Dear Dr. Volker Rath,

We would like to submit our revised manuscript of cp-2016-121.

We believe the paper has been improved very much by taking account all the comments that two reviewers kindly gave to us. Even though we have changed many parts of the manuscript, our final messages remain unchanged.

Hereafter, point-by-point responses to all the reviewer's comments follow. The reviewer's comments in blue italic font, the replies are in black and the changes in red with the line number of the marked-up version.

Best regards,

Atsushi Okazaki and Kei Yoshimura

For Anonymous Referee #1

Note that some of the replies are used in common, since comments 2, 3, 4 are similar to the comments 3, 2, 1 by the anonymous referee #2.

Data assimilation in paleoclimatology is a rapidly growing field. The present paper addresses the model-data comparison step that is critical in every data assimilation scheme. Up to now, proxy records are generally first transformed to obtain a reconstruction of simulated variables such as temperature or precipitation before being assimilated. Simulating the measured quantity using proxy system models and performing the comparison directly for this variable provides in theory many advantages. The present study analyses those advantages and the potential limitations of the methodology based on both idealized and realistic experiments. It demonstrates the ability to directly assimilate isotopic composition of several proxies thanks to the application of forward proxy models. The study also identifies the regions/variables where the skill is already satisfactory and the promising ways of improvement. The authors thus provide very interesting results for methodological developments and the application of data assimilation techniques in paleoclimatology. The study thus deserves publications in Climate of the Past but some modifications are required in the experimental design and in the discussion to reach conclusions that are easier to be interpreted and to be compared with recent work as detailed below.

Thank you very much for the positive and valuable comments.

1. *Several groups are currently working on the direct assimilation of proxy records. The authors could not be blamed for not discussing all the very recent publications in the submitted version but a comparison of the conclusions reached here with the ones of Dee et al. (2016) must at least be included as the latter study is focused on a very close subject. In particular, Dee et al. (2016) compare a direct assimilation of isotopes using an isotope enabled atmospheric model with the assimilation of temperature derived from the proxy records, as in the present paper. The publication of those recent papers also requires to modify some sentences like lines 80-81 and 116-117 where it is said that it is the first time that proxy data are assimilated directly (see also Acevedo et al. 2016).*

   Acevedo W., B. Fallah, S. Reich, and U. Cubasch (2016). Assimilation of PseudoTree-Ring-Width observations into an Atmospheric General Circulation Model. Clim. Past Discuss., doi:10.5194/cp-2016-92, 2016. Available at *http://www.clim-pastdiscuss.net/cp-2016-92/*

Dee, S.G., N.J. Steiger, J. Emile-Geay, and G.J. Hakim (2016): On the utility of proxy system modeling for estimating climate states over the Common Era. Journal of Advances in Modeling Earth Systems. doi:10.1002/2016MS000677. Available at http://onlinelibrary.wiley.com/doi/10.1002/2016MS000677/pdf

We included the Acevedo et al. (2016) and Dee et al. (2016) in Sect. 1 and modified the corresponding sentences. Also, we included Dee et al. (2016) in Sect. 5.1 to discuss the comparison between proxy DA and reconstructed DA.

L85-87, L123-124, L126, L505-507, L595-598

2. *I was surprised that the data assimilation method was not described at all in section 2.1. If I am right an ensemble Kalman filter is applied but this is only stated in the conclusions (the word Kalman is mentioned first line 528). A long description of the method is not required but its main characteristics should at least be mentioned in section 2.1.*

   The description of data assimilation method is included in the revised manuscript (L133-137). We used EnSRF (Whitaker and Mitchell, 2002) with slight modification following the previous studies (Bhend et al., 2012; Steiger et al., 2014).

   L143-150, L155-163, L711-712, L829-830

3. *The interpretation of experiment T2-ASSIM and its comparison with CTRL are not straightforward to me as the conclusions strongly depend on the signal to noise ratio selected and it is not possible from the information given in the paper to compare this signal to noise ratio with the error used in CTRL. One option would be to use the model results to estimate the impact of an error of 0.5 per mil on the isotopic composition, as imposed in CTRL, on a temperature reconstruction based on those isotopic records using simple statistical methods (for instance a regression as often done in paleoclimate reconstructions). Then, additional sensitivity experiments can be performed with such a temperature reconstruction derived from the isotopic composition (and not using the temperature simulated by the model) or alternatively assimilating temperature using the signal to noise ratio of this reconstruction that would be compatible with the error imposed in CTRL.*

   Thank you for the comments. We modified the experimental setting for T2-Assim following your suggestion. In the modified experiment, temperature is reconstructed from the isotopic records which is used in CTRL by simple regression-based method. Proxies whose correlation with local temperature during calibration period (1871-

1950) is not statistically significant ($p < 0.10$) are removed following Mann et al. (2008). This screening process reduced the available data from 94 to 81 grid points. Based on the correlation between isotope ratio and local temperature, SNR can be estimated through the equation (Mann et al., 2007):

$$\mathrm{SNR} = \sqrt{\frac{r^2}{1 - r^2}}$$

where r is the correlation. SNR is shown in Fig. 8. Subsequently, this reconstructed temperature (T_r) is assimilated. The assimilated result is shown in Fig. 7. The result is slightly degraded in T2-Assim compared with CTRL due to relatively large error in T_r (Fig. 8). As shown in Dee et al. (2016), the reconstruction skill is somewhat compensated by the structure of Kalman gain. Figure S1 shows the correlation scale length to show the difference in the structure between CTRL and T2-Assim. The correlation scale length was found by computing point correlation between the prior (temperature) and the prior-estimated observation (temperature and $\delta^{18}O$ for T2-Assim and CTRL, respectively) for the observation grids, binning these correlations by distance, and computing the mean of each bin. The correlation is consistently high in T2-Assim, which means that the observation information is more effectively used to update the analysis. To sum up, the accuracies are not substantially different among proxy DA and reconstructed DA. However, we should note that this is only the case as long as the relation between temperature and isotope remain the same. L288-289, L292-301, L339-340, L461-518, Table 1, Figure 7, Figure 8

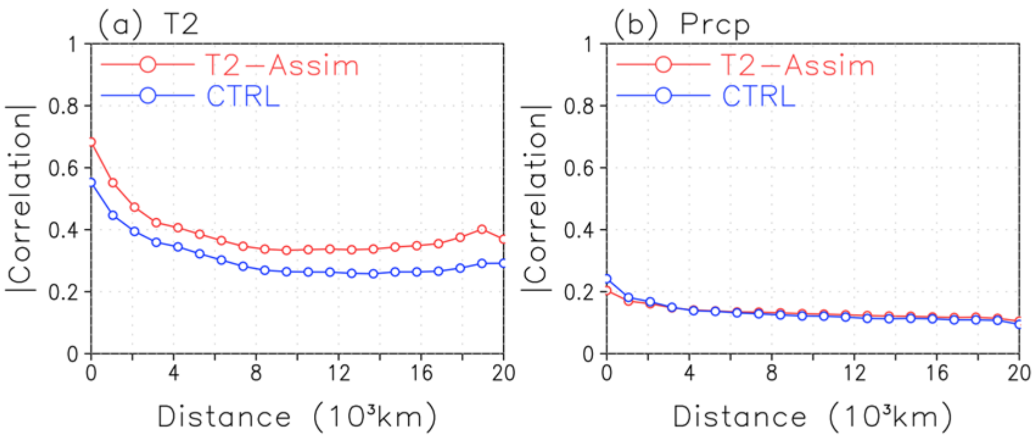


Figure S 1 Mean correlation scale length for T2-Assim (red) and CTRL (blue). The prior is (a) temperature and (b) precipitation. The prior-estimated observation is temperature and $\delta^{18}O$ for T2-Assim and CTRL, respectively for the both panels.

4. *The low skill of experiment REAL can have many origins: biases in climate models, limitations of proxy system models, non-climatic noise in the data, local signal in the records not represented in large-scale models, etc. The present study does not address the relative contribution of each of those elements and this is perfectly fine for me as it is not the goal of the present study. Nevertheless, some recommendations like line 51, line 497, line 502, line 506 on the improvement of models seems relatively vague and not really justified by the results. I would thus recommend to be more careful and to focus on the main results of the study.*
Thank you for the comments. We understand that there are multiple factors other than model errors for the low skill in REAL experiment and that we do not know their relative contribution. Thus, we carefully modified the abstract and Sect. 6 in which we avoided arguing vague explanation.
L50-53, L455-456, L571-593, L595-598

Specific points

1. *Abstract, line 42-43. This sentence is not clear without reading the main text. Please rephrase (see also general comment 2).*
Thank you for the comments. We omitted the sentence for better readability.

2. *Line 100. The data are not erroneous, this is the interpretation that is questionable.*
We reworded that part as "such questionable reconstructed data". Thank you.
L106

3. *Line 143. The 'simplification' is valid for some variables but not for others that change more slowly such as oceanic temperatures.*
We clearly mentioned that the simplification is valid at least for atmospheric variables in the revised manuscript .
L165-166

4. *Line 150-151. What is meant by 'changing the algorithm'. The text should be more explicit and provide a reference if available.*
We rephrased the sentence as "the proxy DA could address non-stationarity if one

uses temporally varying background ensemble”.

L171-172, L623-624, L766-769, L816-817

5. *Line 176. A few words should be given on the version of MIROC5 applied as the reference is not available yet. In particular, it should be stated if only the atmospheric component is applied (as suggested lines 214-215) or if it is coupled to an interactive ocean.*

   Thank you for the comments. The version of the model is five (hence MIROC”5”) and we used only the atmospheric component of the GCM. To make it clearer, we changed the sentence as “we used a newly-developed model based on the atmospheric component of MIROC5”.

   L198

6. *Line 189. Why is the deep ocean composition needed for corals that live in shallow waters?*

   Thank you for the comment. The isotopic ratio in the upper layer of the ocean is determined by the balance of precipitation, evaporation, and vertical mixing from deeper water, not deep water. We modified the term “deep” to “deeper” in the revised manuscript.

   L211, L212

7. *Line 250. I guess the four sensitivity experiments has to be compared to experiment CTRL. This should be already stated at this stage.*

   Two of them (i.e. CGCM and VOBS) were conducted to explain the difference among CTRL and REAL and the experimental settings were changed in a stepwise manner, from idealized way to more realistic way. Thus, CGCM were compared with CTRL, and VOBS were compared with CGCM. The other two were compared with CTRL. We included sentences explaining what experiment was used to evaluate each sensitivity experiment.

   L279, L284-287, L310

8. *Line 322. Is it just a repetition of line 318 with a different sign or new information?*

   No, it is not. The first sentence described the reconstruction skill for temperature and precipitation by comparing the analysis and the truth. On the other hand, the second sentence explained how the high reconstruction skill was achieved by comparing the assimilated variable (δ) and the reconstructed variable (temperature and precipitation) at the site.

9. *Line 333. Why using 'on the other hand' here?*

   The closely correlated area was limited around the observation site for $\delta^{18}O$ in tree-ring cellulose, but the high correlation was not limited around the observation site for $\delta^{18}O$ in coral. Thus, we used the "on the other hand" here. To make the context clearer, we modified the sentence in the revised manuscript (L345-350).

   L367-369

10. *Line 336. The results for temperature should be discussed too.*

    The results for temperature were included in the revised manuscript (L350-351). Thank you.

    L369-370

11. *Line 348. Is this increase noticed in simulation results or in observations? Please be more precise.*

    The temperature has been increased both in observations and simulations. In the manuscript, what we meant was observation. We modified the sentence and put a reference in the revised manuscript (L364-366).

    L383-384, L701-707

12. *Line 411-412. I would suppress this sentence as it does not bring new information.*

    Suppressed.

    L448-450

13. *Line 415-419. I may miss something but I do not see how the low reproducibility of corals could play a role in the perfect model framework of CTRL as it is assumed that the climate and proxy models have no systematic bias (see also line 496).*

    In this chapter, we compared VOBS and REAL, where VOBS is a perfect model experiment assuming that the climate and proxy models have no systematic bias and REAL is not a perfect model experiment. In the REAL, we assimilated observed data in the real world. Thus, models do have biases.

For Anonymous Referee #2

Note that some of the replies are used in common, since general comments 1, 2, 3 are similar to the comments 4, 3, 2 by the anonymous referee #1.

The authors present and analyze a novel approach to directly assimilate proxy information into GCM simulations to reconstruct past climate. They find that while assimilation of isotopic proxies is possible and is clearly beneficial in idealized simulations, the actual benefit of assimilating proxy data is limited due to model errors and the small number of assimilated proxies. Data assimilation in paleoclimatology has attracted a lot of attention recently and the science and methods are developing rapidly. This manuscript represents an important contribution to the field in that for one of the first times, proxy data (rather than reconstructed climatic variables) are assimilated directly for climate reconstructions. Therefore, I recommend this article to be published after the outstanding issues detailed below have been addressed.

Thank you very much for the positive and valuable comments.

General comments:

1. *The sensitivity experiments conducted in this study only 'explain' a small fraction of the difference in correlation between the idealized setup (CTRL) and the application to real proxy data (REAL). The reasons for such a reduction in quality are manifold and include GCM model errors and errors in the proxy forward model that are not quantified in the current analysis. Proxy model errors are shortly discussed at the end of section 4, but it is not clear to me how one could attribute errors to the proxy model or the GCM in the absence of controlled experiments (as also stated by the authors in L504). While performing such controlled experiments with alternative proxy model / GCM combinations is clearly beyond the scope of this paper, I suggest the authors carefully reword the respective paragraphs.*

   Thank you for the comments. We understand that there are multiple factors other than model errors for the low skill in REAL experiment and that we do not know their relative contribution. Thus, we carefully modified the abstract and Sect. 6 in which we clearly mentioned that there remains a lot unexplained and avoided arguing that the model errors are the only reason for the degradation in REAL experiment.

   L50-53, L455-456, L571-593, L595-598

2. *In addition to trying to quantify the limitations of the current proxy DA setup by performing sensitivity experiments, the authors also try to answer a second question: namely whether direct assimilation of proxy data is superior to assimilating climatic variables (here temperature) reconstructed from the proxy data. In contrast to the approach pursued here, it would seem easier to address this question using the REAL experimental setup. Based on this setup, one could derive reconstructed (gridded) temperature data from the exact same proxies that have been used in the REAL experiment and assimilate these reconstructed temperatures instead. Such an experimental framework would be instructive as to whether empirical proxy models (i.e. reconstructed temperatures) outperform the physics-based on-line proxy models. Alternatively, one could devise idealized experiments similar to the ones performed in the study in which one compares assimilations based on the assumption of a perfect proxy model. In contrast to the comparison presented here, one would need to compare the CTRL (or any other of the synthetic proxy experiments) to the corresponding experiment in which the proxy data (+ noise) from the truth run has been used to reconstruct temperatures which are then assimilated. Such analysis, however, may be beyond the scope of this paper and I would be perfectly happy if the authors decide to focus on the main message of the manuscript – the proxy data assimilation and partial attribution of its limited skill to quantifiable sources – only.*

Thank you for the comments. We modified the experimental setting for T2-Assim following your suggestion. In the modified experiment, temperature is reconstructed from the isotopic records which is used in CTRL by simple regression-based method. Proxies whose correlation with local temperature during calibration period (1871-1950) is not statistically significant ($p < 0.10$) are removed following Mann et al. (2008). This screening process reduced the available data from 94 to 81 grid points. Based on the correlation between isotope ratio and local temperature, SNR can be estimated through the equation (Mann et al., 2007):

$$\mathrm{SNR} = \sqrt{\frac{r^2}{1-r^2}}$$

where r is the correlation. SNR is shown in Fig. 8. Subsequently, this reconstructed temperature (T_r) is assimilated. The assimilated result is shown in Fig. 7. The result is slightly degraded in T2-Assim compared with CTRL due to relatively large error in T_r (Fig. 8). As shown in Dee et al. (2016), the reconstruction skill is somewhat compensated by the structure of Kalman gain. Figure S1 shows the correlation scale length to show the difference in the structure between CTRL and T2-Assim. The correlation scale length was found by computing point correlation between the prior (temperature) and the prior-estimated observation (temperature and $\delta^{18}O$ for T2-Assim and CTRL, respectively) for the observation grids, binning these correlations by distance, and computing the mean of each bin. The correlation is consistently high in T2-Assim, which means that the observation information is more effectively used to update the analysis. To sum up, the accuracies are not substantially different among proxy DA and reconstructed DA. However, we should note that this is only the case as long as the relation between temperature and isotope remain the same. L288-289, L292-301, L339-340, L461-518, Table 1, Figure 7, Figure 8

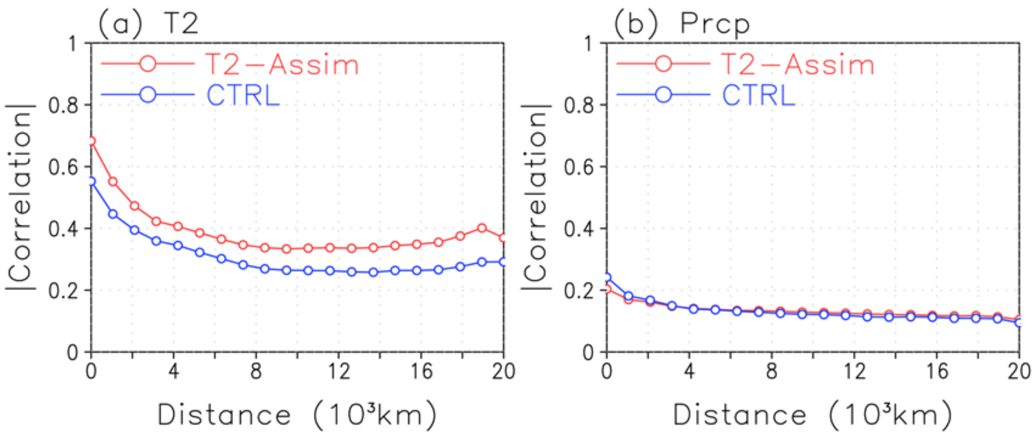


Figure S 2 Mean correlation scale length for T2-Assim (red) and CTRL (blue). The prior is (a) temperature and (b) precipitation. The prior-estimated observation is temperature and $\delta^{18}O$ in coral for T2-Assim and CTRL, respectively for the both panels.

