Proxy-based Reconstructions of Northern Hemisphere Temperature over the Common Era: Emulation and Sensitivity Analysis

Jill Hardy

Department of Earth Sciences, University of Southern California, Los Angeles, CA

Senior Undergraduate Thesis in the Department of Earth Sciences, under the supervision of Dr. Julien Emile-Geay

ABSTRACT

The climate of the Earth is not in a steady state. From ice ages to warm periods, there has been considerable variation in the Earth’s climate throughout its existence. In particular, understanding climate of the Common Era (past 2000 years) is important because it can help lead to a better estimate of the far distant past, as well as, shed light on future variability. Additionally, the climate change debate has brought attention to human impacts on recent climate and whether the current trends are outside the realm of natural variability. While this has created a sense of urgency within the field, the attempts to accurately depict the climate of the Common Era have proven difficult due to the sparse instrumental data available, which reaches no more than a couple centuries. With that, scientists must employ a variety of natural “proxy” data, such as ice cores, corals, and tree rings, to interpret the climate state further back in time. Paleoclimate reconstructions are created which use these proxies to try to determine past temperature variability. For this project, the first objective was to emulate a popular reconstruction of temperature of the Common Era published by Mann et al. (2008). This reconstruction was chosen because of its prominence within the climate change debate; for climate scientists, it has become clear evidence for global warming, but it has been scrutinized by skeptics. Reproducing this study is the starting point because it allowed us to follow the necessary steps that created a well-sought out reconstruction, as well as, give a reputable comparison for later. This emulation was done using the existing reconstruction methods of “composite-plus-scale” (CPS) and “Regularized Expectation-Maximum” (RegEM). As in their study, the “Truncated Total Least Squares” (TTLS) regularization scheme in RegEM is used. Second, we compared these results to those produced by using a new regularization scheme in RegEM, called iTTLS. This comparison will determine whether the data-adaptive property of iTTLS significantly changes the results of the Mann et al. (2008) study, and if so, how. Ultimately, we would like to find that the new scheme improves upon the existing reconstruction, thus, providing more statistically sound results that will help to settle the climate change debate. Lastly, using TTLS, we created individual reconstructions of the different classes of proxies from the Mann study, rather than having one large dataset. It is important to understand the influence of each proxy class on the emulation, so to gauge its effects, pinpoint any problems, and consider any improvements.

1. Introduction

The state of climate over the next century is currently one of the main scientific discussions, with the debate about whether the current temperature trends exceed the natural variability of the planet, leading to many unanswered questions (Folland et al., 1990) (Jansen et al., 2007).
Because of this attention, the scientific community is working towards new methods to create these estimates on centennial timescales. Understanding climate change of the last 2000 years is very useful because better estimates of the recent past are the most reasonable sources of information to improve global circulation model (GCM) performance that predict the near future. The main problem of this field, however, is that the instrumental data, the only known measure for comparing against and calibrating paleoclimate reconstructions, does not exist past roughly 1850AD, thus making it difficult to portray the temperature further back in time with much accuracy. Due to the brevity of this data, we must use natural “proxy” data to supplement what is missing in the instrumental field. These proxies come in the form of ice cores, corals, tree rings, speleothems, and sediment cores, all of which retain interpretable information that can be used to “reconstruct” past climate variability.

A paleoclimate reconstruction is the statistical translation of this proxy data into temperature estimates. However, a major difficulty is that these proxy records become sparser the further back in time, which complicates the statistical estimation. The current approach of using multiproxy databases allows for every proxy in the dataset to be used in conjunction so as to maximize the information available as data becomes sparser with time. The Mann et. al. 2008 (hereby referred to as MANN08) study is one of the first examples of using a multiproxy dataset for northern hemisphere (NH) temperature reconstructions. Covering the Common Era, this study used a large database to create reconstructions with a variety of different criteria. With proxies spanning as far back as 0AD, it is the most comprehensive multiproxy database assembles to date to cover this interval.
This thesis focuses on using the MANN08 multiproxy dataset for two purposes: (1) assessing reproducibility of the original study using entirely different code, and (2) evaluate the sensitivity of the reconstructions to each proxy class.

Emulating previous work is a key part of science. The ability to reproduce a result is the absolute basis of science, and confirms that the process to achieving it is sound. MANN08 created a lot of discussion upon its publication because it had results that many skeptics were ready to attack. But so far, no one has been able to prove it wrong, so the study is regarded by climate scientists as state-of-the-art, and so it is a valued exercise to reproduce it. This prominence is what made MANN08 interesting to emulate and use as a benchmark to compare to our method.

