A Test for Nonlinearity in Temperature Proxy Records

Bård Støve

University of Bergen
Department of Mathematics

Joint work with Fredrik Charpentier Ljungqvist (Stockholm University) and Peter Thejll (Danish Climate Centre)

Forthcoming Journal of Climate

The 5th Trondheim Symposium in Statistics, Selbu 2012

Outline

1. Introduction
   - Temperature reconstruction
   - Proxy records
   - Statistical challenges
2. Nonparametric reconstructions
   - Method
   - Data
   - Results
3. Concluding remarks

Reconstruction (prediction) of historic temperature

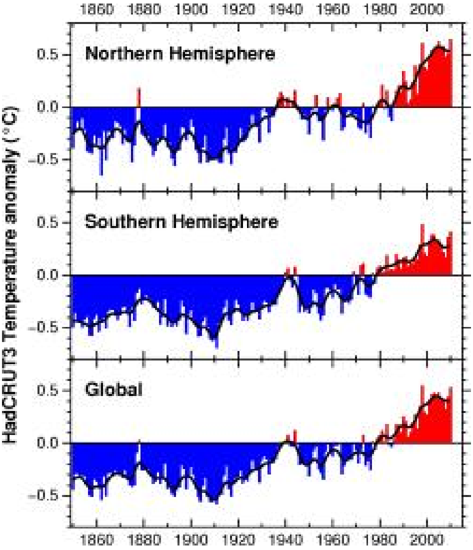
Paleoclimatology: the study of climate and climate change.
Need to know the past climate because:

- Need data descriptions, i.e. statistics for other uses
- Trends
- Likelihood of extreme events
- Temperature variability (e.g. does the recent warming falls outside the range of natural variability)

Need to understand climate physics better:

- Causes of past changes
- Orbital changes
- Solar influence
- Volcanoes
- Internal dynamics
- Geography changes (really long time scales!)

Measured temperature 1850-today



How do you find past climate?

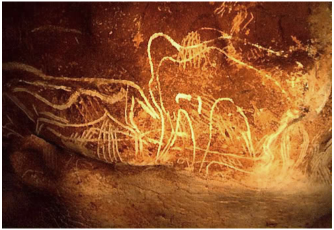
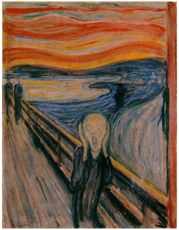
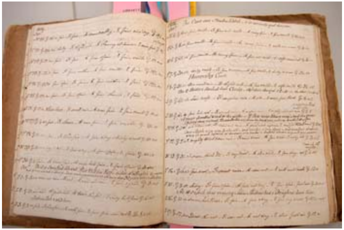
Reliable temperature records exist only for the last 150 years (or so).
Reconstruct from “proxies“:

- Measurements from tree rings, ice sheets and other natural phenomena

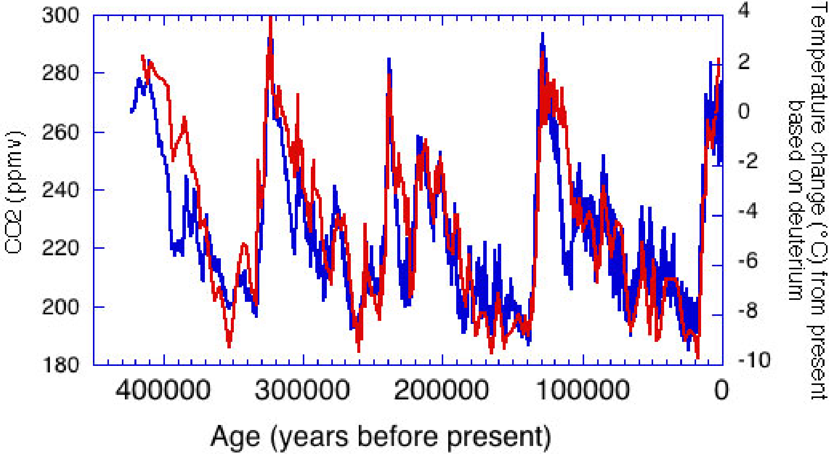
Basic problem:

1. Scientists extract, scale and calibrate this proxy data
2. A training set consisting of the part of the proxy data which overlaps the instrumental temperature period (e.g. 150 years of data) is constructed and used to build a (statistical) model.
3. The model, which maps the proxy record to surface temperature, is used to backcast or reconstruct historical temperatures.

Some proxy records!


A long proxy record from Antartica



Example of reconstruction: Multiple linear regression

Landsberg and Groveman (1978):

$$T_i = \alpha + \beta_1 \cdot P1_i + \beta_2 \cdot P2_i + \beta_3 \cdot P3_i + \ldots \quad (1)$$

at time i temperature (T) can be written as sum of weighted proxies ($P1, P2, P3, \ldots$).
Find coefficients ($\alpha, \beta_1, \beta_2, \beta_3, \ldots$) during 'training period' using OLS and 'significant regressors test'.
Use the model to reconstruct ('backcast') historic temperature based on observations of the proxies.