3. *The data assimilation method is not described at all. Please add a short section on the data assimilation method with the relevant references. I suggest to focus on the choices and setup specific to this study and to provide the appropriate references; an in-depth introduction to the data assimilation method would only be needed if you chose a non-standard assimilation method that is not documented elsewhere. If, as suggested by the final paragraph of the manuscript, an EnKF has been used, then I suggest to also analyse the spread to error ratio or compute rank histograms to get an impression whether the analysis spread matches the analysis error and the*

analysis is well calibrated. Lack of calibration (usually overconfidence) is likely due to a misrepresentation of the observation error matrix (either underestimation of observation error or correlated errors).

The description of data assimilation method is included in the revised manuscript (L133-137). We used EnSRF (Whitaker and Mitchell, 2001) with slight modification following the previous studies (Bhend et al., 2012; Steiger et al., 2014). As described in the manuscript, we did an offline approach, in which the analysis is not cycled to the simulation and the same background is used for every analysis step.

Figure S2 shows the spread and RMSE for surface temperature in REAL experiment. The posterior spread matches with RMSE for the first half of the period but it gradually diverges from the RMSE. This reflects the fact that the system has a difficulty in reconstructing temperature in mid- to high-latitude (Fig. 3), where temperature has been increasing in the period. However, the discrepancy does not necessarily mean that the system is not well calibrated. The relatively scarce observation and short correlation length scale must hamper the reproducibility there. On top of that, we speculate that the metrics such as spread-to-error ratio or rank histogram may not suit for the evaluation of the offline DA. In general, it takes several cycles for the spread to match with RMSE through improving the error covariance matrix in the online DA. Contrarily, because the offline DA uses the same background for every analysis step, the quality of the analysis error covariance remains the same (c.f. $\mathbf{P^a} = [\mathbf{I} - \mathbf{P^f H^t}(\mathbf{HP^f H} + \mathbf{R})^{-1}\mathbf{H}]\mathbf{P^f}$). Therefore, the spread will not tell how well the system is calibrated.

Instead, we show the sensitivity of the system to parameters (observation error and localization scale) in REAL experiment to show that the system is how optimal. The results show that the skill is moderately dependent on both the observation error and the localization scale. For the observation error, the results become better with larger error in the investigated range for the both variables. On the other hand, the sensitivity to the localization scale varies from variable to variable. For temperature, the correlation become better along with the scale in the investigated range. For precipitation, the localization scale of 12000km resulted in the best correlation.

The sensitivity is also different by which metric to be used. For instance, RMSE for precipitation becomes larger along with the observation error (not shown).

Given the results above and because the choice of the parameters does not change the main conclusion of the study, we keep using the original value of the parameters.

L143-150, L155-163, L711-712, L829-830

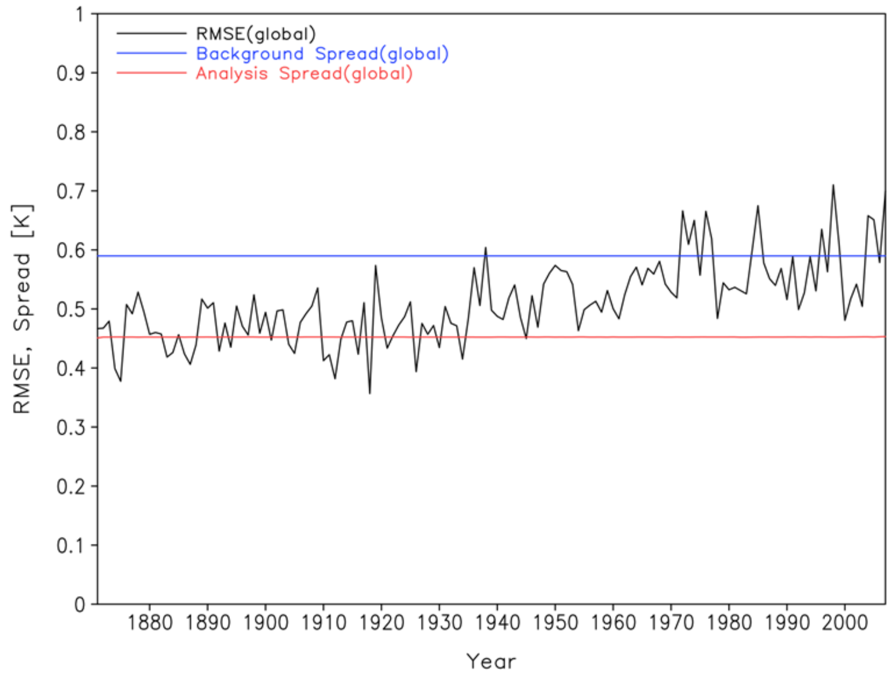


Figure S 3 Global mean of the spread and RMSE for surface temperature in REAL. The spread of background and analysis and RMSE are shown in blue, red, and black, respectively.

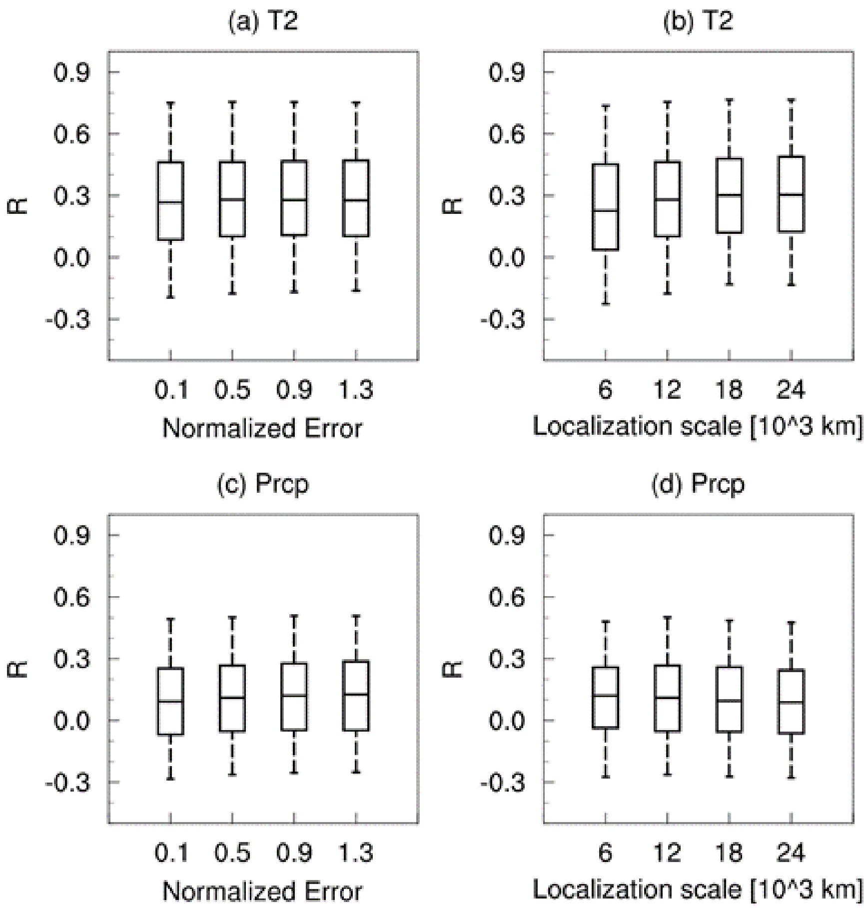


Figure S 4 Box-whisker plot of the distribution of all spatial values for the correlation for (a and b) temperature, and (c and d) precipitation in REAL. (a) and (c) shows the sensitivity to the observation error, and (b) and (d) shows the sensitivity to the localization scale.

4. *Use of the term 'accuracy': The authors repeatedly use the term 'accuracy' to describe the quality of the analysis. This use of language is somewhat misleading, as accuracy in forecast verification has a specific meaning and the appropriate verification score to measure accuracy would be the mean squared or mean absolute error, whereas the correlation is a measure of forecast / analysis association (e.g. Murphy, 1993). I suggest to either rephrase and write of "improved assimilation", "enhanced correlation" etc. or to clearly state that accuracy refers to correlation throughout the manuscript.*

Thank you for the comments. We used the term ‘reconstruction skill’ or ‘skill’ instead of ‘accuracy’ in the revised manuscript. In addition, we clearly stated that we use the correlation coefficient as a measurement of skill (L119-120).
L41, L47, L50, L107, L129-130, L278, L290, L424, L425, L521, L531, L546, L558, L599

Specific comments:

1. *L112: This issue seems important and I think it would be worth revisiting in the conclusions.*
   Thank you for the comment. We revisited the issue in the abstract and the conclusion in the revised manuscript (L48-50, L528-530).
   L54-56, L599-605

2. *L267: stemming from*
   The sentence was deleted because we changed the experimental design for T2-Assim.

3. *L363-365/7: Is this a direct quote from the Xu et al. paper? If so I suggest labelling this as such by using quotation marks.*
   No, it is not. We modified the sentence for better readability (L380-385).
   L398-403

4. *L385: for precipitation*
   Corrected.
   L423

5. *L440: slightly more accurately?*
   Section 5.1 were substantially modified following the general comment #2. Accordingly, the sentence was not used any more in the revised manuscript. Thank you.

6. *L487ff: if the only difference in simulations is observed vs. simulated SSTs, I suggest the authors refrain from using the term forcing in the following lines for better readability.*

We modified Sect. 6 significantly and the corresponding parts were omitted. Thank you.

7. *L499ff: The discussion of the differences of the various sensitivity experiments is hard to read. I suggest to streamline and reword this section along the lines of "Imperfect SST used to drive the CGCM simulation resulted in a slight reduction of correlation compared to the CTRL experiment with perfect SST."*

   The corresponding sentences were rephrased in the revised manuscript following the suggestion (L505-507; 510-512). Thank you.

   L563-564, L568-570

8. *L513: non-climatic factors.*

   Corrected.

   L593

9. *L514: add reference, e.g. Appendix B of Compo et al. 2011*

   We added the reference (Appendix B of Compo et al., 2011) in the revised manuscript.

   L595, L656-661

10. *L525: I suggest to mention that not in all cases direct proxy DA will be beneficial compared to assimilating empirically reconstructed variables. Also, while assimilating more data is expected to increase the quality of the analysis, care has to be taken in assimilating dependent information (e.g. direct assimilation of proxy data and reconstructed variables derived from the same proxy data).*

    The both sentences were included in the revised manuscript (L444-445; 542-543; 544-546). Thank you.

    L489-490, L610-617

11. *Figure 4: The figure labels denote EOF2 whereas only EOF1 is mentioned in the text. Please fix.*

    Thank you for your pointing out. The figure was replaced with the correct figure.

    Figure 4

Development and evaluation of a system of proxy data assimilation for paleoclimate reconstruction

By

Atsushi Okazaki[1] and Kei Yoshimura[2]

[1]RIKEN Advanced Institute for Computational Science, Japan

~~[2]Atmosphere and Ocean Research Institute,~~ [2]Institute of Industrial Science, The University of Tokyo, Japan

Submitted to Climate of the Past

Submitted in November, 2016

Revised in February, 2017

Corresponding author: Atsushi Okazaki, RIKEN Advanced Institute for Computational

Science, 7-1-26 Minatojima-minami-machi, Chuo-ku, Kobe, Hyogo 650-0047, Japan (atsushi.okazaki@riken.jp)

Abstract

Data assimilation (DA) has been successfully applied in the field of paleoclimatology to reconstruct past climate. However, data reconstructed from proxies have been assimilated, as opposed to the actual proxy values. This banned to fully utilize the information recorded in the proxies.

This study examined the feasibility of proxy DA for paleoclimate reconstruction. Isotopic proxies ($\delta^{18}O$ in ice cores, corals, and tree-ring cellulose) were assimilated into models: an isotope enabled general circulation model (GCM) and forward proxy models, using offline data assimilation.

First, we examined the feasibility using an observation system simulation experiment (OSSE). The analysis showed a significant improvement compared with the first guess in the reproducibility of isotope ratios in the proxies, as well as the temperature and precipitation fields, when only the isotopic information was assimilated. The ~~accuracy~~reconstruction skill for temperature and precipitation was especially high at low latitudes. This is due to the fact that isotopic proxies are strongly influenced by temperature and/or precipitation at low latitudes, which, in turn, are modulated by the El Niño-Southern Oscillation (ENSO) on interannual timescales. ~~The proxy temperature DA had comparable or higher accuracy than the reconstructed temperature DA.~~

Subsequently, the proxy DA was ~~compared~~conducted with real proxy data. The reconstruction ~~accuracy~~skill was decreased compared to the OSSE. In particular, the decrease was significant over the Indian Ocean, eastern Pacific, and the Atlantic Ocean where the reproducibility of the proxy model was lower. By changing the experimental design in a stepwise manner, the ~~decrease in accuracy~~decreased skill was ~~found~~suggested to be attributable to the misrepresentation of the atmospheric and proxy models~~. In addition, the accuracy was also dependent on the number~~ and/or ~~distribution~~the quality of the ~~proxies to be assimilated. Thus,~~observations. Although there remains a lot to improve ~~climate~~proxy DA, ~~it~~the result adequately showed that proxy DA is ~~necessary to enhance the performance of models, as well as to increase the number of proxies~~feasible enough to reconstruct past climate.

1. Introduction

Knowledge of past conditions is crucial for understanding long-term climate variability. Historically, two approaches have been used to reconstruct paleoclimate; one based on the empirical evidence contained in proxy data, and the other based on simulation with physically-based climate models. Recently, an alternative approach combining proxy data and climate simulations using a data assimilation (DA) technique has emerged. DA has long been used for forecasting weather and is a well-established method. However, the DA algorithms used for weather forecasts cannot be directly applied to paleoclimate due to the different temporal resolution, spatial extent, and type of information contained within observation data (Widmann et al., 2010). The temporal resolution and spatial distribution of proxy data are significantly lower (seasonal at best) and sparser than the present-day observations used for weather forecasts, and the information we can get does not measure the direct states of climate (e.g., temperature, wind, pressure, etc.), but represents proxies of those states (e.g., tree-ring width, isotopic composition in ice sheets, etc.). Thus, DA applied to paleoclimate is only loosely linked to the methods used in the more mature field of weather forecasting, and it has been developed almost independently from them.

