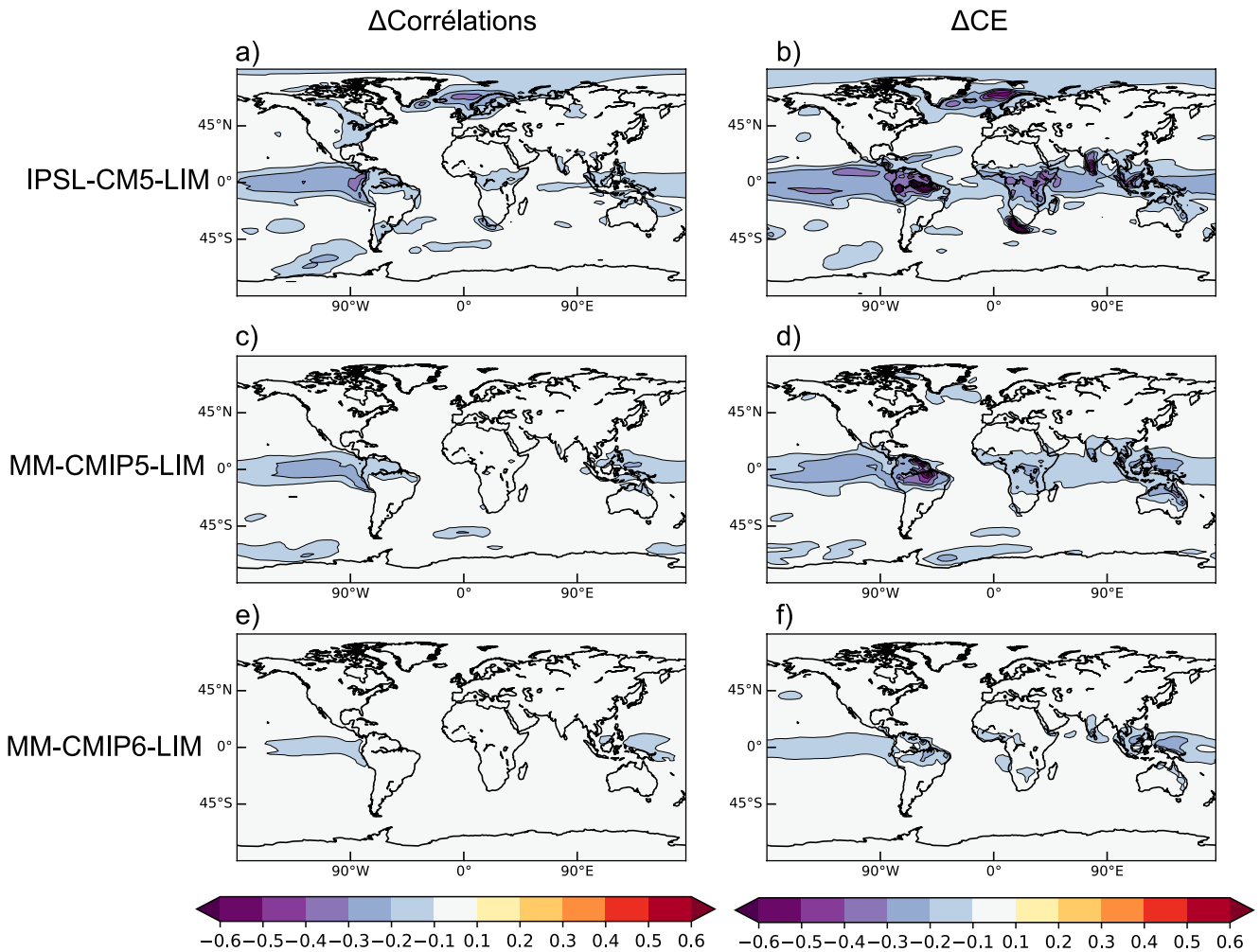

Figure A3. Difference of forecast skill between the IPSL-CM6-LIM and the scores obtained with the others derived LIMs. The difference of correlation (left column) and coefficient of efficiency (CE, right column) at lead-time 12 months are given for the (a and b) IPSL-CM5-LIM, (c and d) MM-CMIP5-LIM and (e and f) MM-CMIP6-LIM.

