An additional step toward comprehensive paleoclimate reanalyses

Hugues Goosse¹

¹Université catholique de Louvain, Earth and Life Institute, Georges Lemaître center for Earth and Climate Research, Place Louis Pasteur, 3, Louvain-la-Neuve, Belgium

Abstract

Although data assimilation in paleoclimatology has shown significant progress, the model data comparison step remains a limiting factor because paleoclimate (proxy) records have generally a complex response to both climatic and nonclimatic factors. In experiments performed in a controlled framework, Dee et al. (2016) have applied proxy system models that simulate tree ring width, isotopic composition of corals and isotopic composition of ice cores from the results of a climate model. The difference between those simulated variables and the value measured on the natural archive can then be computed directly, improving significantly the performance of the data assimilation method.

Data assimilation is a rigorous and self-consistent framework used to estimate the state of a system and the uncertainties of this estimation, by objectively combining information obtained from model results and observations (Figure 1). The method is the foundation for forecasts and also underlies the reanalyses applied so widely in the atmospheric sciences and oceanography [e.g., Kalnay et al., 1996, Dee et al., 2011, Balmaseda et al., 2015].

Data assimilation has attracted increasing interest in the paleoclimate community over the past 15 years [e.g., von Storch et al., 2000, Hughes and Ammann, 2009; Widmann et al., 2010] since the technique has several advantages compared to the empirical-statistical approaches traditionally used to reconstruct temperature or precipitation fields from paleoclimate records. In particular, data assimilation does not rely on the stationarity of the statistical relationship between the observation and the reconstructed field, and the method can also efficiently handle nonlinear relationships. Furthermore, estimates obtained for potentially many different variables are guaranteed to be physically consistent to within known uncertainties.

Several challenges remain to improve paleoclimatological reconstructions using data assimilation. The compilation, interpretation and quality control of available paleoclimate data is a particularly arduous and time consuming task, because of the diversity of the sources and the difficulty in estimating the uncertainties of the observations. State-of-the-art data assimilation methods requires one or two orders of magnitude more computer time than a standard simulation, so computational cost quickly becomes a limiting factor when long periods of the past are investigated. Additionally, specific adaptations to algorithms are needed to deal with the sparse records that represent averages over periods ranging from a few months to several decades and are distributed over strongly spatially biased network.

In this framework, Dee et al. [2016] discuss a major advance in a fundamental element of data assimilation systems: the “observation operator” that transforms model states to fit with observation space. Observations and model results are provided at different locations and often represent different variables. In order to compare those two fields and obtain what is called the innovation (or observational increments, i.e. a vector including the difference between each data and corresponding model results), the observation operator must map from model space to observation space. For example, the radiance measured by a satellite at selected frequencies is related to the sea ice concentration at the ocean surface. Although one could first retrieve the variable simulated by the model (the sea ice concentration in the example) and perform data assimilation in model space, it is almost always more efficient to derive the observed variable (here the radiance) from the model through the observation operator, and perform the model-data comparison on this variable. This implies that the observation operator can become very complex, including a representation of all the processes affecting the signal measured by the instrument.
In paleoclimatology, until now, no assimilation system has included a sophisticated observation operator \[\text{e.g., Goosse et al., 2012; Bhend et al., 2012, Hakim et al., 2016}\]. The observed variables, such as tree ring width or the isotopic composition of ice, were first transformed in an estimate of surface temperature changes. This was generally done using a linear univariate statistical model since it is a simple approach adequate for a first step. The temperature reconstruction was then used as a constraint in the data assimilation exercise. Nevertheless, many paleoclimate records may be influenced by more than one environmental variable. A small ring growth during a particular year, for example, can be due to cold temperatures, or to a drought, or a combination of both. This covariability makes the inversion problem, i.e. estimating the temperature or precipitation for the paleoclimate record, ill-conditioned. It is thus clearly more suitable to directly simulate tree ring width from model results and to compare it to the measurements. This requires the development of so-called forward or proxy system models that directly simulate the response of the natural system from which the climate archive is derived to climate or more generally environmental changes. Several of such models are now available and their application has become easier recently \[\text{e.g., Evans et al., 2013; Dee et al., 2015}\]. For the first time, Dee et al. (2016) have tested the utility of including those proxy system models in the observation operator of a data assimilation process. Working in an idealized framework, they selected models that simulate explicitly three variables as a function of climate: tree ring width, isotopic composition of corals and isotopic composition of ice cores. The results of the experiments clearly demonstrate the benefit of using more sophisticated observational models, especially for tree rings whose growth is influenced by both temperature and precipitation. The improvement is less dramatic for the isotopic composition of corals for which such a univariate model dependent on temperature appears satisfactory.