Individual proxy class reconstructions are also interesting because each class provides unique information to the composite reconstruction. One of the questions considered is the degree of influence a class has on the total reconstruction. For instance, previous Common Era surface temperature reconstructions have been criticized for depending too heavily on tree ring data which is an issue because they are prone to “divergence.” Divergence is when instrumental temperature data disagrees with ring widths, such that known thermometers measure a warming trend, however, the tree ring widths do not pick up this trend. Therefore, the timeseries diverge from one another, with the tree ring proxies producing underestimations of the instrumental data (D'Arrigo et al., 2008). This problem could affect paleoclimate reconstructions because there is concern that if dendrochronological proxies saturate beyond a certain temperature, they may not be able to faithfully reconstruct past warm intervals to their full extent. If we specifically reconstruct tree ring proxy data together, then we can see how much these proxies affect the overall reconstruction and assess methods to get better results. Each proxy class has intrinsic characteristics that cause it to contribute differently to the reconstruction. Systematically analyzing each
class separately will hopefully shed light onto their individual contribution to the overall reconstruction.

Using the “composite-plus-scale” (CPS) (Jones and Mann, 2004) and the “Regularized Expectation-Maximization” (RegEM), in particular, the “Truncated Total Least Squares” (TTLS) regularization scheme in RegEM, MANN08 found that the northern hemisphere (NH) warming associated with the past decade is in fact outside the realm of natural variability, at least back to 700AD (Mann et al., 2008). With the issue of divergence surrounding tree-ring proxies, the inclusion of these data requires careful consideration of their possible effects; there is a trade-off between having the additional data and the bias that may exist from the divergence issue. Therefore, in this study, reconstructions were made both including and excluding tree-ring data (Broecker, 2001; Esper et al., 2002; Mann et al., 1999; Moberg et al., 2005; Rutherford et al., 2005). Without tree-ring data, thus making the reconstructions less biased, the anomalous warming can only be verified for the last 1300 years. Including the data, but subjecting the reconstructions to possible bias, the warming is verified for the last 1700 years. It was found that the TTLS method was less affected by the divergence problem compared to the CPS method during the calibration exercises, and so, its results were more reliable to reconstruct large-scale temperature fluctuations.

The thesis is organized as follows: in section 2, we describe the MANN08 study, including the proxy database and instrumental data; in section 3, we provide the technical background of the reconstruction methods used by MANN08 and our project; section 4 describes the MANN08 emulation procedure and results; section 5 is the description of the iTTLS results; section 6 explains the proxy class reconstructions and their results; and the thesis ends with a Discussion and Conclusion in sections 7 and 8, respectively.
2. Description of Mann et. al. 2008 study

2.1 Proxy Database

Choosing a decent proxy database was the most important decision in this project because it is the basis of the emulation, method comparison, and sensitivity analysis. We chose the MANN08 database for several reasons. For one, it is a very large proxy database, consisting of 1209 total proxies. The study specifically picked proxies which were reported to have good paleoclimate signals. Most of these proxies are annually resolved (1158) and the rest are decadal-ly resolved (51). The decadally resolved records were interpolated and lowpass filtered at a nominal annual resolution. With such a large amount of high resolution proxies, we were more confident that the proxy records would better portray the yearly temperature changes during the validation exercises with the instrumental period, as well as, after the instrumental record does not exist. Second, the proxies were sampled from all over the world (Figure 1). The majority is within the northern hemisphere (1036), favoring land and extratropics; however, there are some proxies in southern hemisphere (173), ocean, and tropical proxies (Mann et al., 2008). For our purposes, we screened for just the NH proxies because the data is more abundant and reliable. Because of this assortment, the database naturally incorporates several classes of proxies: tree rings, ice cores, corals, speleothems, sediment cores, historical records, and a special surface temperature reconstruction based on a combination of proxy, historical, and instrumental data (Luter et al). Utilizing these classes together maximizes the amount of data incorporated for every year of the reconstruction, and thus, provides more statistically sound results.
Figure 1. Top: map showing the distribution of the 1209 proxies from the MANN08 study. Shapes and colors are based on the type of proxy. Bottom: proxy availability over the Common Era, based on proxy class.