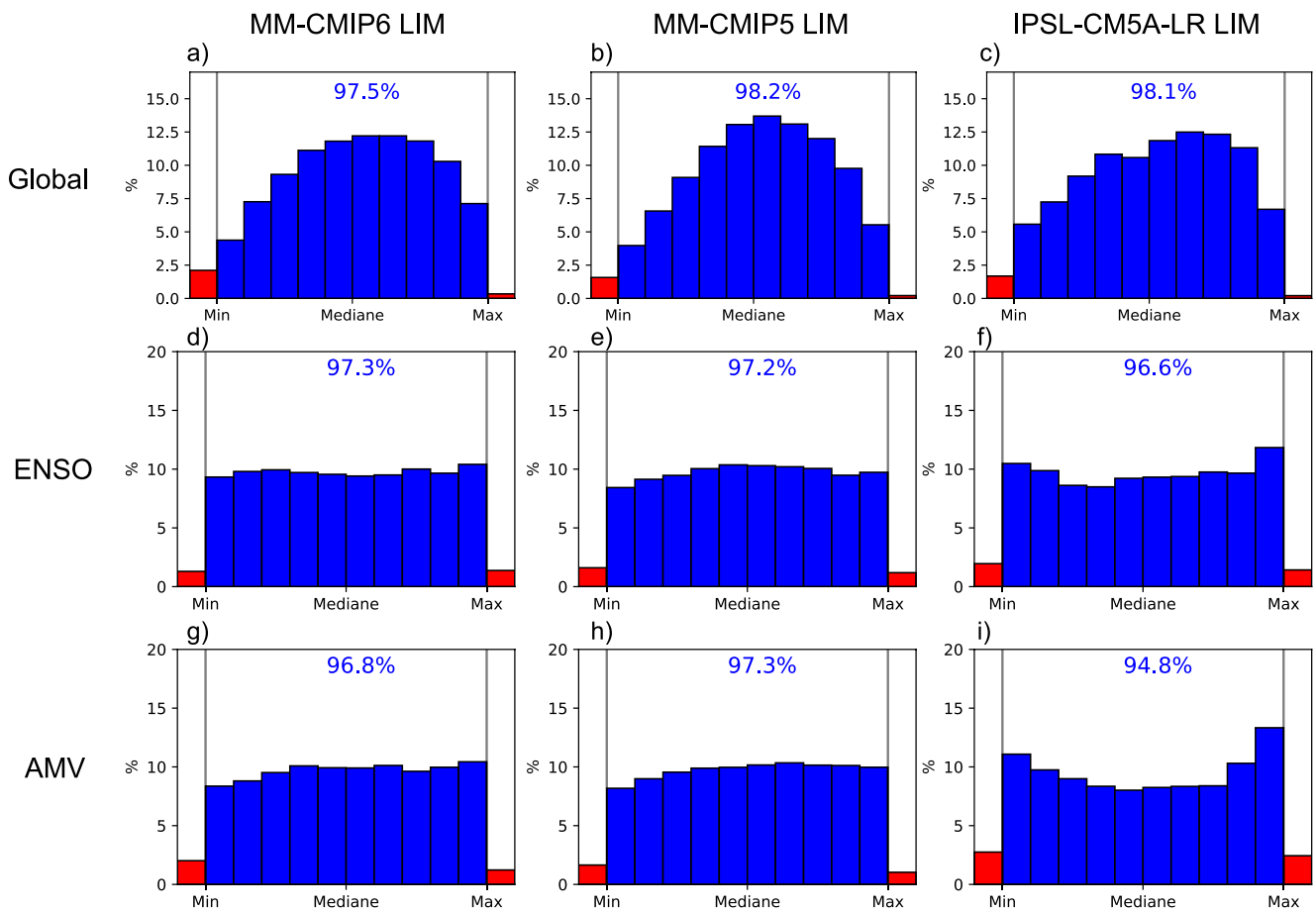

Figure A4. Same as Figure 6 but for the training error obtained with LIMs predicting their own GCMs parents.

