ITS4.3
AS5.2
EGU2020-9038

How many proxy are necessary to reconstruct the temperature of the last millennium?

Fernando Jaume-Santero[1,2], David Barriopedro[2], Ricardo García-Herrera[1,2], Sancho Salcedo-Sanz[3] & Natalia Calvo[1]

[1]Universidad Complutense de Madrid, Facultad de Ciencias Físicas, Departamento de Física de la Tierra y Astrofísica, Madrid, Spain
[2]Instituto de Geociencias (IGEO), (CSIC/UCM), Madrid, Spain
[3]Departamento de Teoría de la Señal y Comunicaciones, Universidad de Alcalá, Alcalá de Henares, Spain

Abstract

Here we show how evolutionary algorithms can be coupled with different reconstruction methods to minimize the spatial bias induced by the non-homogeneous distribution of available proxies. Results from pseudo-proxy experiments indicate that small subsets of records situated over representative locations can outperform the reconstruction skill of the full proxy network. These locations highlight the importance of high-latitude regions and major teleconnection areas to reconstruct annual global temperature fields and their responses to external forcings and internal variability. However, long-term temperature trends such as the transition between the Medieval Climate Anomaly (MCA) and the Little Ice Age (LIA) are better resolved by records situated at lower latitudes.

Introduction

Climate Field Reconstructions (CFRs) generated from different paleoclimate archives are affected by the non-uniform distribution of the observing network. Note that most paleoclimate archives are restricted to land regions (mainly over the middle latitudes of the Northern Hemisphere), while there are extensive unsampled areas in the Southern Hemisphere and high latitudes. This unbalanced distribution of paleoclimate records induces a spatial bias in global CFRs that remains poorly quantified. In this study, an evolutionary algorithm has been coupled with different reconstruction methods to estimate (and minimize) these biases on annually-resolved temperature field reconstructions of the last millennium (850-2005 CE).

Data

13 full-forcing members of the Community Earth System Model Last Millennium Ensemble (CESM-LME) were used as a surrogated reality, where synthetic temperature series from the target simulation (1st ensemble member) matching the 569 locations of the PAGES-2k archive were used as pseudo-proxies. Some of these pseudo-proxy datasets have been perturbed with red noise and a Signal to Noise Ratio (SNR) of 1.

Evolutionary Algorithm

Evolutionary algorithms are inspired by biological processes such as genetic recombinations and mutations that ensure the survival and evolution of best suited individuals within a natural competitive environment. In this study, we used the *Coral Reef Optimization* (CRO) based on the reproduction and survival of different corals within a reef (Fig. 1). It combines several search strategies within a group of different solutions (sets of locations), including a small probability of selecting random locations (mutations). The CRO was coupled with reconstruction methods to search for optimal subsets of N<569 representative locations of the PAGES-2k network that minimize the RMSE between the target field and its reconstruction (Fig. 2).

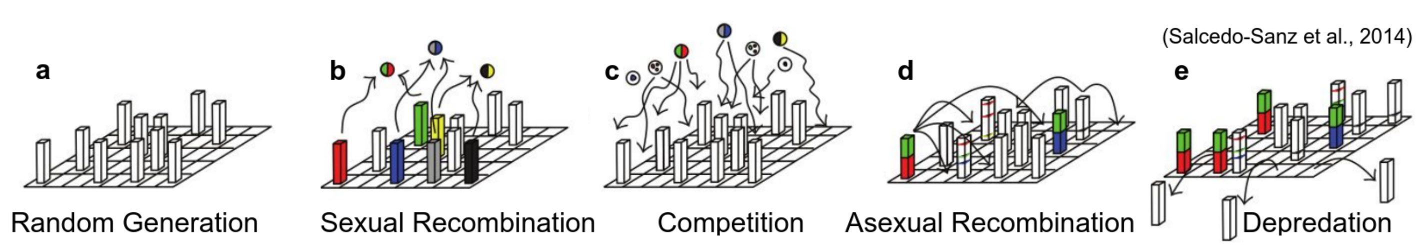

Figure 1: CRO is randomly initialized (**a**), and then steps **b-e** are iterated for a certain number of generations.

European Union
European Social Fund
Gobierno de España
Ministerio de Ciencia e Innovación
Ministerio de Universidades

Reconstruction Methods

Proxy-based reconstructions are obtained using different mathematical techniques such as dimensionality reduction, analogue selection, or Bayesian inference. Two different reconstruction methods were employed to validate the independence of the results (Fig: 2): the *Analogue Method* (AM), and the *Canonical Correlation Analysis* (CCA).