2.2 Instrumental Data

Instrumental data for surface-air temperature are from the University of East Anglia Climatic Research Unit, covering the 1850-2006 period. For the purposes of this project, we only used the combined land and ocean temperature dataset (HadCRUT3v). Also, since we focused on northern hemisphere temperatures, we just used the Had NH means, filtering it at 10-year low and high frequencies (Figure 2).
3. Technical Background

3.1 Composite Plus Scale (CPS)

CPS takes proxy data that is believed to have some degree of predictive power for past temperature fluctuations and first, standardizes and centers it. Then, CPS weighs the data if necessary so that proxies with better sensitivity are considered more strongly. The data is averaged to create a timeseries over the specified region, whether it be hemispheric or global. Finally, the series is scaled against the instrumental record to form a temperature reconstruction (Mann et al., 2008) (Mann PNAS 08 + original references).

There are caveats to the CPS method that has caused it to be phased out for other reconstruction methods. For instance, it tends to smooth too much of the low frequency trends because of its standardization at every “nest.” A nest is an interval of a timeseries that is independently analyzed from all the other intervals and only spliced together at the end. Since many studies are interested in these results, CPS is not the best method to use. Additionally, CPS relies upon a statistical correlation with local temperature. This limits the spatial extent of the composite and may
not reflect large-scale variations as well. However, CPS is a simple method and is, therefore, computationally inexpensive. For large datasets, like MANN08, it is a good method to use as a comparison to the more rigorous method.

3.2 Regularized Expectation-Maximization (RegEM)

The RegEM reconstruction was the main focus of this project. RegEM is useful because it creates estimates for missing values in proxy data and then uses the completed dataset for reconstruction. Very rarely will a proxy dataset that extends far back in time be without missing values because of the difficulty in collecting and analyzing proxy data. However, to exclude far-reaching timeseries that have a few missing values is very limiting and a waste of information. Therefore, by a series of “imputations” into and estimations of the data matrix, RegEM creates a filled matrix that can be further manipulated to create a reconstruction. The process of estimating missing values given available ones is called “imputation” in the statistical literature (Little & Rubin, 2002)

In order to replace the missing values, a regression coefficient must be estimated. A regression coefficient is a value that is multiplied by the available data matrix to form an estimate of the missing value. To compute such regression coefficients, the covariance matrix of the available data must be invertible. This means that the matrix must have an inverse matrix, which is necessary for the linear algebra equations that regression coefficient depends upon. If the matrix does not have an inverse, then the regularization step of RegEM must occur. The basis of regularization is the addition of a regularization parameter that will affect the matrix in such a way that it will become invertible. RegEM looks at a range of parameters, averages their results, and uses it with the matrix. Once the regularization occurs, the regression coefficient is created and then the regression step can occur to fill in the missing values.
The regression begins by imputing random guesses (usually zeros) into the missing values of the matrix. The mean and covariance matrices are computed which are statistics that show the relationship amongst the data. The algorithm moves year-by-year using the other available proxy values to estimate the missing value, given the regression matrix computed over the period of overlap. Once it has moved through the whole matrix, it re-calculates the mean and covariance matrices. The updated statistics are then used as a guide for the next iteration of the process. This time, it will remember where the missing values were replaced and update them again with values that exhibit the improved statistics. Through this method, the missing values become more accurate with each redefinition of the statistics. RegEM stops when a threshold is reached where the data matrix is no longer changing from each iteration of the matrix.

The difference between TTLS and iTTLS (“i” for individual) lies in the regularization scheme of RegEM. During the regularization, principal component analysis (PCA) is done on the covariance matrix. This analysis accounts for all of the variance (a measure of the spread) for the matrix. The variance is organized into principal components (PC), where the first component accounts for the most variability of the data, and each PC thereafter accounts for less. Each component is uncorrelated, so that each bit of variability is uniquely assigned to that PC. However, after a certain PC, the amount of variability in the covariance matrix that it explains is negligible compared to the amount of the previous components. At this point, we can “truncate” the remaining components and assume their contribution to the regularization is insignificant.

This truncation parameter is where TTLS and iTTLS differ. TTLS requires the parameter be specified, whereas, iTTLS has a built-in algorithm that chooses the truncation parameter. The data-adaptive quality of iTTLS is superior to TTLS because it removes the subjectivity of picking the parameter and lets it be completely automated. After the variance is truncated, the im-
proved data is manipulated and put into the covariance matrix. RegEM can then move onto the regression steps with an improved covariance matrix from the last iteration.