Acknowledgments
Resources supporting this work were provided by the ESPRI (Ensemble de Services Pour la Recherche l'IPSL) computing and data centre (https://mesocentre.ipsl.fr), which is supported by CNRS, Sorbonne Université, École Polytechnique, and CNES as well as through national and international grants. This work was undertaken in the framework of the French L-IPSL LABEX and the IPSL Climate Graduate School EUR and benefited from the FNS "SYNERGIA EffeCts of lArge voLcanic eruptions on climate and societies: UnDerstanding impacts of past Events and related subsidence cRises to evaluate potential risks in the future" (CALDERA) project under SNSF Grant agreement number CRSII5_183571—CALDERA. We acknowledge the World Climate Research Programme's Working Group on Coupled Modeling, which is responsible for PMIP, and we thank the climate modeling groups (listed in Table 1) for producing and making available their model output and the Pages2k Consortium for providing the PAGES2k database. Support for the 20th Century Reanalysis Project version 3 data set is provided by the U.S. Department of Energy, Office of Science Biological and

Data Availability Statement

The functions used to produce the analyses as well as the PMIP3 and PMIP4 data, pre-processed data, pseudo-proxies used and the reconstructions obtained with the different LIMs used are available at: https://doi.org/10.5281/zenodo.6364619.

References

Acevedo, W., Reich, S., & Cubasch, U. (2016). Towards the assimilation of tree-ring-width records using ensemble Kalman filtering techniques. *Climate Dynamics*, *46*(5–6), 1909–1920. https://doi.org/10.1007/s00382-015-2683-1

Aumont, O., Éthé, C., Tagliabue, A., Bopp, L., & Gehlen, M. (2015). PISCES-v2: An ocean biogeochemical model for carbon and ecosystem studies. *Geoscientific Model Development*, *8*(8), 2465–2513. https://doi.org/10.5194/gmd-8-2465-2015

Bhend, J., Franke, J., Folini, D., Wild, M., & Brönnimann, S. (2012). An ensemble-based approach to climate reconstructions. *Climate of the Past*, *8*(3), 963–976. https://doi.org/10.5194/cp-8-963-2012

Bianucci, M., & Mannella, R. (2021). On the determination of the optimal parameters in the CAM model. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *31*(3), 033113. https://doi.org/10.1063/5.0032267

Borchert, L. F., Düsterhus, A., Brune, S., Müller, W. A., & Baehr, J. (2019). Forecast-oriented assessment of decadal hindcast skill for North Atlantic SST. *Geophysical Research Letters*, *46*(20), 11444–11454. https://doi.org/10.1029/2019gl084758

Boucher, O., Servonnat, J., Albright, A. L., Aumont, O., Balkanski, Y., Bastrikov, V., et al. (2020). Presentation and evaluation of the IPSL-CM6A-LR climate model. *Journal of Advances in Modeling Earth Systems*, *12*(7), e2019MS002010. https://doi.org/10.1029/2019MS002010

Brune, S., Düsterhus, A., Pohlmann, H., Müller, W. A., & Baehr, J. (2018). Time dependency of the prediction skill for the North Atlantic subpolar gyre in initialized decadal hindcasts. *Climate Dynamics*, *51*(5), 1947–1970. https://doi.org/10.1007/s00382-017-3991-4

Comboul, M., Emile-Geay, J., Hakim, G. J., & Evans, M. N. (2015). Paleoclimate sampling as a sensor placement problem. *Journal of Climate*, *28*(19), 7717–7740. https://doi.org/10.1175/jcli-d-14-00802.1

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., 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. https://doi.org/10.1002/qj.828

d'Orgeval, T., Polcher, J., & De Rosnay, P. (2008). Sensitivity of the West African hydrological cycle in ORCHIDEE to infiltration processes. *Hydrology and Earth System Sciences*, *12*(6), 1387–1401. https://doi.org/10.5194/hess-12-1387-2008

Environmental Research, by the National Oceanic and Atmospheric Administration Climate Program Office, and by the NOAA Physical Sciences Laboratory. We would also like to thank the handling Editor, Jonathan Poterjoy, and the three anonymous reviewers for their insights and comments that helped improve the manuscript.

