https://doi.org/10.5194/cp-2019-80
Preprint. Discussion started: 6 August 2019
© Author(s) 2019. CC BY 4.0 License.

Climate of the Past
Discussions

The importance of input data quality and quantity in climate field reconstructions – results from a Kalman filter based paleodata assimilation method

Franke, Jörg[1,2], Valler, Veronika[1,2], Brönnimann, Stefan[1,2], Neukom, Raphael[1,2], Jaume Santero, Fernando[3]

[1] Institute of Geography, University of Bern, Switzerland.
[2] Oeschger Centre for Climate Change Research, University of Bern, Switzerland.
[3] Universidad Complutense de Madrid.

Correspondence to: Jörg Franke (franke@giub.unibe.ch)

Abstract. Differences between paleoclimatic reconstructions are caused by two main factors, the method and the input data. While many studies compare methods, we will focus in this study on the consequences of the input data choice in a state-of-the-art paleo data assimilation approach. We evaluate reconstruction quality based on three collections of tree-ring records: (1) 54 of the best temperature sensitive tree-ring chronologies chosen by experts; (2) 415 temperature sensitive tree-ring records chosen less strictly by regional working groups and statistical screening; (3) 2287 tree-ring series that are not screened for climate sensitivity. The three data sets cover the range from small sample size, small spatial coverage and strict screening for temperature sensitivity to large sample size and spatial coverage but no screening. Additionally, we explore a combination of these data sets plus screening methods to improve the reconstruction quality.

Neither a large, unscreened collection of proxy data nor the small expert selection leads to the best possible climate field reconstruction. A large collection of unscreened data leads to a poor reconstruction skill. The few best temperature proxies allow for a skillful high latitude temperature reconstruction but fail to provide improved reconstructions for other regions and other variables. We achieve the best reconstruction skill across all variables and regions by combing all available input data but rejecting records with small, insignificant information and removing duplicate records. In case of assimilating tree ring proxies, it appeared to be important to use a tree-ring proxy system model that includes both major growth limitations, temperature and moisture.

1 Introduction

In the past 20 years, a lot of effort has been invested in improving climate reconstructions for the last centuries to millennia based on indirect climate information – so-called "proxies". Focus has been on both, large-scale averages as well as the reconstructions of regional to global fields (Masson-Delmotte et al., 2013; Smerdon and Pollack, 2016). Temporal and spatial resolution varied with the included paleoclimatic archives. However, most reconstructions for the past centuries rely heavily on the most abundant indirect climate archive, tree rings, and specifically on tree-ring width (TRW) and late-wood density (MXD). Differences between reconstructions have mostly been discussed with differences in reconstruction methodology in mind (Christiansen and Ljungqvist, 2017). However, a new study shows good agreement between a wide range of methods, if reconstructions are based on the same input data set (Neukom et al., 2019a; 2019b). Another recent study found that temperature sensitive tree-ring proxies from the PAGES2k database (Emile-Geay et al., 2017) lack multi-centennial trends,

source: https://doi.org/10.7892/boris.133283 | downloaded: 12.10.2024

https://doi.org/10.5194/cp-2019-80
Preprint. Discussion started: 6 August 2019
© Author(s) 2019. CC BY 4.0 License.

which are found in other proxy archives (Klippel et al., 2019). This suggests that the input data probably play a crucial role for differences between different reconstructions. Today, many proxy data archives are available, hence compiling input data for reconstruction is not only a matter of the amount of proxy data, but also of their selection, i.e., screening.

In this study, we therefore aim at evaluating the effect of various tree-ring data collections and their screening on the final reconstructions. Because of the relevance of temperature in the climate change discussion and the fact that many biological proxies react to temperature stress, temperature has so far been the variable of most interest. However, to study the underlying processes a multi-variable perspective is required. Therefore, we evaluate the effects of the input data choice, using a state-of-the-art data assimilation technique, which allows for multi-variable climate reconstructions in form of model simulations that are in optimal agreement with proxy information (Bhend et al., 2012; Franke et al., 2017).

Previous studies based on data assimilation techniques proposed that a higher quantity of input data would always be beneficial. Because the regression-based proxy system models weight each proxy series by their regression residuals, proxies that do not contribute information would be downweighted automatically (Steiger et al., 2018; Tardif et al., 2019). However, this weighting may not work perfectly because of two factors: first, the regression depends on overlapping paleodata and instrumental measurements, which often results in a small sample, uncertain residuals and possible model overfitting. Second, moisture and temperature sensitive proxies may correlate in the period of overlapping data and hence moisture sensitive paleodata will be used to correct temperature and vice versa. However, these two variables probably have very different multi-decadal to centennial variability (Franke et al., 2013). The growth limiting factor may even change over time (Babst et al., 2019).

In this study, we use the Kalman filter based state-of-the-art data assimilation technique introduced in Bhend et al. (2012), which is very similar to the methodology used in the last millennium reanalysis project (Hakim et al., 2016; Tardif et al., 2019). In our experiments, we focus on the effect of the input data choice on the final reconstruction. We compare three published collections of tree-ring records (focusing on TRW and MXD), of which at least two are commonly used for climate reconstructions. These have very different characteristics: (1) The B14 collection of 2287 consistently detrended TRW chronologies from the International Tree Ring Data Base(ITRDB), not screened for climate sensitivity (Breitenmoser et al., 2014); (2) TRW and MXD from the PAGES2K database version 2 (Emile-Geay et al., 2017), with a selection of 415 temperature sensitive records, most selected by a statistical screening for positive correlation with instrumental temperature; and (3) the N-TREND tree-ring collection of 54 TRW, MXD or blended TRW-MXD time series (Wilson et al., 2016), selected by experts to be the best temperature recorders. Thus, the three data sets cover the range from large sample size and spatial coverage but no screening for temperature sensitivity to small sample size and small spatial coverage but strict screening.

In the next section the method and data sets are introduced in greater detail before we show our results. Then we discuss the possible reasons for our results and the differences compared to previous studies. Finally, we draw our conclusion how an optimal proxy selection process should look like.

2 Data and Methods

We use three input data sets for comparison, all consist of annually resolved tree-ring measurements, which have no dating uncertainties:

https://doi.org/10.5194/cp-2019-80
Preprint. Discussion started: 6 August 2019
© Author(s) 2019. CC BY 4.0 License.

