NOTES ON METHODS IN Mann et al. (2009, Science)

Olivier Marchal, Clark 109A, omarchal@whoi.edu

These notes are an attempt to summarize the major methodological steps in the paper of Mann et al. (2009, Science; M09 hereafter), which we read in class (September 20). I do not claim that the description below is accurate; it merely reflects my understanding of the material. I would welcome any comment, particularly on misunderstanding or significant omission on my side. References that provide details about the methods used by M09 are listed at the end of these notes. The paper of Smerdon and Pollack (2016) is an excellent and very accessible introduction to the problem of climate reconstruction over the past 2000 yr. For the mathematically-inclined reader, Schneider (2001) offers a detailed and clear description of the EM and RegEM methods (the latter used in M09) in a climate context.

Two summaries are provided in these notes: a short one and a longer one. Students interested in more details should read the long summary.

1 Short Summary

The major methodological steps in M09 appear to be the following:

Step 1) Data:

a. Instrumental - Annual average temperatures from HadCRUT3, a gridded land surface air temperature and sea surface temperature product. They used one of various regression methods to fill in spatial grid cells for times with no mean annual data.

b. Proxy - Tree rings, marine sediments, lake sediments, speleothems, ice cores, marine corals, and historical documents listed in their supplementary online material (SOM).

Step 2) Screening: For a comparison to results using the full proxy data set, they constructed a subset of the data in which they only retained the records that are linearly correlated to local temperature. Actual p value cutoff used is 0.11-0.12.

Step 3) Standardize the data: Subtract the mean and then divide by the standard deviation.

Step 4) Calibrate proxy data to temperature and Validate. For the calibration, they split the data records in low- and high-frequency components, used a regression method to calibrate proxy data to temperature separately for the two components, and combined the two resulting reconstructions. Note that the predictand in the method is not temperature but its principal components. For the validation, they compared (i) the temperature derived from proxy data using a calibration determined from data for the late half of AD 1850–1995 with (ii) instrumental data for the early half of AD 1850–1995.

Step 6) Re-calibrate and Reconstruct temperature between AD 500–1850: They re-calibrated the proxy data to temperature using the data for the entire instrumental interval AD 1850–1995. From this second calibration, they reconstructed temperature from proxy data between AD 500–1850.

2 Long Summary

It is useful to first read the following excerpt from Tingley et al. (2012), which I found illuminating:

"One common interpretation of the paleoclimate reconstruction problem is to regard instrumental observations of any climate variables before 1850 (and some after this point) as 'missing'. In this view point, popularized by the RegEM algorithm (Schneider, 2001), the paleoclimate reconstruction endeavor reduces to an imputation or missing value problem: the goal is to infer the missing values in available instrumental time series prior to 1850. Labeling the unknown instrumental variables prior to 1850 as missing values is a nomenclature used in the literature (e.g., Schneider, 2001); in practice, the reconstruction proceeds by calibrating the proxy data against available instrumental data (post 1850) and then using that calibration to predict instrumental records prior to 1850."

The expectation maximization (EM) algorithm which is referred to above is an iterative method both for the estimation of mean values and covariance matrices from incomplete datasets and for the imputation of missing values. In contrast to the conventional EM algorithm, the regularized EM algorithm is applicable to datasets in which the number of variables typically exceeds the sample size (Schneider 2001).

Equipped with these ideas, let us move on and consider in more details the apparent steps in M09's methodology.

Step 1: Get the instrumental and proxy data

For the instrumental data, M09 quotes the paper of Brohan et al. (2006, JGR). These instrumental data are referred to as the HadCRUT3 dataset and includes two types of data: data of land surface air temperature (LSAT) and data of sea surface temperature (SST). Both LSAT and SST data are "gridded", i.e., these data have already been interpolated on a regular grid. The instrumental data used by M09 appear to be annual averages. Moreover, M09 stated that "Gaps in the individual annual mean (Jan-Dec) gridbox surface temperature data available from 1850-2006 were infilled". To this end, they used the "Regularized Expectation-Maximimation" method (RegEM) with ridge regression, quoting Schneider

