

Using data assimilation to estimate the consistency between climate proxies and climate model results

AURÉLIEN MAIRESSE AND HUGUES GOOSSE

Georges Lemaître Centre for Earth and Climate Research, Earth and Life Institute, Université catholique de Louvain, Belgium; aurelien.mairesse@uclouvain.be

The data assimilation technique applied to paleoclimate studies is a promising method to highlight the compatibility or the incompatibility between (1) different climate proxies and (2) between climatic information inferred from the proxies and the physics of the climate system represented in models.

The combination of several climate proxies and/or the results from climate models enables us to reconstruct and understand past climate changes. When data and models are used together, the information inferred from the proxies often serves to validate the climate model results while the models allow exploration of the physical processes responsible for the recorded climatic changes. During the past decade, a new statistical tool called data assimilation has been used in paleoclimatology (Widmann et al. 2010). This tool allows us to build a reconstruction of past climate change which is both consistent with the climate computed by a model and that deduced from proxies.

First, we describe how this tool works when it is applied to the paleoclimate research field. Second, we describe a particular application of data assimilation, which can help to elucidate if the hypotheses proposed to explain proxy variations are compatible with the physics of a climate model. This is outlined with a mid-Holocene case study.

How does data assimilation work?

Data assimilation combines the physical laws included in a climate model with the climate information inferred from proxies to produce paleoclimatic reconstructions consistent with both. In our method, this is achieved using a procedure based on a particle filter with resampling (Dubinkina et al. 2011; see Fig. 1 from van Leeuwen 2009 for a graphical representation) applied to the three-dimensional Earth model of intermediate complexity, LOVECLIM (Goosse et al. 2010).

An ensemble of ~100 simulations, also referred to as particles or ensemble members, is initiated in parallel. At the beginning of this procedure, all of

the particles are identical apart from slightly different initial conditions. Due to the chaotic nature of the climatic system each particle will evolve in a different way. After the first assimilation step (which is one year here but it can be any value greater than or equal to the model time resolution) the likelihood of each member of the ensemble is evaluated in order to determine how close the climate state of each particle is compared with the climate inferred from the proxy data. For each variable (e.g. surface air temperature and sea surface temperature; SST) the likelihood is a function of the difference between the values estimated from the proxy records and the values calculated by the climate model. This function is computed for all the locations and months for which paleodata is available. The particles that have the largest likelihood are retained (i.e. the particles whose climate states are the closest to the past climate reconstructed from the proxies). The other ensemble members are rejected. The remaining particles are resampled in order to keep a constant number of particles (i.e. ~100 in this example) and avoid a degenerative issue. We add a small perturbation to the members that have been sampled more than once for the next year of assimilation. The whole procedure is continually repeated until the final year of calculation (i.e. if we perform a 200-year simulation the procedure is repeated 200 times in this example).

The final climate reconstruction obtained by this method is consistent with the LOVECLIM physics, since the LOVECLIM climate model itself is used in the assimilation process. The reconstruction is also as consistent as possible with the climate derived from the proxy data. This is because the method only selects LOVECLIM results that are most compatible with the information

inferred from all the climate proxies, for each time step of the simulated time period.

Our method has produced a reconstruction of surface temperature changes over the past millennium (e.g. Goosse et al. 2006, 2012). The data assimilation in those studies performed well since the LOVECLIM results were efficiently constrained to be close to the surface temperature signal recorded by continental data and, therefore, provided a consistent picture of the climate system during these particular time periods. Recently, we have applied this data assimilation method to the mid-Holocene climate.

Data assimilation applied to the mid-Holocene

A large number of surface air and SST reconstructions are available for the Holocene. A selection from more than 300 published records was performed with the following criteria: (1) each record must come from archives located between 20°N and 90°N and (2) each record must have a mean temporal resolution of at least 250 years for the period of interest (6.5 to 5.5 ka BP).

In accordance with these two principles and restricting selection to only publicly available data, we selected 47 records of surface air temperature and SST for the mid-Holocene. The resulting dataset is heterogeneous for the following reasons: (1) this climatic information was inferred from climate proxies preserved in marine, continental, and ice archives (Fig. 1), (2) for a given archive, different proxies have been used to infer the same type of information (e.g. SST reconstructions in marine cores based on alkenones, Mg/Ca ratio etc., Fig. 1) and (3) the proxies have been measured and interpreted in term of climate variations by different research groups.