1. B14 is a collection Breitenmoser et al. (2014) of 2287 uniformly detrended and standardized TRW measurements from the ITRDB (Zhao et al., 2018). We use the full collection without any further screening for climate/temperature sensitivity. Hence, this represents the data set with the highest quantity of records. However, the weighting of temperature information in the paleodata is completely up to the reconstruction method.
2. PAGES2k is a collection of 415 TRW and MXD series from PAGES2k data base version 2 (Emile-Geay et al., 2017). These are all records that correlate significantly ($p<0.05$) with nearby instrumental temperature measurements and/or have been described by experts to represent temperature variability. This compilation represents a compromise of good quantity, large spatial coverage and good quality paleodata, but experts from various regional groups were differently strict in their screening procedure.
3. N-TREND is a collection of 54 tree-ring reconstructions based on TRW, MXD or a combination of both. They were chosen by experts to be just the best tree-ring paleodata for temperature reconstructions (Wilson et al., 2016). Hence, they are our low quantity, highest quality input data set with least spatial coverage.

The Ensemble Kalman Fitting (EKF) method is a variation of the Ensemble Kalman filter (Evensen, 2003), in which paleodata are assimilated serially (Bhend et al., 2012; Franke et al., 2017). In order to assimilate the paleodata, we need a forward model that simulates them in the model state vector. We use a multiple regression proxy system model (PSM) to simulate tree-ring observations using modeled temperature or precipitation. The regression model is calibrated with gridded instrumental data (CRU TS 3.1, Harris et al., 2014) in the period 1901-1970. It includes monthly temperature (and precipitation) during the growing season April to September. In this study, we limit the analysis to the northern hemisphere because the majority of the tree-ring observations can be found there. In the first four experiments (see Table 1), which only use temperature (T) in the PSM, we have 6 independent variables (i.e., local temperature of the 6 months). If we assume that tree growth was limited by temperature and moisture (TR) variability, we have 12 independent variables. In additional experiments we consider only regression models with consecutive months by fitting all possible combinations of consecutive months and choosing the PSM with the lowest Akaike information criterion (AIC). In the model with temperature and precipitation this can be a different sequence of months for each variable, but both have to be consecutive (e.g. growth is limited by April to June precipitation and June to September temperature). The variance of the regression residuals is used to specify the observation error covariance matrix (assumed diagonal) in the assimilation, i.e. the larger the residuals, the less weight an observation gets and the less the model simulations get corrected.

The same proxy series may exist in several data collections or even within a single collection, possibly in differently treated/detrended versions. We conduct experiments where we prevent single chronologies from being assimilated twice by only choosing the best proxy (smallest regression residuals) in a 0.1ºx0.1º (ca. 10km) grid. Note, for the sensitivity experiments in this study we ignore the length of the proxy records in case there should be records of different length within a grid box.

Background error covariances are calculated from the 30-member ensemble at each time step. This has the advantage of taking time variant covariance structures into account, for instance during El Niño vs La Nina years. The disadvantage is the small 30 ensemble member sample for covariance estimation. We apply a covariance localization to remove random covariances at distant locations, e.g. > 1500 km in case of temperature(Franke et

https://doi.org/10.5194/cp-2019-80
Preprint. Discussion started: 6 August 2019
© Author(s) 2019. CC BY 4.0 License.

al., 2017). A recent comparison of Valler et al. (2018) has shown superior performance when using an improved covariance estimation, which blends 50% of the 30-member time-dependent covariance with 50% of a 250-member "climatological" covariance (Experiment: 50c_PbL_Pc2L in Valler et al. (2018)). In this paper we use both the original setting as in Franke et al. (2017) as well as the improved setting proposed by Valler et al. (2018). Our paleo-reanalysis is based on anomalies from a 71-year period around the current year. Low frequency variability is a function of the models' response to the prescribed external forcings and background conditions, which include sea-surface temperatures. Because low frequency variability is not consistently preserved in paleodata (Franke et al., 2013; Klippel et al., 2019), but reasonably well represented in the model simulations of the last millennium (Franke et al., 2017), this approach is expected to provide consistent skill at all time scales. Note that by subtracting a running mean, model biases are retained. This circumvents a big problem in data assimilation approaches with temporally varying input data networks. Observations that gradually pull the model away from its biased state, can lead to artificial trends or step functions in time-series. However, it must be noted that the final reconstruction is consistent only in the model world.

We evaluate the quality of the reconstruction based on correlation with gridded instrumental observations of temperature, precipitation (Harris et al., 2014) and sea level pressure (Allan and Ansell, 2006) in the period 1901-1990 as a reference (x^{ref}, where x is the state vector). However, rather than analyzing just at correlation itself, we analyze correlation improvements over the original model simulations, because these forced simulations already correlate positively with observations in many locations. Correlation focuses on the co-variability, i.e. the correct sign of the anomaly. Additionally, we use a root-mean-square-error skill score (RE) that describes the improvement of the analysis (x^a) over the original model simulations (background, x^b) over all time steps (i).

$$RE = 1 - \frac{\sum\left(x_i^a - x_i^{ref}\right)^2}{\sum\left(x_i^b - x_i^{ref}\right)^2} \tag{1}$$

It is more difficult to reach positive RE values than correlation improvements, because this score punishes a wrong amplitude of variability (e.g., an uncorrelated reconstruction with correct variance would yield RE = -1). Because it is based on squared errors, too high variability is punished more than little variability, which the ensemble mean of the original model simulations has. We only evaluate correlation improvements and RE of the ensemble mean. Note that reconstructions and validation data are not completely independent, as we estimate PSM regression coefficient from the relationship between them. Hence, our correlation measures and RE skill score may be overestimated, which could be accounted for by using a higher observation error. However, both factors are not crucial for the sensitivity experiments presented in this study, where we are mostly interested in relative quality differences.

To evaluate the influence of the input data on the final reconstruction, we conducted the following set of experiments:

Table 1: Experiments

Name	Proxy system model	Description
NTREND_T	6 regression coeff. for Apr. to Sep monthly temperature (T)	Just using the best tree-ring chronologies for temperature reconstruction, which have been chosen by experts, i.e. very strict selection of few, best records

https://doi.org/10.5194/cp-2019-80
Preprint. Discussion started: 6 August 2019
© Author(s) 2019. CC BY 4.0 License.