Several DA methods have been proposed for paleoclimate reconstruction (von Storch et al., 2000; van der Schrier et al., 2005; Dirren and Hakim, 2005; Goosse et al., 2006; Bhend et al., 2012; Dubinkina and Goosse, 2013; Steiger et al., 2014), and paleoclimate studies using DA have successfully determined the mechanisms behind climate changes (Crespin et al., 2009; Goosse et al., 2010; 2012; Mathiot et al., 2013). In previous studies, the variables used for assimilation have been data reconstructed from proxies (e.g., surface air temperature) because observation operators or forward models for proxies have not been readily available. Hereafter, the DA method that assimilates reconstructed data from proxies is referred to as reconstructed DA. Recently, proxy modelers have developed and evaluated several forward models ~~for stable water isotopic proxies~~ (e.g., Dee et al., 2015 and references therein). ~~In this study, we attempted~~Thanks to that, currently a few studies have started attempting to assimilate proxy data directly ~~for the first time.~~ (Acevedo et al., 2016; Dee et al., 2016).

The main advantage of proxy DA over reconstructed DA is the richness of information used for assimilation. In previous studies, only a single reconstructed field was assimilated. However, proxies are influenced by multiple variables. Hence, the assimilation of a single variable does not use the full information recorded in the proxies.

The reconstruction method itself also limits the amount of information. The most commonly-used climate reconstruction is an empirical and statistical method that relies on the relationships between climate variables and proxies observed in present-day observations. These relationships are then applied to the past climate proxies to reconstruct climate prior to the instrumental period. Most of the studies using this approach assume that the relationship is linear. However, this assumption imposes considerable limitations in which specific climate proxies can be used, and proxies that do not satisfy the assumption have generally been omitted (e.g., PAGES 2k Consortium, 2013). Because information on paleoclimate is scarce, it is desirable to use as much information as possible.

Furthermore, the reconstruction method also limits the quality of information provided. The method also assumes stationarity of the relationship between the climate and the proxies. However, this assumption has been shown to be invalid for some cases (e.g., Schmidt et al. 2007; LeGrande and Schmidt, 2009). In the case of reconstructed DA, the assimilation of such ~~erroneous~~questionable reconstructed data would provide unrealistic results. In the case of proxy DA; however, the ~~accuracy~~skill of the assimilation is expected to be unchanged, provided the model can correctly simulate the non-stationarity.

The concept of proxy data assimilation is not new, and has been proposed in previous studies (Hughes and Ammann, 2009; Evans et al., 2013; Yoshimura et al., 2014; Dee et al., 2015). Yoshimura et al. (2014) demonstrated that the ~~accuracy of the simulation results increased following~~ assimilation of the stable water isotope ratios of vapor improves the analysis for current weather forecasting. They performed an observation system simulation experiment (OSSE) assuming that isotopic observations from satellites were available every six hours. Because the isotope ratio of water is one of the most frequently used climate proxies, this represents a significant first step toward improving the performance of proxy data assimilation in terms of identifying suitable variables for assimilation. However, it is not yet clear whether it is feasible to constrain climate only using isotopic proxies whose temporal resolution and spatial coverage are much longer and sparser than those of the specific study.

This study examined the feasibility of isotopic proxy DA for the paleoclimate reconstruction on the interannual timescale. Because the study represents one of the first ~~attempt~~attempts to assimilate isotopic variables on this timescale, we adopted the framework of an OSSE, as in previous climate data assimilations (Annan and Hargreaves, 2012; Bhend et al., 2012; Steiger et al., 2014~~).~~; Acevedo et al., 2016b; Dee et al., 2016). After the evaluation of proxy DA in the idealized way, we conducted the study with "real" proxy DA. We investigated which factors decreased or increased the ~~accuracy~~skill of the proxy DA. As a measure of skill, we report the correlation coefficient throughout the manuscript.

In this study, we used only oxygen isotopes (^{18}O) as proxies. The isotope ratio is expressed in delta notation ($\delta^{18}O$) relative to Vienna Standard Mean Ocean Water (VSMOW) throughout the manuscript. If the original data were expressed in delta notation relative to Vienna Pee Dee Belemnite (VPDB), they were converted to the VSMOW scale.

This paper is structured as follows. In the following section, the data assimilation algorithm, models, data, and experimental design are presented. Section 3 shows the results of the idealized experiment. Section 4 gives the results of the real proxy DA. The Discussion is presented in Section 5. Finally, we present our conclusions in Section 6.

2. Materials and methods

2.1. Data assimilation algorithm

We used ~~the so-called “offline data assimilation” algorithm to assimilate time-averaged data. In offline data assimilation~~a variant of ensemble Kalman filter (EnKF, see Houtekamer and Zhang, 2016, and references therein); sequential ensemble square root filter (EnSRF; Whitaker and Hamill, 2002). EnSRF updates the ensemble mean and the anomalies from the ensemble mean separately, and processes observations serially one at a time if the observations have independent errors.

To assimilate time-averaged data, slight modification was made for the method following Bhend et al. (2012) and Steiger et al. (2014). In the modified EnSRF, the analysis procedure is not cycled to the simulation (Bhend et al., 2012); thus, the background ensembles can be constructed from existing climate model simulations (Huntley and Hakim, 2010; Steiger et al., 2014). As such, we can assimilate data with any temporal resolution coarser than the model outputs. In this study, we focused on annual ~~data assimilation. Following the procedure proposed by Steiger et al. (2014),~~DA.

There are two ways to construct the background ensemble ~~was taken from part of~~in the approach mentioned above (hereafter offline DA); one using ensemble runs as in weather forecasts (Bhend et al., 2012; Acevedo et al., 2016) and the other using a single ~~climate model simulation~~run (Steiger et al., 2014; Dee et al., 2016). The latter uses the same background ensemble for every analysis step. To reduce computational cost, we chose the latter way, where the ensemble members ~~were~~are individual years ~~instead of independent model simulations. This algorithm was selected to reduce computational costs~~. This simplification was valid because the interannual variability in a single run was inherently indistinguishable from the variability in the annual mean within the ensemble of simulations in which the initial conditions were perturbed~~.~~, at least for atmospheric variables. Thus, the background ensembles were the same for all ~~of~~ the reconstruction years and did not contain any year-specific boundary conditions and forcing information; hence, the background error covariance was constant over time. Therefore, this study did not consider non-stationarity between the proxies and climate. Despite the limitations of the algorithm used in this study, it should be noted that the proxy DA could address non-stationarity ~~by changing the algorithm.~~if one uses temporally varying background ensemble. We return to this point in Section 5.

To control spurious long-distance correlations due to sampling errors, a localization function proposed by Gaspari and Cohn (1999) with a scale of 12,000 km was used. The detailed procedure used for the algorithm is described in Steiger et al. (2014).

2.2. Models

Isotope ratios recorded in ice cores, corals, and tree-ring cellulose were assimilated. To assimilate these variables, forward models for the variables are required. We used the forward model developed by Liu et al. (2013; 2014) for corals, and Roden et al. (2000) for tree-ring cellulose. We assumed that the isotopic composition of ice cores was the same as that of precipitation at the time of deposition. Note that, in reality, the isotope ratio recorded in ice cores is not always equal to that in precipitation due to post-depositional processes (e.g., Schotterer et al., 2004). Because detailed models that explicitly simulate the impact of all the processes involved in determining the value of the ratio are not yet available, we used the isotope ratio in precipitation for that in ice cores to avoid adding unnecessary noise.

The isotopic composition in precipitation was simulated using an atmospheric general circulation model (GCM) into which the isotopic composition of vapor, cloud water, and cloud ice are incorporated as prognostic variables. The model explicitly simulates the isotopic composition with all the details of the fractionation processes combined with atmospheric dynamics and thermodynamics, and hydrological cycles. Hence, the model simulates the isotopic composition consistent with the modeled climate. Although many such models have been developed previously (Joussaume et al., 1984, Jouzel et al., 1987; Hoffmann et al., 1998; Noone and Simmonds, 2002; Schmidt et al., 2005; Lee et al., 2007; Yoshimura et al., 2008; Risi et al., 2010; Werner et al., 2011), we used a newly-developed model (Okazaki et al., in prep.) based on the atmospheric component of MIROC5 (Watanabe et al. 2010). The spatial resolution was set to T42 (approximately 280 km) with 40 vertical layers.

The variability in $\delta^{18}O$ recorded in coral skeleton aragonite ($\delta^{18}O_{coral}$) depends on the calcification temperature and local $\delta^{18}O$ in sea water ($\delta^{18}O_{sw}$) at the time of growth (Epstein and Mayeda, 1953). Previous studies have modeled $\delta^{18}O_{coral}$ as the linear combination of sea surface temperature (SST) and $\delta^{18}O_{sw}$ (e.g., Julliet-Leclerc and Schmidt, 2001; Brown et al., 2006; Thompson et al., 2011), as follows:

$$\delta^{18}O_{coral} = \delta^{18}O_{sw} + aSST \quad (1)$$

where a is a constant which represents the slope between $\delta^{18}O_{coral}$ and SST. In this study, the constant was uniformly set to -0.22‰/°C for all the corals, following Thompson et al. (2011), and we used a model developed by Liu et al. (2013; 2014) to predict $\delta^{18}O_{sw}$. The model is an isotopic mass balance model that considers evaporation, precipitation, and mixing with ~~deep~~deeper ocean water. The coral model uses the monthly output of the isotope-enabled GCM as its input, except for the isotope ratio of ~~deep~~deeper ocean water, which was obtained from observation-based gridded data compiled by LeGrande and Schmidt et al. (2006). After the model calculates the monthly $\delta^{18}O_{coral}$, it is arithmetically averaged to provide the annual $\delta^{18}O_{coral}$.

The isotope ratio in tree-ring cellulose ($\delta^{18}O_{tree}$) was calculated using a model developed by Roden et al. (2000). In this model, $\delta^{18}O_{tree}$ is determined by the isotopic composition of the source water used by trees for photosynthesis, and evaporative enrichment on leaves via transpiration. In this study, the value of the isotopic composition in the source water was arbitrarily assumed to be the moving average, traced three-months backward, of the isotopic composition in precipitation at the site. Again, the model used the monthly output of the isotope-enabled GCM as its input. After performing the tree-ring model calculation, the monthly output was weighted using climatological net primary production (NPP) to calculate the annual average. The NPP data were obtained from the US National Aeronautics and Space Administration (NASA) Earth Observation website (http://neo.sci.gsfc.nasa.gov).

Because the isotopic compositions of the proxies were simulated using the output of the isotope-enabled GCM, their horizontal resolution was the same as that of the GCM.

2.3. Experimental design

2.3.1. Control experiment

The first experiment served as a control (CTRL) experiment, and used the framework of an OSSE. In the experiment, the "simulation" and the "truth" (nature run) were simulated by the same models, with the same forcing, but with different initial conditions. Because the proxy models were driven by the output of the GCM, the modeled proxies were consistent with the modeled climate from the GCM. Thus, here we describe the experimental design for the GCM. The GCM was driven by observed SST and sea-ice data (HadISST; Rayner et al., 2003), and historical anthropogenic (carbon dioxide, methane, and ozone) and natural (total solar irradiance) forcing factors. The simulation covered the period of 1871–2007 (137 years).

Although the simulation period included recent times covered by observational data, we assumed that the only variable that could be obtained was the annual mean of $\delta^{18}O$ in the proxies. We based this assumption on the fact that we wished to perform the DA for a period in which no direct measurements were available, and there were only climate proxies covering the period. Therefore, the temporal resolutions of the “observations” and “simulations” were also annual, considering the typical temporal resolution of the proxies.

Observations were generated by adding Gaussian noise to the truth. The spatial distribution of the observations mimicked that of the proxies. The spatial distributions of each proxy for various periods are mapped in Figure 1. As can be seen from the figure, the distributions and the number of proxies varied with time. However, for the sake of simplicity, the distributions of the proxies were assumed to be constant over time in the CTRL experiment (Figure 1 a). The size of the observation errors will be discussed in Section 2.4.

The state vector consisted of five variables; surface air temperature and amount of precipitation, as well as the isotopic composition in precipitation, coral, and tree-ring cellulose. The first three variables were obtained from the isotope-enabled GCM, and the other two variables were obtained from the proxy models driven by the output of the GCM.

2.3.2. Real proxy data assimilation

The second (REAL) experiment assimilated proxy data sampled in the real world. To mimic realistic conditions, SST and sea-ice concentration data to be used as model forcing were modified from observational to modeled data. In reality, there were no direct observations available for the target period of the proxy DA. Therefore, to reliably evaluate the feasibility of proxy DA, the first estimate should be constructed using modeled SST, as opposed to observed SST. We used SST data from the historical run of the Coupled Model Intercomparison Project Phase 5 (CMIP5; Taylor et al., 2007) from the atmosphere-ocean coupled version of MIROC5 (Watanabe et al., 2010) obtained from the CMIP5 data server (https://pcmdi.llnl.gov/search/cmip5/).

Because the experiment was not an OSSE, nature run was not necessary.

2.3.3. Sensitivity experiments

Four sensitivity experiments were conducted to test the robustness of the results of the proxy DA. In the first sensitivity experiment (CGCM), the simulation run was constructed from the simulation forced by the modeled SST and sea ice as in the REAL experiment. The other settings for the simulation run were the same as those in the CTRL experiment. The nature run was the same as that of the CTRL experiment. Thus, this experiment investigated how the ~~accuracy~~reconstruction skill of the results was decreased by using the simulated SST~~.–~~ compared to the CTRL.

In the second sensitivity experiment (VOBS), the experimental design was the same as that in the CGCM, except for the number of proxies that were assimilated. In the CGCM experiment, the distribution and number of proxies were set to be constant over time, as in the CTRL experiment. In the VOBS experiment, the distribution and number of proxies varied with time ~~to reliably evaluate the results of the REAL experiment relative to those from the CTRL experiment.~~. Thus, this experiment investigated how the reconstruction skill was decreased by changing the number of proxies compared to the CGCM.

In the third sensitivity experiment (T2-Assim), ~~the~~ reconstructed surface temperature ~~added with Gaussian noise~~(T_r) was assimilated. The purpose of the experiment was to compare the ~~accuracy~~skill of the reconstructed DA with that of the proxy DA. The experimental design was the same as that in the CTRL experiment, except for the variables that were assimilated. The ~~noise~~reconstructed temperature was ~~added to~~

~~consider the uncertainties stemmed from the reconstruction. The size~~generated with a linear regression model of ~~error was determined by considering the typical signal-to-noise~~ $T_r = a + b \times \delta^{18}O$ where a and b are coefficients and $\delta^{18}O$ is the observed isotope ratio ~~(SNR) values~~. The coefficients are calibrated with the observed isotope ratio and the true temperature in the CTRL for the period of 1871 to 1950 (80 years). If the correlation between the isotope ratio and the temperature during the calibration period was not statistically significant ($p <$ 0.~~25 and 0.50 (~~10), the data was discarded following Mann et al~~., 2007), as well as a further value of 1.0.~~. (2008). This screening process reduced the available data from 94 to 81 grid points.

The final sensitivity (M08) experiment was used to examine the sensitivity to the observation network. The experimental design was the same as for the CTRL, except for the spatial distribution of the proxy. The proxy network used in the experiment was the same as that of Mann et al. (2008). We assumed that isotopic information was available for all the sites, even when this was not the case. For example, even if only tree-ring width data were available at some of the sites in Mann et al. (2008), in this experiment we assumed that isotopic data recorded in tree-ring cellulose were available at the site. The number of grids containing observations were ~~108~~94 and 250 for the CTRL experiment and M08 respectively. The T2-Assim and the M08 were compared with CTRL.

The experimental designs are summarized in Table 1.

2.4. Observation data

We used paleoclimate data archived at the National Oceanic and Atmospheric Administration (NOAA; https://www.ncdc.noaa.gov/data-access/paleoclimatology-data) and data used in the PAGES 2k Consortium (2013). Additionally, 22 tree-ring cellulose and 7 ice core data sets were collected separately from published papers. We only used oxygen isotopic data (^{18}O) whose temporal resolution was higher than annual; proxies whose resolution was lower than annual were excluded. The full list of proxies used in this study is given in the Appendix. Following Crespin et al. (2009) and Goosse et al. (2010), all proxy records were first normalized, and then averaged onto a T42 grid box to eliminate model bias and produce a regional grid box composite. To compare the results from each experiment effectively, the assimilated variables were all normalized in both the simulation and nature runs, and in the observations in all the experiments.

