

Reconstructing East African rainfall and Indian Ocean sea surface temperatures over the last centuries using data assimilation

François Klein¹ · Hugues Goosse¹

Received: 30 August 2016 / Accepted: 3 August 2017 / Published online: 18 August 2017
© Springer-Verlag GmbH Germany 2017

Abstract The relationship between the East African rainfall and Indian Ocean sea-surface temperatures (SSTs) is well established. The potential interest of this covariance to improve reconstructions of both variables over the last centuries is examined here. This is achieved through an off-line method of data assimilation based on a particle filter, using hydroclimate-related records at four East African sites (Lake Naivasha, Lake Challa, Lake Malawi and Lake Masoko) and SSTs-related records at six oceanic sites spread over the Indian Ocean to constrain the Last Millennium Ensemble of simulations performed by CESM1. Skillful reconstructions of the Indian SSTs and East African rainfall can be obtained based on the assimilation of only one of these variables, when assimilating pseudo-proxy data deduced from the model CESM1. The skill of these reconstructions increases with the number of particles selected in the particle filter, although the improvement becomes modest beyond 99 particles. When considering a more realistic framework, the skill of the reconstructions is strongly deteriorated because of the model biases and the uncertainties of the real proxy-based reconstructions. However, it is still possible to obtain a skillful reconstruction of SSTs over most of the Indian Ocean only based on the assimilation of the six SST-related proxy records selected, as far as a local calibration is applied at all

individual sites. This underlines once more the critical role of an adequate integration of the signal inferred from proxy records into the climate models for reconstructions based on data assimilation.

Keywords Data assimilation · East Africa · Indian Ocean · Model-data comparison

1 Introduction

Equatorial Eastern Africa is characterized by a heterogeneous spatial distribution of precipitation due to local and regional meteorological features, which are influenced by the presence of numerous lakes and a complex topography. Consequently, this region supports a large variety of environments, ranging from deserts with rainfall less than 200 mm per year to tropical rain forests with annual rainfall above 2000 mm (Nicholson 1996). Due to the seasonal migration of the Intertropical Convergence Zone (ITCZ) back and forth across the equator, the annual cycle of rainfall is bimodal over much of East Africa, with a first and main rainy season occurring from March to May (the so-called long rains) and a second from October to December (short rains). Nonetheless, this seasonal distribution of precipitation can change rapidly over short distances and some regions, such as the Ethiopian and the Eastern Highlands, only have one rainy season with a peak during boreal and austral summer, respectively (e.g. Owiti and Zhu 2012; Yang et al. 2015).

Because of the influence of large-scale atmospheric and oceanic factors, the temporal variability of rainfall tends to be more homogeneous than the mean patterns over the region (Nicholson 2014). At the inter-annual scale, numerous observation-based (e.g. Ogallo et al. 1988; Hastenrath et al. 1993; Saji et al. 1999; Webster et al. 1999; Nicholson

Electronic supplementary material The online version of this article (doi:10.1007/s00382-017-3853-0) contains supplementary material, which is available to authorized users.

✉ François Klein
francois.klein@uclouvain.be

¹ Georges Lemaître Centre for Earth and Climate Research (TECLIM), Earth and Life Institute (ELI), Université catholique de Louvain (UCL), Place Louis Pasteur, 3, 1348 Louvain-La-Neuve, Belgium

and Selato 2000; Clark et al. 2003; Schreck and Semazzi 2004; Izumo et al. 2014) and model-based (e.g. Goddard and Graham 1999; Ummenhofer et al. 2009; Rowell 2013; Klein et al. 2016; Schubert et al. 2016) studies have emphasised the teleconnections between the short rains, which are the main driver of the inter-annual variability of the East African rainfall (e.g. Nicholson 1996, 2014), and the Indian and Pacific SSTs. Specifically, precipitation over East Africa during the short rains correlates positively with the SSTs over the Western Indian Ocean and the Central and Eastern Pacific Ocean, while the correlations are negative over the Eastern Indian Ocean. This link appears robust and has been reported using various techniques including correlation coefficients (Ogallo et al. 1988; Goddard and Graham 1999; Saji et al. 1999; Clark et al. 2003; Rowell 2013; Klein et al. 2016; Schubert et al. 2016), visual comparison of different 2D fields (Webster et al. 1999; Ummenhofer et al. 2009), empirical orthogonal functions (Schreck and Semazzi 2004; Goddard and Graham 1999; Saji et al. 1999) and composite (Saji et al. 1999) or harmonic (Nicholson and Selato 2000) analyses. However, such statistical methods do not allow for the identification of the physical nature of this link. For instance, they do not provide insight as to whether the SSTs are responsible for the East African rainfall variability, or whether both variables react to a common forcing.

The goal of this study is to perform reconstructions of the East African rainfall and Indian Ocean SSTs over the last millennium, based on the covariance observed between those two variables, through a data assimilation method. The covariance can indeed be used to potentially obtain better reconstructions of the two variables considered, and can be particularly useful over periods when instrumental observations are lacking and indirect proxy records are scarce. In contrast to more classical statistical methods, data assimilation has the advantage not to rely on the length of a calibration period and on the quality of the records during this period, which could be a problem in Africa. Furthermore, it can handle a potential non-stationarity of the covariance, related for instance to changes in the mean state, and ensure that the reconstruction for both variables is compatible with the physics of the system as represented by the model. Although a link between the East African rainfall and SSTs in the Pacific has also been reported, it seems less robust than the one in the Indian Ocean (Klein et al. 2016), and will thus not be considered in the present study.

Since the early 1990s, data assimilation has been extensively employed in various Earth science disciplines including meteorology, oceanography and hydrology (e.g. Park and Xu 2009; Dee et al. 2011; Balmaseda et al. 2015). In palaeoclimatology, it is still considered as an emerging topic (e.g. Matsikaris et al. 2015), despite growing interest over the past 15 years (e.g. von Storch et al. 2000; Goosse et al. 2006; Crespin et al. 2009; Widmann et al. 2010; Hakim et al.

2016). When applied to palaeoclimatology, data assimilation aims to combine information from model results and proxy-based reconstructions to find estimates of past climate changes. Here, a particle filter method has been preferred over other data assimilation methods such as forcing singular vectors (e.g. Barkmeijer et al. 2003) or pattern nudging (e.g. von Storch et al. 2000). This ensemble method has the advantage of respecting the model physics as much as possible, by selecting a set of model states that are compatible with observations within the range of a finite ensemble, taking into account the uncertainties. While this choice may potentially limit the skill of the reconstructions given that the real climate has a large state space with many degrees of freedom, the method has already led to very satisfying results in conditions similar to the ones of the problem studied here (e.g. Goosse et al. 2012; Mairesse et al. 2013). Furthermore, ensemble particle filters are relatively easy to implement and do not require any information about the model. If a satisfactory reconstruction can be obtained where the constraint is applied, the empirical information will then be propagated to various climate variables and spatially spread, based on the physics and the dynamics of the climate model used.

In the last decade, two types of ensemble-based methods of data assimilation have been applied to study the past millennium climate (e.g. Matsikaris et al. 2015). First, data assimilation can be used to update an ensemble of simulations at regular interval (on-line data assimilation). In this case, the ensemble is generated sequentially, each period depending on the diagnostic made by the data assimilation process on the previous ones. Second, data assimilation can occur a posteriori, making use of an existing ensemble of simulations (off-line data assimilation). The on-line method has advantages compared to the off-line one. Indeed, it maintains the temporal consistency, and it is expected to bring better results than an off-line method as the slow component of the climate system, such as the oceans, can propagate an information forward in time (Pendergrass et al. 2012; Matsikaris et al. 2015). However, if the predictability of the variables of interest is limited between two assimilation steps because of a dominant role of the chaotic nature of the system, an on-line technique would not outperform an off-line one.

The main interest of an off-line method lies in the fact that we can use an existing ensemble of simulations, saving a significant amount of computational time. This allows applying a data assimilation scheme in higher resolution general circulation models (GCMs) and considering larger ensembles, such as in Bhend et al. (2012) or in Steiger et al. (2014). In contrast, on-line data assimilation methods are usually used in palaeoclimatology with simplified models with a relatively coarse resolution such as in Crespin et al. (2009), Goosse et al. (2012), or in Klein et al. (2013), or with

relatively small ensembles (Matsikaris et al. 2015). We do not expect a large predictability of the East African rainfall on the interannual time-scale, meaning that the additional information brought by an online method would be limited, as shown in previous works (e.g. Annan and Hargreaves 2012). Furthermore, an off-line procedure is also much more flexible as many tests can be performed on a single existing ensemble, while new simulations are required each time in the on-line approach. Hence, an off-line method is preferred over an on-line one in the present study, using the ensemble of ten simulations recently performed (Otto-Bliesner et al. 2015) with the Community Earth System Model version 1.1 (CESM1; Hurrell et al. 2013), which is the largest ensemble to date made with a GCM that covers the last millennium.

To our knowledge, reconstructing simultaneously precipitation and SSTs over long periods using data assimilation has never been attempted before. Hence, in order to test the methodology and to identify the most critical issues, this study contains several steps gradually increasing in complexity. The most simple case is the assimilation of pseudo-proxies where the data used to constrain model results are derived from the climate model CESM1 itself. The locations of the sites selected for the assimilation are the ones for which actual proxy-based reconstructions are available. In this framework, the model physics is supposed to be perfect as both the pseudo-observations and the model ensemble originate from the same source. Furthermore, the reconstruction target and the uncertainty of the records are perfectly known, the latter being given by the noise imposed on model results to obtain the pseudo-proxies. This setup is thus ideal to assess precisely the performance of the data assimilation method itself. Two sensitivity analyses are performed using this framework, one assessing the effect of the size of the ensemble, and the other looking at the individual effect of each site on the quality of the reconstructions. In a second step, pseudo-proxy data derived from other climate models as well as from recent observations are considered. This allows studying how sensitive the reconstructions are to biases in the model physics. Lastly, real proxy-based reconstructions are used to constrain the model simulations. The selected hydroclimate-related reconstructions describe the water-balance history during roughly the last millennium of Lake Challa, Lake Naivasha, Lake Masoko and Lake Malawi, and are part of the East African hydroclimate synthesis achieved by Tierney et al. (2013). The six SST-related records are derived from the $\delta^{18}\text{O}$ content of coral archives spread over the Indian Ocean, the choice of the records being based on the compilation of Tierney et al. (2015).

This study is structured as follows. The next section describes the climate model used (Sect. 2.1), the data assimilation method (Sect. 2.2), and the different data assimilated, i.e. the pseudo-proxies (Sect. 2.3) and the proxy-based reconstructions (Sect. 2.4). Section 3 describes the reconstructions

of the East African precipitation and Indian Ocean SSTs based on the assimilation of the pseudo-proxy data and Sect. 4 is focussed on the reconstructions performed by assimilating the proxy-based reconstructions. Finally, Sect. 5 presents the discussion and conclusions.

2 Methodology

2.1 Model results

Data assimilation is based on the ensemble of simulations performed over the period 850–2005 AD with CESM1 (the CESM1 Last Millennium Ensemble or CESM-LME; Otto-Bliesner et al. 2015). This model has a horizontal resolution of about two degrees for the atmosphere and land components, and of about one degree for the ocean and sea-ice components. The ensemble contains ten simulations, which only differ from slightly different atmospheric states at the start of the experiments. They are driven through the last millennium by changes in both natural and anthropogenic climate forcings. The solar irradiance variations follow the reconstruction of Vieira et al. (2011), with the spectral variations and ‘11-year’ solar cycle from Schmidt et al. (2012). Changes in the forcings related to volcanic aerosols are derived from Gao et al. (2008). The evolution of the major greenhouse gases (CO_2 , CH_4 and N_2O) is based for the pre-industrial period (850–1850 AD) on Flückiger et al. (2002) and MacFarling Meure et al. (2006), and for the recent period on Hansen and Sato (2004). Anthropogenic changes in land use/land cover are first derived from the reconstruction of Pongratz et al. (2009) until 1850, and follow Hurtt et al. (2011) thereafter. Finally, Earth’s orbital parameters are updated every year based on Berger et al. (1993).

CESM1 is able to simulate relatively well the major features of the East African rainfall, the Indian Ocean SSTs, and the link between those two variables. For instance, Klein et al. (2016) have shown that both the unimodal cycle of rainfall observed at Lake Masoko and Lake Malawi and the bimodal cycle observed at Lake Challa and Lake Naivasha are well reproduced by the model, its results being the closest to observations compared to five other GCMs. This model is able to simulate correctly the seasonal cycle of the SSTs over the eastern and the western Indian Ocean, although the amplitude of the seasonal cycle is slightly underestimated in the western Indian Ocean, as is the case for all GCMs tested (Fig. S1). The recent trend in the Indian Ocean SSTs is also relatively well reproduced by CESM1 and the other GCMs (Fig. S2). Finally, the pattern of correlations between the East African rainfall and the Indian Ocean SSTs in CESM1 is the one that best matches the observations compared to the other GCMs (Klein et al. 2016), making CESM1 adequate for the present analysis.

2.2 Data assimilation method

The off-line ensemble-based method of data assimilation applied here uses a particle filter (e.g. van Leeuwen 2009), as described in Dubinkina et al. (2011). This method has the advantage to be very general. It does not rely on hypotheses such as the Gaussian distribution of the error, as is the case for other ensemble-based methods such as the Kalman filter (van Leeuwen 2009), and it enables the experiments to be performed off-line, which is needed here. The frequency of the data assimilation is annual, meaning that annual mean values from every members of the ensemble of simulations, called particles, are compared to data each year of the reconstruction period. Based on this comparison, the likelihood of every particle is computed, taking into account the uncertainties of the data. The likelihood is a measure of the ability of the different ensemble members to simulate the signal showed in the data. The model-data comparison is performed using anomalies in order to remove the potential model mean biases, although CESM1 simulates quite well the East African rainfall and the Indian Ocean SSTs (and previous section Klein et al. 2016). Depending on the likelihood values, a weight is then attributed to the particles. When the weights of each ensemble member are known for a given year, a weighted mean is computed, which provides the reconstruction for that year.