PAGES_T	Same as above	Using a selection of temperature sensitive proxies, selected by the regional PAGES working groups, i.e. mostly statistical screening for temperature signal. Therefore, more records but less strictly screened than NTREND. Probably, included some moisture or partly moisture sensitive proxies, too.
B14_T	Same as above	Consistently detrended tree-ring data from the ITRDB by B14. This proxy set includes the largest amount of proxy series. However, many of them do not include any climate signal.
ALL_T	Same as above	All three data sets together, largest data set with greatest spatial coverage. However, duplicate proxies cannot be excluded
ALL_TR	12 regression coeff. For Apr. to Sep. monthly temperature and precipitation (R)	Same as above
ALL_TR_scr0.05	Same as above	Same as above but with additional screening, i.e. only records with a climate signal (p-value < 0.05) will be assimilated. In the experiments above these series got little weight due to large errors (regression residuals) but were still assimilated.
ALL_TR_scr0.05_ AIC_NOdup	Max. 12 regression coeff. but only consecutive months are allowed, still mixed temperature and precipitation signals are possible	Same as above but we chose with the AIC the regression model under the precondition that only climate from consecutive months can influence tree growth, which is more realistic due to local growing season length. Additionally, we remove duplicate proxies by only considering the best proxy (lowest regression residuals) within a 0.1ºx0.1º (ca. 10 km) grid. In each grid box we keep both, the best mainly temperature limited and the best mainly moisture sensitive proxy if both exist.
ALL_TR_scr0.05_ AIC_NOdup_ClimCovar	Same as above	Same as above but with background error-covariance estimate not only from the 30 ensemble members of the current year. Instead we use a mix of 50% error covariance coming from 250 random ensemble members and years.

https://doi.org/10.5194/cp-2019-80
Preprint. Discussion started: 6 August 2019
© Author(s) 2019. CC BY 4.0 License.

3 Results

We start with comparing experiments NTREND_T, PAGES_T and B14_T against observations, i.e., we compare the role of the choice of the three input data sets assuming only temperature dependence and no constraint on the regression model structure. In terms of correlation improvement over the background (i.e., the model simulations) in temperature (Fig. 1A,B,C) the highest local improvements are reached with the NTREND data set, however the largest spatial coverage of improvement is found with the B14 data set. Note that temperature correlation improves with all data sets and decreases nowhere, although some proxy records in the B14 data set do not contain any temperature signal. In terms of correlation the data assimilation scheme appears to weight the input data appropriately.

Although these first three experiments only use a temperature PSM, information can spread to other variables through the covariance matrix. Looking at precipitation correlation improvements (Fig. 2A,B,C), we find hardly any improvements with the NTREND collection. In contrast, the B14 data set leads to some precipitation correlation improvements over North America, where no NTREND series are located. B14 provides temperature information in places where temperature is correlated with precipitation.

The correct sign of the anomaly, measured by correlation, is only telling us one aspect of the reconstruction quality. To see if the amplitude of the anomaly is also reconstructed correctly, we look at the RE skill score (see methods). Here, we find large differences between the proxy collections. With NTREND_T we find improvements everywhere, whereas B14_T shows more regions with negative than positive skill (note that we use PSM with only temperature). The PAGES data set is again in the middle. This suggest that using moisture sensitive proxies to reconstruct temperature as in B14_T, which works just because temperature and precipitation are correlated at a given location, is not ideal. Hence, we would like to take the proxies' temperature or moisture sensitivity better into account and to find an option to use the PAGES and B14 collection at locations, where no expert selected proxies are available but rather keep the quality of the expert selected data, where it is available.

Before we come to a more sophisticated PSM and more sophisticated input data screening, we simply combine all three data sets using still a model with only temperature (ALL_T). This experiment performs well. Temperature correlation now reaches levels of the NTREND data set, where it is available and covers the entire region, where we only have data in PAGES or B14. RE values are positive in most regions, too. However, around India and the Himalaya as well as in the US southwest, skill is negative. Precipitation correlations improved only marginally (Fig. 4D) and precipitation RE (Fig. 5D) is mostly negative.

The obvious change to improve precipitation reconstruction skill is to use a PSM that includes precipitation, i.e. a multiple regression model with 12 coefficients for temperature and precipitation influence during the 6-months growing season (experiment ALL_TR). Temperature skill remains at the same high level, but precipitation skill clearly improves (Fig. 4E and 5E). Correlations improve everywhere and RE values become positive in most region with the exception of the Himalaya region and most northeast of Russia.

So far, we have not excluded any proxies from the data assimilation. We trust that proxies with no or a weak climate signal simply have regression coefficient close to zero and large residuals. This way they hardly affect the analysis. However, in a regression model with 12 independent variables and only 70 years of overlapping data, some records may just by chance get more weight than they deserve. Therefore, our next step is the introduction of a weak screening. In a first step, we only assimilate proxies with p-values < 0.05 for the full regression model

https://doi.org/10.5194/cp-2019-80
Preprint. Discussion started: 6 August 2019
© Author(s) 2019. CC BY 4.0 License.

(ALL_TR_scr0.05). This removes ca. 16% of the proxies and hardly affects correlations (Fig. 1F, 2F, 3F) but removes most of the negative RE values in both, temperature and precipitation (Fig. 4F and 5F).

This result appears good, but this could also be a result of overfitting the regression model. In addition to this statistical argument, allowing all possible combinations of the 12 variables makes not much physiological sense either. Hence, the next step is to constrain the model. The tree growth should be affected by climate conditions in a locally varying growing season of consecutive months. We fit all possible combinations of temperature and precipitation influences in consecutive months and chose the model with the lost AIC (see methods, note that additionally, duplicates are removed; experiment ALL_TR_scr0.05_AIC_NOdup). As a result of this more physically based growth model, reconstruction skill decreases slighty (Fig. 4G and 5G). This suggests that the previously noted improvement in RE was indeed due to overfitting. Nevertheless, correlations remain on the same high level everywhere (Fig. 1G, 2G, 3G). Only RE decreases in some regions with a high number of paleodata such as parts of China and parts of North America.

Recently, Valler et al. (2018) could show that major improvements of the method used in this study can be achieved by using a background error covariance matrix, which is not only calculated from the 30 ensemble members for the current year (Franke et al. 2017) but blended with a climatological error covariance matrix based on random years and ensemble members from the original model simulations (see methods, experiment ALL_TR_scr0.05_AIC_NOdup_ClimCovar). Using improved covariance information increases RE values again and only very few grid boxes with negative skill remain. Moreover, the largest effects of the better error covariance estimation appear in variables that have not been assimilated such as sea level pressure (Fig. 3H). This is very important because one of the reasons for using data assimilation instead to traditional statistical reconstruction techniques is the possibility to gain knowledge about further variables in a physically consistent way, which allows for a better dynamic interpretation of the identified climatic variations.