(2001) and Rutherford et al. (2003). Here is a short description of of EM and RegEM provided to us by Andy Solow (MPC, WHOI): "The EM algorithm is a standard method of performing maximum likelihood estimation when there are missing data. The regularized EM algorithm (RegEM) is a version of the EM algorithm that exploits information in the observations about the missing values. There are various ways to do this." In ridge regression, regularization is accomplished through the use of a ridge parameter h. Whereas M09 do not describe how h is calculated, it is generally determined in such a way as to minimize, approximately, the expected mean squared error of the imputed values (Golup et al. 1979; Schneider 2001).

For the proxy data, M09 seem to rely on data used in Mann et al. (2008; M08). These data are listed in the supplementary online material (SOM) of M09. As shown in class by Katie, they come from tree rings, marine sediments, lake sediments, speleothems, ice cores, marine corals, and historical documents. M09 stated that they did not include the instrumental data used by M08 "so that gridbox level assessments of skill would be entirely independent of information from the instrumental record." I take this to mean that M09 did not want instrumental data to bias the results from the validation process (more on this below).

Step 2: Screening

M09 clarify in their SOM that the full proxy dataset is "emphasized", but that results from a screened proxy dataset are also reported for comparison. Screening means, in effect, the removal of a fraction of those proxy records that do not seem to be significantly related to local temperature variations. It seems that M09 retained proxy records for which the (linear?) correlation with the 'co-located' instrumental record leads to a p-value less than 0.11 – 0.12 (see SOM of M09).

Step 3: Standardize the data

For their climate field reconstruction (CFR), M09 state that they used the "hybrid frequency-domain RegEM CFR procedure" of M07. For conciseness, this procedure is referred here to as the "M07 method". Based on M07, it seems that M09 standardized both the instrumental data and the proxy data. Standardization signifies subtracting the mean of the time series from each value of the time series, and then divide the difference by the standard deviation of the time series, as clarified in class by Jimmy.

Step 4: Calibrate proxy data to temperature and Validate

Consider first the calibration. M09 reported that they applied the M07 method to calibrate the proxy data in terms of temperature. The calibration interval is a period of overlap between instrumental and proxy data (more on this below). The M07 method seems to operate in three parts (cf. M07, but also Rutherford et al., 2005, and Mann et al., 2005, both quoted in M07):

First, the instrumental and proxy datasets are each split in two subsets through application of a low-pass filter. The low-pass filter is intended to isolate the low-frequency component from the high-frequency component in each data record, where 'low-frequency' means here any period greater than 20 yr. I could not find a description of the specific filter used in M09 (perhaps in MATLAB code?). In any event, this part of the M07 method would explain why the method is labelled as 'hybrid frequency-domain'.

Second, RegEM is applied to calibrate the proxy data (which are now apparently both standardized and filtered) in terms of temperature, separately for the two frequency components. According to Rutherford et al. (2005), "... the results of the two independent reconstructions are then combined to yield a complete reconstruction". I could not find a description, neither in M07 nor in M09, about how the combination is actually done. Note that, in the present application of RegEM, M09 did not seem to have used ridge regression, but another form of regularization which is called 'truncated total least-squares' (TTLS). M07 wrote "There are a number of possible ways to regularize the EM algorithm, including principal component (PC) regression, truncated total least squares regression (TTLS) ..., and ridge regression .... Both ridge regression and TTLS account for observational error in available data (i.e., represent errors-in-variables approaches) ... In TTLS ... regularization is accomplished through a choice of the truncation parameter K". This description would explain the origin of the parameter K in M09, which is set equal to 3 back to AD 1600 and to 2 prior to AD 1600 (see their SOM).