Errors were added to the truth in a normalized manner to provide the observation~~.~~ for all the experiment other than REAL. The normalized error was uniformly set to 0.50 for all the proxies. This was based on the measurement error of δ^{18}O in ice cores being reported to range from 0.05 to 0.2‰ (e.g., Rhodes et al., 2012; Takeuchi et al., 2014), and the corresponding normalized error (measurement error divided by standard deviation of proxy) then ranges from 0.03 to 0.1, with an average of 0.06. Similarly, the measurement error of $\delta^{18}O$ in coral ranges from 0.03 to 0.11‰ (e.g., Asami et al., 2004; Goodkin et al., 2008), and the corresponding normalized error ranges from 0.24 to 1.1, with an average of 0.53. The measurement error of $\delta^{18}O$ in tree-ring cellulose ranges from 0.1 to 0.3‰ (e.g., Managave et al, 2011; Young et al, 2015), and the corresponding normalized error ranges from 0.08 to 0.55, with an average of 0.28. In practice, due to the error of representativeness and that in observation operator, it is common to increase the observation errors to ensure that the analysis functions effectively (Yoshimura et al., 2014). Furthermore, the measurement errors were not always available; therefore, a uniform value of 0.5 was used for all the proxies. The corresponding signal-to-noise ratio (SNR) is 2.0. The errors are assumed to be independent for all the experiments.

3. Results from the OSSE

The time series of the first estimation, the analysis, and the real values for $\delta^{18}O$ in corals are compared as an example in Figure 2 at a location where observational data were available (1°N, 157°W). Because the first estimate was the same for all reconstruction years, it is drawn as horizontal lines. After the assimilation, the analysis agreed well with the real values (R = 0.96, $p < 0.001$). This confirmed that the assimilation performed well. We then examined how accurately the other variables were reconstructed by assimilating isotopic information. Figure 2 also shows the time series of surface air temperature and precipitation for the same site. There was a clear agreement between the analysis and the truth for both variables (R = 0.92 and 0.88 respectively for temperature and precipitation). This indicated that temperature and precipitation were effectively reconstructed by assimilating isotopic variables at this site. This was because the isotope ratio in corals has a signature not only from temperature as given in Eq. 1, but also precipitation (Liu et al., 2013); the correlation with $\delta^{18}O_{coral}$ was -0.88 (p <0.001) for both temperature and precipitation, respectively. This example shows that the isotopic proxy records more than one variable.

Figure 3 maps the correlation coefficients between the analysis and the truth for the isotope ratio, temperature, and precipitation for 1970–1999. Because the first estimate was constant over time, the temporal correlation between the first estimate and the real value was zero everywhere. Thus, a positive correlation indicated that the DA improved the simulation.

The correlation for $\delta^{18}O$ in precipitation were high at the observation sites, regardless of the proxy type. This was because $\delta^{18}O$ in both corals and trees is affected by the isotopic composition in precipitated water derived from sea water or soil water. The correlation for $\delta^{18}O$ in tree-ring cellulose were also high at the observation sites. On the other hand, the high correlation for $\delta^{18}O$ in corals were not limited around the observation sites but were generally high at low- to mid-latitudes~~, and~~. Similarly, the ~~spatial pattern~~correlation was ~~similar~~high at low- to ~~that of~~mid-latitudes for surface temperature. The correlation was also statistically significant ($p < 0.05$) around the observation sites in high latitude. In contrast, closely correlated areas were restricted to low-latitude for precipitation.

How can the spatial distribution of the correlation pattern be explained; i.e., what do the proxies represent? To investigate this question, empirical orthogonal function (EOF) analysis was conducted for the simulated $\delta^{18}O$ in precipitation, corals, and tree-ring cellulose. Only grids that contained observations were included in the analysis. The variables were centered around their means before the analysis. The data covered the period 1871–2007. The EOF patterns and temporal correlations between surface temperature and the characteristic evolution of EOF, or the principal components (PCs) of the first mode of each proxy are shown in Figure 4.

The first mode of $\delta^{18}O$ in ice core explains 14.3% of the total variance ant it is the only significant mode according to the Rule of Thumb (North et al., 1982) (the first and the second mode were indistinguishable). The maximum loadings were in Greenland and

Antarctica where temperature increase has been ~~increasing significantly~~observed for the past hundred years~~.~~ (e.g. Hartmann et al., 2013). Indeed, the PC1 shows the significant trend and is correlated with global mean surface temperature (R=0.44, p < 0.001). Therefore, it is legitimate to regard ice core data as a proxy of global temperature as revealed from observation (Schneider and Noone, 2007).

The first modes of δ^{18}O in corals, and tree-ring cellulose represent ENSO. The explained variance of the first modes of δ^{18}O in corals, and tree-ring cellulose was 44.2, and 19.0%, respectively. The maximum loadings occurred in the central Pacific for corals, and Tibet for tree-ring cellulose. The temporal correlation between the PC1s and NINO3 index were 0.95, and 0.37 for corals and tree-ring cellulose, respectively. Because the isotopic composition in corals is influenced by sea temperature, it is expected that the δ^{18}O in corals from the central Pacific records the ENSO signature. Interestingly, the analysis revealed that the δ^{18}O in tree-ring cellulose was also influenced by ENSO; hence, this proxy contributes to the reconstruction of temperature and precipitation over the tropical Pacific. Indeed, many previous studies have reported the link between δ^{18}O in tree-ring cellulose and ENSO (Sano et al. 2012; Xu et al. 2011; 2013; 2015). ~~The link was explained as follows by Xu et al. (2011): Numerous studies have associated Indian monsoon rainfall with ENSO (e.g., Rasmusson and Carpenter 1983), albeit the~~

~~relationship was found to be non-stationary over time (Kumar, 1999).~~Xu et al. (2011) inferred the link is caused by the association between ENSO and Indian monsoon rainfall (e.g. Rasmusson and Carpenter, 1983). The positive phase of ENSO results in a decrease in summer monsoon rainfall in India, which leads to dry conditions in summer. The decrease in precipitation leads to isotopically-enriched precipitation, and the dry conditions enhance the enrichment of water in leaves. Correspondingly, the $\delta^{18}O$ in tree-ring cellulose becomes heavier than normal in the positive phase of ENSO. Due to the relationships between the coral and tree-ring cellulose data and ENSO, the correlation coefficient between the analysis and ~~real values~~the truth for the NINO3 index was as high as 0.95 ($p < 0.001$).

Although EOF analysis did not reveal any other significant correlation between PCs and climate indices, climate indices for the North Atlantic Oscillation and Southern Annular Mode calculated using the reconstructed data were significantly correlated with the truth (0.59 and 0.46, respectively).

4. Real proxy data assimilation

Based on the results of the idealized experiment described in the previous section, we performed a “real” proxy DA, in which sampled and measured data in the real world were assimilated.

The temporal correlation between the analysis and observations for temperature and precipitation are shown in Figure 5 (d, h). The observations were obtained from HadCRUT3 (Brohan et al., 2006) for temperature, and GHCN-Monthly Version 3 (Peterson and Vose, 1997~~).~~) for precipitation.

Although the real proxy DA had reasonable ~~accuracy~~skill, it was inferior relative to the CTRL experiment. We investigated the cause of the decreased ~~accuracy~~skill using the outputs of the sensitivity experiments. The design of the experiments was changed in a stepwise fashion to more realistic conditions of proxy data assimilation from the idealized conditions. The correlations between the analysis and the truth, or the observation, for the experiments are shown in Figure 5. The truths for the CGCM and VOBS experiments were the same as those for the CTRL experiment. The global mean correlation coefficients for temperature~~ and~~, precipitation, and NINO3 in the experiments are summarized in Figure 6. Note that the correlation was averaged in the same domain for all the experiments to take into account the differences in representativeness.

In the CGCM experiment, the temporal correlations between the analysis and the ~~real values~~truth were similar to those in the CTRL experiment for both temperature and precipitation (Figure 5 b, f). This indicates that ENSO and its impacts were well represented in the modeled SST used to construct the “simulation”. Watanabe et al. (2010) reported similar modeled SST and observational values for the amplitude of ENSO measured by the NINO3 index, and the spatial patterns of the temperature and precipitation fields regressed on the NINO3 time series (see Figures 13 and 14 in their report).

Because the number of proxies for assimilation differed from that in the CGCM experiment, it was not straightforward to compare the results of the REAL experiment with those of the CGCM experiment. To enable an effective comparison of the results, the same number of proxies were assimilated in the VOBS experiment as in the REAL experiment and the same settings were used as in the CGCM experiment for the other variables. Consequently, the performance of the assimilation of the VOBS experiment was similar to that of the CGCM experiment for 1970–1999. ~~Because the number of proxies for assimilation was similar for this period, the assimilation of the VOBS experiment performed well.~~

When the REAL and VOBS experiments were compared, the correlation coefficients for temperature were significantly decreased over the Indian Ocean, eastern Pacific, and Atlantic Ocean. These areas corresponded to areas of low reproducibility in the coral model (Liu et al, 2014). The effects of sea current and river flow in these areas, which were not included in the coral model, were deemed to be considerable. Although we cannot attribute all the decreased skill to the coral model, the reproducibility of $\delta^{18}O$ in corals in these areas requires improvement to enhance the performance of the assimilation.

5. Discussion

5.1. Comparison with the reconstructed temperature assimilation

Hughes and Ammann (2009) recommended assimilating measured proxy data, as opposed to reconstructed data derived from the proxy data. This subsection compares the results from the CTRL and T2-Assim experiments~~ with three different SNR values. Both experimental frameworks were OSSE, and the observations and reconstructed temperature were assumed to be available for the same sites as in the CTRL experiment. To account for the uncertainty derived from the statistical reconstruction, Gaussian noise was added to the temperature from the nature run to generate the observational values in the T2-Assim experiment in a similar fashion to the CTRL experiment. The SNR of the reconstructed temperature was set to 0.25 and 0.50, which are typical values for proxy records (e.g., Mann et al., 2007). Additionally, we also considered an SNR value of 1.0.~~.

Figure 7 shows the spatial distribution of the correlation coefficients for temperature and precipitation between the truth and the analysis for each experiment. ~~The global mean~~

~~correlation coefficients for temperature (precipitation) were 0.49 (0.29), 0.50 (0.22), 0.39 (0.16), and 0.25 (0.10) for the experiments assimilating $\delta^{18}O$ in proxies, and those assimilating temperature with SNR values of 1.0, 0.50, and 0.25, respectively (Figure 8). The values were higher for the assimilated $\delta^{18}O$ in proxy than for assimilated temperature, with SNR values of 0.25 and 0.50 for both precipitation and temperature. The temperature was reconstructed slightly accurately by assimilation of temperature with a low noise value (SNR = 1.0) than by assimilation of $\delta^{18}O$ in the proxies. Although using an SNR = 1.0 produced more accurate reconstructed field than the ordinal statistical reconstruction, the superior accuracy of the assimilation of proxy data relative to the assimilation of reconstructed temperature was dependent on the magnitude of the SNR; i.e., the accuracy of assimilation of the reconstructed values was dependent on the quality of the reconstructed data. The quality of the reconstructed data was in turn dependent on the stationarity between the proxies and climate, and the degree to which the proxy was affected by factors other than the variable of interest.~~As a whole, the reconstruction skill was slightly degraded in T2-Assim compared with CTRL with the global mean correlation coefficients for temperature (precipitation) of 0.50 (0.30), 0.45 (0.23), for CTRL and T2-Assim, respectively. On the other hand, the skill of proxy DA was not always better than that of T2-Assim (e.g. temperature in tropical Atlantic Ocean). Those pros and cons can be explained by the difference in the observation error and the structure of Kalman gain. Figure 8 shows the SNR of the T_r ranging from 0.22 to 1.6 with the average of 0.65. Accordingly, the observation error is larger than that of CTRL everywhere, and this resulted in the reduction of the reconstruction skill. On the other hand, the better skill in T2-Assim should be owing to the difference in Kalman gain. The Kalman gain determines analysis increments by spreading the information in observations through the covariance between the prior and the prior-estimated observations. We found that the correlations between the prior (temperature) and the prior-estimated observation (temperature and $\delta^{18}O$ for T2-Assim and CTRL, respectively) were consistently high in T2-Assim than in CTRL (not shown) as Dee et al. (2016) showed. Thus, the information in the observations were more effectively spread to the analysis in T2-Assim, and this resulted in the improved skill. Note that the screening process hardly hampered the reconstruction skill, because even if the reconstructed temperature was fully used (i.e. not screened), the skills were almost the same as T2-Assim.

Conducting similar experiments, Dee et al. (2016) also concluded that the reconstruction skills were almost the same among proxy DA and reconstructed DA if the relation between the reconstructed variable and the proxy is linear. As isotope-enabled GCMs (Schmidt et al. 2007; LeGrande and Schmidt. 2009) and observations and models for tree-rings width (D'Arrigo et al. 2008; Evans et al. 2014; Dee et al., 2016) have demonstrated ~~non-stationarity and non-linearity~~, however, the relations between the proxies and climate~~.~~ are non-linear and non-stationary as well. Thus, ~~we cannot~~ it is difficult to expect that ~~a high SNR~~the skill of reconstructed DA will be ~~maintained over time. However, stationarity and linearity do not~~ the same as that of proxy DA if we have ~~to be considered if~~ the well-defined forward proxy ~~model is well-defined~~ models (Hughes and Ammann, 2009). ~~Therefore~~Although the current models are far from perfect as implicated in Sect. 4.2, the assimilation of proxy data ~~offers~~will offer a useful tool for the reconstruction of paleoclimate, in which the relationship between the proxies and climate constructed with the present-day conditions does not apply.

5.2. Sensitivity to the distribution of the proxies

The ~~accuracy~~skill of the proxy DA was relatively low over Eurasia and North America, even in the idealized experiment. It was unclear whether this was because of limitations in the proxy data assimilation or the scant distribution of the proxies. This subsection investigates the reasons for the relatively low reproducibility in these areas by comparing the results of the CTRL and M08 experiments, focusing on North America. The number of grids for which proxy data were available over North America was 11 and 126 for the

CTRL and M08, respectively.

The results for North America are shown in Figure 9. The figure shows the temporal correlation coefficients between the analysis and the truth for surface air temperature and precipitation. The correlation coefficients were calculated for 1970–1999. The ~~accuracy~~skill was high in the area in which the proxies were densely distributed for both variables. The values of the coefficients averaged over the United States (30–50°N, 80–120°W) were 0.~~68~~69 and 0.~~52~~53 for temperature and precipitation, respectively. Compared to the ~~CTRL experiment, the accuracy was enhanced for both variables. The values of the~~ coefficients ~~were~~of 0.~~17~~23 and 0.~~24~~26, respectively, in the CTRL experiment~~.~~, the skill was enhanced for both variables. This implies that the performance of the reconstruction was strongly dependent on the distribution of the proxy data. Taking into consideration that proxy DA can assimilate not only proxy data~~,~~ but also reconstructed data, proxy DA can take advantage of the use of increasingly large amounts of data. Although it is beyond the scope of this study, the combined use of these data is expected to improve the performance of proxy DA.

6. Conclusion and summary

The feasibility of using proxy DA for paleoclimate reconstruction was examined in both idealized and real conditions experiments. The idealized (CTRL) experiment had high ~~accuracy~~skill at low latitudes due to the dependency of coral data on temperature and precipitation in these regions, and the correlation between ENSO and $\delta^{18}O$ in corals in Pacific and tree-ring cellulose in Tibet. ~~We performed additional experiments to examine the robustness of proxy DA. In the first experiment, the simulation run was constructed from a simulation forced by modeled SST and sea ice (CGCM experiment). The experiment examined the extent to which the accuracy of the results was decreased using the simulated forcings. The results showed little difference between the performance of the reconstruction for both the temperature and precipitation fields~~Encouraged by the results~~. This was because ENSO, which is the most important mode for the reconstruction, was well represented in the modeled SST. Finally~~, real proxy DA was performed, where the simulation run was constructed from the simulation forced by the modeled SST, and the real (observed) proxy data were assimilated into the simulation (REAL experiment). The ~~accuracy~~skill of the reconstruction decreased ~~over the Indian Ocean, eastern Pacific, and the Atlantic Ocean, where the reproducibility of the proxy model was lower.~~ compared to CTRL.

~~The results indicated the need~~To investigate the reason for the relatively low skill in REAL compared to CTRL, we performed additional experiments; CGCM and VOBS.

The imperfect SST used to drive the CGCM experiment resulted in a slight reduction of the skill compared to the CTRL experiment with perfect SST. This was because ENSO, which is the most important mode for the reconstruction, was well represented in the modeled SST. ~~improve~~ The result is encouraging because to apply the DA system to reconstruct ages where no instrumental observation is available, we must rely on SST simulated by coupled GCM. Similarly, assimilating the unfixed number of the observation only slightly decreased the reconstruction skill as shown in the comparison between CGCM and VOBS.