4 Discussion

The assimilation results with the three data sets and a temperature PSM in terms of temperature correlation differences are as expected. We calculate the regression coefficients based on instrumental temperature. Hence, all proxies that correlate in some way with instrumental temperature will be used to update the analysis temperature. The analysis has highest correlations improvements with instrumental temperature if the proxies themselves had highest correlations, which is the case for the NTREND data set with the best temperature proxies only. Correlations improvements are lower but cover a larger area with the B14 collection.

Note that correlation improvements can be a result of a negative relationship between tree-ring width and instrumental temperature if local growth is moisture limited and growing season temperature and precipitation are negatively correlated. This can be a benefit because through the covariance we use the extra information that dry summers are also warm and vice versa. Hence, we find much better precipitation correlation with the B14 collection than with the NTREND data set. However, using moisture sensitive trees to update temperature fields may cause problems. Precipitation variability shows high inter-annual variability in many locations but neither the same inter- to multi-decadal variability as temperature nor its centennial trend (Hartmann et al., 2013; Landrum et al., 2013). Hence, updating other than the assimilated variables through the covariance matrix can cause problems on longer than inter-annual scale (Tardif et al., 2019).

https://doi.org/10.5194/cp-2019-80
Preprint. Discussion started: 6 August 2019
© Author(s) 2019. CC BY 4.0 License.

The regression model is calibrated on the interannual time scale assuming that TRW limitations remain the same. However, this may not be the case (Babst et al. 2019), and therefore decadal-to-multidecadal variability may be less well represented. A similar argument holds for the update introduced by the model covariance matrix, which, although state dependent, may yield optimal estimates only for seasonal and not decadal time scales. However, our approach avoids these pitfalls in two ways. First, at multidecadal and longer time-scales, the model takes over, and therefore relations in our reconstructions are not constrained to be stationary across time scales. Furthermore, with our approach, the stationarity assumption is restricted to the regression model, thus it is a local stationarity - no further stationarity assumption concerning spatial variability is introduced except for experiment ALL_TP_scr0.05_AIC_NOdup_ClimCovar, where 50% of the background error covariance matrix is climatological and thus stationary. Most other approaches assume stationary spatial covariances.

Theoretically, it would it be optimal to assimilate all available data and let each record be weighted based by its error. However, the true observation error is unknown and the estimation uncertain. In our case, we use a multiple regression proxy-system model with 6/12 variables (six months of temperature and optionally six months of precipitation) in a 70-year period of overlapping instrumental data and proxy measurements to estimate regression coefficients. This rather short period and large number of independent variables can lead to overfitting the model and thus underestimating the observation error, which is defined by the regression residuals. Together with the low signal-to-noise ratio of many tree-ring chronologies, this can lead to an over- or under-correction of the model field in the assimilation step. An additional experiment with doubled observation error (not shown) increases RE values clearly. This suggests that PSM overfitting is part of the reason for the negative RE skill scores in the B14_T experiment in contrast to the NTREND_T experiment (Fig. 2A and C).

In the following experiments (ALL_TR_scr0.05, ALL_TR_scr0.05_AIC_NOdup, ALL_TR_scr0.05_AIC_NOdup_ClimCovar) we tried to reduce the consequences of uncertain error estimates step by step. Excluding proxies without a significant climate signal ($p<0.05$) for the full regression model, clearly improves the RE skill score for temperature and precipitation in large parts of Asia (Fig. 4F and 5F). This highlights the negative effects of spurious correlation – even if it is very weak – on the analysis. Hence, screening the data appears to be important, especially in data sparse regions, where there is no chance for better records with smaller errors to correct errors introduced due to spurious covariances. In other reconstruction methods, for instance principle component regression or the search for the best analogs, screening of records will additionally be necessary to avoid spatial biases due to non-homogeneous proxy distributions (Bradley, 1996; Rutherford et al., 2005). However, this is negligible in the data assimilation framework because the number of assimilated records has a regional instead of global impact and because the method provides a measure of uncertainty in form of ensemble spread at each grid cell.

In the experiment, in which we only allow for a single growing season (ALL_TR_scr0.05_AIC_NOdup) per year instead of a statistically optimal selection of months and by removing duplicate records that are in more than one of the data collections, correlations improve slightly but RE decreases slightly. Obviously, we continue with this more realistic setup, but note that the choice what is “best” depends on the chosen statistic or the reconstruction characteristics that are wished by the user. For instance, correlation just measures covariance whereas RE is based on squared errors and hence punishes especially large biases, i.e. it favors an underestimation variability over an overestimation.

https://doi.org/10.5194/cp-2019-80
Preprint. Discussion started: 6 August 2019
© Author(s) 2019. CC BY 4.0 License.

Finally, we introduce an improved background error covariance estimation scheme (ALL_TR_scr0.05_AIC_NOdup_ClimCovar, Valler et al. 2019). Because assimilated information is spread in space and in between variables through the covariance matrix, it is important to estimate covariances well. Estimating covariance from both, the 30 members at the current time step and from climatology and then blending both information, especially improves our results for variables, which have not been assimilated such as sea level pressure (Fig. 3H).

In reality, climate signals in tree-ring proxies may be even more complicated than a function of moisture availability and growing season temperature. Limiting factors may change over time (Babst et al., 2019) or light availability may be important and not always be highly correlated with temperature, i.e. more diffuse light after volcanic eruptions may stimulate growth (Stine and Huybers, 2014). More sophisticated proxy system forward models such as VS-Lite (Tolwinski-Ward et al., 2011) could be used in data assimilation (Acevedo et al., 2016). In fact, we have applied VS-lite to all TRW records in B14 (Breitenmoser et al., 2014). However, addressing the effects of using VS-lite rather than a regression model would require a dedicated paper.

Finally, we tested the order of assimilated data, because we assimilate data serially. In combination with using covariance localization, the order could influence the final reconstruction (Greybush et al., 2011). Assimilating the data from the best to worst record in terms of regression residuals and in opposite order from worst to best, hardly influenced correlation and RE skill scores at all (not shown). Hence, we continue to assimilate records starting with the best ones, similar to traditional reanalysis, which sort observations from the largest to smallest expected variance reduction in the reanalysis (Slivinski et al., 2019; Whitaker et al., 2008).