The third part of the M07 method as described in M09 relates to the principal component analysis (PCA). That M09 relied on PCA is suggested in their SOM: “The number of surface temperature modes, i.e., Empirical Orthogonal Function (EOF)/Principal Component (PC) pairs, retained in the analysis (M) ...”. The motive for using PCA in M09 seems to reside in the following statement in M07: “To further insure regularization of the procedure, the predictand [personal note: in the current study, the predictands represent annual mean temperatures, for which instrumental data are available over the calibration interval but missing prior that interval] is represented in the data matrix $\mathbf{X}$ by its leading M PC time series, where M is small compared to the total number of nonzero eigenvalues of the ... covariance matrix” (for $\mathbf{X}$). Later M07 wrote “This step is performed only once, at initiation of the RegEM procedure ... The predictand in the end is then reconstructed through the appropriate eigenvector expansion, using the M reconstructed PC series.” Thus, it seems that M09 applied PCA to the (now apparently standardized and filtered) instrumental data, and used the PCs as predictands in RegEM. According to their SOM, M07 set M equal to 6 for back to AD 1600 and to values in the range [2–5] prior to AD 1650, “depending on how far back”.

Consider now the validation of the temperature reconstruction approach. Instrumental data for the late half of the interval AD 1850–1995 are used to calibrate the proxy data in terms of temperature, and instrumental data for the early half of this interval are used to test the temperature values predicted from proxy data using the calibration. Thus, the late half of the interval 1850–1995 is a ‘calibration period’ and the early half of this interval is a ‘validation period’. Three statistics are used to measure the skill of the proxy data to reconstruct temperature during the validation period: the reduction of error (RE), the coefficient of efficiency (CE), and the squared correlation (r^2). A formal definition and a clear illustration of the relevance of these three measures can be found on pp. 92–95 of the National Research Council Report (2006).

Step 6: Re-calibrate and Reconstruct temperature between AD 500–1850

M09 produced a second calibration of proxy data to temperature using the operations detailed in step (5) but now with the data available for the entire instrumental interval AD 1850–1995. The proxy data for the pre-instrumental interval AD 500–1850 are then used to reconstruct temperature during this interval, using the calibration established from data over the instrumental interval AD 1850-1995. Specifically, the PCs of temperature from AD 500-1850 are reconstructed, and these PCs, together with their corresponding spatial patterns derived from instrumental data between AD 1850–1950, are used to calculate the spatial distribution of temperature at different times between AD 500–1850. From these distributions, the times series in Fig. 1 and the spatial patterns in Figs. 2–3 of the paper of M09 have been produced.

REFERENCES

Brohan et al., Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850, Journal of Geophysical Research, vol. 111, doi:10.1029/2005JD006548, 2006

Golup et al., Generalized cross-validation as a method for choosing a good ridge parameter, Technometrics, 21, 215-223, 1979

Mann et al., Testing the fidelity of methods used in proxy-based reconstructions of past climate, Journal of Climate, 18, 4097-4107, 2005

Mann M. et al., Robustness of proxy-based climate field reconstruction methods, Journal of Geophysical Research, vol. 112, doi:10.1029/2006JD008272, 2007

Mann et al., Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia, Proceedings of the National Academy of Sciences, 105, 13252-13257, 2008

Mann et al., Global signatures and dynamical origins of the Little Ice Age and Medieval Climate Anomaly, Science, 326, 1256-1260, 2009

National Research Council Report, Surface temperature reconstructions for the last 2,000 years, 160 pages, 2016
(available at http://www.nap.edu/catalog/11676.html)

Rutherford et al., Proxy-based Northern Hemisphere surface temperature reconstructions: Sensitivity to method, predictor network, target season, and target domain, Journal of Climate, 18, 2308-2329, 2005

Schneider T., Analysis of incomplete climate data: Estimation of mean values and covariance matrices and imputation of missing values, Journal of Climate, 14, 853-871, 2001

Smerdon J. and Pollack H., Reconstructing Earth's surface temperature over the past 2000 years: The science behind the headlines, WIREs Climate Change, doi: 10.1002/wcc.418, 2016
(available at http://www.ldeo.columbia.edu/ jsmerdon/)

Tingley M., Piecing together the past: Statistical insights into paleoclimatic reconstructions, Quaternary Science Reviews, 35, 1-22, 2012