Statistical challenges

- Regression on non-stationary time series (spurious regression, i.e. nonsense regression)
- Auto-correlated (proxy) time series
- Co-Integrated time series
- Weak relationship between proxies and temperature (is it linear!?)
- Many proxies as many as (or even more than) data observations (proxy selection!)
- Measurement errors on proxies
- Missing observations of proxies
- How to estimate uncertainties of a temperature reconstruction

These challenges should be considered when building a reconstruction model.

Other reconstruction methods

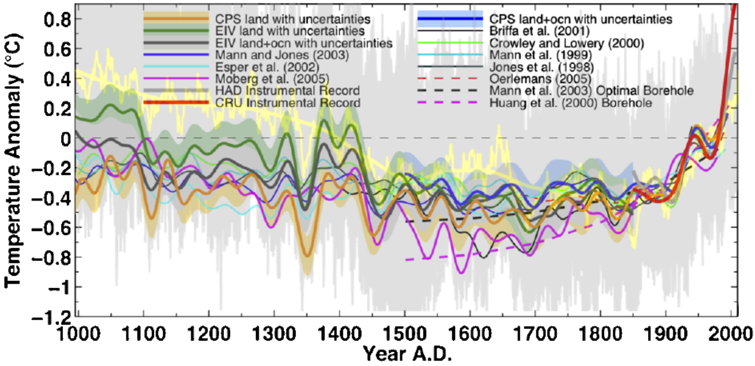
- Principal component regression: purpose is dimension reduction. Transform proxies to PCs, and fit temperature to the first k PCs.
- RegEM (Regularized Expectation-Maximization): allow for missing data and protect against overfitting by regularization (e.g. by Ridge regression), see Mann et al (2007)
- Bayesian hiearachical models (Li et al (2010))

Other methods exist, see e.g. Christiansen et al (2009) for a review.
Note there is a difference between global and field methods:

- global: the hemispheric-mean temperature is reconstructed directly
- field: the global temperature field is reconstructed followed by a calculation of the hemispheric-mean temperature

Temperature reconstructions

Mann et al (1998, 99): NH temperature warmer during the late 20th century than at any other time during the past millennium → hockey stick controversy. Note the yellow reconstruction with estimated uncertainty in grey from McShane and Wyner (2011).



Our research question

Most current temperature reconstruction methods used are linear, i.e. a linear relationship between the proxy and the temperature is assumed.
But;

- Do we know apriori whether the relationship is linear?
- Evidence that high-latitude tree-ring records show a lessened or negative response to higher temperature in the twentieth century (Andreu-Hayles et al 2011, Loehle 2009)
- Other proxies (e.g. pollen) seem also to have a nonlinear response to temperature (Birks et al 2010)

We use nonparametric regression methods and related tests that can reveal if there is a nonlinear relationship between proxies and temperature. Such methods only occasionally utilized in paleoclimatic research.

Nonparametric regression models I

The aim of nonparametric models is to relax assumptions on the form of a regression function, and to let the data search for a suitable function that describes the available data well.

These approaches are powerful in exploring fine structural relationships and provide very useful diagnostic tools for parametric models.

Additive nonparametric regression model:

$$Y = \alpha + \sum_{j=1}^{d} m_j(X_j) + e, \qquad (2)$$

where $m_1, \ldots, m_d$ are unknown uni-variate functions, $\mathrm{E}(e) = 0$, $\mathrm{Var}(e) = \sigma^2$ and e is independent of the vector of co-variates $\mathbf{X}$.

Nonparametric regression models II

To ensure identifiability, $m_1, ..., m_d$ are required to satisfy

$$\mathrm{E}[m_j(X_j)] = 0, \quad j = 1, ..., d, \tag{3}$$

which implies $\mathrm{E}(Y) = \alpha$.
If the additive model, (5), is correct then

$$\mathrm{E}[Y - \alpha - \sum_{j \neq k} m_j(X_j) | X_k] = m_k(X_k), \quad k = 1, ..., d. \tag{4}$$

This relationship suggests an iterative procedure for the estimation of the unknown functions. Thus for a known constant α and given functions m_j, $j \neq k$, the function m_k can be estimated by a uni-variate regression fit based on the observations (X_k^i, Y_i), $i = 1, ..., n$, where X_k^i is the ith observation of the kth additive variable. Denote the univariate smoother of m_k by S_k.
Estimation of the unknown functions $m_1, ..., m_d$ is done by the backfitting algorithm, introduced by Breiman and Friedman (1985) and Buja et al (1989).