4. **Emulation of Mann et. al. 2008 study**

4.1 **Motivation**

Scientific progress relies on reproducibility (climatecode.org). Without being able to recreate the process of getting results, it is difficult to trust that they are sound, let alone be able to improve upon them. The accessibility of the MANN08 code and results enabled us to have almost everything necessary to reproduce the data, and thus, corroborate the scientific progress of their study. In addition to this, for this project, the emulation provides a learning experience for how paleoclimate reconstructions work. Since one of the goals is to compare how iTTLS does against TTLS, we need to be sure that the way both are done are the same in every aspect except the regularization chosen. With these reasons in mind, the first part of this project was to follow all of the steps done in the MANN08 study to create an emulation of their northern hemisphere reconstruction.

4.2 **Methods**

In order for a proxy series to be deemed usable for predicting temperature, it was required to pass a screening process. A proxy was compared to its closest temperature point during the full calibration interval of 1850-1995AD (146 years), correlation statistics were produced, and their significance assessed. The significance thresholds were evaluated at the P=0.10 level through a one-sided T-test; threshold values for the sample correlation were $|r| = 0.11$ and $|r| = 0.34$, for each the annually and decadally resolved proxies respectively (MANN08 SI). The degrees of
freedom were adjusted for autocorrelation. If it was found to be statistically significant, then the proxy passed the screening process and was used towards creating the long term reconstruction. For the northern hemisphere, 421 proxies passed the screening process.

In MANN08, 60 different reconstructions were created using TTLS depending on a variety of factors, including the hemisphere (NH, SH, or global), the type of instrumental data (land-only, land-ocean, infilled), and the proxies (screened or full) desired. Based on the reconstruction we wanted to make, we needed to figure out which of the many versions was the comparable reconstruction. Ultimately, we decided to reconstruct the northern hemisphere temperature using the land-ocean annual mean NH HadCRUT3v data, and incorporate the screened proxies alone. This decision was made because the NH temperatures are the most reliable reconstructions due to the vast amount of proxy data available in the NH. SH and global reconstructions are less dependable due to the sparseness of proxy data in the southern hemisphere. Additionally, since oceans cover the majority of the NH, the best estimate of NH temperature should incorporate available data from oceans. Therefore, we chose the instrumental dataset that provided data to do so. Finally, the screening process is worthwhile because it narrows down the large database to one where we can be more confident that we are not including extremely noisy proxies. However, the screening does have downsides, including uncertainty in the set-up of the significance test of which the whole process relies upon. Also, the process requires a correlation with local temperature, which might not be the only way to assess the relevance of the proxy data.

4.3 Frozen Network Analysis

An important prerequisite for generating long reconstructions is creating verification statistics on a subset of the data where there is overlap between proxy and instrumental data. These validation exercises help to assess the effectiveness of the reconstruction methods. Using the
span of the 146-year instrumental period, some of the data is used to create a reconstruction just over the chosen timeframe, while the remaining data is withheld as a comparison. Two 100-year calibration intervals were chosen from 1850-1949 (early) and 1896-1995 (late). The remaining 46-year periods in each were used as verification intervals. This means that the proxy data made available during the 100-year calibration interval was used to “reconstruct” the corresponding 46-year verification interval. The verification interval data is then unmasked and compared to see how well the reconstruction method predicted what we have measured, thus creating verification statistics.

To do this, each calibration/verification set was analyzed separately and then averaged in order to produce the most accurate representation possible. “Nests” were created in 100-year time periods from 1850-0AD, with the first nest being 50 years long (1850-1800AD). There were 19 total nests. In each nest, the proxies with available data during that time period were noted (e.g. all of the proxies with data from 1499-1400AD were in one nest). If no data was present during a particular nest, the reconstruction method would not choose that proxy to help with the reconstruction at that time slice. Naturally, the nests contained less data the further back in time, as fewer proxies spanned the entire 1996 years. The reconstruction method moved nest-by-nest through the proxy database and only incorporated their data from 1995AD to 1850AD for the reconstructions; the data past 1850AD was only used to acknowledge whether a proxy contained data in a particular nest. This process emulates how the full reconstruction works because it only incorporates proxies available at the given time and ignores the rest. Figure 3 exhibits these reconstructions for low and high frequency trends, as well as, the unfiltered data. The vertical dashed line denotes the boundary between the calibration and verification intervals. The gray line is the withheld verification data, thus showing where the ideal reconstruction would trend. None
of the reconstructions do a great job, but the late calibration/early verification set is overall better than the other set.