Doucet, A., de Freitas, J., & Gordon, N. (2001). *Sequential Monte Carlo methods in practic*e. Springer. https://doi.org/10.1007/978-1-4757-3437-9

Dubinkina, S., & Goosse, H. (2013). An assessment of particle filtering methods and nudging for climate state reconstructions. *Climate of the Past*, *9*(3), 1141–1152. https://doi.org/10.5194/cp-9-1141-2013

Dubinkina, S., Goosse, H., Sallaz-Damaz, Y., Crespin, E., & Crucifix, M. (2011). Testing a particle filter to reconstruct climate changes over the past centuries. *International Journal of Bifurcation and Chaos*, *21*(12), 3611–3618. https://doi.org/10.1142/s0218127411030763

Dufresne, J. L., Foujols, M. A., Denvil, S., Caubel, A., Marti, O., Aumont, O., et al. (2013). Climate change projections using the IPSL-CM5 Earth system model: From CMIP3 to CMIP5. *Climate Dynamics*, *40*(9), 2123–2165. https://doi.org/10.1007/s00382-012-1636-1

Folland, C. K., Parker, D. E., Colman, A. W., & Washington, R. (1999). Large scale modes of ocean surface temperature since the late nineteenth century. In *Beyond El Niño* (pp. 73–102). Springer.

Gent, P. R., Danabasoglu, G., Donner, L. J., Holland, M. M., Hunke, E. C., Jayne, S. R., et al. (2011). The community climate system model version 4. *Journal of Climate*, *24*(19), 4973–4991. https://doi.org/10.1175/2011jcli4083.1

Goosse, H. (2017). Reconstructed and simulated temperature asymmetry between continents in both hemispheres over the last centuries. *Climate Dynamics*, *48*(5–6), 1483–1501. https://doi.org/10.1007/s00382-016-3154-z

Goosse, H., Crespin, E., de Montety, A., Mann, M. E., Renssen, H., & Timmermann, A. (2010). Reconstructing surface temperature changes over the past 600 years using climate model simulations with data assimilation. *Journal of Geophysical Research*, *115*(D9), D09108. https://doi.org/10.1029/2009jd012737

Goosse, H., Renssen, H., Timmermann, A., Bradley, R. S., & Mann, M. E. (2006). Using paleoclimate proxy-data to select optimal realizations in an ensemble of simulations of the climate of the past millennium. *Climate Dynamics*, *27*(2), 165–184. https://doi.org/10.1007/s00382-006-0128-6

Gray, S. T., Graumlich, L. J., Betancourt, J. L., & Pederson, G. T. (2004). A tree-ring based reconstruction of the Atlantic multidecadal oscillation since 1567 AD. *Geophysical Research Letters*, *31*(12), L12205. https://doi.org/10.1029/2004gl019932

Hakim, G. J., Emile-Geay, J., Steig, E. J., Noone, D., Anderson, D. M., Tardif, R., et al. (2016). The last millennium climate reanalysis project: Framework and first results. *Journal of Geophysical Research: Atmospheres*, *121*(12), 6745–6764. https://doi.org/10.1002/2016jd024751

Hargreaves, J., & Annan, J. (2002). Assimilation of paleo-data in a simple Earth system model. *Climate Dynamics*, *19*(5), 371–381. https://doi.org/10.1007/s00382-002-0241-0

Henley, B. J., Gergis, J., Karoly, D. J., Power, S., Kennedy, J., & Folland, C. K. (2015). A tripole index for the interdecadal Pacific oscillation. *Climate Dynamics*, *45*(11), 3077–3090. https://doi.org/10.1007/s00382-015-2525-1

Hourdin, F., Rio, C., Grandpeix, J. Y., Madeleine, J. B., Cheruy, F., Rochetin, N., et al. (2020). LMDZ6A: The atmospheric component of the IPSL climate model with improved and better tuned physics. *Journal of Advances in Modeling Earth Systems*, *12*(7), e2019MS001892. https://doi.org/10.1029/2019ms001892

Houtekamer, P. L., & Zhang, F. (2016). Review of the ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, *144*(12), 4489–4532. https://doi.org/10.1175/mwr-d-15-0440.1

Jebri, B., Khodri, M., Echevin, V., Gastineau, G., Thiria, S., Vialard, J., & Lebas, N. (2020). Contributions of internal variability and external forcing to the recent trends in the Southeastern Pacific and Peru–Chile upwelling system. *Journal of Climate*, *33*(24), 10555–10578. https://doi.org/10.1175/jcli-d-19-0304.1