Only the first nine simulations of the CESM-LME are used in the data assimilation process, because pseudo-proxy data are, in some experiments, derived from the tenth one. When mentioning the CESM-LME later in the text, it thus refers to the first nine members of the ensemble only. An ensemble of nine particles may be too small to obtain a good reconstruction since the range of possible climate states covered by those simulations may not be large enough to reproduce the signal present in the data assimilated (e.g. Goosse et al. 2006). However, it is not possible to perform additional simulations with CESM1 over this period. Hence, in order to increase the size of the ensemble and thus the probability to find a good match between model results and observations for a given year, the natural variability is sampled by also selecting other years for the ensemble of model states, that will thus differ from the year in observations.

This is only valid if the forcings play a marginal role compared to natural variability, as appears to be the case for interannual changes in East African precipitation and, to a lesser extent, in Indian Ocean SSTs (Klein et al. 2016). There is thus a priori no reason why the timing of events in observations and in a simulation should be similar, meaning that any model year can be included in the ensemble. For the CESM-LME, this means that the potential size of the ensemble can rise up to 9×1156 members. Here, six ensemble sizes are considered containing 9, 27, 99, 207, 1035 and 2079 particles. In the first case, the data for a

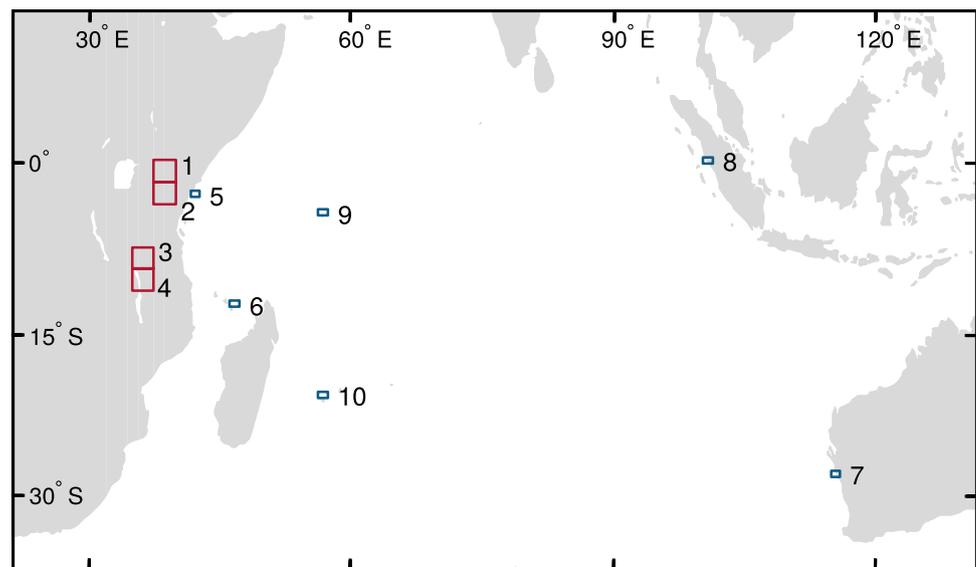
given year are only compared to the nine ensemble member results available for this actual year. When 27 particles are selected, the model results of one year out of 500 are added to the ensemble, besides the nine results of the actual year of the process. To increase the size of the ensemble to 99, 207, 1035 and 2079, one year out of 100, 50, 10 and 5 are considered, besides the actual year of the experiments. Note that we keep using only the first nine ensemble members of the CESM-LME even when real proxy records or results of other models are assimilated, in order to have a meaningful comparison between all experiments.

2.3 Pseudo-proxy data

This study starts with the assimilation of the results of the tenth ensemble member of the CESM-LME over the last millennium, at the same locations as the real proxy-based reconstructions that are described in the next section. The cells of the atmospheric and oceanic grids containing the records sites are shown in Fig. 1. The variables measured or reconstructed from the proxy data, ie. $\delta^{18}\text{O}$ for the oceanic ones and different hydroclimate-related variables for the East African lakes, are not explicitly simulated by the climate model. Coral $\delta^{18}\text{O}$ changes are directly linked to variations in SSTs (e.g. Juillet-Leclerc and Schmidt 2001; Stevenson et al. 2013), and East African hydroclimate changes are mostly dependent on precipitation rather than on evaporation in models (Klein et al. 2016). Thus, in a first step, the variable assimilated at the locations of the six oceanic records is yearly averaged SST, while it is yearly averaged precipitation at the locations of the four continental grid cells containing the East African hydroclimate records. The pseudo-proxy time series are generated by the addition to the results of the tenth ensemble member of the CESM-LME of a white Gaussian noise with a standard deviation of $5 \text{ mm}\cdot\text{month}^{-1}$ for rainfall results and of $0.25 \text{ }^\circ\text{C}$ for SST results, these values corresponding to the data error estimates applied in the data assimilation process. This produces time series with signal-to-noise ratios of a similar magnitude to previous studies (e.g. Bhend et al. 2012; Steiger and Hakim 2016), with values ranging from 0.16 to 0.57 for SST and from 0.14 to 0.31 for rainfall. Moreover, changing these estimates by 20% positively or negatively does not significantly affect the results.

In order to estimate the uncertainty associated to the model biases, the second step of this study consists in assimilating pseudo-proxy data, but taken from other GCMs, ie. MPI-ESM-P (Stevens et al. 2013) and GISS-E2-R (Schmidt et al. 2014), as well as from instrumental observations, using the gridded precipitation data set GPCC-v7 (version 7 of the Global Precipitation Climatology Centre data set; Schneider et al. 2014) and the SSTs data set ERSST-v4 (version 4 of the Extended Reconstructed Sea Surface Temperature data

Fig. 1 Grid cells of CESM1 containing the sites of the proxy-based reconstructions selected for data assimilation. The red cells belong to the atmospheric grid while blue ones to the oceanic grid. The numbers correspond to the identifier of the reconstructions as shown in Table 1



set; Smith et al. 2008). The model MPI-ESM-P displays a strong teleconnection between the East African rainfall and Indian SSTs, that is relatively similar as the one simulated by CESM1, while GISS-E2-R shows a very weak link between both variables (Klein et al. 2016). These two models are thus good candidates to assess the effect of uncertainties in model physics and dynamics on the quality of the reconstructions. Regarding the instrumental observations, the data assimilation process is applied only over the period 1901–2005. The same noise as above is added to the model time series to obtain pseudo-proxies. These different models and gridded data sets do not have the same spatial resolution as CESM1. All results are thus interpolated into the CESM1 continental and oceanic grids before applying the data assimilation scheme.

2.4 Proxy-based reconstructions

The last, and most complex, step of this study consists in assimilating real proxy-based reconstructions of the East African hydroclimate and Indian Ocean coral $\delta^{18}\text{O}$. The selection of the hydroclimate reconstructions is based on the compilation achieved in Tierney et al. (2013), which contains seven lake-based East African proxy records covering the last millennium with a time resolution of at least 50 years. As in Klein et al. (2016), the reconstructions originating from Lake Challa, Lake Naivasha, Lake Masoko and Lake Malawi have been selected here. The records from Lake Tanganyika and Lake Edward are discarded because they are located far from the Indian Ocean, which has thus potentially less influence on their local hydroclimate. The reconstruction from Lake Victoria is also not included because the representation of this lake is different from one model to another. Indeed, CESM1

and GISS-E2-R simply ignore it, while it is present in the model MPI-ESM-P as a single grid point lake, meaning that the specific conditions imposed by the presence of the lake (e.g. Thiery et al. 2015) cannot be reproduced in the simulations.

The hydroclimate reconstructions are based on different proxies, but the four of them can be qualitatively seen as smoothed versions of the local moisture-balance of the areas in which they originate. For Lake Challa, the hydroclimate variation is derived in Tierney et al. (2013) from the first principal component of composite variation in three moisture-balance proxies, that accounts for 40% of the variance in the data (supplementary material of Tierney et al. 2013). The three proxies are a presumed indicator of catchment runoff (the branched and isoprenoidal tetraether index (BIT); Verschuren et al. 2009), an isotopic proxy for rainfall source and intensity (δD in the leaf waxes of terrestrial plants; Tierney et al. 2011), and a proxy for variation in dry-season length and windiness (varve thickness; Wolff et al. 2011). Regarding Lake Naivasha, the time series is a lake-level reconstruction derived from the sediment lithostratigraphy (Verschuren 2001), supported by salinity reconstructions based on fossil diatom and midge assemblages (Verschuren et al. 2000). In Lake Masoko, the hydroclimate reconstruction is inferred from the low-field magnetic susceptibility of the sediment, which is a proxy for lake-level changes and/or wind stress. Two such records are available for this lake, one that goes back to -43,300 AD (Garcin et al. 2006) and one that starts around 1500 AD (Garcin et al. 2007). The Masoko time series chosen here is obtained from Tierney et al. (2013), who used the last millennium of the longer record but with age-depth tie-points translated from the shorter one (Anchukaitis and Tierney 2013). The hydroclimate reconstruction in Lake Malawi is based on the mass accumulation rate of the

terrigenous sediment fraction, suggested to be a runoff proxy (Brown and Johnson 2005; Johnson and McCave 2008).

The Indian Ocean coral $\delta^{18}\text{O}$ time series selected in this study are taken from the work of Tierney et al. (2015) who compiled 57 marine coral-based archives in order to reconstruct SST in the tropics over the past four centuries. Out of these 57 records, 13 belong to the Indian Ocean domain, including 12 oxygen isotopic composition of coral carbonate. However, six of them are discarded because located on land in the grid of the CESM model. Thus, six time series of coral $\delta^{18}\text{O}$ remain available for our analysis. They originate from Malindi (Cole 2000), Mayotte (Zinke et al. 2008, 2009), Houtman Abrolhos (Kuhnert et al. 1999), Mentawai (Abram et al. 2008), Seychelles (Charles 1997) and La Réunion (Pfeiffer et al. 2004). The exact geographic coordinates of those reconstructions, as well as the period they cover, are given in Table 1. Note that the comparison of model results and proxy-based reconstructions during the process of data assimilation is made at the model grid cell scale, the related potential representativeness error being included in the data error considered in the data assimilation procedure.

Data assimilation involves an objective comparison between model results and the data assimilated, meaning that both have to represent the same physical quantity. Since the variables reconstructed from proxy records are not included as such in the climate models, a processing is applied to the simulated variables in order to emulate coral $\delta^{18}\text{O}$ and East African hydroclimate. This may potentially affect the quality of the reconstructions. Hence, before assimilating the real proxy-based reconstructions, the impact of assimilating the measured variables, ie. coral $\delta^{18}\text{O}$ and East African hydroclimate, instead of SSTs and precipitation, is investigated. This is done by considering once more the results of the tenth ensemble member of the CESM-LME as the “reality”.

For each ensemble member of the CESM-LME, the model coral $\delta^{18}\text{O}$ is computed based on the linear function of SST and sea surface salinity (SSS) derived by Thompson et al. (2011):

$$\delta^{18}\text{O}_{\text{CESM1-LME}^{1-10}} = -0.22 \times \text{SST}_{\text{CESM1-LME}^{1-10}} + 0.16 \times \text{SSS}_{\text{CESM1-LME}^{1-10}} \quad (1)$$

where SST is expressed in $^{\circ}\text{C}$ and SSS in PSU. The error chosen to obtain the pseudo-proxies is 0.05‰ .

It is not possible to apply the same approach for the four hydroclimate proxy-based reconstructions because of the complex and not well understood relationship between those proxies and the East African hydroclimate-related variables, including lake level, catchment runoff or seasonal drought severity depending on the site. In this final test, the model variable that is compared to the data is annually averaged precipitation minus evaporation (P–E). The data assimilation process is performed annually. However, to qualitatively match the temporal resolution of the proxy-based reconstructions, the values considered each year are smoothed by performing averages over sub-periods of ten years. Furthermore, since the reconstructions only provide common relative changes in moisture balance, both the data assimilated (pseudo-proxy data or real-world proxy-based reconstructions) and the model simulations time series are standardized before the comparison by subtracting their mean over the whole period and by dividing by their standard deviation. The data error taken into account is here estimated to be 0.5, whatever the data considered for the assimilation. The resulting signal-to-noise ratios of the pseudo-proxy time series range from 0.14 to 0.28 and from 0.56 to 0.61 for $\delta^{18}\text{O}$ and hydroclimate time series, respectively. These are thus typical values (e.g. Bhend et al. 2012; Steiger and Hakim 2016) and

Table 1 Geographic coordinates and period covered by the proxy time series used in this study

Id	Location	Lat	Long	Period	References
Hydroclimate proxy-based reconstructions					
1	Naivasha	−0.77	36.35	884–1993	Verschuren et al. (2000)
2	Challa	−3.32	37.7	1031–2005	Time series computed in Tierney et al. (2013) based on Verschuren et al. (2009), Tierney et al. (2011) and Wolff et al. (2011)
3	Masoko	−9.33	33.76	452–1999	Time series computed in Anchukaitis and Tierney (2013) based on Garcin et al. (2006, 2007)
4	Malawi	−10	34.22	1270–1978	Brown and Johnson (2005); Johnson and McCave (2008)
Coral $\delta^{18}\text{O}$					
5	Malindi	−3.20	40.10	1887–2002	Cole (2000)
6	Mayotte	−12.65	45.10	1865–1994	Zinke et al. (2008, 2009)
7	Houtman Abrolhos	−28.47	113.77	1794–1994	Kuhnert et al. (1999)
8	Mentawai	−0.13	98.52	1858–1998	Abram et al. (2008)
9	Seychelles	−4.62	55	1846–1995	Charles (1997)
10	La réunion	−21.03	55.25	1832–1995	Pfeiffer et al. (2004)

once again, the results do not change substantially if these error estimates vary by 20%.

3 Reconstructions using pseudo-proxy data

3.1 Pseudo-proxies derived from CESM1

In this section, we focus on the experiments performed in the idealized setup, with the assimilation of the tenth ensemble member of the CESM1-LME.