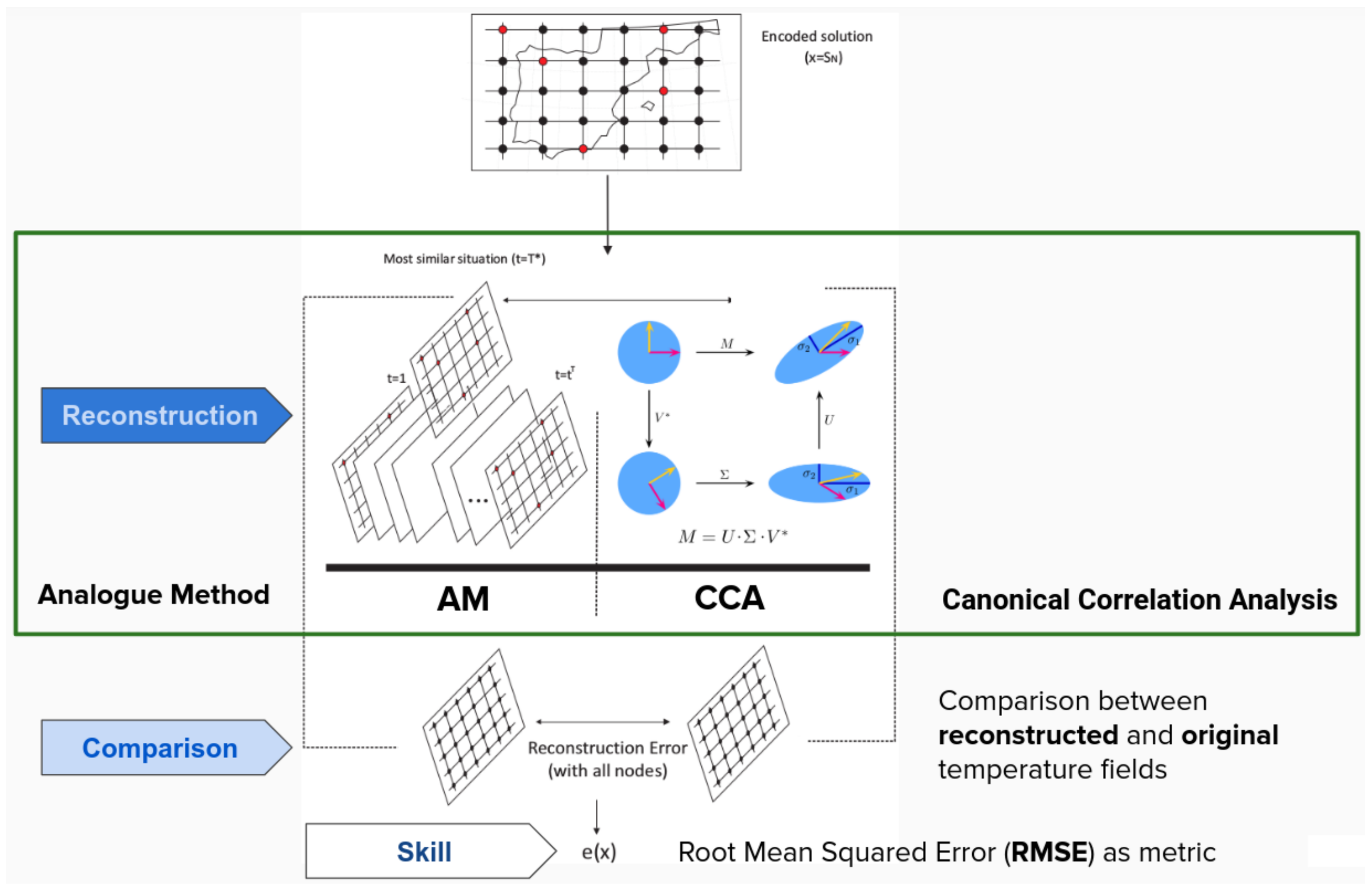

Figure 2: Temperature field reconstruction procedure. Top: Each coral represents an encoded solution (red locations in the map). Middle: Reconstructions are generated with those locations using either the AM or the CCA (green box). Bottom: The temperature field reconstruction is compared against the target to determine the strength of the solution (the smaller the RMSE, the stronger the solution).