Johnson, S. D., Battisti, D. S., & Sarachik, E. S. (2000). Empirically derived Markov models and prediction of tropical Pacific sea surface temperature anomalies. *Journal of Climate*, *13*(1), 3–17. https://doi.org/10.1175/1520-0442(2000)013<0003:edmmap>2.0.co;2

Jungclaus, J. H., Bard, E., Baroni, M., Braconnot, P., Cao, J., Chini, L. P., et al. (2017). The PMIP4 contribution to CMIP6–Part 3: The last millennium, scientific objective, and experimental design for the PMIP4 past1000 simulations. *Geoscientific Model Development*, *10*(11), 4005–4033. https://doi.org/10.5194/gmd-10-4005-2017

Jungclaus, J. H., Fischer, N., Haak, H., Lohmann, K., Marotzke, J., Matei, D., et al. (2013). Characteristics of the ocean simulations in the Max Planck Institute Ocean Model (MPIOM) the ocean component of the MPI-Earth system model. *Journal of Advances in Modeling Earth Systems*, *5*(2), 422–446. https://doi.org/10.1002/jame.20023

Kageyama, M., Albani, S., Braconnot, P., Harrison, S. P., Hopcroft, P. O., Ivanovic, R. F., et al. (2017). The PMIP4 contribution to CMIP6–Part 4: Scientific objectives and experimental design of the PMIP4-CMIP6 Last Glacial Maximum experiments and PMIP4 sensitivity experiments. *Geoscientific Model Development*, *10*(11), 4035–4055. https://doi.org/10.5194/gmd-10-4035-2017

Keeler, R. J., & Ellis, S. M. (2000). Observational error covariance matrices for radar data assimilation. *Physics and Chemistry of the Earth - Part B: Hydrology, Oceans and Atmosphere*, *25*(10–12), 1277–1280. https://doi.org/10.1016/s1464-1909(00)00193-3

Kuptametee, C., & Aunsri, N. (2022). A review of resampling techniques in particle filtering framework. *Measurement*, *193*, 110836. https://doi.org/10.1016/j.measurement.2022.110836

Kurosawa, K., & Poterjoy, J. (2021). Data assimilation challenges posed by nonlinear operators: A comparative study of ensemble and variational filters and smoothers. *Monthly Weather Review*, *149*(7), 2369–2389. https://doi.org/10.1175/mwr-d-20-0368.1

Liu, H., Liu, Z., & Lu, F. (2017). A systematic comparison of particle filter and EnKF in assimilating time-averaged observations. *Journal of Geophysical Research: Atmospheres*, *122*(24), 13–155. https://doi.org/10.1002/2017jd026798

Liu, J. S., & Chen, R. (1998). Sequential Monte Carlo methods for dynamic systems. *Journal of the American Statistical Association*, *93*(443), 1032–1044. https://doi.org/10.1080/01621459.1998.10473765

Lurton, T., Balkanski, Y., Bastrikov, V., Bekki, S., Bopp, L., Braconnot, P., et al. (2020). Implementation of the CMIP6 forcing data in the IPSL-CM6A-LR model. *Journal of Advances in Modeling Earth Systems*, *12*(4), e2019MS001940. https://doi.org/10.1029/2019ms001940

Mann, M. E., Zhang, Z., Hughes, M. K., Bradley, R. S., Miller, S. K., Rutherford, S., & Ni, F. (2008). 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), 13252–13257. https://doi.org/10.1073/pnas.0805721105

Martinez-Villalobos, C., Newman, M., Vimont, D. J., Penland, C., & David Neelin, J. (2019). Observed El Niño-La Niña asymmetry in a linear model. *Geophysical Research Letters*, *46*(16), 9909–9919. https://doi.org/10.1029/2019gl082922

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

Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., et al. (2019). Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1. 2) and its response to increasing CO2. *Journal of Advances in Modeling Earth Systems*, *11*(4), 998–1038. https://doi.org/10.1029/2018ms001400