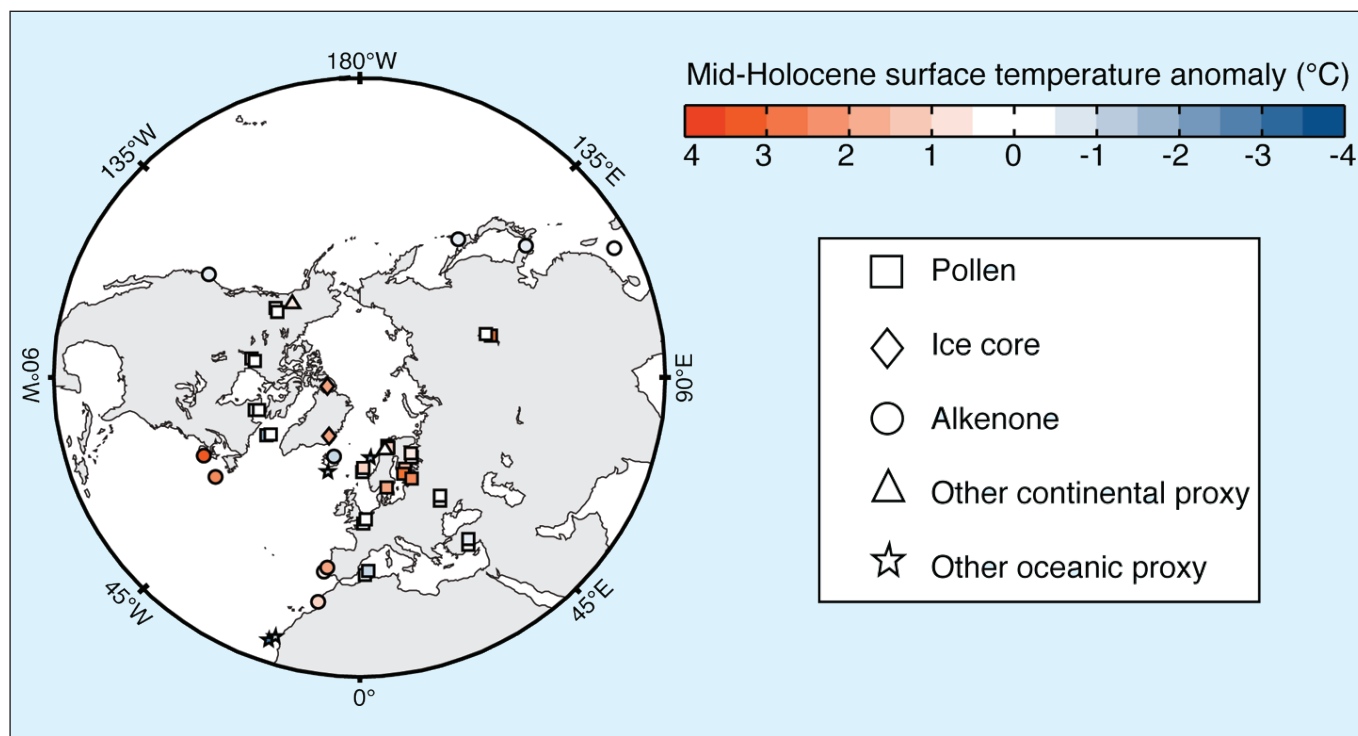


Figure 1: Inferred mid-Holocene surface temperature anomalies (°C) compared to a reference period (1000 to 1500 AD). Where there is more than one proxy record at the same location, the markers representing the proxies are slightly shifted for improved readability.

We have performed the first mid-Holocene data assimilation with the selected dataset and the LOVECLIM model. For this 200-year snapshot experiment, the constraint provided by data assimilation is weak and the disagreement between the climate proxies and model results based on this data assimilation method is still large. For all the locations and the months for which proxy information is available, the LOVECLIM results with data assimilation are on average only 10% closer to the climate signal extracted from the proxies than with the LOVECLIM results produced without data assimilation. In other words, because of the heterogeneous nature of the proxy dataset, the simulations with data assimilation mainly highlight incompatibilities between the proxies and with the model physics rather than producing a shift of the model state that results in a better agreement with the proxy-based climatic reconstructions.

Incompatibilities between proxy and model

First, some variations observed in the climate reconstructions inferred from the proxies cannot be explained by LOVECLIM because they are related to phenomena occurring at a scale smaller than the model grid resolution. For example, this is the case with some SST reconstructions from marine cores retrieved in coastal margins such as the Tagus Estuary (Portugal) where

Holocene SST variations are partly influenced by the Tagus River input (Rodrigues et al. 2009). Such a regional influence in this area is not represented in the LOVECLIM model.

Second, incompatibilities exist between reconstructions based on different types of proxies (see Fig. 1). Future work will aim at identifying these inconsistencies by performing additional experiments with data assimilation. We will run several ensemble simulations, each constrained by climate records from only one type of proxy at a time (e.g. pollens). Each set of simulations will enable the identification of the processes that could explain the recorded signal used according to the climate model physics. Subsequently, it will be also possible to analyze the results of these experiments at locations where other proxies, not selected to drive this set of simulations, are available. For instance, we will compare the results from an assimilation which includes only pollen data, with SSTs inferred from alkenones. This comparison could aid in deciphering whether the SST signal deduced from alkenones should be interpreted as an annual or a summer signal to improve the compatibility between the pollen and the alkenone-based climate records, according to the LOVECLIM physics. This procedure may lead to a tentative revised interpretation of the climate proxy. Even if this proves too challenging, the uncertainty in

model-data comparison associated with incompatibilities between proxy-based reconstructions could at least be estimated.

Outlook

We highlight the potential use of the data assimilation method for paleoclimate studies. This method enables us to assess compatibilities and/or incompatibilities between different climate proxy records for the mid-Holocene time interval. In the future, we could use data assimilation to suggest a revised interpretation of the proxies in order to have a better consistency between different climate proxies and enable more accurate model-data comparisons.

Selected references

Full reference list online under:
http://www.pages-igbp.org/products/newsletters/ref2013_1.pdf

Dubinkina S et al. (2011) *International Journal of Bifurcation and Chaos* 21: 3611-3618

Goosse H et al. (2006) *Climate Dynamics* 27(2): 165-184

Goosse H et al. (2010) *Geoscientific Model Development* 3(2): 603-633

Goosse H et al. (2012) *Global and Planetary Change* 84-85: 35-47

van Leeuwen PJ (2009) *Monthly Weather Review* 137(12): 4089-4114