3.1.1 Skill of the reconstructions

As expected, rainfall and SSTs after data assimilation get closer to the pseudo-proxy at the grid points where data are available (Fig. S9). However, an ensemble sufficiently large is necessary to propagate the information spatially over the whole Indian Ocean and eastern equatorial region of Africa, and to spread it to the other variable (rainfall or SSTs), as described in the supplementary materials in Section S2. Since the best reconstructions are achieved using 2079 particles, and since the computational cost of these experiments is relatively modest thanks to the use of an off-line data assimilation method, this ensemble size will be used in all the experiments shown in this study. Overall, 27% (38%) of these 2079 particles have a non-negligible

weight in the experiment with data assimilation of rainfall (SSTs), with the first particle representing in average 3.5% of the total weight of the reconstructions in both experiments (Figs. S3, S4).

Figure 2 shows the coefficient of efficiency (CE) for all the grid cells of our study area, computed from rainfall and SST results, for three types of experiments performed (assimilation of rainfall, of SST and of both variables), using 2079 particles. The CE (Lorenz 1956) is a diagnostic classically used to measure the skill of reconstructions (e.g. Steiger et al. 2014). It has the advantage, compared to the root-mean-square-error (RMSE) for instance, not to be influenced by the variance of the time series on the diagnostic, and thus to provide a more meaningful comparison of the quality of the reconstruction at different grid cells. It is defined for a time series including n samples as:

$$CE = 1 - \frac{\sum_{i=1}^n (x_i - \hat{x}_i)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (2)$$

where x is the “true” time series, \bar{x} is the “true” time series mean over a climatological reference period, and \hat{x} is the reconstructed time series, ie. the output after data assimilation. CE ranges from one, corresponding to a perfect fit between the “true” and the reconstructed time series, to $-∞$. It is positive when the reconstruction is more skillful than

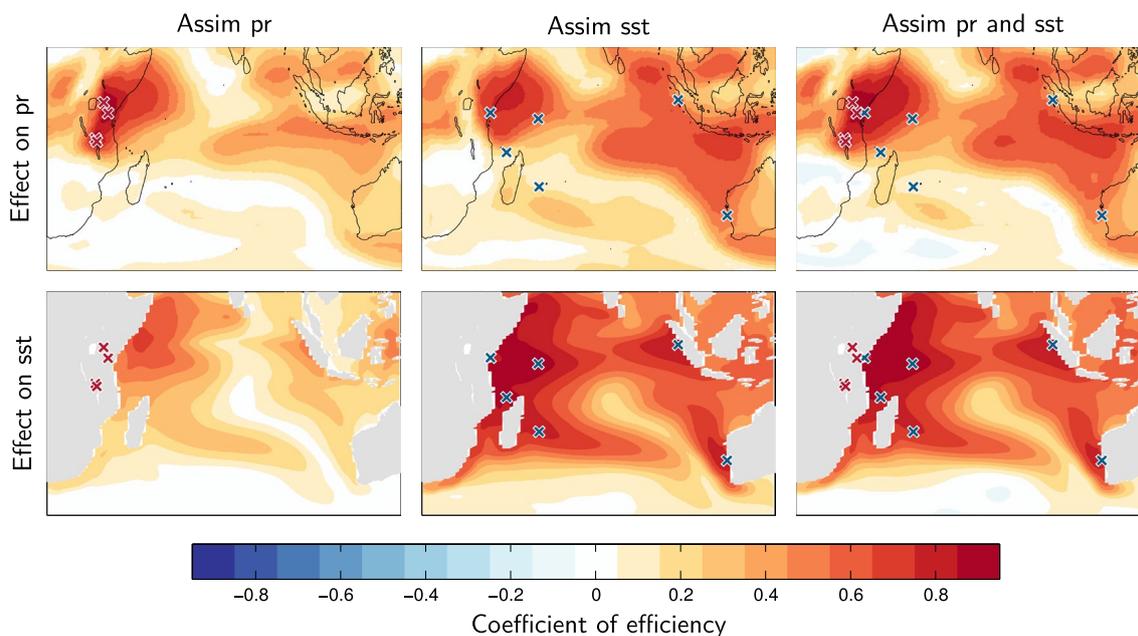


Fig. 2 Coefficients of efficiency in the reconstructions based on the CESM-LME with data assimilation of the tenth ensemble member of the CESM-LME over the whole last millennium period (850–2005 AD), using an ensemble size of 2079 particles. For each experiment [assimilation of the four rainfall results only (*left*), assimilation of

the six SST results only (*middle*), and assimilation of both (*right*)], the coefficients of efficiency for annual mean rainfall and SSTs are shown. The *crosses* indicate the locations where data are assimilated, and are *red* in case of rainfall and *blue* in case of SST

the climatological mean and negative when the opposite is true.

With an ensemble of 2079 particles, the assimilation of rainfall (first column in Fig. 2) provides a good reconstruction of rainfall as well as of SSTs over the western and the eastern equatorial Indian Ocean. Assimilating SSTs (second column in Fig. 2) delivers even better results, with an improved representation of SSTs and of rainfall all over the Indian Ocean, although the consistency between the pseudo-observations and our reconstruction is weaker for the precipitation in the central Indian Ocean. It also leads to a skillful reconstruction of rainfall in East Africa, over the south of the Horn of Africa and east of Lake Victoria. In contrast, values of CE are close to zero further south, including at the locations of the lakes Masoko and Malawi. The reconstructions obtained when assimilating both variables basically combines the positive effects of both single-variable assimilation experiments (third column in Fig. 2).

3.1.2 Time consistency of the data-assimilated model output

A sufficient number of particles is needed to obtain skillful reconstructions. Besides the model results for the actual year considered, results of other years must thus also be included to have a large ensemble and increase the range of available climate states. However, the drawback is that the time may be disrupted in the reconstruction, i.e. the reconstruction for a given year may be based on climate

states of earlier and/or later periods. Here, we analyze if the filter preferentially select the model year corresponding to data or if the time consistency has only a weak influence.

Figure 3 shows the years of the particles selected in the reconstructions for each year of experiment, grouped by periods of ten years for a better readability. In the case of the assimilation of rainfall alone (Fig. 3a), the particles chosen for each year are distributed very homogeneously over the whole last millennium. Every year appears thus interchangeable, which means that no period is characterized by a substantial modification of annual precipitation throughout the last millennium compared to others. In particular, the impact of volcanic eruptions, whose the ten most important are represented by red bars with a length proportional to the quantity of sulphate aerosols released, is not visible.

The picture is different when the Indian Ocean SSTs are assimilated (Fig. 3b). In this case, the particles selected by the filter come more often from corresponding years of simulation than from other years, as depicted by the dark diagonal line. Moreover, the impact of volcanic activity is visible. When a volcanic eruption is observed, the probability to select a particle characterized by a high aerosol load is higher, although it does not necessarily correspond to the same period in the observation. Indeed, for the particle filter, the different eruptions are interchangeable, meaning that their effect on Indian SSTs goes beyond the natural variability of SSTs over the last millennium but is relatively similar for each major eruption.

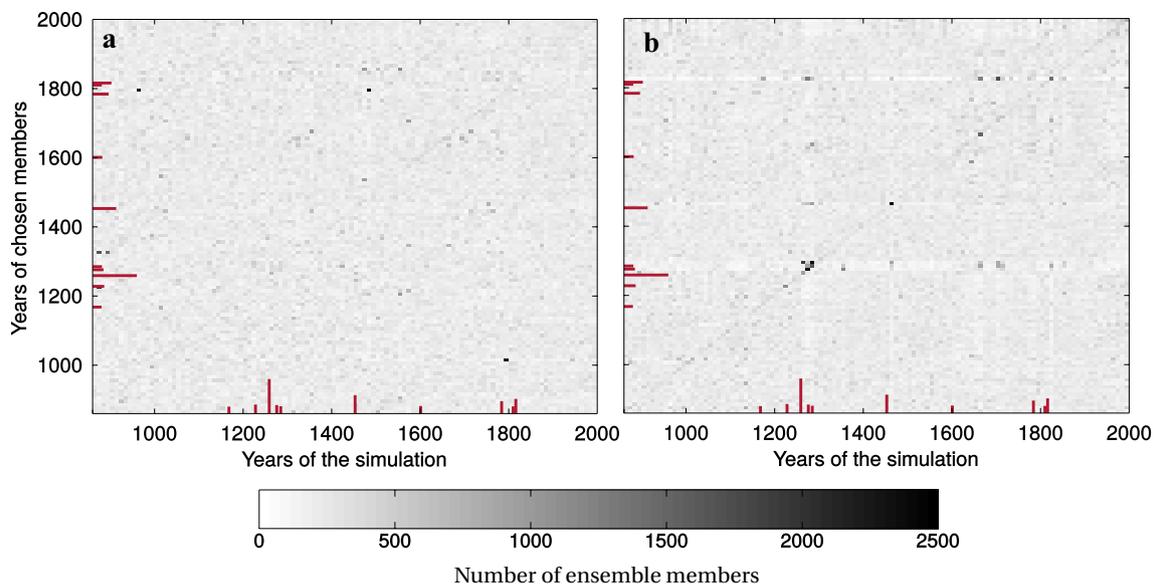


Fig. 3 Years of the particles selected in the data-assimilated model output versus reconstructed years, for data assimilation constrained by precipitation (**a**) and SSTs (**b**), using 2079 particles. The years are grouped by periods of 10 years for a better readability. The *red*

bars are the ten largest volcanic eruptions of the last millennium, with length proportional to the amount of sulphate aerosol released, based on Gao et al. (2008)

Finally, in the case of the assimilation of both rainfall and SSTs, the situation becomes very noisy, the forced variability of the SSTs appearing masked by the natural variability of the East African rainfall (Fig. S5). Overall, the reconstructions are thus characterized by a strong time inconsistency, ie. each year of the reconstructions is based on the results of numerous other years, although to a lesser extent in the reconstruction with data assimilation of SSTs.

3.1.3 Sensitivity analysis to the pseudo-proxy data assimilated

Some of the pseudo-proxies have more influence on the skill of the reconstructions at the large-scale than others. To investigate the contribution of individual records, we have performed data assimilation experiments constrained by each pseudo-proxy time series separately, using 2079 particles. The assimilation of rainfall over Naivasha and over Challa (id 1 and 2 in Fig. 1) yields very similar results (Fig. 4). It allows obtaining a virtually perfect match between the simulation with data assimilation and the pseudo-proxy assimilated over those two grid cells. In contrast, the reconstruction has only little or even no skill farther south, in the grid cells containing the lakes Masoko and Malawi (id 3 and 4 in Fig. 1). This situation is valid the other way around, when assimilating precipitation in Masoko and Malawi. This emphasizes the strong heterogeneous nature of the East African hydroclimate, as already shown in previous studies (e.g. Tierney et al. 2013; Klein et al. 2016). Assimilating rainfall over Challa and Naivasha improves the representation of Indian SSTs over the six grid cells considered, the best skill being obtained over the Seychelles (id 9 in Fig. 1) while the worse one, but still with a positive CE, in the location which is the farthest away from East Africa, ie. in the Houtman Abrolhos, located west of Australia (id 7). In contrast, rainfall over Masoko and Malawi brings nearly no skill to

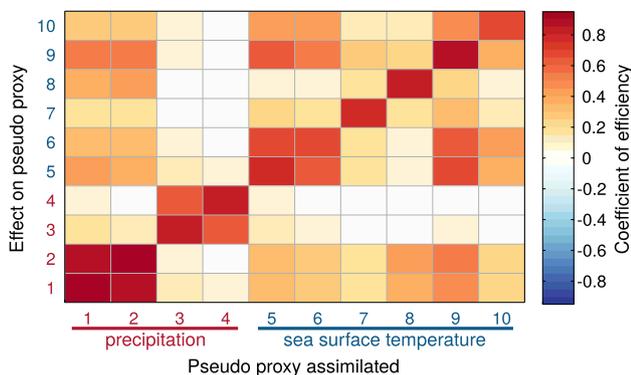


Fig. 4 Coefficient of efficiency at the ten grid cells considered in this study (see locations in Fig. 1), for each sensitivity experiment in which only one pseudo data is assimilated, using 2079 particles

reconstruct Indian Ocean SSTs, as shown by the CE close to zero in every oceanic grid cell.

The skill brought by the assimilation of individual SST records appears more homogeneous than when considering the East African rainfall: the representation of SSTs over all oceanic grid cells and of rainfall over Challa and Naivasha grid cells is improved in every sensitivity experiments, as depicted by the positive CE. In contrast, there is no experiment with data assimilation of SST that provides skillful reconstruction of rainfall over Masoko and Malawi. Despite these similarities, some differences are also visible among the six experiments. The Houtman Abrolhos and Mentawai sites (id 7 and 8) seem to be the most isolated. Indeed, the values of the CE of the reconstructions achieved by the assimilation of these results at the other sites are the lowest, and the improvements lead by the assimilation of the results of the other sites is weaker there than elsewhere. This can be easily explained by the fact that Houtman Abrolhos and Mentawai are located very far from the other records (id 7 and 8 in Fig. 1). On the contrary, out of the six oceanic sites considered, the Seychelles is characterized by the largest impact on the reconstruction skill, which can be explained by its central localization within the Indian Ocean.

In summary, the SST reconstruction performed by the assimilation of rainfall in Challa and in Naivasha are skillful in all the oceanic sites considered, although the quality of the reconstructions decreases with the distance. Symmetrically, rainfall over Challa and Naivasha are improved when assimilating Indian Ocean SSTs. In contrast, rainfall over Masoko and Malawi appears to be isolated from the rainfall farther north but also from the Indian Ocean SSTs.