McPhaden, M. J., Zebiak, S. E., & Glantz, M. H. (2006). ENSO as an integrating concept in Earth science. *Science*, *314*(5806), 1740–1745. https://doi.org/10.1126/science.1132588

Miller, R. L., Schmidt, G. A., Nazarenko, L. S., Tausnev, N., Bauer, S. E., DelGenio, A. D., et al. (2014). CMIP5 historical simulations (1850–2012) with GISS ModelE2. *Journal of Advances in Modeling Earth Systems*, *6*(2), 441–478. https://doi.org/10.1002/2013ms000266

Nadiga, B. T. (2021). Reservoir computing as a tool for climate predictability studies. *Journal of Advances in Modeling Earth Systems*, *13*(4), e2020MS002290. https://doi.org/10.1029/2020ms002290

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of Hydrology*, *10*(3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

Neukom, R., Gergis, J., Karoly, D. J., Wanner, H., Curran, M., Elbert, J., et al. (2014). Inter-hemispheric temperature variability over the past millennium. *Nature Climate Change*, *4*(5), 362–367. https://doi.org/10.1038/nclimate2174

Newman, M. (2013). An empirical benchmark for decadal forecasts of global surface temperature anomalies. *Journal of Climate*, *26*(14), 5260–5269. https://doi.org/10.1175/jcli-d-12-00590.1

Ohgaito, R., Yamamoto, A., Hajima, T., O'ishi, R., Abe, M., Tatebe, H., et al. (2021). PMIP4 experiments using MIROC-ES2L Earth system model. *Geoscientific Model Development*, *14*(2), 1195–1217. https://doi.org/10.5194/gmd-14-1195-2021

Okazaki, A., Miyoshi, T., Yoshimura, K., Greybush, S. J., & Zhang, F. (2021). Revisiting online and offline data assimilation comparison for paleoclimate reconstruction: An idealized OSSE study. *Journal of Geophysical Research: Atmospheres*, *126*(16), e2020JD034214. https://doi.org/10.1029/2020jd034214

PAGES2k Consortium. (2017). A global multiproxy database for temperature reconstructions of the Common Era. *Scientific Data*, *4*.

PAGES2k Consortium. (2019). Consistent multi-decadal variability in global temperature reconstructions and simulations over the Common Era. *Nature Geoscience*, *12*(8), 643–649. https://doi.org/10.1038/s41561-019-0400-0

Parsons, L. A., & Hakim, G. J. (2019). Local regions associated with interdecadal global temperature variability in the last millennium reanalysis and CMIP5 models. *Journal of Geophysical Research: Atmospheres*, *124*(17–18), 9905–9917. https://doi.org/10.1029/2019jd030426

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

Penland, C., & Matrosova, L. (1994). A balance condition for stochastic numerical models with application to the El Nino-Southern Oscillation. *Journal of Climate*, *7*(9), 1352–1372. https://doi.org/10.1175/1520-0442(1994)007<1352:abcfsn>2.0.co;2

Penland, C., & Matrosova, L. (1998). Prediction of tropical Atlantic sea surface temperatures using linear inverse modeling. *Journal of Climate*, *11*(3), 483–496. https://doi.org/10.1175/1520-0442(1998)011<0483:potass>2.0.co;2

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

Perkins, W. A., & Hakim, G. (2020). Linear inverse modeling for coupled atmosphere-ocean ensemble climate prediction. *Journal of Advances in Modeling Earth Systems*, *12*(1), e2019MS001778. https://doi.org/10.1029/2019ms001778

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

Planton, Y., Vialard, J., Guilyardi, E., Lengaigne, M., & Izumo, T. (2018). Western Pacific oceanic heat content: A better predictor of La Niña than of El Niño. *Geophysical Research Letters*, *45*(18), 9824–9833. https://doi.org/10.1029/2018gl079341

Power, S., Casey, T., Folland, C., Colman, A., & Mehta, V. (1999). Inter-decadal modulation of the impact of ENSO on Australia. *Climate Dynamics*, *15*(5), 319–324. https://doi.org/10.1007/s003820050284

Richter, I., Chang, P., & Liu, X. (2020). Impact of systematic GCM errors on prediction skill as estimated by linear inverse modeling. *Journal of Climate*, *33*(23), 10073–10095. https://doi.org/10.1175/jcli-d-20-0209.1