Figure 3. Frozen network analysis reconstructions compared to the ideal (original) data. Left: late verification interval (1950-1995). Right: early verification interval (1850-1895). A perfect reconstruction would follow the original series (gray) exactly and have an RE score of one.

Once the reconstructions were created, they were compared with the actual withheld verification interval data to gauge the accuracy of the method. The “reduction of error” (RE) statistic is what comes out of this exercise. The RE score compares the mean squared error (MSE) of the reconstruction (using the verification interval data) with the MSE of the calibration interval data which is constant in time. Therefore, a perfect RE score would be 1, assuming that the reconstruction method can perfectly predict the sample mean of the withheld data. -∞ is the worst
score, indicating there is no predictive power within the reconstruction method; however, any negative RE scores are generally considered not predictive. The RE score is the preferred measure of skill because it best explains the changes in variance and mean for the reconstructed period relative to the calibration period (Rutherford et al., 2005). Other verification scores are created, including the “coefficient of efficiency” (CE) score and the squared Pearson correlation coefficient ($r^2$). However, their analysis is second to the RE score as chosen by the previous studies (Mann et al., 2008) (Rutherford et al., 2005).

Figure 4 shows the RE scores for both the low and high frequency frozen network analyses. The high frequency scores are negative for every nest until 1400AD for the early calibration/late verification and 1300AD for the late calibration/early verification. This explains how poorly the high frequency network trends predict the temperature further back in time. While the low frequency scores are not very impressive, they have more positive scores than the high frequency, which is important because low frequency trends are what we are most interested in.

After the full RegEM method has created a reconstruction, it executes the nest splicing based on the calculated RE scores. A nest’s reconstruction was only added to the total reconstruction if its RE score is greater than the RE score of the previous nest. Therefore, the data was only included if believed to improve the outcome of the reconstruction. Poor RE scores denote bad predictive power, so it is not beneficial to incorporate that data.
Figure 4. RE scores for low frequency (green) and high frequency (yellow) for the frozen network analyses. (a) Early calibration (1850-1949AD)/Late verification (1950-1995). (b) Late calibration (1896-1995AD)/Early verification (1850-1895).

Figure 5. Low frequency RE scores from the frozen network analysis. Top: the average of the two calibration/verification sets. Bottom: RE scores from the top plot that only help improve the reconstruction. Note that the same RE scores can be used for several nests if the following nests do not have a better score.
4.4 Results

The actual process of doing the emulation proved to be more difficult than originally expected. Many trials were done attempting to recreate the version of the reconstruction available in the MANN08 results; but we were never able to fully emulate it. The reason is still unknown, as we tried several different strategies to fix the problem. The main goal was to make sure that our RE scores matched the MANN08 scores. From the start, they were quite different, so the attempts we made to alleviate the issue were based around at least getting the scores to switch nests at the same time. Since not every century had RE scores which improved upon the previous century, not every nest was included. If we were able to get the RE score improvements to align, then at least we were sure that the same proxies were being used for each time step. But we could not complete this. It is clear in Figure 5 that certain segments of the reconstruction follow very similar patterns, while others are quite different. The nests from 600-1000AD use similar data and the two reconstructions almost exactly align. Table 1 shows the differences in RE scores from what MANN08 reported and what we found. Not only are the scores dissimilar which is an issue for an emulation, but the table also shows that the choice of when to switch nests (i.e. when the RE scores improve) do not line up.