From the suite of experiments, more than half of the difference between CTRL and REAL remained unexplained. This remained difference can have a lot of origins: e.g. errors in the isotope~~-enabled~~ incorporated atmospheric GCM~~ and~~, the proxy models~~. The differences between the CTRL and CGCM experiments were due to the use of misrepresented SST values by the coupled GCM. The differences between the CGCM and VOBS experiments were due to the large number of observations for assimilation. Finally, the differences between the VOBS and REAL experiments were due to the misrepresentation of the atmospheric GCM incorporating isotope and~~, the proxy data and so on. The errors in the models~~. The differences were largest between the VOBS and REAL experiments (Figure 6). Although it is difficult at this stage to conclude which~~

include such as model ~~caused the decrease in accuracy, it is necessary to improve the reproducibility of models in these regions, and we will investigate the reproducibility of each~~biases and too simplified or totally lacked model ~~in future studies.~~components. For instance, the coral model does not take into account the impact of ocean current or river runoff in this study. Furthermore, ~~accurate models for ice cores that incorporate the entire~~ post-depositional processes for simulating isotope ratio in ice core were not included at all. Those processes should be ~~developed~~included to enable more efficient utilization of all ~~of~~ the data.

~~In addition to model reproducibility, the~~ The errors in proxy data ~~may have contributed to the decrease in the accuracy of the proxy DA results by transferring erroneous values. It is possible that the data might not have been representative~~include such as misrepresentation of the targeted temporal and/or spatial scales. ~~Furthermore,~~ It is also possible that the data were highly distorted by non-climatic ~~factor(s).~~factors. Thus, a thorough quality control, similar to the procedures used in weather forecasting, should be conducted before assimilation~~.~~ (e.g. Appendix B of Compo et al., 2011). At this stage, it is difficult to show the relative contributions of each factor to the degraded skill in REAL, it is necessary to estimate the impact of structural errors in models as done in Dee et al. (2016).

Although the ~~accuracy~~skill of proxy DA is dependent on the ~~REAL experiment was decreased compared with~~reproducibility of the ~~CTRL experiment,~~models and the number and quality of the observations, the results suggest that it ~~may still be possible~~is feasible to ~~reliably reconstruct~~ constrain climate using only proxies. Especially, ENSO and ENSO-related variations in temperature and precipitation ~~with this~~should be reliably reconstructed even with the current proxy DA system and proxy network used in this study because the correlation coefficient between the analysis and the observations was as high as 0.83 in the REAL experiment. Although the reconstruction of ENSO is dependent on data from corals, and the time span covered by corals is relatively short (a few hundred years), ENSO can still be reliably reconstructed due to its global impact, as was demonstrated in the relationship between isotopes in tree-ring cellulose from Tibet.

Moreover, ~~because~~ we expect that the reproducibility will increase as more proxy data become available because it was heavily dependent on the spatial distribution~~, we expect that it will increase as more proxy data become available. In this sense, because~~. Given that proxy DA can assimilate both proxy data and data reconstructed ~~data~~from proxy, and that the reconstruction skill in reconstructed DA is partly superior to proxy DA, the combined use of the two types of data is ~~expected to improve~~beneficial for the performance~~ of the assimilation.~~. In that case, care must be taken not to assimilate dependent information (e.g. proxy data and reconstructed data from the same proxy).

The DA algorithm used in this study did not consider non-stationarity among proxies and climate variables because the Kalman gain was constant over time. To address non-stationarity, the Kalman gain for a specific reconstruction year should be constructed for several tens of years before and after that year. ~~Furthermore, an ensemble Kalman filter (EnKF)~~Nevertheless, EnKF can only capture linear relationships between observations and the modeled state. The use of other algorithms, such as particle filter (e.g. van Leeuwen, 2009), or four-dimensional variational assimilation (e.g. Rabier et al., 2000), should be investigated in future studies for scenarios where non-linearity is not negligible. Thus, it is important in future studies to investigate non-stationarity and non-linearity among proxies and climate variables to identify suitable algorithms for proxy DA.

7. Acknowledgements

The first author was supported by the Japan Society for the Promotion of Science (JSPS) via a Grant-in-Aid for JSPS Fellows. This study was supported by the Japan Society for the Promotion of Science Grants 15H01729, 26289160, and 23226012, the SOUSEI Program, the ArCS project of MEXT, Project S-12 of the Japanese Ministry of the Environment, and the CREST program of the Japan Science and Technology Agency.

8. References

Acevedo, W., Reich, S., and Cubasch, U., Towards the assimilation of tree-ring-width records using ensemble Kalman filtering techniques, Clim. Dyn., 46, 1909-1920, 2016a.

Acevedo, W., Fallah, B., Reich, W., and Cubasch, U., Assimilation of pseudo-tree-ring-width observations into an atmospheric general circulation model, Clim. Past Discuss., 2016.

Annan, J. D. and Hargreaves, J. C., Identification of climatic state with limited proxy data, Clim. Past, 8, 1141-1151, 2012.

Asami, R., Yamada, T., Iryu, Y., Meyer, C. P., Quinn, T. M., and Paulay, G., Carbon and oxygen isotopic composition of a Guam coral and their relationships to environmental variables in the western Pacific, Palaeogeogr Palaeocl, 212, 15, 1-22, 2004.

Bhend. J., Franke, J., Folini, D., Wild, M., and Brönnimann, S., An ensemble-based approach to climate reconstructions, Clim. Past, 8, 963-976, 2012.

Brohan, P., Kennedy, J. J., Harris, I., Tett, S. F. B., and Jones, P. D., Uncertainty estimates in regional and global observed temperature changes: A new data asset from 1850, J. Geophys. Res. 111, D12106, 2006.

Brown, J., Simmonds, I., and Noone, D., Modeling $\delta^{18}O$ in tropical precipitation and the surface ocean for present-day climate, J. Geophys. Res., 111, D05105, 2006.

Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., Gleason, B. E, Jr., Vose, R. S., Rutledge, G., Bessemoulin, P., Brönnimann, S., Brunet, M., Crouthamel, R. I., Grnt, A. N., Groisman, P. Y., Jones, P. D., Kruk, M. C., Kruger, A. C., Marshall, G. J., Maugeri, M., Mok, H. Y., Nordli, Ø., Ross, T. F., Trigo, R. M., Wang, X. L., Woodruff, S. D., and Worley, S. J., The twentieth Century Reanalysis Project, Q. J. Roy. Meteor. Soc., 137, 1-28, 2011.

Crespin, E., Goosse, H., Fichefet, T., and Mann, M., The 15th century Arctic warming in coupled model simulations with data assimilation, Clim. Past, 5, 389-401, 2009.

D’Arrigo, R., Wilson, R., Liepert, B., and Cherubini, P., On the ‘Divergence Problem’ in Northern Forests: A review of the the tree-ring evidence and possible causes, Global Planet. Change, 60, 289-305, 2008.

Dee, S., Emile-Geay, J., Evans, M., Allam, A., Steig, E., and Thompson, D., PRYSM: An open-source framework for PRoxY System Modeling, with applications to oxygen-isotope systems, Journal of Advances in Modeling Earth Systems, 7, 1220-1247, 2015.

Dee, S., Steiger, N. J., Emile-Geay, J., and Hakim, G. J., On the utility of proxy system models for estimating climate states over the common era, Journal of Advances in Modeling Earth Systems, 8, 1164-1179, 2016.

Dirren, S. and Hakim, C., Toward the assimilation of time-averaged observations, Geophys. Res. Lett., 32, L04804, 2005.

Dubinkina, S. and Goosse, H., An assessment of particle filtering methods and nudging for climate state reconstructions, Clim. Past, 9, 1141-1152, 2013.

Epstein, S., and Mayeda, T., Variation of O^{18} content of waters from natural sources, Geochimica Cosmochim. Ac., 4, 213-224, 1953.

Evans, M. N., Tolwinski-Ward, S. E., Thompson, D. M., and Anchukaitis, K. J., Applications of proxy system modeling in high resolution paleoclimatology, Quaternary Sci. Rev., 76, 16-28, 2013.

Evans, M. N., Smerdon, J. E., Kaplan, A., Tolwinski-Ward, S. E., and González-Rouco, J. F., Climate field reconstruction uncertainty arising from multivariate and nonlinear properties of predictors, Goephys. Res. Lett., 41, doi:10.1002/2014GL062063, 2014.

Gaspari, G., and Cohn, S., Construction of correlation functions in two and three dimensions, Q. J. Roy. Meteor. Soc., 125, 723-757, 1999.

Goodkin, N. F., Hughen, K. A., Curry, W. B., Doney, S. C., and Ostermann, D. R., Sea surface temperature and salinity variability at Bermuda during the end of the Little Ice Age, Paleoceanography, 23, PA3203, 2008.

Goosse, H., Renssen, H., Timmermann, A., Bradley, R., and Mann, M., Using paleoclimate proxy-data to select optimal realisations in an ensemble of simulations of the climate of the past millennium, Clim. Dyn., 27, 165-184, 2006.

Goosse, H., Crespin, E., de Montety, A., Mann, M., Renssen, H., and Timmermann, A., Reconstructing surface temperature changes over the past 600 years using climate model simulations with data assimilation, J. Geophys. Res., 115, D09108, 2010.

Goosse, H., Crespin, E., Dubinkina, S., Loutre, M., Mann, M., Renssen, H., Sallaz-Damaz, Y., Shindell, D., The role of forcing and internal dynamics in explaining the "Medieval Climate Anomaly", Clim. Dyn., 39, 2847-2866, 2012.

Hartmann, D. L., Klein Tank, A. M. G., Rusticucci, M., Alexander, L. V., Brönnimann, S., Charabi, F., Dentener, F. J., Dlugokencky, E. J., Easterling, D. R., Kaplan, A., Soden, B. J., Thorne, P. W., Wild, M., and Zhai, P. M., Observations: Atmosphere and Surface. In: Climate Change 2013: The physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, United Kingdom and

New York, NY, USA, 2013.

Hoffmann, G., Werner, M., Heimann, M., Water isotope module of the ECHAM atmospheric general circulation model: A study on timescales from days to several years, J. Geophys. Res., 103, D1427, 16871-16896, 1998.

Houtekamer, P. L., and Zhang, F., Review of the ensemble Kalman filter for atmospheric data assimilation, Mon. Weather Rev., 144, 4489-4532, 2016.

Hughes, M., and Ammann, C., The future of the past -an earth system framework for high resolution paleoclimatology: editorial essay, Climatic Change, 94, 247-259, 2009.

Huntley, H., and Hakim, G., Assimilation of time-average observations in a quasi-geostrophic atmospheric jet model, Clim. Dyn., 35, 995-1009, 2010.

Joussaume, S., Sadourny, R., and Jouzel, J., A general circulation model of water isotope cycles in the atmosphere, Nature, 311, 24-29, 1984

Jouzel, J., Russell, G. L., Suozzo, R. J., Koster, R. D., White, J. W. C., and Broecker, W. S., Simulations of the HDO and $H_2{}^{18}O$ Atmospheric cycles using the NASA GISS General Circulation Model: The seasonal cycle for present-day conditions, J. Geophys. Res., 92, D12, 14739-14760, 1987.

Julliet-Leclerc, A., and Schmidt, G., A calibration of the oxygen isotope paleothermometer of coral aragonite from Porites, Geophys. Res. Let., 28, 21, 4135-413, 2001.

~~Kumar, K., Rajagopalan, B., Cane, M. A., On the weakening relationship between the Indian Monsoon and ENSO, Science, 284, 2156, 1999.~~

Lee, J.-E., Fung, I., DePaolo, D., and Henning, C., Analysis of the global distribution of water isotopes using the NCAR atmospheric general circulation model, J. Geophys, Res., 112, D16306, 2007.

LeGrande, A., and Schmidt, G., Global gridded data set of the oxygen isotopic composition in seawater, Geophys. Res. Lett., 33, L12604, 2006.

LeGrande, A., and Schmidt, G., Sources of Holocene variability of oxygen isotopes in paleoclimate archives, Clim. Past, 441-455, 2009.

Liu, G., Kojima, K., Yoshimura, K., Okai, T, Suzuki, A., Oki, T., Siringan, F., Yoneda, M., and Kawahata, H., A model-based test of accuracy of seawater oxygen isotope ratio record derived from a coral dual proxy method at southeastern Luzon Island, the Philippines, J. Geophys. Res-Biogeo., 118, 853-859, 2013.

Liu, G., Kojima, K., Yoshimura, K., and Oka, A., Proxy interpretation of coral-recorded seawater ^{18}O using 1-D model forced by isotope-incorporated GCM in tropical oceanic regions, 119, doi: 10.1002/2014JD021583, 2014.

Managave, S. R., Sheshshayee, M. S., Ramesh, R., Borgaonkar, H. P., Shad, S. K., Bhattacharyya, A., Response of cellulose oxygen isotope values of teak trees in differing monsoon environments to monsoon rainfall, Dendrochronologia, 29, 89-97, 2011.

Mann, M., Rutherford, S., Wahl, E., and Ammann, C., Robustness of proxy-based climate field reconstruction methods, 112, D12109, 2007.

Mann, M., Zhang, Z., Hughes, M., Bradley, R., Miller, S., Rutherford, S., Ni, F., Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia, P. Natl. Acad. Sci. USA., 105, 13252-13257, 2008.

Mathiot, P., Goosse, H., Crosta, X., Stenni, B., Braida, M., Renssen, H., Van Meerbeeck, C. J., Masson-Delmotte, V., Mairesse, A., and Dubinkina, S., Using data assimilation to investigate the causes of Southern Hemisphere high latitude cooling from 10 to 8 ka BP, Clim. Past, 9, 887-901, 2013.

Noone, D. and Simoonds, I., Associations between $\delta^{18}O$ of water and climate parameters in a simulation of atmospheric circulation for 1979-95, J. Climate, 15, 3150-3169, 2002.

North, G., Bell, T. L., and Cahalan, R. F., Sampling errors in the estimation of empirical orthogonal functions, Mon. Weather Rev., 110, 699-706, 1982.

Okazaki, A., and Yoshimura, K., Development of stable water isotope incorporated atmosphere-land coupled model MIROC5, in prep.

PAGES 2k Consortium, Continental-scale temperature variability during the past wto millennia, Nat. Geosci., 6, 339-346, 2013.

Peterson, T. C., and Vose, R. S., An overview of the global historical climatology network temperature database, B. Am. Meteorol. Soc., 78, 2837-2849, 1997.

Rabier, F., Järvinen, H., Klinker, E., Mahfouf, J.-F. and Simmons, A. The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics. Q.J.R. Meteorol. Soc., 126: 1143–1170, 2000.

Rasmusson, E. M., and Capenter, T. H., The relationship between eastern Equatorial Pacific sea surface temperatures and rainfall over India and Sri Lanka, Mon. Weather Rev., 111, 517-528, 1983.

Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., Kent, E. C., Kaplan, A., Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, J. Geophys, Res., 108, D144407, 2003.