Richter, I., & Tokinaga, H. (2020). An overview of the performance of CMIP6 models in the tropical Atlantic: Mean state, variability, and remote impacts. *Climate Dynamics*, *55*(9), 2579–2601. https://doi.org/10.1007/s00382-020-05409-w

Rousset, C., Vancoppenolle, M., Madec, G., Fichefet, T., Flavoni, S., Barthélemy, A., et al. (2015). The Louvain-La-Neuve sea ice model LIM3.6: Global and regional capabilities. *Geoscientific Model Development*, *8*(10), 2991–3005. https://doi.org/10.5194/gmd-8-2991-2015

Schmidt, G. A., Jungclaus, J. H., Ammann, C. M., Bard, E., Braconnot, P., Crowley, T. J., et al. (2012). Climate forcing reconstructions for use in PMIP simulations of the last millennium (v1. 1). *Geoscientific Model Development*, *5*(1), 185–191. https://doi.org/10.5194/gmd-5-185-2012

Shin, J., Park, S., Shin, S. I., Newman, M., & Alexander, M. A. (2020). Enhancing ENSO prediction skill by combining model-analog and linear inverse models (MA-LIM). *Geophysical Research Letters*, *47*(1), e2019GL085914. https://doi.org/10.1029/2019gl085914

Slivinski, L. C., Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Giese, B. S., McColl, C., et al. (2019). Towards a more reliable historical reanalysis: Improvements for version 3 of the Twentieth Century Reanalysis system. *Quarterly Journal of the Royal Meteorological Society*, *145*(724), 2876–2908. https://doi.org/10.1002/qj.3598

Smerdon, J. E., & Pollack, H. N. (2016). Reconstructing Earth's surface temperature over the past 2000 years: The science behind the headlines. *Wiley Interdisciplinary Reviews: Climate Change*, *7*(5), 746–771. https://doi.org/10.1002/wcc.418

Snyder, C., Bengtsson, T., Bickel, P., & Anderson, J. (2008). Obstacles to high-dimensional particle filtering. *Monthly Weather Review*, *136*(12), 4629–4640.

Stevenson, S., Otto-Bliesner, B. L., Brady, E. C., Nusbaumer, J., Tabor, C., Tomas, R., et al. (2019). Volcanic eruption signatures in the isotope-enabled last millennium ensemble. *Paleoceanography and Paleoclimatology*, *34*(8), 1534–1552. https://doi.org/10.1029/2019pa003625

Thirumalai, K., Quinn, T. M., Okumura, Y., Richey, J. N., Partin, J. W., Poore, R. Z., & Moreno-Chamarro, E. (2018). Pronounced centennial-scale Atlantic Ocean climate variability correlated with Western Hemisphere hydroclimate. *Nature Communications*, *9*(1), 1–11. https://doi.org/10.1038/s41467-018-02846-4

Tian, B., & Dong, X. (2020). The double-ITCZ bias in CMIP3, CMIP5, and CMIP6 models based on annual mean precipitation. *Geophysical Research Letters*, *47*(8), e2020GL087232. https://doi.org/10.1029/2020gl087232

Ting, M., Kushnir, Y., Seager, R., & Li, C. (2009). Forced and internal twentieth-century SST trends in the North Atlantic. *Journal of Climate*, *22*(6), 1469–1481. https://doi.org/10.1175/2008jcli2561.1

van Dijk, E., Jungclaus, J., Lorenz, S., Timmreck, C., & Krüger, K. (2022). Was there a volcanic-induced long-lasting cooling over the Northern Hemisphere in the mid-6th–7th century? *Climate of the Past*, *18*(7), 1601–1623. https://doi.org/10.5194/cp-18-1601-2022

Van Leeuwen, P. J. (2009). Particle filtering in geophysical systems. *Monthly Weather Review*, *137*(12), 4089–4114. https://doi.org/10.1175/2009mwr2835.1

Wang, J., Emile-Geay, J., Guillot, D., McKay, N. P., & Rajaratnam, B. (2015). Fragility of reconstructed temperature patterns over the Common Era: Implications for model evaluation. *Geophysical Research Letters*, *42*(17), 7162–7170. https://doi.org/10.1002/2015gl065265