3.2 Pseudo-proxies derived from other GCMs and recent observational data

This section focuses on the assimilation of time series derived from the GCMs MPI-ESM-P and GISS-E2-R, and from instrumental observations over the 20th century, which is an intermediary step before assimilating the real proxy-based reconstructions throughout the last millennium. The physics of MPI-ESM-P, GISS-E2-R and of the reality being different than the one in CESM1, the ensemble of particles may not be able to adequately reproduce the covariance present in the data. The resulting reconstructions are thus expected to be less skillful than when assimilating the tenth ensemble member of the CESM-LME. Only the reconstructions using 2079 particles, ie. the best ones (Figs. S6, S7, S8), are shown. A study of the time distribution of the particles selected for each year of the reconstruction has been done for these experiments. The results are roughly similar to the ones obtained with the assimilation of the tenth ensemble member of the CESM-LME (Fig. 3), although the effect of forcings is more marked here in the simulations

with data assimilation of SSTs. These results are available in the supplementary materials in Section S3.

Despite the similarities in the teleconnections between the East African rainfall and Indian SSTs simulated by MPI-ESM-P and CESM1 Klein et al. (2016), the reconstruction based on the assimilation of pseudo data derived from MPI-ESM-P has a much lower performance than when assimilating pseudo data derived from the tenth ensemble member of the CESM-LME (Fig. 5). The assimilation of precipitation alone leads only to a very local improvement in the representation of the East African rainfall. The skill of the reconstruction rapidly decreases away from the data locations, to reach values of CE close to zero in most of the regions displayed in Fig. 5, or even negative, for instance in the south of the Horn of Africa. Furthermore, the assimilation of precipitation has virtually no positive effect on Indian SSTs. In contrast, assimilating SSTs over the six grid cells considered allows having a good reconstruction of the SSTs over all the Indian Ocean, but does not improve the representation of rainfall, with often negative CE values. Finally, assimilating both the East African rainfall and SSTs basically leads to the addition of the effects of the single-variable assimilations.

The situation is relatively similar when considering the simulation with data assimilation of the results derived from the model GISS-E2-R (Fig. 6), characterized by a very weak relationship between the East African rainfall and Indian SSTs (Klein et al. 2016). Using the SSTs from

six grid cells only as a constraint for the assimilation is sufficient to have a good reconstruction of SSTs over the whole Indian Ocean, but this does not lead to any improvement in the representation of rainfall, some regions being even characterized by reconstructions worse than a climatological mean. Once again, the assimilation of East African rainfall does not lead to a propagation of the information away from the locations where the constraint is applied. There is, however, a difference with the results of the model MPI-ESM-P. When assimilating pseudo-proxies of both rainfall and SSTs, the quality of the SST reconstruction decreases compared to the case in which only SST is assimilated. This illustrates the inconsistency of the representation of the link between rainfall and SSTs in the models CESM1 and GISS-E2-R.

GPCC-v7, the data set selected for rainfall, only covers the land-surface. This prevents any assessment of the reconstruction skill of precipitation over the Indian Ocean. Nevertheless, we can see in Fig. 7 that the results obtained when assimilating instrumental observations over the last century are very similar to those when assimilating GISS-E2-R time series. Indeed, only taking into account the four rainfall results does not improve the representation of precipitation elsewhere than where the constraint is applied, while, in contrast, the six assimilated time series of SSTs appear to be enough to reconstruct SSTs over the whole of the Indian Ocean. However, the

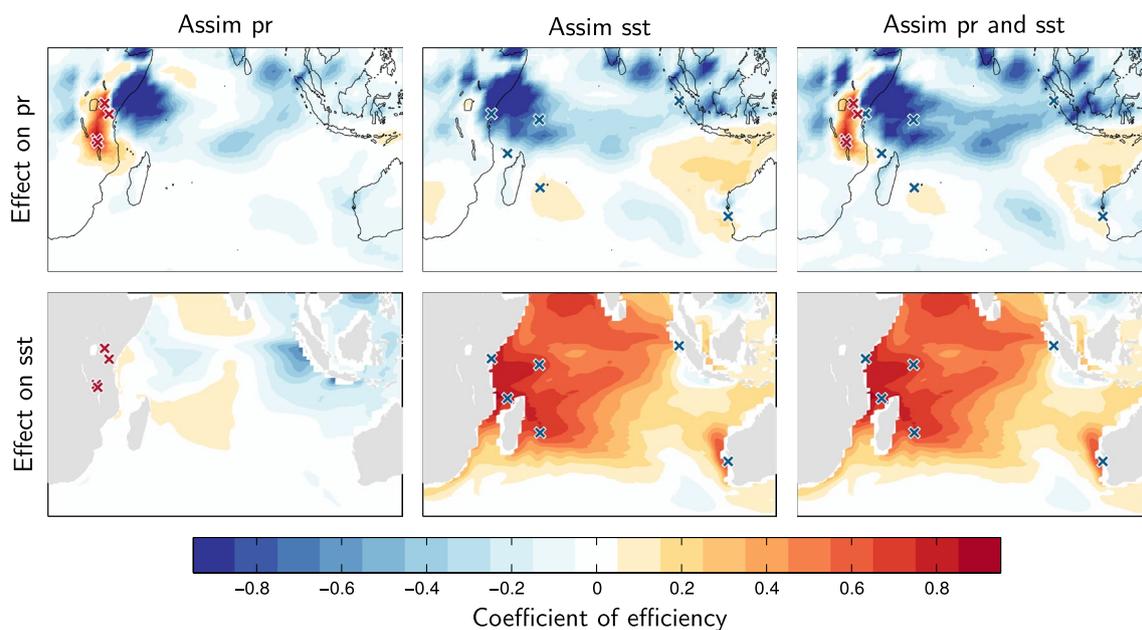


Fig. 5 Coefficients of efficiency in the reconstructions based on the CESM-LME with data assimilation of pseudo-proxies derived from the last millennium simulation performed by MPI-ESM-P over the whole last millennium period (850–2005 AD), using 2079 particles. For each experiment [assimilation of the four rainfall results only

(left), assimilation of the six SST results only (middle), and assimilation of both (right)], the coefficients of efficiency for annual mean rainfall and SSTs are shown. The crosses indicate the locations where data are assimilated, and are red in case of rainfall and blue in case of SST

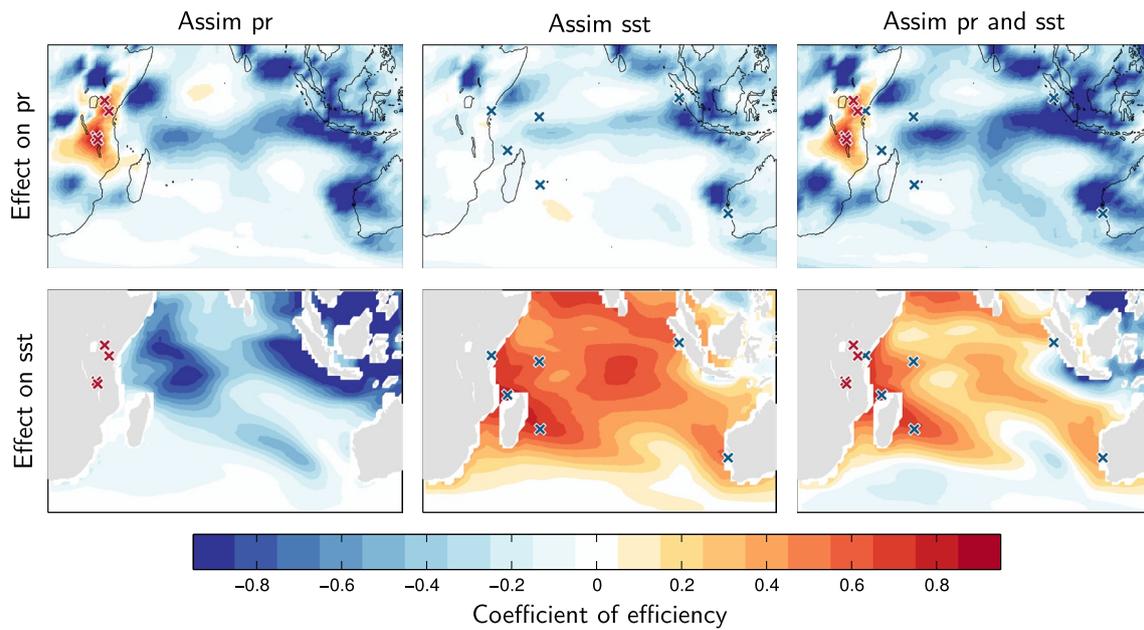


Fig. 6 Coefficients of efficiency in the reconstructions based on the CESM-LME with data assimilation of pseudo-proxies derived from the last millennium simulation performed by GISS-E2-R over the whole last millennium period (850–2005 AD), using 2079 particles. For each experiment [assimilation of the four rainfall results only

(left), assimilation of the six SST results only (middle), and assimilation of both (right)], the coefficients of efficiency for annual mean rainfall and SSTs are shown. The crosses indicate the locations where data are assimilated, and are red in case of rainfall and blue in case of SST

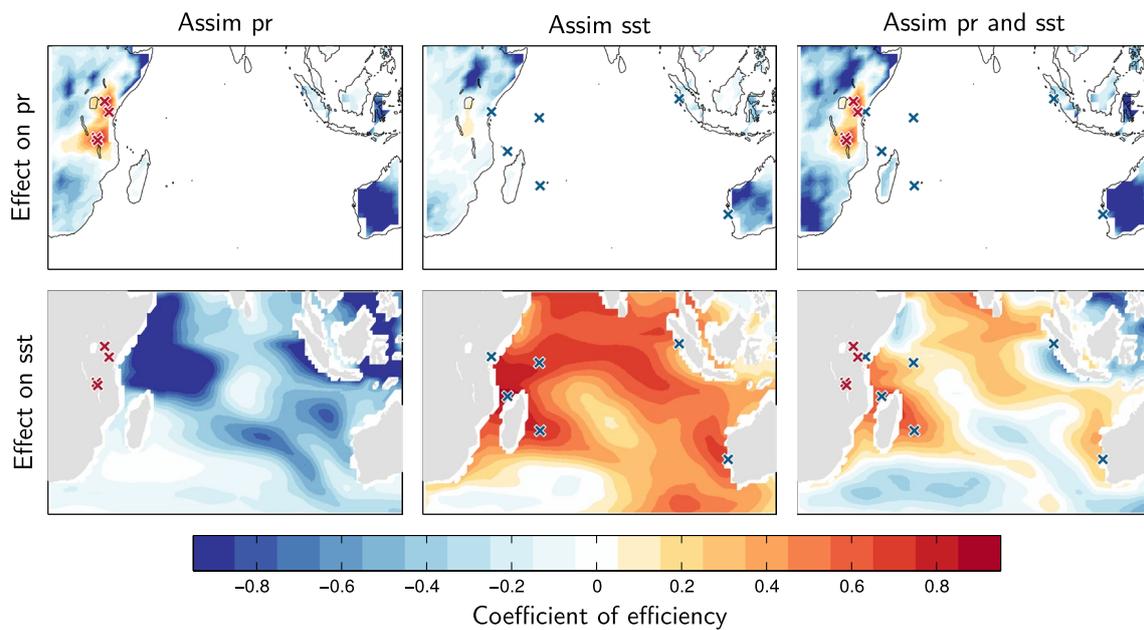


Fig. 7 Coefficients of efficiency in the reconstructions based on the CESM-LME with data assimilation of the time series derived from the data sets ERSST-v4 (for SSTs) and GPCC-v7 (for rainfall) over the period 1900–2005 AD, using 2079 particles. For each experiment [assimilation of the four rainfall results only (left), assimilation of

the six SST results only (middle), and assimilation of both (right)], the coefficients of efficiency for annual mean rainfall and SSTs are shown. The crosses indicate the locations where data are assimilated, and are red in case of rainfall only covers the land-surface

quality of this reconstruction is substantially decreased when also assimilating rainfall time series.

Compared to the perfect model experiments, assimilating rainfall time series from other sources than CESM1 no longer provides valuable reconstructions of rainfall and of SSTs over the Indian Ocean. In fact, the reconstructions are often worse than a climatological mean, characterizing contradictions in the physics and the dynamics represented in the different data sets. Furthermore, assimilating both rainfall and SST tends to worsen the reconstructions obtained when assimilating the SSTs alone, illustrating the inconsistent covariances of rainfall and SSTs compared to CESM1. In contrast, it is still possible to obtain a skillful reconstruction of the Indian SSTs, based on six time series only.

4 Reconstructions using real proxy-based reconstructions

Before using available proxy-based reconstructions over the last millennium, the next section investigates the impact of assimilating the reconstructed variables as defined in Sect. 2.4, ie. the East African hydroclimate and Indian Ocean coral $\delta^{18}\text{O}$, instead of rainfall and SSTs, respectively.

4.1 From the simulated towards the reconstructed variables

In order to focus on the loss in reconstruction skill due only to the assimilation of different variables, the results from the tenth ensemble member of the CESM-LME are used for the generation of the pseudo-proxy data, such as in Sect. 3.1. Decadal mean precipitation minus evaporation is the model version of what is reconstructed from hydroclimate-related records. Assimilating this variable instead of annual mean precipitation does not allow obtaining skillful reconstructions of annual rainfall or SST (not shown), which is not surprising giving the strong interannual variability of the East African rainfall that is smoothed at the decadal time scale. However, the resulting decadal reconstructions of rainfall and of Indian Ocean SSTs have some skill (Fig. 8a, b), with spatial pattern of CE similar to the ones found in the annual reconstructions achieved with the assimilation of annual mean rainfall. Nevertheless, the skill is somewhat diminished, with a mean CE (averaged over the region shown in Fig. 8) of 0.12 instead of 0.19 for rainfall, and of 0.11 instead of 0.19 for SSTs (Fig. 8a and b to be compared to Fig. 2). The assimilation of annual mean coral $\delta^{18}\text{O}$ instead of annual mean SST does not lead to any substantial decline in the reconstruction skill, the mean CE even being unchanged with values of 0.30 and of 0.48 for precipitation and for SST, respectively (Fig. 8c, d).