Rhodes, R. H., Bertler, N. A. N., Baker, J. A., Steen-Larsen, H. C., Sneed, S. B., Morgenstern, U., and Johnsen, S. J., Little Ice Age climate and oceanic conditions

of the Ross Sea, Antarctica from a coastal ice core record, Clim. Past, 8, 1223-1238, 2012.
Risi, C., Bony, S., Vimeux, F., and Jouzel, J., Water-stable isotopes in the LMDZ4 general circulation model: Model evaluation for present-day and past climates and applications to climatic interpretations of tropical isotopic records, J. Geophys. Res., 115, D12118, 2010.
Roden, J., Lin, G., and Ehleringer, J., A mechanistic model for interpretation of hydrogen and oxygen isotope ratios in tree-ring cellulose, Geochim. Cosmochim. Ac., 64, 21-35, 2000.
Sano, M., Xu, C., and Nakatsuka, T., A 300-year Vietnam hydroclimate and ENSO variability record reconstructed from tree ring $\delta^{18}O$, J. Geophys. Res. 117, D12115, 2012.
Schmidt, G., Hoffmann, G., Shindell, D., and Hu, Y., Modeling atmospheric stable isotopes and the potential for constraining cloud processes and staratosphere-troposphere water exchange, J. Geophys. Res., 110, D21314, 2005.
Schmidt, G., LeGrande, A., and Hoffmann, G., Water isotope expressions of intrinsic and forced variability in coupled ocean-atmosphere model, J. Geophys. Res., 112, D10103, 2007.
Schneider, D. P. and Noone, D. C., Spatial covariance of water isotope records in a global netweok of ice cores spanning twentieth-century climate change, J. Geophys. Res., 112, D18105, 2007.
Schotterer, U., Stichler, W., Ginot, P., The influence of post-depositional effects on ice core studies: Examples from the Alps, Andes, and Altai, in Earth Paleoenvironments: Records Preserved in Mid- and Low-Latitude Glaciers, pp.39-59, Kluwer Acad, Dordrecht, The Netherlands, 2004
Steiger, N., Hakim, G., Steig, E., Battisti, D., and Roe, G., Assimilation of Time-Averaged Pseudoproxies for Climate Reconstruction, 27, 426-441, 2014.
Taylor, K. E., Stouffer, R. J., Meehl, G., An overview of CMIP5 and the experiment design, B. Am. Meteor. Soc., 93, 485-498, 2007.
Takeuchi, N., Fujita, K., Aizen, V. B., Narama, C., Yokoyama, Y., Okamoto, S., Naoki, K., and Kobota, J., The disappearance of glaciers in the Tien Shan Mountains in Central Asia at the end of Pleistocene, Quaternary Sci. Rev., 103, 26-33, 2014.
Thomposon, D. M., Ault, T. R., Evans, M. N., Cole, J. E., and Emile-Geay, J., Comparison of observed and simulated tropical climate trends using a forward model of coral $d^{18}O$, Geophys. Res. Let., 38, L14706, 2011.
van der Schrier, G. and Barkmeijer, J., Bjerknes' hypothesis on the coldness during

AD1790-1820 revisited, Clim. Dyn., 25, 537-553, 2005.

van Leeuwen, P. J., Particle filtering in geophysical systems, Mon. Weather Rev., 137, 4089-4114, 2009.

von Storch, H., Cubasch, U., Gonzalez-Rouco, J. F., Jones, J. M., Voss, R., Widmann, M., and Zorita, E., Combining paleoclimatic eviedence and GCMs by means of data assimilation though upscaling and nudging (DATUN), Proc. 11th Symposium on Global Climate Change Studies, AMS Long Beach, CA, 2000.

Watanabe, M., Suzuki, T., O'ishi, R., Komuro, Y., Watanabe, S., Emori, S., Takemura, T., Chikira, M., Ogura, T., Sekiguchi, M., Takata, K., Yamazaki, D., Yokohota, T., Nozawa, T., Hasumi, H., Tatebe, H., and Kimoto, M., Improved climate simulation by MIROC5: Mean States, Variability, and Climate Sensitivity, 23, 6312-6335, 2010.

Werner, M., Langebroek, P., Carlsen, T., Herold, M., and Lohmann, G., Stable water isotopes in the ECHAM5 general circulation model: Toward high-resolution isotope modeling on a global scale, J. Geophys. Res., 116, D15109, 2011.

Whitaker, J. S., and Hamill, T. M., Ensemble data assimilation without perturbed observations, Mon. Weather Rev., 130, 1913-1924, 2002.

Widmann, M., Goosse, H., van der Schrier, G., Schnur, R., and Barkmeijer, J., Using data assimilation to study extratropical Northern Hemisphere climate over the last millennium, Clim. Past, 6, 627-644, 2010.

Xu, C., Sano, M., and Nakatsuka, T., Tree ring cellulose $\delta^{18}O$ of Fokienia hodginsii in northern Laos: A promising proxy to reconstruct ENSO?, J. Goephys. Res. 116, D245109, 2011.

Xu, C., Zheng, H., Nakatsuka, T., and Sano, M., Oxygen isotope signatures preserved in tree ring cellulose as a proxy for April-September precipitation in Fujian, the subtropical region of southeast China, J. Geophys. Res-Atmos., 118, 12805-12815, 2013.

Xu, C., Pumijumnong, N., Nakatsuka, T., Sano, M., Li, Z., A tree-ring cellulose $\delta^{18}O$-based July-October precipitation reconstruction since AD 1828, northwest Thailand, J. Hydrol., 529, 422-441, 2015.

Yoshimura, K., Kanamitsu, M., Noone, D., and Oki, T., Historical isotope simulation using Reanalysis atmospheric data, J. Geophys, Res., 113, D19108, 2008.

Yoshimura, K., Miyoshi, T., and Kanamitsu, M., Observation system simulation experiments using water vapor isotope information, J. Goephys, Res., 119, 7842-7862, 2014.

Young, G. H. F., Loader, N. J., McCarroll, D., Bale, R. J., Demmler, J. C., Miles, D., Nayling, N., Rinne, K. T., Robertson, I., Watts, C., and Whitney, M., Oxygen stable isotope ratios from British oak tree-rings provide a strong and consistent record of past changes in summer rainfall, Clim. Dyn., 45, 3609-3622, 2015.

Tables

Table 1. Experimental designs. The observation network used in the CTRL experiment is denoted as Orig.

	SST data to drive simulation run	SST data to drive truth run	Assimilated variable	Observation network	Missing data
CTRL	HadISST	HadISST	Simulated $\delta^{18}O$	Orig	w/o missing
CGCM	Modeled SST	HadISST	Simulated $\delta^{18}O$	Orig	w/o missing
VOBS	Modeled SST	HadISST	Simulated $\delta^{18}O$	Orig	w/ missing
REAL	Modeled SST	-	Observed $\delta^{18}O$	Orig	w/ missing
T2-Assim	HadISST	HadISST	Reconstructed T2 from simulated $\delta^{18}O$	Orig	w/o missing
M08	HadISST	HadISST	Simulated $\delta^{18}O$	M08	w/o missing

Figures

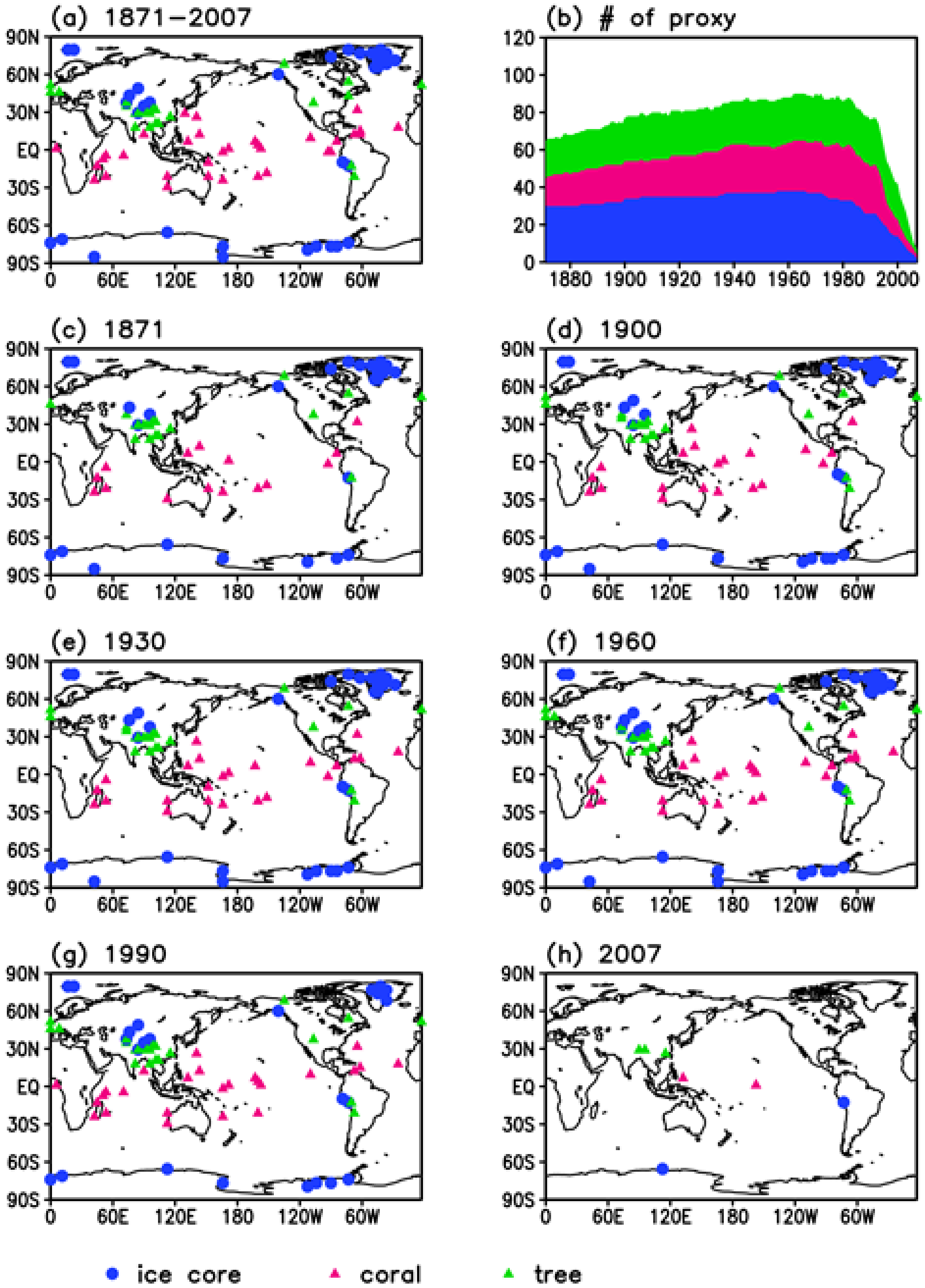


Figure 1

Spatial distribution of proxies ($\delta^{18}O$ in ice cores, corals, and tree-ring cellulose, denoted by blue, pink, and green, respectively). (a) Proxies spanning at least one year during 1871–2000 are mapped (b) The number of proxies is depicted as a function of time. (c–

h) The spatial distributions of the proxies are mapped for (c) 1871, (d) 1900, (e) 1930, (f) 1960, (g) 1990, and (h) 2007.

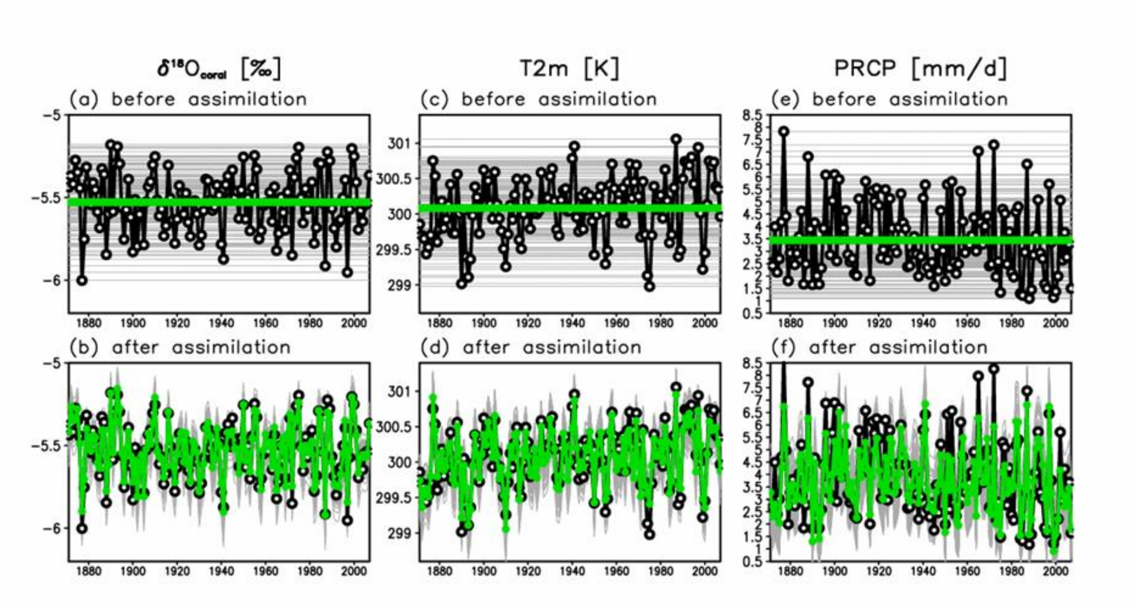
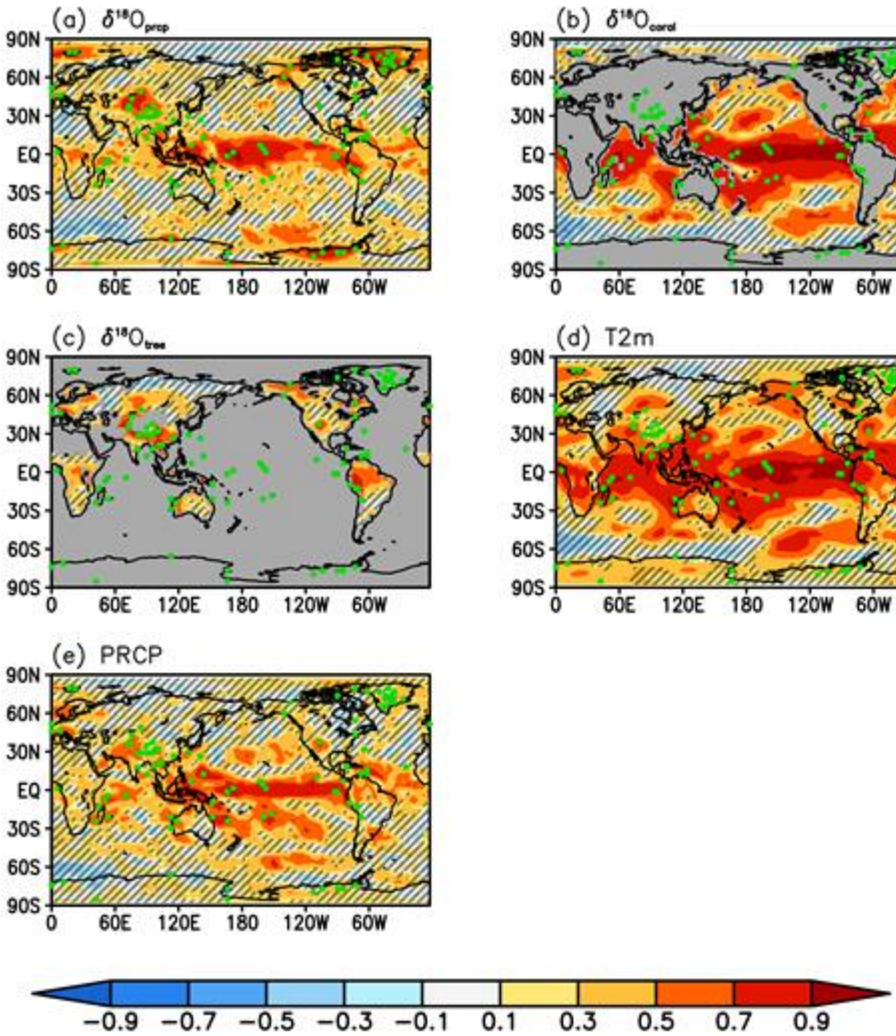


Figure 2

Annual mean $\delta^{18}O$ in corals at a location where observational data were available (1°N, 157°W) for (a) background and (b) analysis. The black line indicates the truth, gray lines indicate ensemble members, and green line indicates the ensemble mean.