![Figure 5. Reconstruction using TTLS (red) compared to the MANN08 results (purple) using the same criteria. Note that the interval from 600-1000AD match quite closely.](image-url)
<table>
<thead>
<tr>
<th>Nest</th>
<th>MANN08</th>
<th>This study, TLS</th>
<th>MANN08</th>
<th>This study, TLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1800-1855</td>
<td>0.31</td>
<td>0.6302</td>
<td>0.76</td>
<td>0.7064</td>
</tr>
<tr>
<td>1700-1799</td>
<td>0.59</td>
<td>0.6487</td>
<td>0.76</td>
<td>0.7064</td>
</tr>
<tr>
<td>1600-1699</td>
<td>0.65</td>
<td>0.4764</td>
<td>0.76</td>
<td>0.7064</td>
</tr>
<tr>
<td>1500-1599</td>
<td>0.67</td>
<td>0.7064</td>
<td>0.76</td>
<td>0.7064</td>
</tr>
<tr>
<td>1400-1499</td>
<td>0.52</td>
<td>0.4849</td>
<td>0.76</td>
<td>0.4849</td>
</tr>
<tr>
<td>1300-1399</td>
<td>0.76</td>
<td>0.4022</td>
<td>0.76</td>
<td>0.4032</td>
</tr>
<tr>
<td>1200-1299</td>
<td>0.45</td>
<td>-0.3521</td>
<td>0.71</td>
<td>0.4032</td>
</tr>
<tr>
<td>1100-1199</td>
<td>0.24</td>
<td>-1.1424</td>
<td>0.71</td>
<td>0.4032</td>
</tr>
<tr>
<td>1000-1099</td>
<td>0.12</td>
<td>0.4032</td>
<td>0.71</td>
<td>0.4032</td>
</tr>
<tr>
<td>900-999</td>
<td>0.71</td>
<td>-0.0272</td>
<td>0.71</td>
<td>0.1375</td>
</tr>
<tr>
<td>800-899</td>
<td>0.67</td>
<td>0.1375</td>
<td>0.68</td>
<td>0.1375</td>
</tr>
<tr>
<td>700-799</td>
<td>0.68</td>
<td>0.1375</td>
<td>0.68</td>
<td>0.1375</td>
</tr>
<tr>
<td>600-699</td>
<td>0.57</td>
<td>-3.4216</td>
<td>0.57</td>
<td>-0.8456</td>
</tr>
<tr>
<td>500-599</td>
<td>0.56</td>
<td>-5.0505</td>
<td>0.56</td>
<td>-0.8456</td>
</tr>
<tr>
<td>400-499</td>
<td>0.55</td>
<td>-5.0505</td>
<td>0.55</td>
<td>-0.8456</td>
</tr>
<tr>
<td>300-399</td>
<td></td>
<td>-8.6052</td>
<td></td>
<td>-0.8456</td>
</tr>
<tr>
<td>200-299</td>
<td></td>
<td>-8.6052</td>
<td></td>
<td>-0.8456</td>
</tr>
<tr>
<td>100-199</td>
<td></td>
<td>-8.6052</td>
<td></td>
<td>-0.8456</td>
</tr>
<tr>
<td>0-99</td>
<td></td>
<td>-0.8456</td>
<td></td>
<td>-0.8456</td>
</tr>
</tbody>
</table>

**Table 1.** Low Frequency RE scores for both MANN08 and this study. The right two columns show only the scores that increase, and therefore, contribute to the improvement of the reconstruction. Note: MANN08 excluded RE scores if they were negative, hence, the missing data past 400AD.

There were several attempts that we made to fix the problem, but all were unsuccessful. The first effort was changing the lowpass filter frequency when creating the RE scores. Originally, we had the frequency cutoff at 20 years. However, MANN08 used a 10-year filter instead. Next, we found that MANN08 did not rescale the short term reconstructions that are made during the validation exercises, whereas, we did add this process. Also within these exercises, we
changed our code to not only perform a principal component analysis on the proxy data, but also on the instrumental data as described by MANN08. Finally, we replaced our method for creating truncation parameters with the parameters that are available from MANN08. And while this change made the most significant difference, it still was not enough to make the emulation exact. Along with other known NH temperature reconstructions, Figure 6 shows our reconstruction (red) at a 30-year lowpass filter with a 95% jackknife confidence interval. It is easy to see that our study tends to have higher temperatures in the later centuries compared to the other results, with a quite large uncertainties. However, it does catch similar variations as the others. Note that the MANN08 composite TTLS result, as well as, the MANN08 “flavor” that we were trying to emulate, are both shown.

![Graph of NH temperature reconstructions](image)

**Figure 6.** TTLS reconstruction (red) compared to other NH temperature reconstructions at a 30-year lowpass filter.
5. Sensitivity Analysis

5.1 Motivation

While the current reconstruction methods of CPS and TTLS produce decent results, there are characteristics of each that can be improved upon. CPS is a simple method that does not produce good low frequency reconstructions, which is the focus of our study. At each nest, CPS standardizes the proxies which smoothes too much of the information, thus eliminating much of the trends. RegEM TTLS is an improvement upon CPS because it preserves much of the spatial patterns, as well as, it does not assume any local connection between the proxy and the instrumental record. It simultaneously calibrates all of the data for the proxy and temperature series during that nest, which provides better information for a hemispheric-wide reconstruction (Mann 2002). But, as described earlier, TTLS requires a designated truncation parameter in order to carry out the reconstruction. This makes the process subjective, allowing for unnecessary differences in the result depending on the specifications given. The new regularization scheme, iTTLS, improves upon this issue by introducing an automated way to choose the truncation parameter, thus taking away the problematic subjectivity. As new methods continue to be created that attempt to improve upon the previous results, they deserve to be examined for successfulness. Here, our goal was just that: to use the MANN08 results, that apply the previous reconstruction methods, as a baseline for comparison with the iTTLS scheme of RegEM, in hopes of advancing our knowledge of creating paleoclimate reconstructions.