Conclusion

How to choose input data for paleo data assimilation? We address this question by comparing three paleodata compilations of different sizes as well as using all data set together in combination with various screening approaches.

Just using a large collection of proxy data (B14) does not lead to a skillful reconstruction. In contrast, just using a small expert selection of the best temperature proxies (NTREND) leads to a good high latitude temperature reconstruction but wastes the potential of modern data assimilation technique to reconstruct the 4-dimensional multi-variate state of the atmosphere. However, simply combing all available input data and leaving the weighting completely to a statistical model does not lead to optimal results, either. Rejecting records without a clear climatic signal, removing duplicates and using a physically plausible PSM altogether lead to a better reconstruction.

Hence the answer to our research question if it is better to assimilate all available proxy data or just the best expert selection has to be answered with: neither of the two is optimal.

We achieve the best results in terms of correlation and RE, if we use a large collection of proxy records. However, to make proper use of input data, which was not screened by experts, it is crucial to:

1. use proxy system models that properly represent the paleodata, here taking possible temperature and moisture limitations of tree growth into account.
2. use correct physical assumptions, in our case about tree growth, to avoid statistical overfitting.
3. remove input data with random, not significant climate signals.

https://doi.org/10.5194/cp-2019-80
Preprint. Discussion started: 6 August 2019
© Author(s) 2019. CC BY 4.0 License.

4. care about overfitting (underestimation of errors)

For a future project, it would be very interesting to study how different reconstruction methods handle these three differently screened data sets to see, if these results are valid for other reconstructions methods, too?

Author contribution

JF had the initial idea for this paper and performed most of the analysis and drafted the manuscript. VV contributed to the code development. SB helped to shape the manuscript and experimental design. RN contributed additional analysis and all authors provided critical feedback and contributed to the writing of the manuscript.

Competing interests

There are no competing interests present.

Acknowledgements

This project was supported by the Swiss National Science Foundation project 162668 (RE-USE) and EU ERC project 787574 (PALAEO-RA). We like to thank CSCS for their support in conducting the ECHAM simulations.

References

Acevedo, W., Reich, S. and Cubasch, U.: Towards the assimilation of tree-ring-width records using ensemble Kalman filtering techniques, Climate Dynamics, 46(5), 1909–1920, doi:10.1007/s00382-015-2683-1, 2016.

Allan, R. and Ansell, T.: A New Globally Complete Monthly Historical Gridded Mean Sea Level Pressure Dataset (HadSLP2): 1850–2004, JCLI, 19, 5816–5842, doi:10.1175/JCLI3937.1, 2006.

Babst, F., Bouriaud, O., Poulter, B., Trouet, V., Girardin, M. P. and Frank, D. C.: Twentieth century redistribution in climatic drivers of global tree growth, Science Advances, 5(1), doi:10.1126/sciadv.aat4313, 2019.

Bhend, J., Franke, J., Folini, D., Wild, M. and Brönnimann, S.: An ensemble-based approach to climate reconstructions, Climate of the Past, 8, 963–976, doi:10.5194/cp-8-963-2012, 2012.

Bradley, R. S.: Are there optimum sites for global paleotemperature reconstruction?, in Climatic Variations and Forcing Mechanisms of the Last 2000 Years, vol. 3, pp. 603–624, Springer, Berlin, Heidelberg, Berlin, Heidelberg. 1996.

Breitenmoser, P., Brönnimann, S. and Frank, D.: Forward modelling of tree-ring width and comparison with a global network of tree-ring chronologies, Climate of the Past, 10(2), 437–449, doi:10.5194/cp-10-437-2014, 2014.

Christiansen, B. and Ljungqvist, F. C.: Challenges and perspectives for large-scale temperature reconstructions of the past two millennia, RG, 55(1), 40–96, doi:10.1002/2016RG000521, 2017.

Emile-Geay, J., McKay, N. P., Kaufman, D. S., Gunten, von, L., Wang, J., Anchukaitis, K. J., Abram, N. J., Addison, J. A., Curran, M. A. J., Evans, M. N., Henley, B. J., Hao, Z., Martrat, B., McGregor, H. V., Neukom, R., Pederson, G. T., Stenni, B., Thirumalai, K., Werner, J. P., Xu, C., Divine, D. V., Dixon, B. C., Gergis, J., Mundo, I. A., Nakatsuka, T., Phipps, S. J., Routson, C. C., Steig, E. J., Tierney, J. E., Tyler, J. J., Allen, K. J.,

https://doi.org/10.5194/cp-2019-80
Preprint. Discussion started: 6 August 2019
© Author(s) 2019. CC BY 4.0 License.

Bertler, N. A. N., Björklund, J., Chase, B. M., Chen, M.-T., Cook, E., de Jong, R., DeLong, K. L., Dixon, D. A., Ekaykin, A. A., Ersek, V., Filipsson, H. L., Francus, P., Freund, M. B., Frezzotti, M., Gaire, N. P., Gajewski, K., Ge, Q., Goosse, H., Gornostaeva, A., Grosjean, M., Horiuchi, K., Hormes, A., Husum, K., Isaksson, E., Kandasamy, S., Kawamura, K., Kilbourne, K. H., Koç, N., Leduc, G., Linderholm, H. W., Lorrey, A. M., Mikhalenko, V., Mortyn, P. G., Motoyama, H., Moy, A. D., Mulvaney, R., Munz, P. M., Nash, D. J., Oerter, H., Opel, T., Orsi, A. J., Ovchinnikov, D. V., Porter, T. J., Roop, H. A., Saenger, C., Sano, M., Sauchyn, D., Saunders, K. M., Seidenkrantz, M.-S., Severi, M., Shao, X., Sicre, M.-A., Sigl, M., Sinclair, K., St George, S., St Jacques, J.-M., Thamban, M., Thapa, U. K., Thomas, E. R., Turney, C., Uemura, R., Viau, A. E., Vladimirova, D. O., Wahl, E. R., White, J. W. C., Yu, Z. and Zinke, J.: Data Descriptor: A global multiproxy database for temperature reconstructions of the Common Era, Sci. Data, 4, doi:10.1038/sdata.2017.88, 2017.