Watanabe, S., Hajima, T., Sudo, K., Nagashima, T., Takemura, T., Okajima, H., et al. (2011). MIROC-ESM 2010: Model description and basic results of CMIP5-20c3m experiments. *Geoscientific Model Development*, *4*(4), 845–872. https://doi.org/10.5194/gmd-4-845-2011

Whitaker, J. S., & Hamill, T. M. (2002). Ensemble data assimilation without perturbed observations. *Monthly Weather Review*, *130*(7), 1913–1924. https://doi.org/10.1175/1520-0493(2002)130<1913:edawpo>2.0.co;2

Widmann, M., Goosse, H., van der Schrier, G., Schnur, R., & Barkmeijer, J. (2010). Using data assimilation to study extratropical Northern Hemisphere climate over the last millennium. *Climate of the Past*, *6*(5), 627–644. https://doi.org/10.5194/cp-6-627-2010

Winkler, C. R., Newman, M., & Sardeshmukh, P. D. (2001). A linear model of wintertime low-frequency variability. Part I: Formulation and forecast skill. *Journal of Climate*, *14*(24), 4474–4494. https://doi.org/10.1175/1520-0442(2001)014<4474:almowl>2.0.co;2

Xu, Z., Chang, P., Richter, I., & Tang, G. (2014). Diagnosing southeast tropical Atlantic SST and ocean circulation biases in the CMIP5 ensemble. *Climate Dynamics*, *43*(11), 3123–3145. https://doi.org/10.1007/s00382-014-2247-9

Yukimoto, S., Kawai, H., Koshiro, T., Oshima, N., Yoshida, K., Urakawa, S., et al. (2019). The meteorological research institute Earth system model version 2.0, MRI-ESM2. 0: Description and basic evaluation of the physical component. *Journal of the Meteorological Society of Japan. Ser. II*, *97*(5), 931–965. https://doi.org/10.2151/jmsj.2019-051

Zanchettin, D., Rubino, A., Matei, D., Bothe, O., & Jungclaus, J. H. (2013). Multidecadal-to-centennial SST variability in the MPI-ESM simulation ensemble for the last millennium. *Climate Dynamics*, *40*(5), 1301–1318. https://doi.org/10.1007/s00382-012-1361-9

References From the Supporting Information

Hoerling, M. P., Kumar, A., & Zhong, M. (1997). El Niño, La Niña, and the nonlinearity of their teleconnections. *Journal of Climate*, *10*(8), 1769–1786. https://doi.org/10.1175/1520-0442(1997)010<1769:enolna>2.0.co;2

Jourdain, N. C., Lengaigne, M., Vialard, J., Izumo, T., & Gupta, A. S. (2016). Further insights on the influence of the Indian Ocean dipole on the following year's ENSO from observations and CMIP5 models. *Journal of Climate*, *29*(2), 637–658. https://doi.org/10.1175/jcli-d-15-0481.1

Larkin, N. K., & Harrison, D. E. (2002). ENSO warm (El Niño) and cold (La Niña) event life cycles: Ocean surface anomaly patterns, their symmetries, asymmetries, and implications. *Journal of Climate*, *15*(10), 1118–1140. https://doi.org/10.1175/1520-0442(2002)015<1118:ewenoa>2.0.co;2

Menkes, C. E., Lengaigne, M., Vialard, J., Puy, M., Marchesiello, P., Cravatte, S., & Cambon, G. (2014). About the role of westerly wind events in the possible development of an El Niño in 2014. *Geophysical Research Letters*, *41*(18), 6476–6483. https://doi.org/10.1002/2014gl061186

Okumura, Y. M., & Deser, C. (2010). Asymmetry in the duration of El Niño and La Niña. *Journal of Climate*, *23*(21), 5826–5843. https://doi.org/10.1175/2010jcli3592.1

Puy, M., Vialard, J., Lengaigne, M., Guilyardi, E., DiNezio, P. N., Voldoire, A., et al. (2019). Influence of westerly wind events stochasticity on El Niño amplitude: The case of 2014 vs. 2015. *Climate Dynamics*, *52*(12), 7435–7454. https://doi.org/10.1007/s00382-017-3938-9