5.2 Sensitivity to Regularization Scheme

While the iTTLS method has been shown to produce good results in previous studies, we found that the scheme is currently unable to work with large, noisy datasets, like the one used in this study. When the number of variables (i.e. the number of proxies) becomes larger than the
number of instrumental observations, the scheme has difficulty maneuvering through the matrix. Through much trial and error, the procedure continually could not handle the matrix; in particular, it is the Cholesky decomposition that fails. For this project, the Cholesky decomposition was used to break down the sample covariance matrix of the available data in such a way that we could use the resulting matrix as the data matrix that is then put through the rest of the reconstruction procedure. However, this decomposition requires a positive-definite matrix; otherwise, it will have problems creating a stable matrix. In our case, the proxy matrix is ill-conditioned, with many one or more eigenvalues numerically close to zero, hence hindering the decomposition. When this occurs, it is possible to use incomplete Cholesky factorization, which is not as stringent about the condition of the matrix. But even with this slightly different form, the sample covariance matrix was so badly conditioned that it compromised algorithm stability. This result created numeric instabilities, in particular, negative square roots that produce complex numbers, and thus, were not able to make the positive data matrix needed throughout the rest of the process. Thus, an important finding of this study is that the iTTLS algorithm cannot yet be applied to the MANN08 database, and will need to be reformulated for this purpose.

6. Influence of Proxy Class on NH Temperature Reconstructions

Since the MANN08 study compiled a large dataset with many different proxy classes, it is worthwhile to note the effects that these classes have on the reconstruction of past temperature. The total database had 1209 proxies, but with the screening process used by MANN08, it was narrowed down to 421 proxies. Of these, there were tree-rings (305), ice cores (10), corals (6), speleothems (8), historical documents (10), sediment cores (10), and a special composite record (71) using proxy, historical, and instrumental data (Luterbacher et al 2004). There was one sepa-
rate composite record; however, it was not used in this analysis. The 305 tree-ring proxies were
grouped together from the maximum latewood density composite series (MXD), the Internation-
al Tree Ring Data Bank, and a few regional composite series.

From this breakdown alone, it is interesting to see whether the screening process prefer-
entially removed certain classes. While there were the most tree rings, it was also the most repre-
sented class globally (1032); so, 305 proxies is only 20% of the total tree ring data. Does this in-
dicate whether tree ring data is reliable? A lot of it was screened out, yet there are still far more
tree rings than any other class. In addition to this tree ring question, another useful outcome of
reconstructing by class is being able to see which proxies contribute to the reconstruction further
back in time. There might be a class that has a large number of short term proxies; whereas, a
different class may only have a few proxies, but each one extends the entire Common Era. The
latter class would be very helpful to the reconstruction, as it provides a lot of information for the
part of the reconstruction that is most difficult to produce accurately. This places a big amount of
reliability on this particular class, so we would want to make sure that its predictive power is rel-
atively good. The outcome of this finding would give information for future studies where this
proxy class could be incorporated.

For each of these reconstructions, the dataset was quite small compared to the total and
NH screened networks we had previously used with TTLS. When there are a very small number
of proxies, TTLS can have difficulty working with the matrix. As the nests go further back in
time, proxies have more missing values, until eventually the dataset could be empty. In this pro-
ject, the majority of the classes had this occur and was not able to produce full-length reconstruc-
tions. But when the available data matrix thinned out to below three proxies available, TTLS
would already begin to encounter problems. To alleviate this issue, we stopped TTLS as soon as
the number of proxies went below three. With this specification, no errors occurred and we were able to get reconstructions for every class.