Evensen, G.: The Ensemble Kalman Filter: theoretical formulation and practical implementation, Ocean Dynam, 53(4), 343–367, doi:10.1007/s10236-003-0036-9, 2003.

Franke, J., Brönnimann, S., Bhend, J. and Brugnara, Y.: A monthly global paleo-reanalysis of the atmosphere from 1600 to 2005 for studying past climatic variations, Sci. Data, 4, 170076, doi:10.1038/sdata.2017.76, 2017.

Franke, J., Frank, D., Raible, C. C., Esper, J. and Brönnimann, S.: Spectral biases in tree-ring climate proxies, Nature Climate change, 3(4), 360–364, doi:10.1038/nclimate1816, 2013.

Greybush, S. J., Kalnay, E., Miyoshi, T., Ide, K. and Hunt, B. R.: Balance and Ensemble Kalman Filter Localization Techniques, Monthly Weather Review, 139(2), 511–522, doi:10.1175/2010MWR3328.1, 2011.

Hakim, G. J., Emile-Geay, J., Steig, E. J., Noone, D., Anderson, D. M., Tardif, R., Steiger, N. and Perkins, W. A.: The last millennium climate reanalysis project: Framework and first results, J. Geophys. Res. Atmos., 121(1), 6745–6764, doi:10.1002/2016JD024751, 2016.

Harris, I., Jones, P. D., Osborn, T. J. and Lister, D. H.: Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset, Int. J. Climatol., 34(3), 623–642, doi:10.1002/joc.3711, 2014.

Hartmann, D. L., Tank, A. K., Rusticucci, M., Alexander, L. V., Brönnimann, S., Charabi, Y., Dentener, F. J., Dlugokencky, E. J., Easterling, D. R., Kaplan, A., Soden, B. J., Thorne, P. W. and Wild, M.: Observations: atmosphere and surface. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, in Climate Change 2013 - The Physical Science Basis, edited by T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. M. Midgley, pp. 159–254, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. 2013.

Klippel, L., St George, S., Büntgen, U., Krusic, P. J. and Esper, J.: Differing pre-industrial cooling trends between tree-rings and lower-resolution temperature proxies, Clim. Past Discuss, 1–21, doi:10.5194/cp-2019-41, 2019.

Landrum, L., Otto-Bliesner, B. L., Wahl, E. R., Conley, A., Lawrence, P. J., Rosenbloom, N. and Teng, H.: Last Millennium Climate and Its Variability in CCSM4, J Climate, 26(4), 1085–1111, doi:10.1175/JCLI-D-11-00326.1, 2013.

Masson-Delmotte, V., Schulz, M., Abe-Ouchi, A., Beer, J., Ganopolski, A., Gonzalez-Rouco, F. J., Jansen, E., Lambeck, K., Luterbacher, J., Naish, T., Osborn, T., Otto-Bliesner, B., Quinn, T., Ramesh, R., Rojas, M., Shao, X. and Timmermann, A.: Information from paleoclimate archives, in Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by T. F. Stocker, D. Qin, G.-K.

https://doi.org/10.5194/cp-2019-80
Preprint. Discussion started: 6 August 2019
© Author(s) 2019. CC BY 4.0 License.

Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. M. Midgley, pp. 383–464, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. 2013.

Neukom, R., Barboza, L. A., Erb, M. P., Shi, F., Emile-Geay, J., Evans, M. N., Franke, J., Kaufman, D. S., Lücke, L., Rehfeld, K., Schurer, A., Zhu, F., Brönnimann, S., Hakim, G. J., Henley, B. J., Ljungqvist, F. C., McKay, N., Valler, V. and Gunten, von, L.: Consistent multidecadal variability in global temperature reconstructions and simulations over the Common Era, Nature Geosci, 536(8), 411, doi:10.1038/s41561-019-0400-0, 2019a.

Neukom, R., Steiger, N., Gómez-Navarro, J. J., Wang, J. and Werner, J. P.: No evidence for globally coherent warm and cold periods over the preindustrial Common Era, Nature, 571(7766), 550–554, doi:10.1038/s41586-019-1401-2, 2019b.

Rutherford, S., Mann, M. E., Osborn, T. J., Bradley, R. S., Briffa, K. R., Hughes, M. K. and Jones, P. D.: Proxy-based Northern Hemisphere surface temperature reconstructions: Sensitivity to method, predictor network, target season, and target domain, J Climate, 18(13), 2308–2329, 2005.

Slivinski, L. C., Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Giese, B. S., McColl, C., Allan, R., Yin, X., Vose, R., Titchner, H., Kennedy, J., Spencer, L. J., Ashcroft, L., Brönnimann, S., Brunet, M., Camuffo, D., Cornes, R., Cram, T. A., Crouthamel, R., Castro, F. D., Freeman, J. E., Gergis, J., Hawkins, E., Jones, P. D., Jourdain, S., Kaplan, A., Kubota, H., Le Blancq, F., Lee, T. C., Lorrey, A., Luterbacher, J., Maugeri, M., Mock, C. J., Moore, G. W. K., Przybylak, R., Pudmenzky, C., Reason, C., Slonosky, V. C., Smith, C., Tinz, B., Trewin, B., Valente, M. A., Wang, X. L., Wilkinson, C., Wood, K. and Wyszyn'ski, P.: Towards a more reliable historical reanalysis: Improvements for version 3 of the Twentieth Century Reanalysis system, Quarterly Journal of the Royal Meteorological Society, qj.3598, doi:10.1002/qj.3598, 2019.

Smerdon, J. E. and Pollack, H. N.: Reconstructing Earth's surface temperature over the past 2000 years: the science behind the headlines, Wiley Interdisciplinary Reviews: Climate Change, 7(5), 746–771, doi:10.1002/wcc.418, 2016.

Steiger, N. J., Smerdon, J. E., Cook, E. R. and Cook, B. I.: A reconstruction of global hydroclimate and dynamical variables over the Common Era, Sci. Data, 5, 180086–15, doi:10.1038/sdata.2018.86, 2018.

Stine, A. R. and Huybers, P.: Arctic tree rings as recorders of variations in light availability, Nature Communications, 5(1), 3836, doi:10.1038/ncomms4836, 2014.

Tardif, R., Hakim, G. J., Perkins, W. A., Horlick, K. A., Erb, M. P., Emile-Geay, J., Anderson, D. M., Steig, E. J. and Noone, D.: Last Millennium Reanalysis with an expanded proxy database and seasonal proxy modeling, Climate of the Past, 15(4), 1251–1273, doi:10.5194/cp-15-1251-2019, 2019.