Results

-Selection of representative proxy locations

A set of 17 representative perfect pseudo-proxies (CRO-MIN) is enough to obtain temperature fields with the same RMSE as the full-proxy reconstruction (Fig. 3). The minimum RMSE is achieved with an optimal subset (CRO-OPT) of 120 locations, which improves the reconstruction skill in almost all regions (Fig. 4). For SNR=1, optimized subsets of noisy pseudo-proxies can outperform the skill of the full network of perfect pseudo-proxies. Therefore, biases induced by the non-homogeneous distribution of records are comparable to those arising from noisy proxies perturbed with the same variance as the temperature signal.

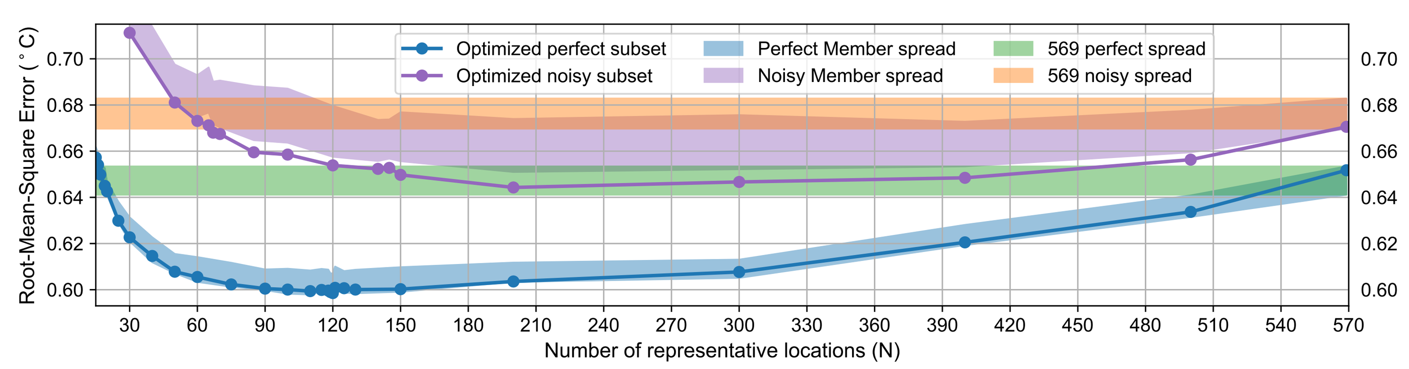

Figure 3: RMSE of CRO reconstructions as a function of the number of selected perfect (blue) and noisy pseudo-proxies (purple). The green and orange shades represent the 13-member reconstruction skill spread obtained with all perfect and noisy (SNR=1) pseudo-proxies. The blue-shaded area represents the spread obtained by using the optimized first member subset of N pseudo-proxies to reconstruct the remaining members of the ensemble. The purple-shaded area is the same as the blue-shaded one but for reconstructions using noisy pseudo-proxies.

-Optimal temperature field reconstructions

The algorithm tends to prioritize those records situated over climate teleconnections, explaining a large fraction of the global variance. It also highlights the importance of polar regions to reconstruct annual global temperature patterns, and captures the global responses to external forcings and internal variability (Fig. 5).