(a) $\delta^{18}O_{prcp}$
(b) $\delta^{18}O_{coral}$
(c) $\delta^{18}O_{tree}$
(d) T2m
(e) PRCP
90N 60N 30N EQ 30S 60S 90S
0 60E 120E 180 120W 60W
−0.9 −0.7 −0.5 −0.3 −0.1 0.1 0.3 0.5 0.7 0.9

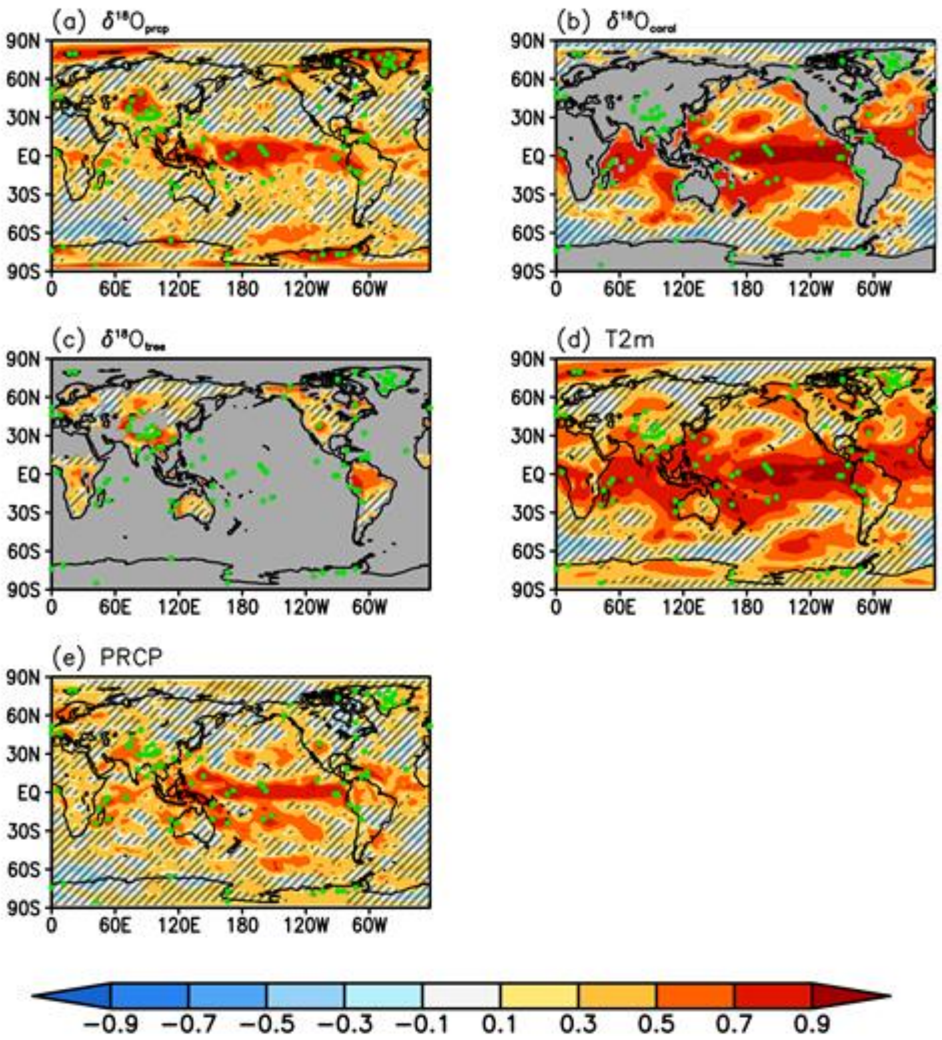
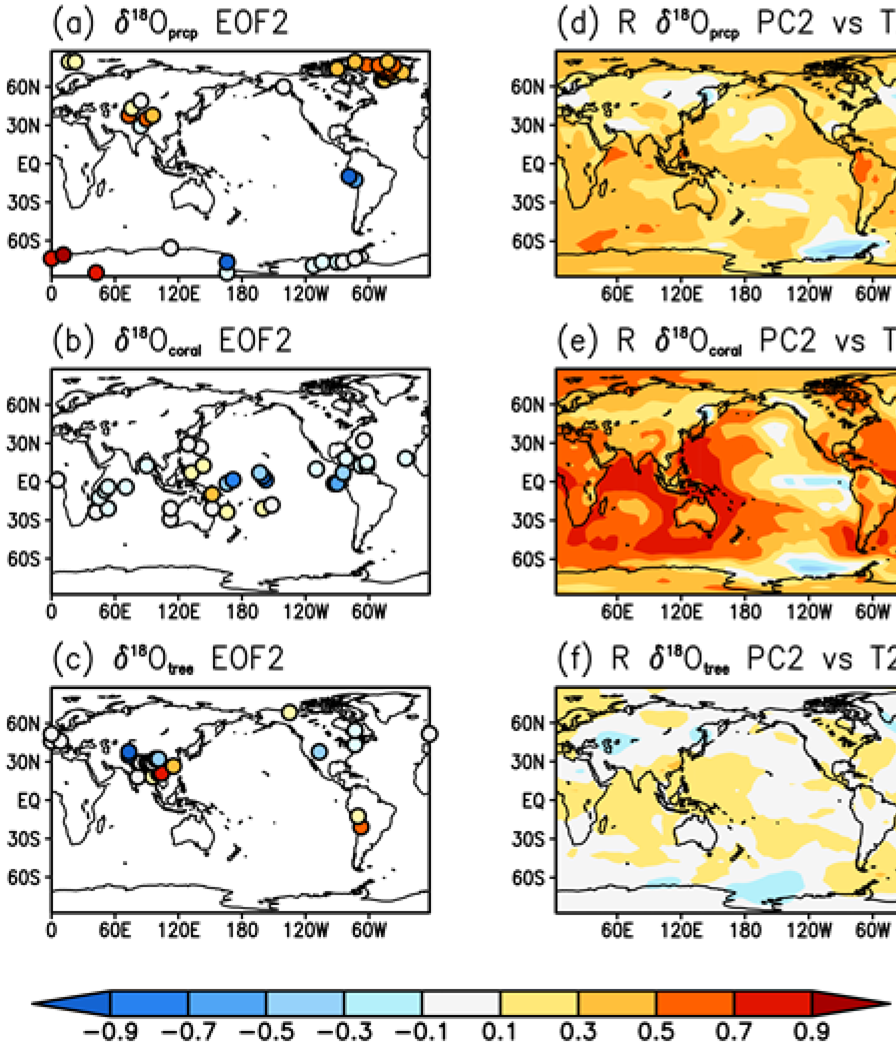


Figure 3

Temporal correlation between the analysis and the truth. The green dot represents the location of the proxy sampling site. The hatched area indicates where the correlation is not statistically significant ($p > 0.05$).


(a) δ18Oprcp EOF2
(b) δ18Ocoral EOF2
(c) δ18Otree EOF2
(d) R δ18Oprcp PC2 vs T2
(e) R δ18Ocoral PC2 vs T2
(f) R δ18Otree PC2 vs T2
60N
30N
EQ
30S
60S
0
60E
120E
180
120W
60W
−0.9 −0.7 −0.5 −0.3 −0.1 0.1 0.3 0.5 0.7 0.9

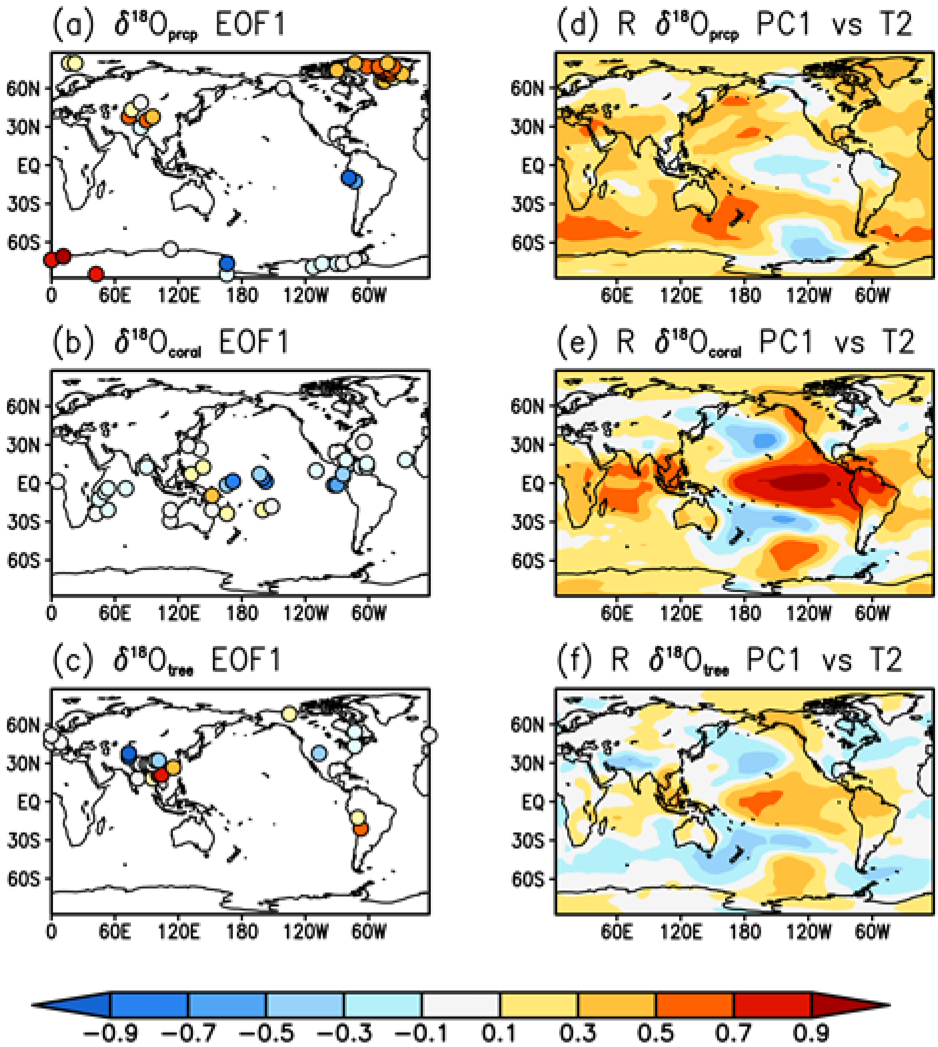
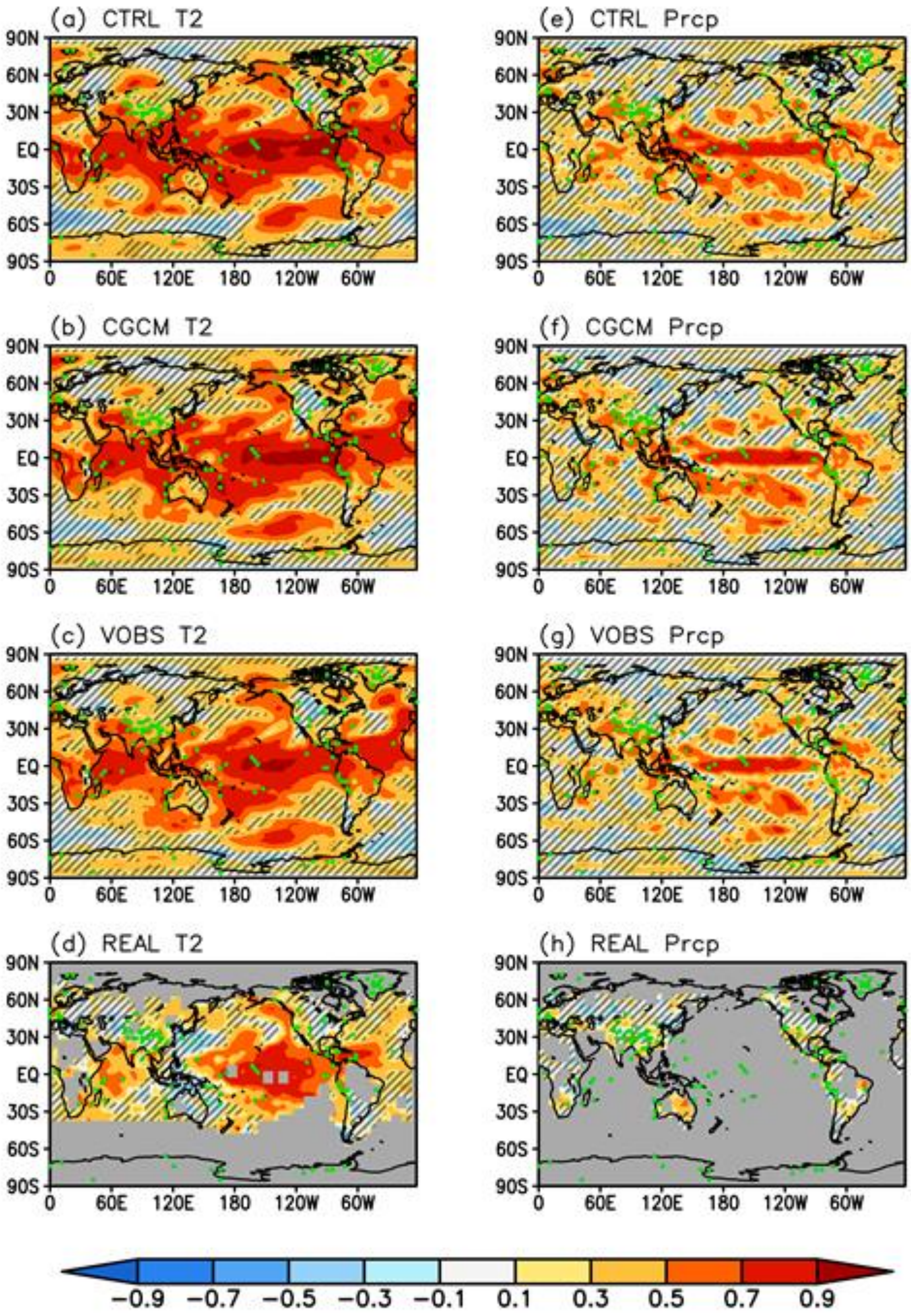


Figure 4

First mode of EOF and the correlation between PC1 and temperature for (a and d) ice cores, (b and e) corals, and (c and f) tree-ring cellulose.


(a) CTRL T2
(e) CTRL Prcp
(b) CGCM T2
(f) CGCM Prcp
(c) VOBS T2
(g) VOBS Prcp
(d) REAL T2
(h) REAL Prcp
90N 60N 30N EQ 30S 60S 90S
0 60E 120E 180 120W 60W
−0.9 −0.7 −0.5 −0.3 −0.1 0.1 0.3 0.5 0.7 0.9

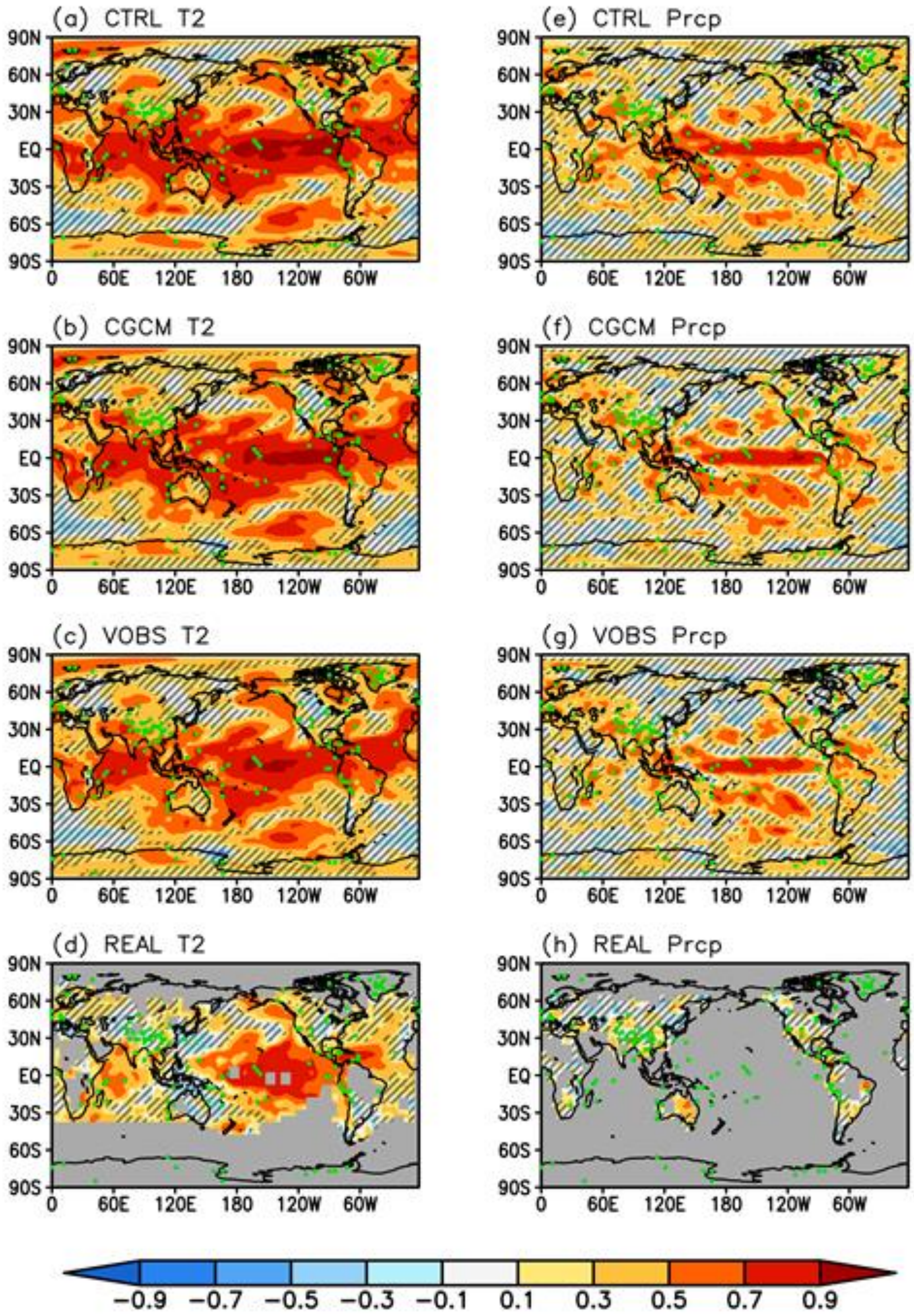
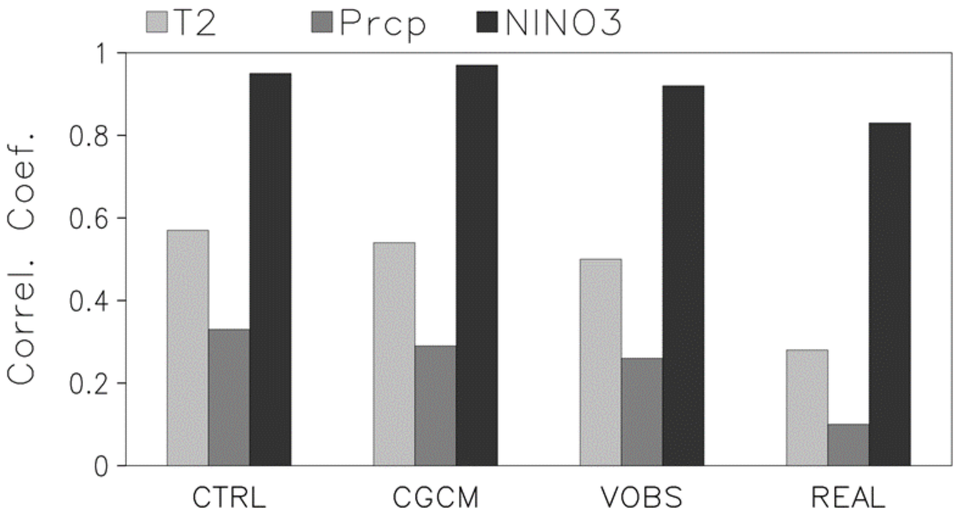