Tolwinski-Ward, S. E., Evans, M. N., Hughes, M. K. and Anchukaitis, K. J.: An efficient forward model of the climate controls on interannual variation in tree-ring width, Climate Dynamics, 36(1), 2419–2439, doi:10.1007/s00382-010-0945-5, 2011.

Valler, V., Franke, J. and Brönnimann, S.: Impact of different estimations of the background-error covariance matrix on climate reconstructions based on data assimilation, Clim. Past Discuss, 1–27, doi:10.5194/cp-2018-168, 2018.

Whitaker, J. S., Hamill, T. M., Wei, X., Song, Y. and Toth, Z.: Ensemble data assimilation with the NCEP Global Forecast System, Monthly Weather Review, 136(2), 463–482, doi:10.1175/2007MWR2018.1, 2008.

Wilson, R., Anchukaitis, K., Briffa, K. R., Büntgen, U., Cook, E., D'arrigo, R., Davi, N., Esper, J., Frank, D., Gunnarson, B., Hegerl, G., Helama, S., Klesse, S., Krusic, P. J., Linderholm, H. W., Myglan, V., Osborn, T. J.,

https://doi.org/10.5194/cp-2019-80
Preprint. Discussion started: 6 August 2019
© Author(s) 2019. CC BY 4.0 License.

Rydval, M., Schneider, L., Schurer, A., Wiles, G., Zhang, P. and Zorita, E.: Last millennium northern hemisphere summer temperatures from tree rings: Part I: The long term context, Quaternary Science Reviews, 134, 1–18, doi:10.1016/j.quascirev.2015.12.005, 2016.

Zhao, S., Pederson, N., D'Orangeville, L., HilleRisLambers, J., Boose, E., Penone, C., Bauer, B., Jiang, Y. and Manzanedo, R. D.: The International Tree-Ring Data Bank (ITRDB) revisited: Data availability and global ecological representativity, J Biogeogr, 46(2), 355–368, doi:10.1111/jbi.13488, 2018.

Whitaker, J. S., Hamill, T. M., Wei, X., Song, Y. and Toth, Z.: Ensemble data assimilation with the NCEP Global Forecast System, Monthly Weather Review, 136(2), 463–482, doi:10.1175/2007MWR2018.1, 2008.

Wilson, R., Anchukaitis, K., Briffa, K. R., Büntgen, U., Cook, E., D'arrigo, R., Davi, N., Esper, J., Frank, D., Gunnarson, B., Hegerl, G., Helama, S., Klesse, S., Krusic, P. J., Linderholm, H. W., Myglan, V., Osborn, T. J.,
Rydval, M., Schneider, L., Schurer, A., Wiles, G., Zhang, P. and Zorita, E.: Last millennium northern hemisphere summer temperatures from tree rings: Part I: The long term context, Quaternary Science Reviews, 134, 1–18, doi:10.1016/j.quascirev.2015.12.005, 2016.

Zhao, S., Pederson, N., D'Orangeville, L., HilleRisLambers, J., Boose, E., Penone, C., Bauer, B., Jiang, Y. and Manzanedo, R. D.: The International Tree-Ring Data Bank (ITRDB) revisited: Data availability and global ecological representativity, J Biogeogr, 46(2), 355–368, doi:10.1111/jbi.13488, 2018.

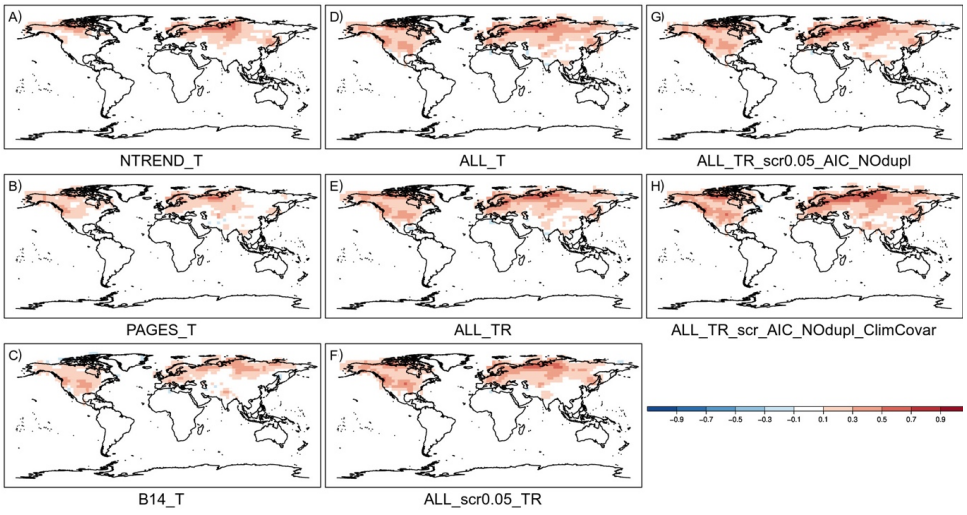

Figure 1: Temperature correlation improvement of the analysis over the original model simulations, i.e. correlation between analysis and CRU TS minus correlation between simulations and CRU TS, where red colors indicate an improvement of the analysis. All maps show the Apr. to Sep. growing season of the northern hemisphere.

https://doi.org/10.5194/cp-2019-80
Preprint. Discussion started: 6 August 2019
© Author(s) 2019. CC BY 4.0 License.