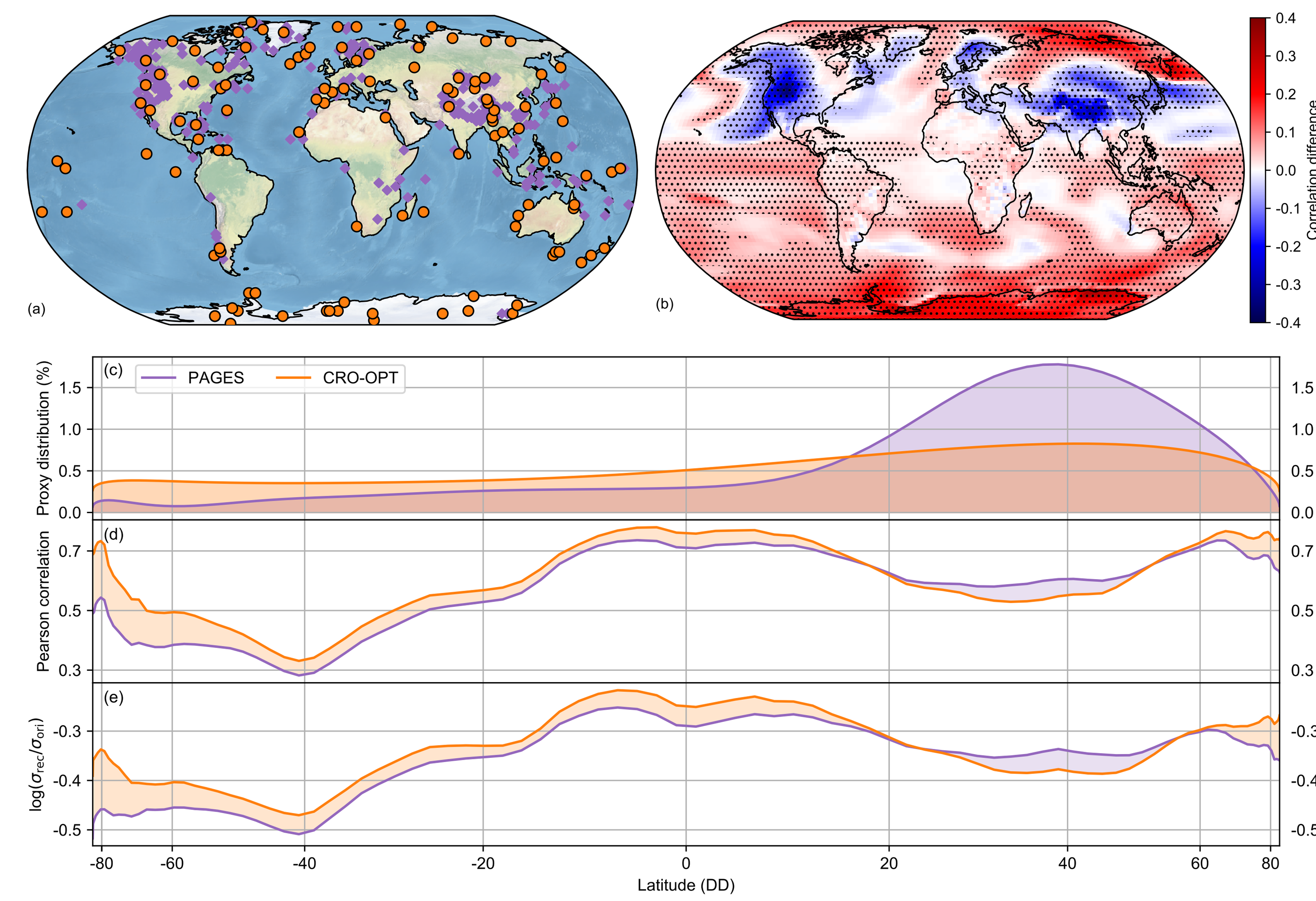

Figure 4: **a**, Spatial distribution of CRO-OPT records (orange dots) obtained from the full PAGES-2k network (purple diamonds). **b**, Correlation difference between temperature reconstructions with CRO-OPT and all perfect proxies. **c**, Normalized latitudinal distribution of records (in % with respect to the total number of pseudo-proxies) for the CRO-OPT subset (orange) and the full-proxy PAGES-2k network (purple). **d**, Latitudinal mean correlations for the CRO-OPT (orange) and full-proxy (purple) reconstructions. **e**, Latitudinal variability ratio for the CRO-OPT (orange) and full-proxy (purple) reconstructions (σ_{rec}) compared with the target simulation (σ_{ori}).