From Figure 7, we see that only the speleothem and sediment core classes have proxies that extend the entire two millennia. Combined, those classes started with 18 proxies, which is already a small dataset. However, there were at least three records in each class that retained data into the last nest, which is an accomplishment that even the huge tree ring class could not achieve. The speleothems have the largest fluctuations in their temperature, while the sediment cores have the smallest variation amongst all of the classes. These could be factors to consider when analyzing the results of the composite reconstruction. For instance, looking at the composite, there appears to be a relationship between the positive temperature anomaly around 1500AD and the spike in the document class data. Even though most proxy classes still had data at that
time, the documentary data seems to dominate the resulting composite. This brings back the idea of certain classes being weighted more than others and whether those are reliable weights. Many people would not consider historical documents as strong of a proxy class compared to the “natural” proxies that are not created by humans. But the analysis of a matrix of regression coefficients would provide this answer, and gauge the helpfulness of each class.

7. Discussion

Even though the main goal of the project, which was to have a comparison between TTLS and iTTLS, was not able to be completed, there are still some valuable results that came out of this thesis. For one, accomplishing an exact emulation of the MANN08 data was not an easy task. While we performed many different changes on the code, our efforts could not produce an exact replica. This brings back the question of reproducibility which, as mentioned previously, is critical to science. There are still many other factors we could explore and checks we could perform before coming to a conclusion, but our initial attempts at an emulation were only partially successful. Another important result is in fact that we could not use iTTLS to complete the reconstruction. This shows two things: there is room for improvement in iTTLS (and other reconstruction methods) and current proxy databases have limitations due to their high-dimensionality and amount of non-climatic noise. The former is obviously not easily fixed and the errors encountered could be unavoidable. But the focus should lie on both. We need better methods and more selective databases made of high-quality proxies. Although the MANN08 database was considered large, the screening process removed two-thirds of the proxies as having insignificant predictive power. A crucial goal should be to not only create large databases, but also ones that retain better interpretable data. With improvements in one or both of these results, paleoclimate reconstruction will better estimate past temperature variability. In addition to these
findings, investigating more about the influence of proxy classes on datasets is a valuable avenue for research. Much of the current research focuses on the actual climate change. However, if we learn more about which proxies are most influential (for better or worse) then we will have a better grasp on the temperatures that the proxies predict.

8. Conclusion

Climate change is such an important topic right now because its science is a relatively new field where there is still a lot of understanding to be had. Multiproxy paleoclimate reconstructions are the best interpretable estimates of past temperature variability, yet, the methods still have plenty of caveats. In this project, for one, the emulation using TTLS was not exact. Though several trials were completed, we were not able to solve the differences, with the main issue being the inability to get the nests to align and use the same proxy sets at the same time throughout the reconstruction. Second, iTTLS would not function properly. This came from properties of the proxy matrix and sensitivities of iTTLS. The matrix was large and noisy which caused iTTLS to have trouble manipulating it, thus causing instabilities that proved fatal to creating a full reconstruction for analysis. Third, the majority of the proxies in our dataset were thrown out as insignificant. This poses the issue of choosing the best screening process and how to collect proxies that can pass significance tests. Finally, proxy classes individually played big roles in the composite reconstruction, and factors like divergence may cause the results to be untrustworthy and not representative of the true variability.

However, even with all of these concerns, this thesis had worthwhile information come out of it that can lead to future research. The inexact emulation leaves room for continued efforts to complete it. This is important due to the idea of reproducibility in science and how it is the crutch of progress. Learning how iTTLS handles large, noisy datasets may lead to improvements
of the method. If this occurs, it will broaden the abilities of reconstruction methods to include not-so-ideal datasets and expand the wealth of knowledge about paleoclimate temperature variability. Ideally, however, datasets input into iTTLS would be less noisy. How to achieve this may require changes in the screening of proxies in significance testing. Or, there needs to be more proxy records that have predictive capabilities. Finally, the degree of importance of each proxy class in a reconstruction is very necessary to understand because good or bad characteristics of a proxy class will provide information on its worth to the composite. Determining how to fix the problems or where to reward the positives will only better the results.

If such a prominent study like MANN08 is subject to all of these concerns and there is already so much to learn about what else can be improved upon, it shows the necessity of scientific progress in the field of paleoclimate reconstructions. With these improvements, hopefully the climate change debate will have enough evidence to be put to rest.

Acknowledgements

Thank you to my advisor for all of his dedication in helping me through this project. Thank you to my friends and family who supported me throughout the four years I worked towards this degree. Thank you to the USC Provost’s Office for funding my research in its entirety. Thank you to the USC Geological Sciences Department for all of their help and support throughout my four years in the program.
References