Figure 5

Temporal correlation between the analysis and the truth for (a–d) temperature and (e–h) precipitation, for each experiment. The green dot represents the location of the proxy sampling site. The hatched area indicates where the correlation is not statistically significant ($p > 0.05$).

895


T2
Prcp
NINO3
Correl. Coef.
0
0.2
0.4
0.6
0.8
1
CTRL
CGCM
VOBS
REAL

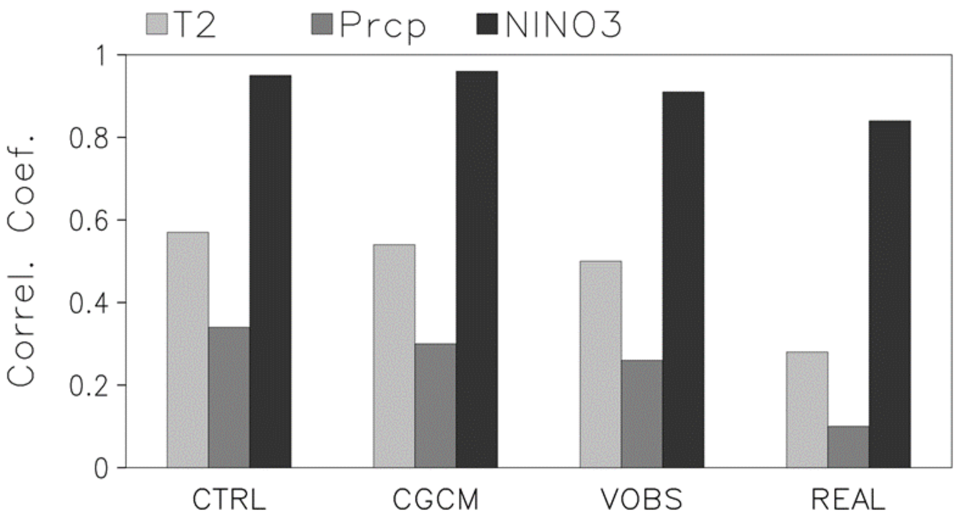
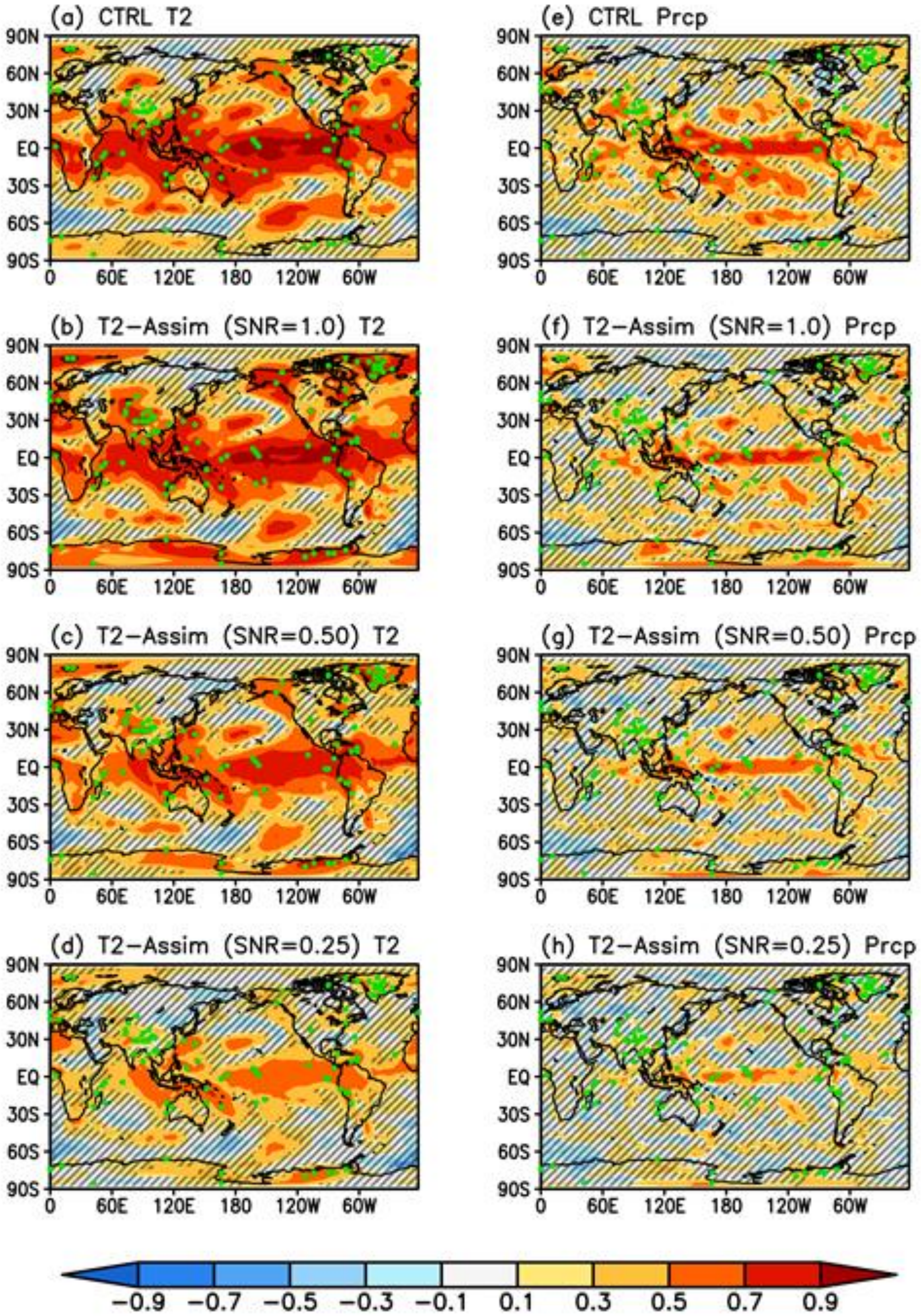


Figure 6

Temporal correlation between the analysis and the truth for each experiment for 1970–1999. The values for temperature and precipitation are the global mean of the temporal correlations.


(a) CTRL T2
(e) CTRL Prcp
(b) T2−Assim (SNR=1.0) T2
(f) T2−Assim (SNR=1.0) Prcp
(c) T2−Assim (SNR=0.50) T2
(g) T2−Assim (SNR=0.50) Prcp
(d) T2−Assim (SNR=0.25) T2
(h) T2−Assim (SNR=0.25) Prcp
90N
60N
30N
EQ
30S
60S
90S
0
60E
120E
180
120W
60W
−0.9 −0.7 −0.5 −0.3 −0.1 0.1 0.3 0.5 0.7 0.9

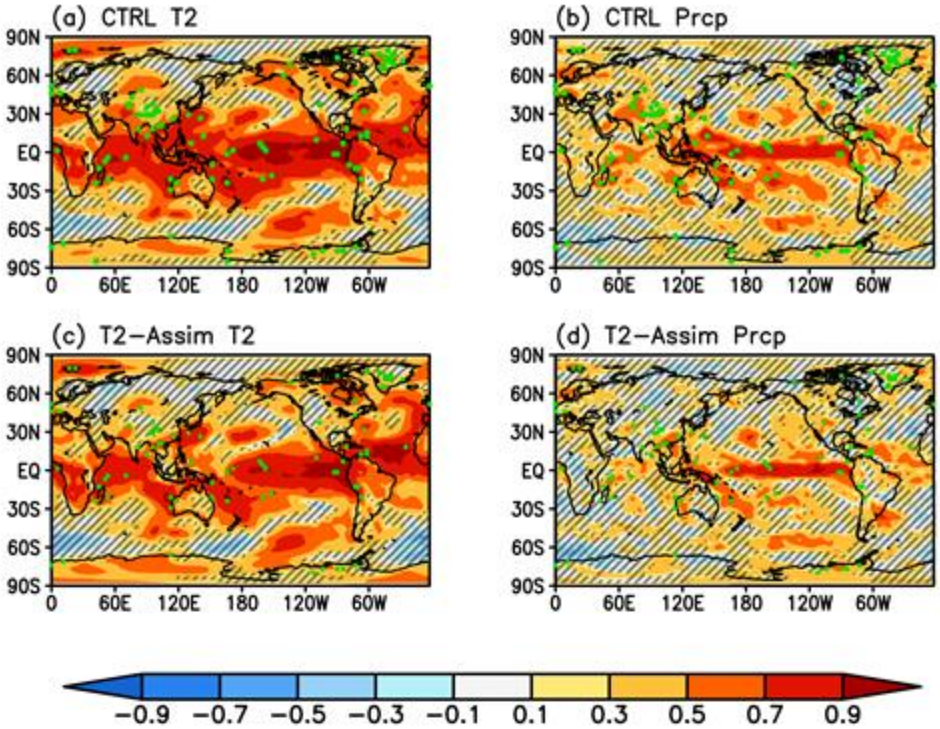
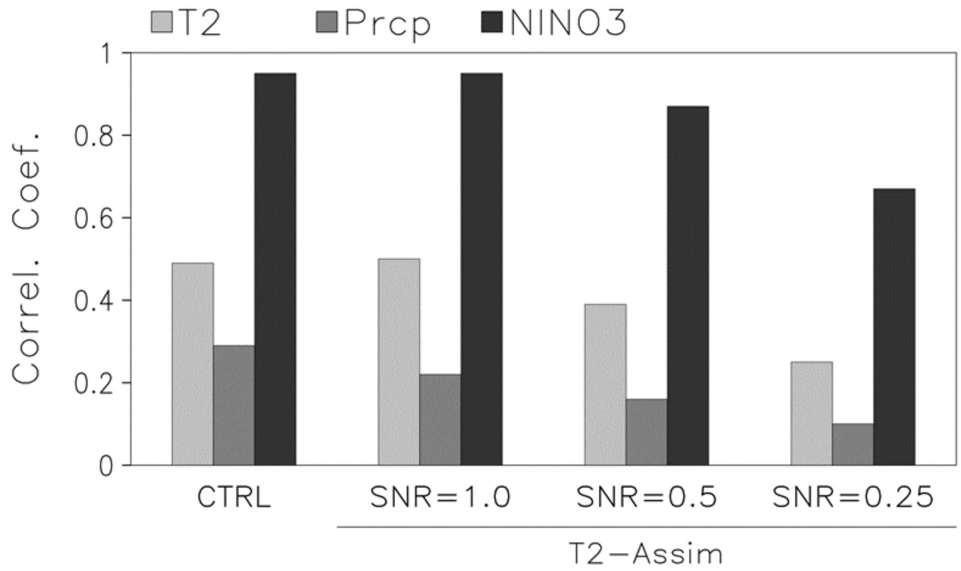


Figure 7

Temporal correlations between the analysis and the truth for (a~~–d~~, c) temperature and (~~e–h~~b, d) precipitation, for (a~~ and e~~, b) CTRL and (b~~-~~, d~~ and f-h~~) T2-Assim. The green dot represents the location of the proxy sampling site. The hatched area means that the correlation is not statistically significant (p > 0.05).

910



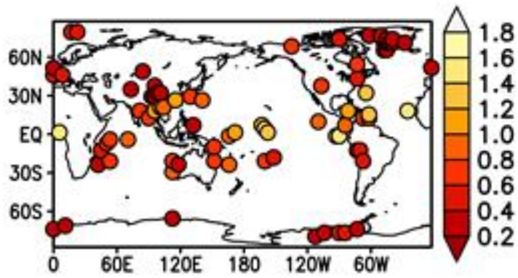


Figure 8

~~Temporal correlation between the analysis and the truth for each experiment for 1970–1999. The values for temperature and precipitation are the global mean of the temporal correlations.~~

Signal to noise ratio (SNR) of the reconstructed temperature from the observation used in CTRL.

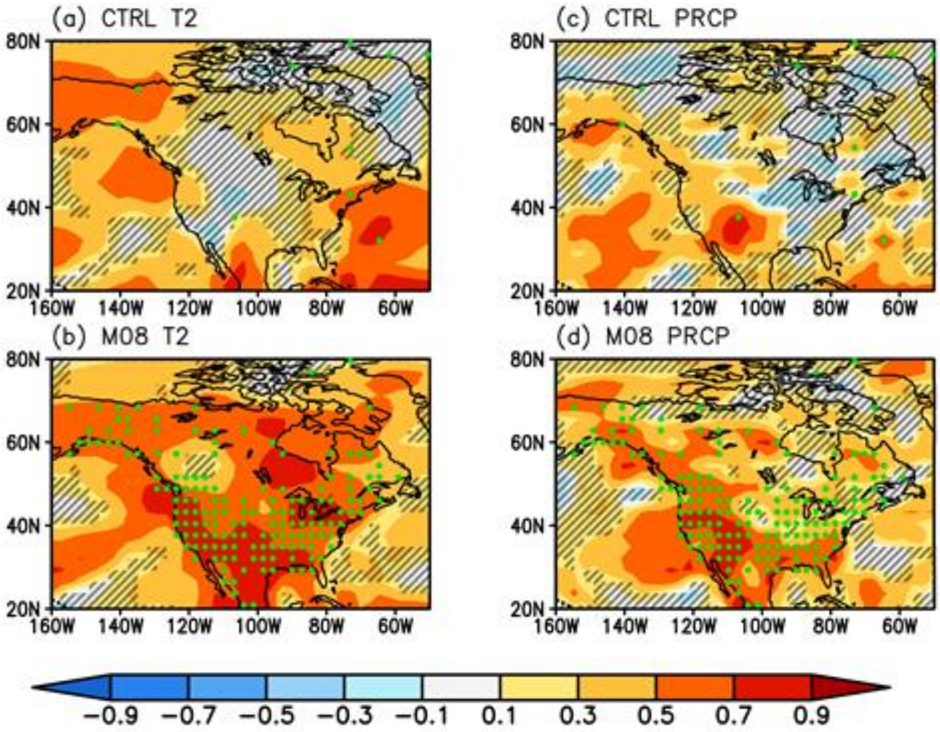


Figure 9

Temporal correlations in North America between the analysis and the truth for (a–d) temperature, and (e–h) precipitation, for experiments using different proxy networks. The green dot represents the location of the proxy sampling site. The hatched area indicates where the correlation is not statistically significant ($p > 0.05$).