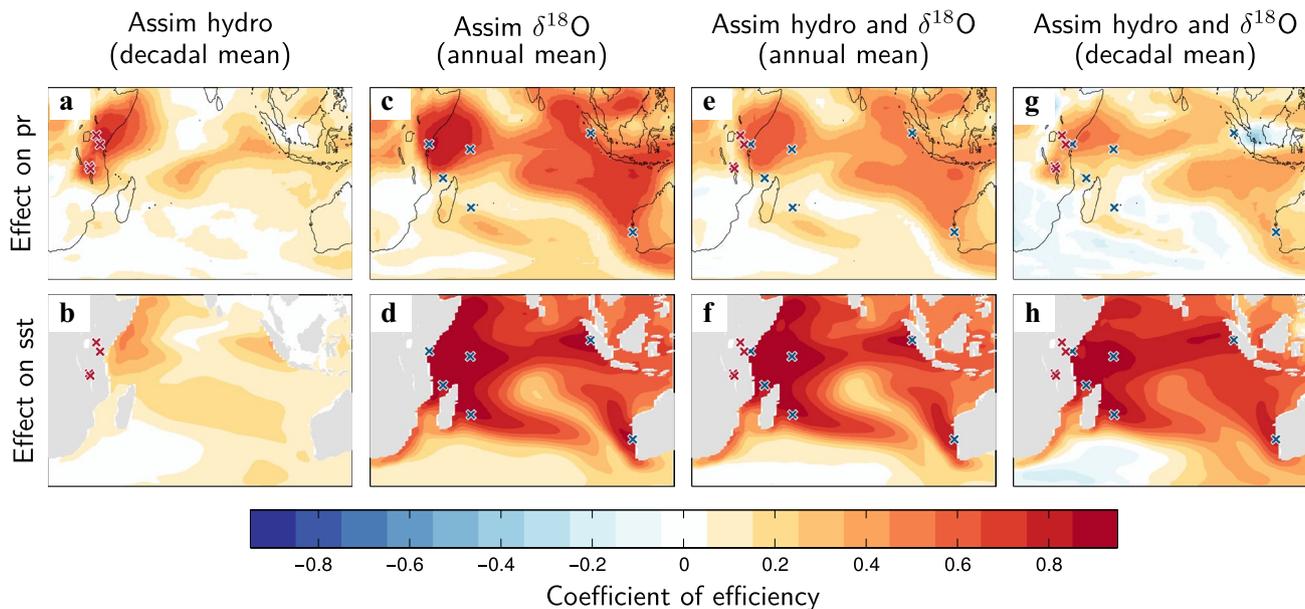


Fig. 8 Coefficients of efficiency in the reconstructions based on the CESM-LME with data assimilation of pseudo-proxies derived from the tenth ensemble member of the CESM-LME over the whole last millennium period (850–2005 AD), using 2079 particles. From the left to the right, the panels show the coefficients of efficiency based on rainfall and SSTs from the experiments with data assimilation of

the four hydroclimate results only, of the six coral $\delta^{18}\text{O}$ results only, and of both for the last two columns. The temporal frequency on which the coefficients of efficiency are computed is shown in brackets. The crosses indicate the locations where data are assimilated, and are red in case of hydroclimate and blue in case of coral $\delta^{18}\text{O}$

As expected given the previous results, there is a loss of skill in the annual reconstruction of rainfall when assimilating together coral $\delta^{18}\text{O}$ and hydroclimate instead of SST and precipitation, with a mean CE of 0.21 (Fig. 8e) instead of 0.27. This is due to the lack of information on annual rainfall brought by the assimilation of hydroclimate variables. Actually, also considering hydroclimate in addition to coral $\delta^{18}\text{O}$ for the assimilation tends to decrease uniformly the CE of the reconstruction of rainfall compared to the ones obtained with the assimilation of coral $\delta^{18}\text{O}$ alone. By contrast, the skill of the reconstruction of annual SST is very similar when also assimilating hydroclimate variables together with coral $\delta^{18}\text{O}$ (Fig. 8f), which was also the case when assimilating both annual rainfall and annual SST instead of annual SST alone. Finally, assimilating model-versions of both reconstructed variables also allows having a good reconstruction of decadal mean SSTs (Fig. 8h), with a mean CE of 0.24, and a relatively good reconstruction of decadal mean rainfall (Fig. 8g), with a mean CE of 0.12. In the latter reconstructions, however, the CE over the East African lakes are not as high as in the experiment with data assimilation of hydroclimate only (Fig. 8a).

Overall, no loss of skill in reconstructing rainfall and SSTs is observed when assimilating coral $\delta^{18}\text{O}$ instead of SST. There is a difference in the reconstructions of annual rainfall and SSTs when assimilating hydroclimate instead of rainfall, but which lies on the difference in the time resolution of the records.

4.2 Assimilation of proxy-based reconstructions

In this section, we first check if the reconstruction based on data assimilation is in agreement with the data that is assimilated in order to identify potential incompatibilities between the proxy records and model results, before looking at spatial reconstructions. We only consider the experiments with the assimilation of hydroclimate and coral $\delta^{18}\text{O}$ separately,

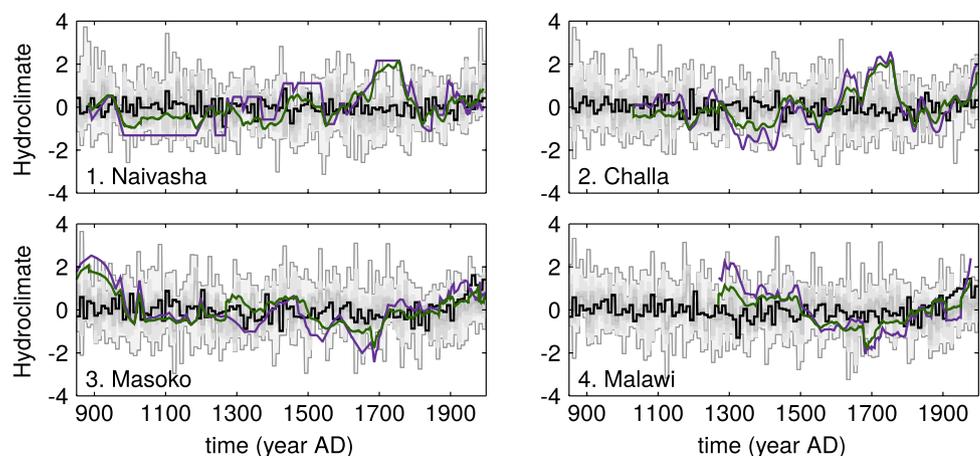
given the negative impact on the reconstructions brought by the assimilation of the combination of those two variables illustrated in Sect. 3.2. The large-scale skill of the reconstructions of rainfall and SSTs is assessed hereafter from the difference to recent observational data sets. As before, we show the reconstructions using 2079 particles.

Despite the discrepancies between Naivasha and Challa hydroclimate proxy-based reconstructions over the first four centuries of the last millennium, both time series agree from 1400 AD onward on relatively dry conditions followed by a wetting trend peaking between about 1700 and 1750 AD (violet curves in Fig. 9-1, 2). This wet period is followed in both hydroclimate reconstructions by a short dry period before some smaller scale fluctuations. The hydroclimate time series inferred from the Masoko and Malawi records are also characterized by some similar long term changes over the last millennium, but that differ from the Naivasha/Challa pattern (violet curves in Fig. 9-3, 4). Indeed, they agree on a drying trend until about 1700, that is followed by a gradual increase in humidity towards the present time.

Except during the last two centuries at Masoko and Malawi where the model mean shows a trend towards a wetter climate that is consistent with the reconstructed signal, there is no common signal between the proxy-based reconstructed and the simulated hydroclimate without data assimilation. With data assimilation, however, the model time series get much closer to the four reconstructions (green curves to be compared to violet curves in Fig. 9), leading to a significant reduction of the RMSE (Table 2).

Out of the six coral $\delta^{18}\text{O}$ proxy-based reconstructions, only the one from Houtman Abrolhos goes back to the early 1800s. It shows that this period was characterized by high values of coral $\delta^{18}\text{O}$ relative to the last two centuries (violet curve in Fig. 10-7). This is associated with low temperatures, likely due to the multiple tropical volcanic eruptions that occurred during this period (Tierney et al. 2015), as can be seen in Fig. 3. The values of $\delta^{18}\text{O}$ then gradually decreases

Fig. 9 Evolution of lake-based hydroclimate throughout the last millennium at the four continental sites considered (see Fig. 1), according to the first nine ensemble members of the CESM-LME (black curve is the mean surrounded by the ensemble range in grey), the proxy-based reconstructions described in Sect. 2.4 (in violet), and the data-assimilated model output (in green). The axes are oriented such that wetter conditions point upward



to reach their lowest level at the end of the 20th century, following the warming of the Indian Ocean. This decreasing trend is observed in all reconstructions, although it is less marked in Mentawai. The ensemble mean of the simulations without data assimilation (black curves in Fig. 10) usually agrees with these reconstructed downward trends, although the simulated trend is less pronounced. Both reconstructed and simulated changes in coral $\delta^{18}\text{O}$ show a relatively high variability at the interannual time scales, but as can be expected, the temporal changes do not coincide.

Data assimilation improves the consistency between reconstructed and simulated time series (green curves in

Fig. 10), as shown by the smaller values of RMSE after assimilation (Table 3). The biggest improvement is achieved at Malindi, where data assimilation allows matching very well the observed interannual changes. Data assimilation also makes the simulated trends over the last two centuries closer to the reconstructed ones, although the slope is still too weak in La Réunion.

Data assimilation provides thus skillful local reconstructions compared to the time series that are assimilated, which is the minimum requirement. The goal here is to obtain spatial reconstructions of precipitation and SSTs, whose skill will depend on the propagation of the information contained

Table 2 RMSE between simulated and reconstructed time series without (third column) and with (fourth column) data assimilation, at each East African site considered

Id	Site	RMSE: data vs model output (ensemble mean)	RMSE: data vs data-assimilated model output	Difference (in %)
1	Naivasha	1.08	0.62	-43.05
2	Challa	1.10	0.57	-57.08
3	Masoko	0.62	0.47	-47.40
4	Malawi	0.43	0.58	-57.74

Fig. 10 Evolution of coral $\delta^{18}\text{O}$ throughout the last two centuries at the six oceanic sites considered (see Fig. 1), according to the first nine ensemble members of the CESM-LME (black curve is the mean surrounded by the ensemble range in grey), the proxy-based reconstructions described in Sect. 2.4 (in violet), and the data-assimilated model output (in green). Results are shown as anomalies with respect to the whole period covered

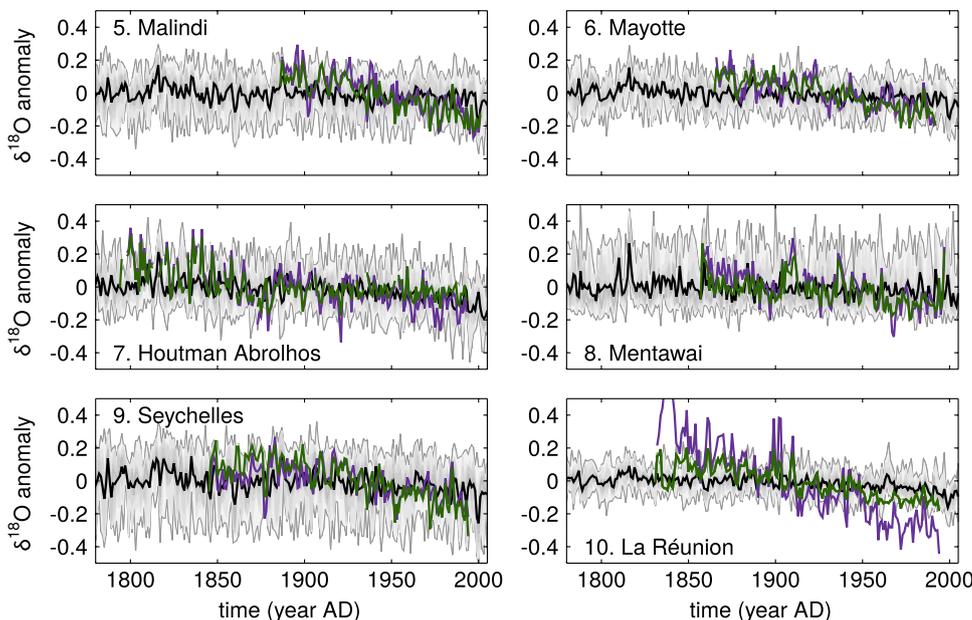


Table 3 RMSE between simulated and reconstructed time series without (third column) and with (fourth column) data assimilation, at each oceanic site considered

Id	Site	RMSE: data vs model output (ensemble mean)	RMSE: data vs data-assimilated model output	Difference (in %)
5	Malindi	0.123	0.052	-58.00
6	Mayotte	0.096	0.055	-42.51
7	Houtman Abrolhos	0.124	0.053	-56.92
8	Mentawai	0.120	0.054	-54.89
9	Seychelles	0.106	0.071	-32.78
10	La réunion	0.222	0.168	-24.36

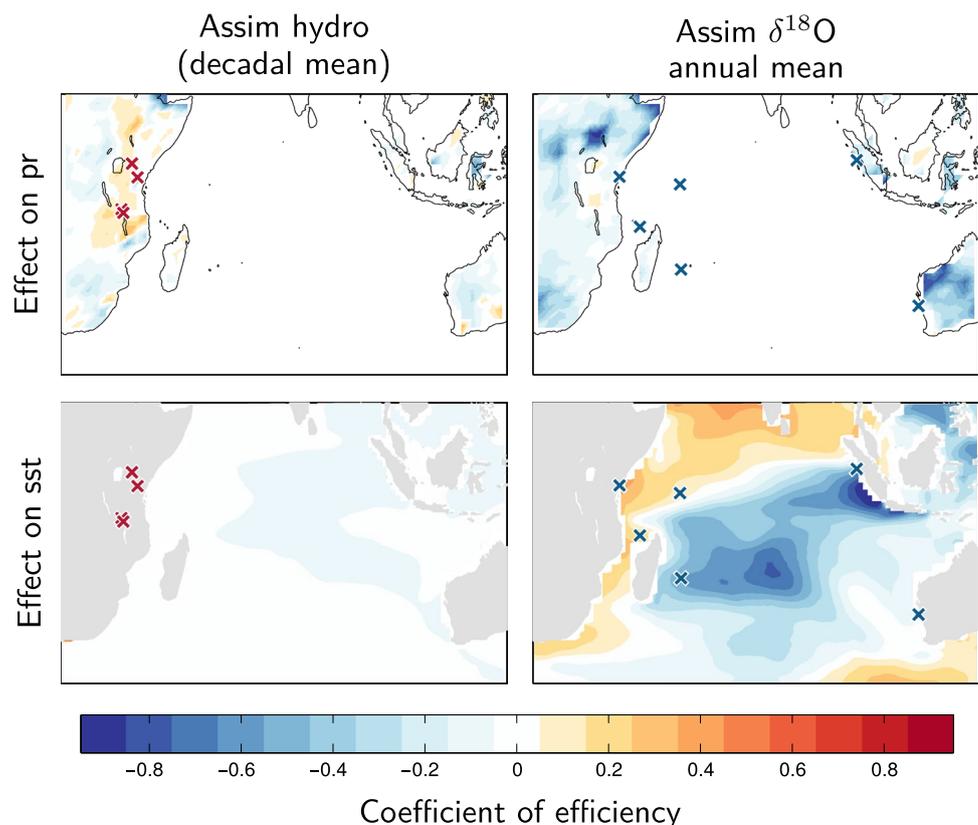
in the ten proxy-based reconstructions considered. There is no spatial proxy-based reconstruction available over the full period investigated that would allow assessing the skill of our reconstructions. Hence, it is achieved here by comparing our results with the observational data sets ERSST-v4 and GPCC-v7, over the 20th century (Fig. 11).