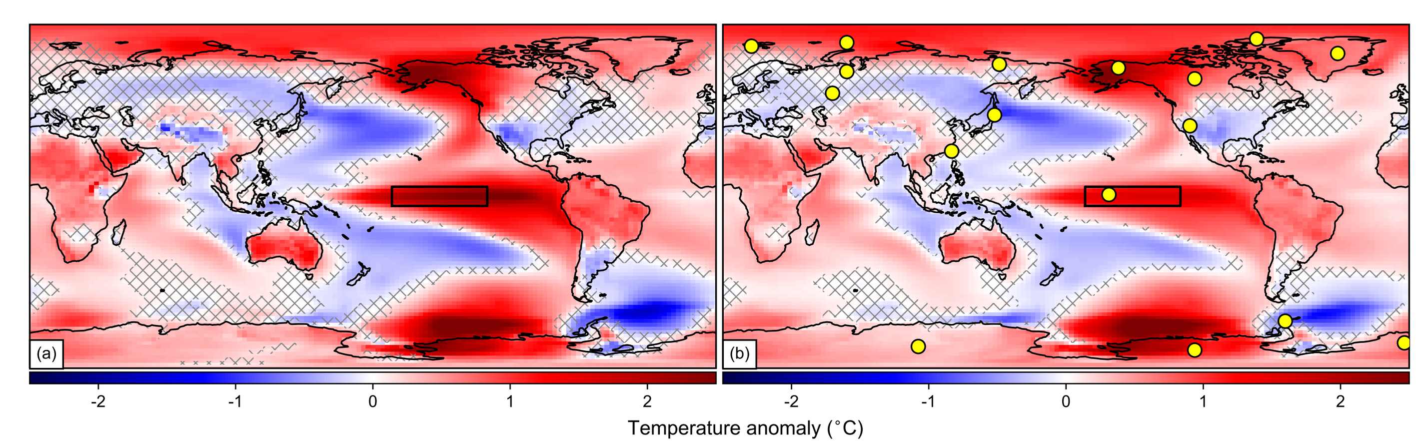

Figure 5: **a**, El Niño-Southern Oscillation (ENSO) pattern of the target simulation. **b**, Reconstruction of the ENSO pattern with the optimal subset of 17 representative records (yellow dots).

-Global Mean Temperatures (GMT) on real datasets

Global Mean Temperature estimations (GMTg) of the HadCRUT 4.2 and LMR datasets have been obtained using 120 representative locations from CRO-OPT. In both cases, they are consistent with the real area-weighted GMT obtained with all grid-points (Fig. 6). Moreover, the coefficients of determination (R^2) are higher for CRO-OPT than for the complete PAGES-2k network, and they are also significantly higher than selecting random sets of locations from the PAGES-2k archive. This evidences the representativeness of CRO-OPT in observations and real-world reconstructions.

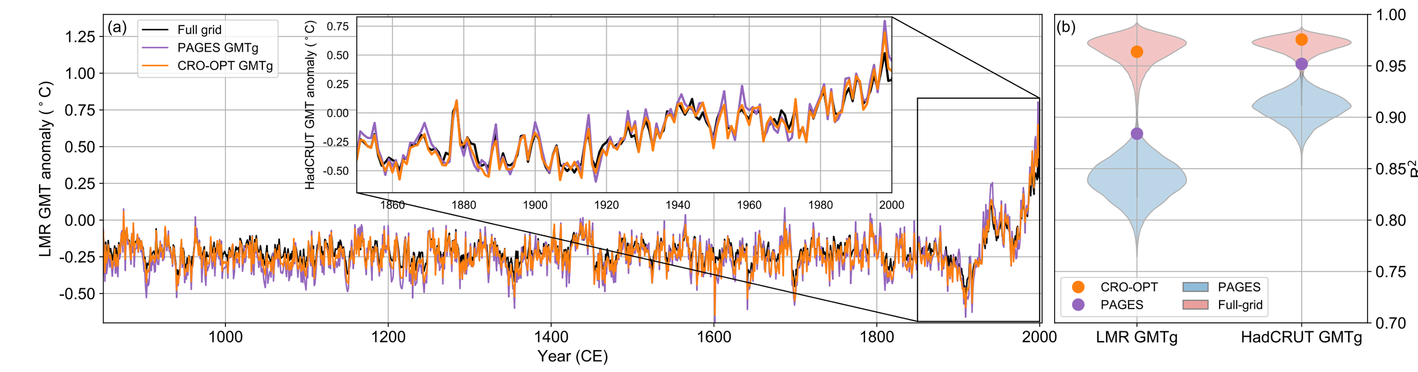

Figure 6: **a**, GMT anomalies from the LMR for 850-2000 CE (**Inset a**, GMT anomalies from HadCRUT4 for 1850-2000 CE). Purple and orange lines show the GMTg of these datasets using the grid points matching the PAGES-2k and CRO-OPT locations. All anomalies are computed with respect to the 1961-1990 baseline. **b**, R^2 between GMT and GMTg from PAGES-2k (purple) and CRO-OPT (orange) locations. Violins are distributions obtained for 10^4 subsets of random locations from the PAGES-2k network (blue) and the full global grid (red).

-Long-term temperature reconstructions

While annual temperature fields are well described by high-latitude records, multi-centennial temperature variations such as the MCA (950-1250) - LIA (1450-1850) transition are better represented by records situated at lower latitudes (Fig. 7), suggesting that representative locations depend on the temporal resolution of the target.