In paleoclimatology, until now, no assimilation system has included a sophisticated observation operator \[\text{e.g., Goosse et al., 2012; Bhend et al., 2012, Hakim et al., 2016}\]. The observed variables, such as tree ring width or the isotopic composition of ice, were first transformed in an estimate of surface temperature changes. This was generally done using a linear univariate statistical model since it is a simple approach adequate for a first step. The temperature reconstruction was then used as a constraint in the data assimilation exercise. Nevertheless, many paleoclimate records may be influenced by more than one environmental variable. A small ring growth during a particular year, for example, can be due to cold temperatures, or to a drought, or a combination of both. This covariability makes the inversion problem, i.e. estimating the temperature or precipitation for the paleoclimate record, ill-conditioned. It is thus clearly more suitable to directly simulate tree ring width from model results and to compare it to the measurements. This requires the development of so-called forward or proxy system models that directly simulate the response of the natural system from which the climate archive is derived to climate or more generally environmental changes. Several of such models are now available and their application has become easier recently \[\text{e.g., Evans et al., 2013; Dee et al., 2015}\]. For the first time, Dee et al. (2016) have tested the utility of including those proxy system models in the observation operator of a data assimilation process. Working in an idealized framework, they selected models that simulate explicitly three variables as a function of climate: tree ring width, isotopic composition of corals and isotopic composition of ice cores. The results of the experiments clearly demonstrate the benefit of using more sophisticated observational models, especially for tree rings whose growth is influenced by both temperature and precipitation. The improvement is less dramatic for the isotopic composition of corals for which such a univariate model dependent on temperature appears satisfactory.

The study of Dee et al. (2016) is thus an important step toward comprehensive paleoclimate reanalyses and it will certainly stimulate new work and developments in the community. It is now necessary to measure the improvements brought by proxy system models in realistic conditions, using existing data compilations. This implies additional challenges. The signals recorded in natural archives are complex and often not well understood combinations of the response to several climatic and nonclimatic factors. Consequently, current proxy system models still have clear limitations. In particular, some of them are well adapted to specific regions but are not able to simulate adequately the observed changes everywhere \[\text{e.g., Breitenmoser et al., 2014}\]. Applying them in some areas may then be worse than considering simpler and more robust statistical assumptions. Furthermore, climate model results are biased and those biases must generally be removed before used them to drive the proxy system model, leading to additional assumptions on the validity and stability of those corrections through time. Finally, the observations represent a region that is often much smaller than the model grid size and a downscaling procedure may be required. The way those elements impact the quality of the reconstruction must be evaluated before obtaining a fully operational system. Nevertheless, the results presented by Dee et al. (2016) are very encouraging. Including proxy system models in the data assimilation framework appears both technically possible and beneficial. They have thus convincely demonstrated that their application should be more systematic in data assimilation in paleoclimatology.

---

**Figure 1.** Schematic representation of a classical data assimilation scheme. The analysis, i.e., the estimate of the state of the system, is based on model results and observations, taking into account the uncertainties on both sources of information. In this framework, model results have to be compared to observations to compute the so-called innovation, which is a measure of the model-data difference. Rough observations and model results have thus to be processed to ultimately represent the same physical quantity, using various methods as highlighted on the figure. The analysis may then serve as an initial state for subsequent steps, leading to a sequential state estimation or to improvements of the previous estimate.
References