The skill of the reconstruction of rainfall performed by assimilating hydroclimate is limited, with only a thin strip of slightly positive CE observed in East Africa, containing the lakes Masoko and Malawi (Fig. 11). However, these positive CE are actually due to the common recent trend between simulated and observed hydroclimate rather than to the data assimilation process. The data set for rainfall covers only the land-surface, which prevents the estimation of the reconstruction skill over the Indian Ocean. Nevertheless, it is likely close to zero or even negative as deduced from the experiments performed in an idealized framework. As expected again from the pseudo-proxy experiments, the assimilation of hydroclimate does not lead to any skill in reconstructing SSTs. The skill of the reconstruction of Indian Ocean SSTs is, to some extent, higher when assimilating coral $\delta^{18}\text{O}$, with positive CE in the north of the ocean as well as along the east coast of Africa, an area that includes three sites where data is assimilated, ie. Malindi (id 5), Mayotte (id 6) and the Seychelles (id 9). However, the latter experiment does

not allow skillful reconstructions of rainfall over Africa, and the CE of the reconstructions of SSTs are negative in a large part of the Central and Eastern Indian Ocean.

The lack of skill even at some of the record sites may seem surprising at first sight. Indeed, the data assimilation process is able to bring the simulated hydroclimate and coral $\delta^{18}\text{O}$ close to the proxy time series (Figs. 9, 10), as shown by the decreased RMSE. Furthermore, measured coral $\delta^{18}\text{O}$ and hydroclimate-related proxy records should be, by construction, closely related to local SST and rainfall, respectively. However, the correlation coefficients between observed decadal rainfall and proxy-based reconstructed hydroclimate at the four East African sites are weak (not shown). This is probably due to the quite indirect nature of the link between the hydroclimate-related reconstructed and simulated variables. Indeed, the model variable is smoothed precipitation minus evaporation while the reconstructed hydroclimate variables depend on numerous other elements such as the inflow and outflow from and to rivers, surface runoff and ground water, the evaporation from the lakes, as well as interactions with the aquifer (e.g. Becht and Harper 2002), that may obscure the signal brought by actual precipitation minus evaporation. Actually, the skillful reconstructions of rainfall in a few locations of East Africa is attributable to a common trend that exists between the observed and the simulated rainfall

Fig. 11 Coefficients of efficiency computed from the difference between the CESM-LME with data assimilation of the real-world proxy-based reconstructions and the data sets ERSST-v4 (for SSTs) and GPCC-v7 (for rainfall, only covering the land-surface) over the period 1900–2005 AD, using 2079 particles. The two left panels show the coefficients of efficiency based on the reconstructions of decadal mean rainfall and SSTs from the experiments with data assimilation of the four hydroclimate results only, and the two right panels show the coefficients of efficiency based on the reconstructions of annual mean rainfall and SSTs from the experiments with data assimilation of the six coral $\delta^{18}\text{O}$ time series only. The crosses indicate the locations where data are assimilated, and are red in case of hydroclimate and blue in case of coral $\delta^{18}\text{O}$



time series (not shown), rather than to the data assimilation process.

The negative CE at some of the oceanic records cannot be explained by a low correlation between SSTs and coral $\delta^{18}\text{O}$. Indeed, although other environmental variables also play a role, the relationship between those two variables is robust and well-known (e.g. Brown et al. 2006; Stevenson et al. 2013). However, the coefficient binding the two variables in Eq. 1 may vary significantly from site to site (Evans et al. 2000). In our data assimilation experiment, we have applied at each location the value of $-0.22 \text{ } ^\circ/\text{oo } ^\circ\text{C}^{-1}$, as proposed in Thompson et al. (2011) as a mean estimate. Ideally, the forward model should be calibrated at each individual site. However, this requires to take into account changes in SSS, while there is no reliable instrumental data set of salinity for the whole period considered. The changes in SST dominate on changes in SSS to produce the coral $\delta^{18}\text{O}$. Hence, the impact of a local calibration is investigated here by estimating coral $\delta^{18}\text{O}$ from model results using a linear regression based on local observed SSTs from the instrumental data set ERSST-v4, instead of the forward model proposed in Thompson et al. (2011). In this case, the error considered in the data assimilation process is different at the six records sites depending on the regression model. This was not the case in the experiments using the forward model where only the error of the measurement is considered, while the error related to the relationship between coral $\delta^{18}\text{O}$ on the one hand and SST and SSS on the other was neglected.

This approach allows having skillful reconstructions of SSTs over the Indian Ocean (Fig. 12). However, as might be expected, the skill is not as high as when assimilating

instrumental observations (Fig. 7), with a mean CE over the Indian Ocean of 0.10 instead of 0.32. Also, as in the previous experiments, the representation of precipitation on land is not improved.

5 Discussion and conclusions

The focus of this paper is to improve reconstructions of the East African rainfall and Indian Ocean SSTs based on their covariance. This is achieved by means of an off-line method of data assimilation, using the recently available ensemble of last millennium simulations performed with the model CESM1. Hydroclimate-related records at four East African sites, Lake Naivasha, Lake Challa, Lake Malawi and Lake Masoko, as well as SSTs-related records at six oceanic sites spread over the Indian Ocean, are selected as constraints for the assimilation.

The assimilation of pseudo-proxy data derived from the tenth ensemble member of the CESM-LME allows having skillful reconstructions of the Indian SSTs based on the East African rainfall, and vice versa. The reconstructions obtained when assimilating both variables basically combine the benefits obtained in each single-variable assimilation experiment. The quality of the reconstructions is however strongly dependent on the number of particles considered. Only using 9 particles is usually not sufficient to spread spatially the information contained in the record sites selected. Hence, most reconstructions based on data assimilation showed in this paper use more particles than actually available for each year of assimilation, by also selecting particles belonging

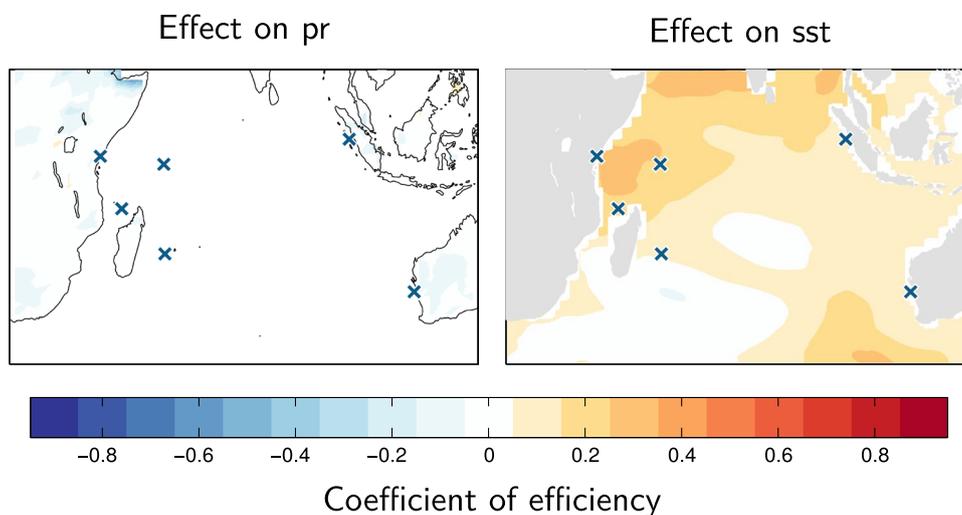


Fig. 12 Coefficients of efficiency computed from the difference between rainfall (a) and SSTs (b) of CESM-LME with data assimilation of measured coral $\delta^{18}\text{O}$ and the data sets GPCP-v7 (for rainfall, only covering the land-surface) and ERSST-v4 (for SSTs). The period covered is 1900–2005 AD. The ensemble size is 560. During the data

assimilation process, the model coral $\delta^{18}\text{O}$ is induced from the parameters of the linear regression computed between coral $\delta^{18}\text{O}$ and local observed SSTs (data set ERSST-v4). The CE are computed from annual mean results. The crosses indicate the locations where data are assimilated

to other years than the one considered. This increases the range of possible climate states and thus allows improving the agreement between the data-assimilated model output and the data. Beyond 99 particles, however, the improvement brought by further increasing the size of the ensemble becomes modest in our experimental set up.

The four East African records selected for the assimilation do not contribute equally in providing skillful reconstructions of the East African rainfall and of Indian SSTs. Indeed, precipitation over the lakes Masoko and Malawi appears to be isolated from the large-scale since neither the assimilation of Indian SSTs nor the assimilation of rainfall further north in East Africa provides skillful reconstruction of rainfall over these sites, and conversely. Hence, Challa and Naivasha are much better candidates to reconstruct large-scale patterns of climate than Masoko and Malawi, according to the physics of CESM1. This heterogeneity is consistent with the very different patterns of correlations between the East African rainfall and Indian Ocean SSTs obtained when considering the rainfall averaged, on the one hand, over the region including Challa and Naivasha, and on the other hand, over the region containing Masoko and Malawi (Klein et al. 2016). This may appear contradictory with the recent study of Nicholson (2014), which shows that changes in the East African rainfall over the last two centuries are relatively coherent over the region. The latter study is based on a principal component analysis, the objective of such tool being to highlight common patterns of variability. Here, we suggest that the common signals in the East African rainfall found in Nicholson (2014) do not explain a fraction of the total variance that is large enough to obtain skillful reconstruction of rainfall over an area of East Africa based on the rainfall over another area.

Unfortunately, reconstructing Indian SSTs based on the East African rainfall or the East African rainfall based on Indian ocean SSTs is only possible in idealized conditions when the model physics is supposed to be perfect. If the pseudo-proxies are derived from another model, this ability is lost because of the differences in model physics. This is also the case if real observations are used. Furthermore, assimilating precipitation no more improves regional rainfall, but only local precipitation. Taking into account a framework closer to reality, in which the model covariances are different from the real-world climate, thus drastically reduces the skill of the reconstructions. This clearly highlights the fact that using the same model for producing the pseudo-proxy time series and the ensemble of simulations is a strong simplification of reality, that likely artificially inflates the skill of the data assimilation approach (e.g., as shown in Dee et al. 2016). Hence, the results derived from such a framework should be interpreted with caution.

Still, the experiments using pseudo-proxies derived from another model than the one that has provided the

ensemble of simulations or from recent observations show that a reconstruction of the whole Indian Ocean SSTs can be obtained by only assimilating the SST results of six sites spread over the Indian Ocean. However, this does not constrain enough the evolution of the East African rainfall. The representation of the covariance of the Indian Ocean SSTs seems thus to be relatively consistent among models and between models and recent instrumental data sets, unlike the one between the East African rainfall at different locations and of the teleconnection between the two variables. In addition to the different patterns associated with the modes of natural variability in models and observations, this may be related to the differential impact of the volcanic events and, to a lesser extent, of the anthropogenic forcings, on the variables assimilated. In the reconstructions based on the assimilation of SST results, most large eruptions are followed by similar changes in the pattern of SSTs that go beyond the natural variability. In contrast, if there is an effect of the eruptions on the East African hydroclimate, it is not visible in our experiments, meaning that the effect of the volcanism on the East African hydroclimate does not overwhelm the internal variability.

Assimilating real proxy-based reconstructions is not straightforward, given that the physical quantity that is reconstructed from proxy records is not directly simulated by the climate model. Here, the model variables are processed to match the reconstructed variables as closely as possible. For oceanic records, it is done through the use of a linear bivariate model that computes the model coral $\delta^{18}\text{O}$ based on simulated SSS and SSTs (Thompson et al. 2011). This is harder for the East African hydroclimate proxy records, since the variables reconstructed as well as the lakes from which they originate are different. However, all proxy-based reconstructions can be qualitatively interpreted as smoothed versions of local-moisture balance, that is translated in the model by smoothed precipitation minus evaporation. Within the ideal framework considering the assimilation of the results of the tenth ensemble member of the CESM-LME, assimilating these model versions of the reconstructed variables has only little impact on the quality of the reconstructions of rainfall and of SSTs, the overall spatial patterns of the reconstructions skill being similar than when assimilation rainfall and SSTs.