Climate of the Past
Discussions

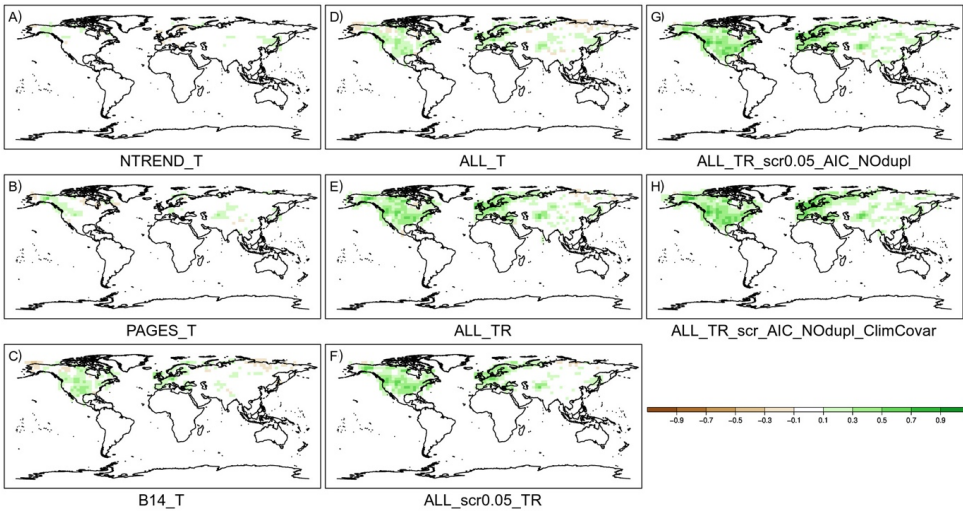

Figure 2: Same as Fig. 1 for precipitation correlation, where green colors indicate an improvement of the analysis.

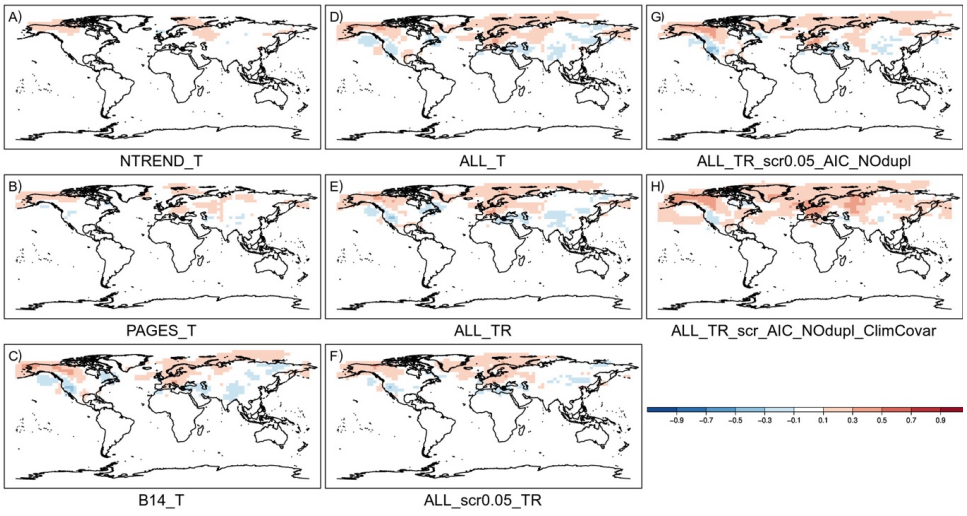

Figure 3: Same as Fig. 1 for SLP correlation, where red colors indicate an improvement of the analysis.

https://doi.org/10.5194/cp-2019-80
Preprint. Discussion started: 6 August 2019
© Author(s) 2019. CC BY 4.0 License.

Climate of the Past
Discussions

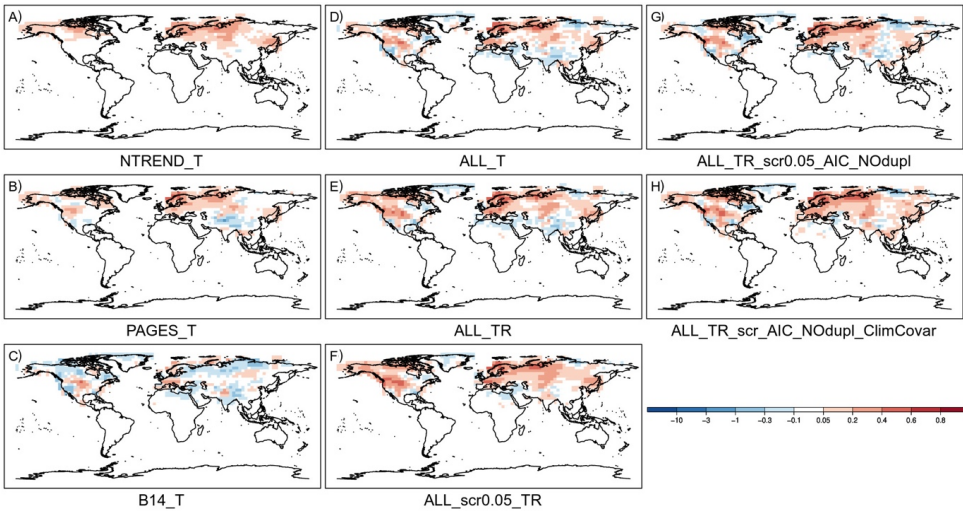

Figure 4: Temperature RE skill score, where red colors indicate an improvement of the analysis.

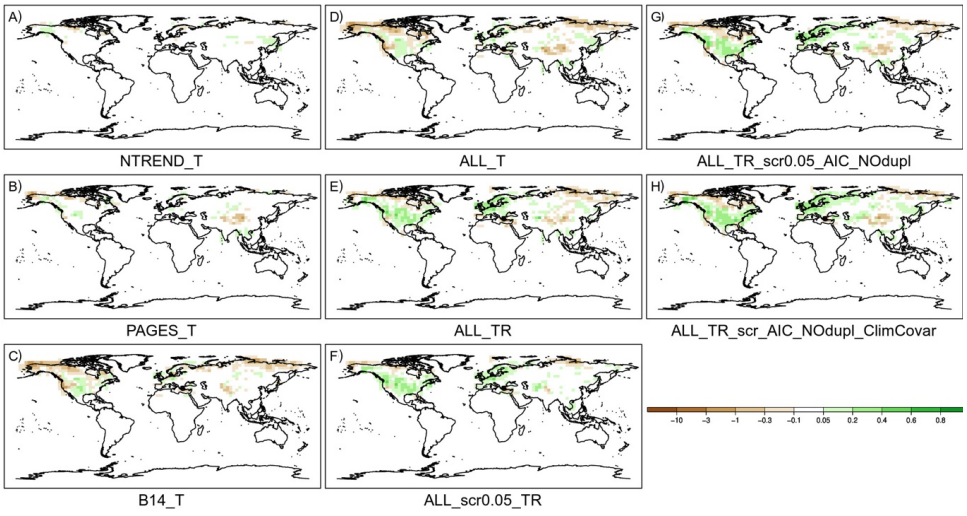

Figure 5: Precipitation RE skill score, where greens colors indicate an improvement of the analysis.