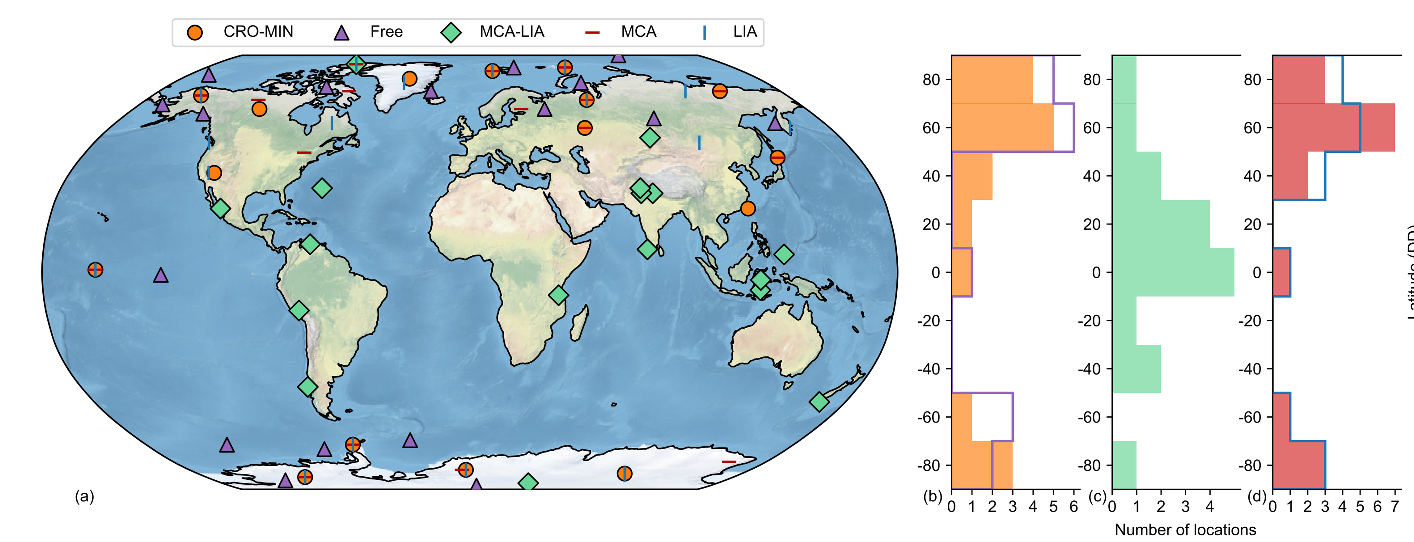

Figure 7: Distribution of representative locations for different experiments with perfect pseudo-proxies. **a**, 2-D and **b-d**, latitudinal distribution of optimized subsets of perfect pseudo-proxies (with the same size as CRO-MIN). Optimized reconstruction of the global annual temperature fields from locations constrained to the PAGES-2k network (CRO-MIN, orange) and from an unconstrained selection (Free, purple). Optimized subsets of the PAGES-2k network for the reconstruction of the global annual temperature fields of the MCA (red) and LIA (blue) periods separately, and the spatial pattern of the mean temperature difference between the MCA and LIA (MCA-LIA, green). Latitudinal distribution of **b**, CRO-MIN (orange shading) and Free (purple line). **c**, MCA-LIA. d, MCA (red shading) and LIA (blue line).

Conclusions

- Under idealized conditions, non-homogeneous proxy distributions can debase the reconstruction skill at the same level than using proxies with red noise and SNR=1.
- A reduced subset of 17 representative perfect pseudo-proxies generates reconstructions with better skill than using the entire PAGES-2k network (569).
- Temperature fields at annual (centennial) resolution are better reconstructed with locations in higher (lower) latitudes and teleconnection regions.
- Representative locations depend on the temporal resolution of the target.

References

- Jaume-Santero F. et al. Selection of optimal proxy locations for temperature field reconstructions using evolutionary algorithms., *Sci. Rep.*, (Accepted, 2020).
- Salcedo-Sanz, S., et al. Near-optimal selection of representative measuring points for robust temperature field reconstruction with the CRO-SL and analogue methods., *Glob. Planet. Change*, **178**, 15-34, 10.1016/j.gloplacha.2019.04.013 (2019).