The situation is different with real proxy-based reconstructions. The assimilation technically works: compared to the model ensemble mean, the model results with data assimilation are characterized by strongly decreased RMSE with the reconstructed or measured time series at all record sites selected. However, the skill of the resulting reconstructions, as derived from recent instrumental data set, is limited. A first issue for the experiments with data assimilation of hydroclimate-related records is the very indirect link between the proxy-based reconstructed variables and

the model version of these variables. Furthermore, the spatial representativeness of the proxy-based reconstructions is assumed to match the surface covered by the grid cells that contain the proxy sites. This is clearly a strong assumption, given the very high spatial variability associated with rainfall, which can further alter the skill of the reconstructions. However, even if it was possible to directly assimilate the reconstructed variable using a precise forward model for each lake, and even if the spatial representativeness of the proxy-based reconstructions was perfectly consistent with the grid cell of the model, assimilating hydroclimate is not expected to bring skillful spatial reconstructions of SSTs or rainfall, as deduced from the results of the pseudo-proxy experiments.

Assimilating coral $\delta^{18}\text{O}$ is a priori more straightforward. A simple and robust forward proxy model can be used to estimate this variable from model results, which allows a meaningful comparison with proxy records. However, using the forward model with standard (spatially constant) coefficients provides SST reconstructions with only modest skill. The issue here is that the relationship between coral $\delta^{18}\text{O}$ and SSTs varies between the individual sites. When this spatial variability is taken into account, a skillful reconstruction of SSTs over most of the Indian Ocean can be achieved with the assimilation of only six records, but this does not lead to a better representation of the East African rainfall compared to the climatological mean.

Two critical points that tend to decrease the capacity to obtain skillful reconstructions have thus been highlighted in this study. First, the biases in the physics of the model do not allow to take advantage of the correlation between SSTs and precipitation. By construction, the data assimilation method employed in this study provides reconstructions based on fields consistent with the physics of the model used, i.e. CESM1. However, all climate models show very different teleconnections between the East African rainfall and tropical SSTs. Although CESM1 is one of the closest to the observations regarding this diagnostic over the recent past, the differences between the observed and simulated teleconnection is still large, as demonstrated by the much poorer reconstructions achieved when assimilating recent observations compared to the assimilation of the tenth ensemble member of the CESM-LME. The results of this study highlight the importance of the model errors, showing the strong relevance to continue improving the GCMs. This appears particularly critical for precipitation, the simulation of SST covariance being already good enough to provide useful information for reconstructing this variable. The second critical point is the interpretation of the proxy-records in order to have a coherent model-data comparison. The variables that are included in climate models are often different from the ones reconstructed from proxy records, and the link between simulated variables and proxy may

vary significantly between locations. Hence, the effort to reconcile both sources of information, for instance through forward modelling, has to be pursued.

Acknowledgements We thank the two anonymous reviewers for their careful reading and the constructive comments that helped improving the manuscript. This work is supported by the BRAIN-be programme of the Belgian Federal Science Policy Office (BelSPO) through project BR/121/A2 “Patterns and mechanisms of climate extremes in East Africa” (PAMEXEA). We acknowledge the World Climate Research Programme’s Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modelling groups for producing and making available their model output. For CMIP the US Department of Energy’s Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. Hugues Goosse is research director with the FRS/FNRS, Belgium.

References

- Abram NJ, Gagan MK, Cole JE, Hantoro WS, Mudelsee M (2008) Recent intensification of tropical climate variability in the Indian Ocean. *Nat Geosci* 1(12):849–853. doi:10.1038/ngeo357
- Anchukaitis KJ, Tierney JE (2013) Identifying coherent spatiotemporal modes in time-uncertain proxy paleoclimate records. *Clim Dyn* 41(5):1291–1306. doi:10.1007/s00382-012-1483-0. <http://www.springerlink.com/index/10.1007/s00382-012-1483-0>
- Annan JD, Hargreaves JC (2012) Identification of climatic state with limited proxy data. *Clim Past* 8(4):1141–1151. doi:10.5194/cp-8-1141-2012
- Balmaseda M, Hernandez F, Storto a, Palmer MD, Alves O, Shi L, Smith GC, Toyoda T, Valdivieso M, Barnier B, Behringer D, Boyer T, Chang YS, Chepurin G, Ferry N, Forget G, Fujii Y, Good S, Guinehut S, Haines K, Ishikawa Y, Keeley S, Köhl a, Lee T, Martin MJ, Masina S, Masuda S, Meyssignac B, Mogensen K, Parent L, Peterson K, Tang YM, Yin Y, Vernieres G, Wang X, Waters J, Wedd R, Wang O, Xue Y, Chevallier M, Lemieux JF, Dupont F, Kuragano T, Kamachi M, Awaji T, Caltabiano a, Wilmer-Becker K, Gaillard F (2015) The Ocean Reanalyses Intercomparison Project (ORA-IP). *J Oper Oceanogr* 8(sup1):s80–s97. doi:10.1080/1755876X.2015.1022329. <http://www.tandfonline.com/doi/full/10.1080/1755876X.2015.1022329>
- Barkmeijer J, Iversen T, Palmer TN (2003) Forcing singular vectors and other sensitive model structures. *Q J R Meteorol Soc* 129(592):2401–2423. doi:10.1256/qj.02.126
- Becht R, Harper DM (2002) Towards an understanding of human impact upon the hydrology of Lake Naivasha, Kenya. *Hydrobiologia* 488:1–11
- Berger A, Loutre MF, Tricot C (1993) Insolation and Earth’s orbital periods. *J Geophys Res* 98(D6):10,341–10,362. doi:10.1029/93JD00222
- Bhend J, Franke J, Folini D, Wild M, Brönnimann S (2012) An ensemble-based approach to climate reconstructions. *Clim Past* 8(3):963–976. doi:10.5194/cp-8-963-2012
- Brown ET, Johnson TC (2005) Coherence between tropical East African and South American records of the Little Ice Age. *Geochem Geophys Geosyst* 6:1–11. doi:10.1029/2005GC000959
- Brown J, Simmonds I, Noone D (2006) Modeling $\delta^{18}\text{O}$ in tropical precipitation and the surface ocean for present-day climate. *J Geophys Res Atmos* 111(5):1–16. doi:10.1029/2004JD005611

- Charles CD (1997) Interaction Between the ENSO and the Asian Monsoon in a Coral Record of Tropical Climate. *Science* 69(1):27–42. doi:[10.1126/science.277.5328.925](https://doi.org/10.1126/science.277.5328.925)
- Clark CO, Webster PJ, Cole JE (2003) Interdecadal variability of the relationship between the Indian Ocean zonal mode and East African coastal rainfall anomalies. *J Clim* 16:548–554. doi:[10.1175/1520-0442\(2003\)016<0548:IVOTRB>2.0.CO;2](https://doi.org/10.1175/1520-0442(2003)016<0548:IVOTRB>2.0.CO;2)
- Cole JE (2000) Tropical Pacific Forcing of Decadal SST Variability in the Western Indian Ocean over the past two centuries. *Science* 287(5453):617–619. doi:[10.1126/science.287.5453.617](https://doi.org/10.1126/science.287.5453.617)
- Crespin E, Goosse H, Fichefet T, Mann ME (2009) The 15th century Arctic warming in coupled model simulations with data assimilation. *Clim Past* 5(3):389–401. doi:[10.5194/cp-5-389-2009](https://doi.org/10.5194/cp-5-389-2009)
- Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda MA, Balsamo G, Bauer P, Bechtold P, Beljaars ACM, van de Berg L, Bidlot J, Bormann N, Delsol C, Dragani R, Fuentes M, Geer AJ, Haimberger L, Healy SB, Hersbach H, Hólm EV, Isaksen L, Kallberg P, Köhler M, Matricardi M, McNally AP, Monge-Sanz BM, Morcrette JJ, Park BK, Peubey C, de Rosnay P, Tavolato C, Thépaut JN, Vitart F (2011) The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q J R Meteorol Soc* 137(656):553–597. doi:[10.1002/qj.828](https://doi.org/10.1002/qj.828)
- Dee SG, Steiger NJ, Emile-Geay J, Hakim GJ (2016) On the utility of proxy system models for estimating climate states over the common era. *J Adv Model Earth Syst* 8(3):1164–1179. doi:[10.1002/2016MS000677](https://doi.org/10.1002/2016MS000677)
- Dubinkina S, Goosse H, Sallaz-Damaz Y, Crespin E, Crucifix M (2011) Testing a particle filter to reconstruct climate changes over the past centuries. *Int J Bifurc Chaos* 21(12):3611–3618. doi:[10.1142/S0218127411030763](https://doi.org/10.1142/S0218127411030763)
- Evans MN, Kaplan A, Cane MA (2000) Intercomparison of coral oxygen isotope data and historical sea surface temperature (SST): potential for coral-based SST field reconstructions. *Paleoceanography* 15(5):551–563. doi:[10.1029/2000PA000498](https://doi.org/10.1029/2000PA000498)
- Flückiger J, Monnin E, Stauffer B, Schwander J, Stocker TF, Chappellaz J, Raynaud D, Barnola JM (2002) High-resolution Holocene N₂O ice core record and its relationship with CH₄ and CO₂. *Glob Biogeochem Cycles* 16(1):1010. doi:[10.1029/2001GB001417](https://doi.org/10.1029/2001GB001417). <http://www.agu.org/pubs/crossref/2002/2001GB001417.shtml>
- Gao C, Robock A, Ammann C (2008) Volcanic forcing of climate over the past 1500 years: an improved ice core-based index for climate models. *J Geophys Res Atmos* 113(December):1–15. doi:[10.1029/2008JD010239](https://doi.org/10.1029/2008JD010239)
- Garcin Y, Williamson D, Taieb M, Vincens A, Mathé PE, Majule A (2006) Centennial to millennial changes in maar-lake deposition during the last 45,000 years in tropical Southern Africa (Lake Masoko, Tanzania). *Palaeogeogr Palaeoclimatol Palaeoecol* 239:334–354. doi:[10.1016/j.palaeo.2006.02.002](https://doi.org/10.1016/j.palaeo.2006.02.002)
- Garcin Y, Williamson D, Bergonzini L, Radakovich O, Vincens A, Buchet G, Guiot J, Brewer S, Mathé PE, Majule A (2007) Solar and anthropogenic imprints on Lake Masoko (southern Tanzania) during the last 500 years. *J Paleolimnol* 37:475–490. doi:[10.1007/s10933-006-9033-6](https://doi.org/10.1007/s10933-006-9033-6)
- Goddard L, Graham NE (1999) Importance of the Indian Ocean for simulating rainfall anomalies over eastern and southern Africa. *J Geophys Res* 104(D16):19,099. doi:[10.1029/1999JD900326](https://doi.org/10.1029/1999JD900326)
- Goosse H, Renssen H, Timmermann A, Bradley RS, Mann ME (2006) Using paleoclimate proxy-data to select optimal realisations in an ensemble of simulations of the climate of the past millennium. *Clim Dyn* 27(2–3):165–184. doi:[10.1007/s00382-006-0128-6](https://doi.org/10.1007/s00382-006-0128-6)
- Goosse H, Crespin E, Dubinkina S, Loutre MF, Mann ME, Renssen H, Sallaz-damaz Y, Shindell D (2012) The role of forcing and internal dynamics in explaining the “Medieval Climate Anomaly”. *Clim Dyn*
- Hakim GJ, Emile-Geay J, Steig EJ, Noone D, Anderson DM, Tardif R, Steiger N, Perkins WA (2016) The Last Millennium Climate Reanalysis Project: Framework and First Results. *J Geophys Res Atmos*. doi:[10.1002/2016JD024751](https://doi.org/10.1002/2016JD024751)
- Hansen J, Sato M (2004) Greenhouse gas growth rates. *Proc Natl Acad Sci USA* 101(46):16,109–16,114. doi:[10.1073/pnas.0406982101](https://doi.org/10.1073/pnas.0406982101)
- Hastenrath S, Nicklis A, Greischar L (1993) Atmospheric-hydrospheric mechanisms of climate anomalies in the western equatorial Indian Ocean. *J Geophys Res* 96(C11):20,219–20,235
- Hurrell JW, Holland MM, Gent PR, Ghan S, Kay JE, Kushner PJ, Lamarque JF, Large WG, Lawrence D, Lindsay K, Lipscomb WH, Long MC, Mahowald N, Marsh DR, Neale RB, Rasch P, Vavrus S, Vertenstein M, Bader D, Collins WD, Hack JJ, Kiehl J, Marshall S (2013) The community earth system model: a framework for collaborative research. *Bull Am Meteorol Soc* 94(9):1339–1360. doi:[10.1175/BAMS-D-12-00121.1](https://doi.org/10.1175/BAMS-D-12-00121.1)
- Hurttt GC, Chini LP, Frolking S, Ra Betts, Feddema J, Fischer G, Fisk JP, Hibbard K, Houghton RA, Janetos A, Jones CD, Kindermann G, Kinoshita T, Klein Goldewijk K, Riahi K, Shevliakova E, Smith S, Stehfest E, Thomson A, Thornton P, van Vuuren DP, Wang YP (2011) Harmonization of land-use scenarios for the period 1500–2100: 600 years of global gridded annual land-use transitions, wood harvest, and resulting secondary lands. *Clim Change* 109:117–161. doi:[10.1007/s10584-011-0153-2](https://doi.org/10.1007/s10584-011-0153-2)
- Izumo T, Lengaigne M, Vialard J, Luo JJ, Yamagata T, Madec G (2014) Influence of Indian Ocean Dipole and Pacific recharge on following year’s El Niño: interdecadal robustness. *Climate Dynamics* 42(1–2):291–310. doi:[10.1007/s00382-012-1628-1](https://doi.org/10.1007/s00382-012-1628-1)
- Johnson TC, McCave IN (2008) Transport mechanism and paleoclimatic significance of terrigenous silt deposited in varved sediments of an African rift lake. *Limnol Oceanogr* 53(4):1622–1632. doi:[10.4319/lo.2008.53.4.1622](https://doi.org/10.4319/lo.2008.53.4.1622)
- Juillet-Leclerc A, Schmidt G (2001) A calibration of the oxygen isotope paleothermometer of coral aragonite from Porites. *Geophys Res Lett* 28(21):4135–4138. doi:[10.1029/2000GL012538](https://doi.org/10.1029/2000GL012538)
- Klein F, Goosse H, Mairesse A, de Vernal A (2013) Model-data comparison and data assimilation of mid-Holocene Arctic sea-ice concentration. *Clim Past Discuss* 9:6515–6549. doi:[10.5194/cpd-9-6515-2013](https://doi.org/10.5194/cpd-9-6515-2013)
- Klein F, Goosse H, Graham NE, Verschuren D (2016) Comparison of simulated and reconstructed variations in East African hydroclimate over the last millennium. *Clim Past* 12(7):1499–1518. doi:[10.5194/cp-12-1499-2016](https://doi.org/10.5194/cp-12-1499-2016). <http://www.clim-past.net/12/1499/2016/>
- Kuhnert H, Pätzold J, Hatcher B, Wyrwoll KH, Eisenhauer A, Collins LB, Zhu ZR, Wefer G (1999) A 200-year coral stable oxygen isotope record from a high-latitude reef off Western Australia. *Coral Reefs* 18(1):1–12. doi:[10.1007/s003380050147](https://doi.org/10.1007/s003380050147)
- van Leeuwen PJ (2009) Particle filtering in geophysical systems. *Mon Weather Rev* 137(12):4089–4114. doi:[10.1175/2009MWR2835.1](https://doi.org/10.1175/2009MWR2835.1)
- Lorenz EN (1956) Empirical orthogonal functions and statistical weather prediction. Tech. rep., Statistical Forecasting Scientific Rep. 1, Department of Meteorology, Massachusetts Institute of Technology, Cambridge, MA
- MacFarling Meure C, Etheridge D, Trudinger C, Steele P, Langenfelds R, Van Ommen T, Smith A, Elkins J (2006) Law Dome CO₂, CH₄ and N₂O ice core records extended to 2000 years BP. *Geophys Res Lett* 33:2000–2003. doi:[10.1029/2006GL026152](https://doi.org/10.1029/2006GL026152)
- Mairesse A, Goosse H, Mathiot P, Wanner H, Dubinkina S (2013) Investigating the consistency between proxy-based reconstructions and climate models using data assimilation: a mid-Holocene case study. *Clim Past* 9:2741–2757. doi:[10.5194/cp-9-2741-2013](https://doi.org/10.5194/cp-9-2741-2013)
- Matsikaris A, Widmann M, Jungclaus J (2015) On-line and off-line data assimilation in palaeoclimatology: a case study. *Clim Past* 11(1):81–93. doi:[10.5194/cp-11-81-2015](https://doi.org/10.5194/cp-11-81-2015)

- Nicholson S (1996) A review of climate dynamics and climate variability in Eastern Africa. In: Johnson TC, Odada EO (eds) *The limnology. Climatology and Paleoclimatology of the East African Lakes*, Gordon and Breach, pp 25–56
- Nicholson SE (2014) Spatial teleconnections in African rainfall: a comparison of 19th and 20th century patterns. *Holocene* 24(12):1840–1848. doi:10.1177/0959683614551230
- Nicholson SE, Selato JC (2000) The influence of La Nina on African rainfall. *Int J Climatol* 20:1761–1776. <http://128.186.98.10/people/nicholson/papers/lanina.pdf>
- Ogallal LJ, Janowiak JE, Halpert MS (1988) Teleconnection between Seasonal Rainfall over East Africa and Global Sea surface temperature anomalies. *J Meteorol Soc Jpn* 66(6):807–822
- Otto-Bliesner BL, Brady EC, Fasullo J, Jahn A, Landrum L, Stevenson S, Rosenbloom N, Mai A, Strand G (2015) Climate variability and change since 850 C.E.: an ensemble approach with the community earth system model (CESM). *Bull Am Meteorol Soc*. doi:10.1175/BAMS-D-14-00233.1 (in press)
- Owiti Z, Zhu W (2012) Spatial distribution of rainfall seasonality over East Africa. *J Geogr Region Plan* 5(15):409–421. doi:10.5897/JGRP12.027
- Park SK, Xu L (2009) *Data assimilation for atmospheric, oceanic and hydrologic applications*. Springer, Berlin. doi:10.1007/978-3-540-71056-1
- Pendergrass AG, Hakim GJ, Battisti DS, Roe G (2012) Coupled air-mixed layer temperature predictability for climate reconstruction. *J Clim* 25(2):459–472. doi:10.1175/2011JCLI4094.1
- Pfeiffer M, Timm O, Dullo WC, Podlech S (2004) Oceanic forcing of interannual and multidecadal climate variability in the southwestern Indian Ocean: Evidence from a 160 year coral isotopic record (La Réunion, 55degE, 21degS). *Paleoceanography* 19(4):1–14. doi:10.1029/2003PA000964
- Pongratz J, Reick CH, Raddatz T, Claussen M (2009) Effects of anthropogenic land cover change on the carbon cycle of the last millennium. *Glob Biogeochem Cycles* 23:1–13. doi:10.1029/2009GB003488
- Rowell DP (2013) Simulating SST teleconnections to Africa: what is the state of the art? *J Clim* 26(15):5397–5418. doi:10.1175/JCLI-D-12-00761.1
- Saji NH, Goswami BN, Vinayachandran PN, Yamagata T (1999) A dipole mode in the tropical Indian Ocean. *Nature* 401(6751):360–363. doi:10.1038/43854. <http://www.ncbi.nlm.nih.gov/pubmed/16862108>
- Schmidt GA, Jungclauss JH, Ammann CM, Bard E, Braconnot P, Crowley TJ, Delaygue G, Joos F, Krivova Na, Muscheler R, Otto-Bliesner BL, Pongratz J, Shindell DT, Solanki SK, Steinhilber F, Vieira LEA (2012) Climate forcing reconstructions for use in PMIP simulations of the Last Millennium (v1.1). *Geosci Model Dev* 5(2011):185–191. doi:10.5194/gmd-5-185-2012
- Schmidt GA, Kelley M, Nazarenko L, Ruedy R, Russell GL, Aleinov I, Bauer M, Bauer SE, Bhat MK, Bleck R, Canuto V, Chen Yh, Cheng Y, Clune TL, Genio AD, Fainchtein RD, Faluvegi G, Hansen JE, Healy RJ, Kiang NY, Koch D, Aa Lacis, Legrande AN, Lerner J, Lo KK, Matthews EE, Menon S, Miller RL, Oinas V, Olosou AO, Perlwitz JP, Puma MJ, Putman WM, Rund D, Romanou A, Sato M, Shindell DT, Sun S, Syed RA, Tausnev N, Tsigaridis K, Unger N, Voulgarakis A, Yao MS, Zhang J (2014) Configuration and assessment of the GISS ModelE2 contributions to the CMIP5 archive. *J Adv Model Earth Syst* 6(1):141–184. doi:10.1002/2013MS000265
- Schneider U, Becker A, Finger P, Meyer-Christoffer A, Ziese M, Rudolf B (2014) GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle. *Theoret Appl Climatol* 115:15–40. doi:10.1007/s00704-013-0860-x
- Schreck CJ, Semazzi FHM (2004) Variability of the recent climate of eastern Africa. *Int J Climatol* 24(6):681–701. doi:10.1002/joc.1019
- Schubert SD, Stewart RE, Wang H, Barlow M, Berbery EH, Cai W, Hoerling MP, Kanikicharla KK, Koster RD, Lyon B, Mariotti A, Mechoso CR, Müller OV, Rodriguez-Fonseca B, Seager R, Senevirante SI, Zhang L, Zhou T (2016) Global meteorological drought: a synthesis of current understanding with a focus on SST drivers of precipitation deficits. *J Clim* 29(11):3989–4019. doi:10.1175/JCLI-D-15-0452.1
- Smith TM, Reynolds RW, Peterson TC, Lawrimore J (2008) Improvements to NOAA's historical merged land-ocean surface temperature analysis (1880–2006). *J Clim* 21(10):2283–2296. doi:10.1175/2007JCLI2100.1
- Steiger N, Hakim G (2016) Multi-timescale data assimilation for atmosphere–ocean state estimates. *Clim Past* 12(6):1375–1388. doi:10.5194/cp-12-1375-2016
- Steiger NJ, Hakim GJ, Steig EJ, Battisti DS, Roe GH (2014) Assimilation of time-averaged pseudoproxies for climate reconstruction. *J Clim* 27(1):426–441. doi:10.1175/JCLI-D-12-00693.1
- Stevens B, Giorgetta M, Esch M, Mauritsen T, Crueger T, Rast S, Salzmann M, Schmidt H, Bader J, Block K, Brokopf R, Fast I, Kinne S, Kornbluh L, Lohmann U, Pincus R, Reichler T, Roeckner E (2013) The atmospheric component of the MPI-M earth system model: ECHAM6. *J Adv Model Earth Syst* 5:1–27. doi:10.1002/jame.20015
- Stevenson S, McGregor HV, Phipps SJ, Fox-Kemper B (2013) Quantifying errors in coral-based ENSO estimates: toward improved forward modeling of d18O. *Paleoceanography* 28(4):633–649. doi:10.1002/palo.20059
- von Storch H, Cubasch U, González-Rouco J, Jones J, Voss R, Widmann M, Zorita E (2000) Combining paleoclimatic evidence and GCMs by means of data assimilation through upscaling and nudging. 11th Symposium on global climate change studies. AMS Long Beach, CA, pp 28–31
- Thiery W, Davin E, HJ Panitz, Demuzere M, Lhermitte S, Van Lipzig N (2015) The impact of the African Great Lakes on the regional climate. *J Clim* 28(10):4061–4085. doi:10.1175/JCLI-D-14-00565.1
- Thompson DM, Ault TR, Evans MN, Cole JE, Emile-Geay J (2011) Comparison of observed and simulated tropical climate trends using a forward model of coral d18O. *Geophys Res Lett* 38(14):1–6. doi:10.1029/2011GL048224
- Tierney JE, Russell JM, Sinninghe Damsté JS, Huang Y, Verschuren D (2011) Late quaternary behavior of the East African monsoon and the importance of the Congo Air Boundary. *Quat Sci Rev* 30(7–8):798–807. doi:10.1016/j.quascirev.2011.01.017. <http://linkinghub.elsevier.com/retrieve/pii/S0277379111000321>
- Tierney JE, Smerdon JE, Anchukaitis KJ, Seager R (2013) Multi-decadal variability in East African hydroclimate controlled by the Indian Ocean. *Nature* 493(7432):389–392. doi:10.1038/nature11785. <http://www.ncbi.nlm.nih.gov/pubmed/23325220>
- Tierney JE, Abram NJ, Anchukaitis KJ, Evans MN, Giry C, Kilbourne KH, Saenger CP, Wu HC, Zinke J (2015) Tropical sea surface temperatures for the past four centuries reconstructed from coral archives. *Paleoceanography* 30:226–252. doi:10.1029/2014PA002717
- Ummenhofer CC, Sen Gupta A, England MH, Reason CJC (2009) Contributions of Indian Ocean sea surface temperatures to enhanced East African Rainfall. *J Clim* 22(4):993–1013. doi:10.1175/2008JCLI2493.1
- Verschuren D (2001) Reconstructing fluctuations of a shallow East African lake during the past 1800 yrs from sediment stratigraphy in a submerged crater basin. *J Paleolimnol* 25(3):297–311. doi:10.1023/A:1011150300252

- Verschuren D, Laird KR, Cumming BF (2000) Rainfall and drought in equatorial east Africa during the past 1,100 years. *Nature* 403(6768):410–414. doi:[10.1038/35000179](https://doi.org/10.1038/35000179)
- Verschuren D, Sinninghe Damsté JS, Moernaut J, Kristen I, Blaauw M, Fagot M, Haug GH (2009) Half-precessional dynamics of monsoon rainfall near the East African Equator. *Nature* 462(7273):637–641. doi:[10.1038/nature08520](https://doi.org/10.1038/nature08520). <http://www.ncbi.nlm.nih.gov/pubmed/19956257>
- Vieira LEA, Solanki SK, Krivova NA, Usoskin I (2011) Evolution of the solar irradiance during the Holocene. *Astron Astrophys* 531(A6):1–20. doi:[10.1051/0004-6361/201015843](https://doi.org/10.1051/0004-6361/201015843). [arXiv:1103.4958](https://arxiv.org/abs/1103.4958)
- Webster PJ, Moore aM, Loschnigg JP, Leben RR (1999) Coupled ocean-atmosphere dynamics in the Indian Ocean during 1997–98. *Nature* 401(6751):356–360. doi:[10.1038/43848](https://doi.org/10.1038/43848). <http://www.ncbi.nlm.nih.gov/pubmed/16862107>
- 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. *Clim Past* 6(5):627–644. doi:[10.5194/cp-6-627-2010](https://doi.org/10.5194/cp-6-627-2010). <http://www.clim-past.net/6/627/2010/>
- Wolff C, Haug GH, Timmermann A, Sinninghe Damsté JS, Brauer A, Sigman DM, Cane MA, Verschuren D (2011) Reduced interannual rainfall variability in East Africa during the last ice age. *Science (New York, NY)* 333(August):743–747. doi:[10.1126/science.1203724](https://doi.org/10.1126/science.1203724)
- Yang W, Seager R, Ma Cane, Lyon B (2015) The annual cycle of the east African precipitation. *J Clim* 28(6):2385–2404. doi:[10.1175/JCLI-D-14-00484.1](https://doi.org/10.1175/JCLI-D-14-00484.1)
- Zinke J, Pfeiffer M, Timm O, Dullo WC, Kroon D, Thomassin BA (2008) Mayotte coral reveals hydrological changes in the western Indian ocean between 1881 and 1994. *Geophys Res Lett* 35(23):1–5. doi:[10.1029/2008GL035634](https://doi.org/10.1029/2008GL035634)
- Zinke J, Pfeiffer M, Timm O, Dullo WC, Brummer GJA (2009) Western Indian Ocean marine and terrestrial records of climate variability: a review and new concepts on land-ocean interactions since AD 1660. *Int J Earth Sci* 98(1):115–133. doi:[10.1007/s00531-008-0365-5](https://doi.org/10.1007/s00531-008-0365-5)