Estimating the unknown functions m_j (I)

Observations (X_k^i, Y_i), $i = 1, \dots, n$. X_k^i is the ith observation of the kth additive variable. The algorithm works as follows:

Step 1. Initialization: $\hat{\alpha} = n^{-1} \sum_{i=1}^n Y_i$, $\hat{m}_k = m_k^0$ for $k = 1, \dots, d$.
Step 2. Find new transformations: For $k = 1, \dots, d$:

$$\hat{m}_k = S_k[Y - \hat{\alpha} - \sum_{j \neq k} \hat{m}_j(X_j) | X_k];$$

$$\text{centre the estimator to obtain } \hat{m}_k^* = \hat{m}_k - n^{-1} \sum_{i=1}^n \hat{m}_k(X_k^i),$$

$$\text{and } \hat{\alpha}^* = \hat{\alpha} + n^{-1} \sum_{i=1}^n \hat{m}_k(X_k^i).$$

Step 3. Repeat step 2 until convergence.
The idea behind this algorithm is to carry out a fit, calculate partial residuals from that fit, and refit again ⇒ backfitting

Estimating the unknown functions m_j (II)

The starting functions $m_1^0, ..., m_d^0$ can be obtained in various ways,e.g. from a linear regression fit of Y on the co-variates X_k. The smoothing operator S_k can be other nonparametric regression estimators such as kernel methods.

Overfitting can be a problem for such models. May consider a semiparametric model instead (some linear terms and some nonparametric terms)

$$Y = \sum_{j=1}^{p} m_j(X_j) + \sum_{j=p+1}^{d} \beta_j X_j + e, \qquad (5)$$

Generalized additive model

The additive model (2) is a version of a wider model, called generalized additive model (GAM), introduced by Hastie and Tibshirani (1990). Here the conditional mean $(m(\mathbf{X}))$ of a response Y is related to an additive function of the predictors via a link function g:

$$g[m(\mathbf{X})] = \alpha + m_1(X_1) + \ldots + m_p(X_d). \tag{6}$$

An alternative to the backfitting algorithm is the marginal integration method, see Tjøstheim and Auestad (1994)

Proxy records I

Use 30 proxies with annual resolution

- one speleothem micro-layer thickness record,
- three ice-core $\delta^{18}O$ records (one of which a stack of several others)
- seven varved thickness sediment records
- twelve tree-ring width records
- five tree-ring maximum latewood density records
- one tree-ring height-increment record
- one tree-ring $\delta^{13}C$ record

We test the set of 30 proxies for relevance by screening them in terms of how well they correlate to both the global mean temperature and the local temperatures (not unproblematic!) $\Rightarrow$ ends up with 15 proxies in our models

Proxy records II

R1, R2 correlation to NH mean T (with and without linear trends) for 1850-1969. R3 and R4 the same but correlated against the local gridpoint T, R5 from the original author. Δ is dating uncertainty

Proxy record	Season	Lon (°)	Lat (°)	Extent	R1	R2	R3	R4	R5	Δ	Proxy type	Reference
Agassiz Ice Cap	Annual	−73.10	80.70	800–1972	**0.37**	**0.25**	0.13	−0.01	—	±0	Ice-core δ^{18}O	Vinther et al. (2008)
Avam–Taimyr	July	93.00	70.00	800–2003	**0.49**	**0.30**	**0.29**	**0.25**	0.39	±0	Tree-ring width	Briffa et al. (2008)
Big Round Lake	July–September	−68.50	69.83	971–2003	**0.41**	**0.21**	0.03	−0.06	0.46	±1–20	Varved lake sediment	Thomas and Briner (2009)
Blue Lake	June–August	−150.46	68.08	800–2005	−0.02	−0.15	−0.02	0.01	0.56	±12	Varved lake sediment	Bird et al. (2009)
Boreal/Upper Wright	June–August	−118.46	36.54	800–1992	**0.36**	0.08	0.06	0.06	—	±0	Tree-ring width	Lloyd and Graumlich (1997)
Central Europe	June–August	8.00	46.30	800–2003	0.12	0.08	**0.53**	**0.54**	0.72	±0	Tree-ring width	Büntgen et al. (2011)
Columbia Icefield	May–August	−117.15	52.15	950–1998	**0.18**	0.03	**0.40**	**0.38**	0.73	±0	Tree-ring density	Luckman and Wilson (2005)
Donard Lake	June–August	−61.35	66.66	800–1992	−0.09	−0.01	−0.11	−0.08	0.57	±1–20	Varved lake sediment	Moore et al. (2001)
Eastern Carpathians	July–August	25.10	47.10	994–2005	0.05	0.11	**0.34**	**0.33**	0.42	±0	Tree-ring width	Popa and Kern (2009)
Finnish Lapland	June–August	25.00	69.00	800–2005	**0.35**	**0.25**	**0.54**	**0.50**	0.64	±0	Tree-ring width	Helama et al. (2010)
French Alps	June–August	7.00	45.50	800–2008	**0.31**	**0.18**	**0.59**	**0.59**	0.39	±0	Tree-ring width	Corona et al. (2011)
Greenland composite	Annual	−40.00	70.00	800–1973	**0.23**	0.13	**0.31**	**0.21**	0.56	±0	Stacked ice-core δ^{18}O	Vinther et al. (2010)
Gulf of Alaska	January–September	−145.00	60.00	800–1999	**0.22**	0.00	**0.36**	**0.32**	0.48	±0	Tree-ring width	D'Arrigo et al. (2006)
Iceberg Lake	May–June	−142.95	60.78	800–1998	0.17	0.11	−0.13	−0.16	0.23	±32	Varved lake sediment	Loso (2009)
Indigirka	June–August	148.15	70.53	800–1993	**0.32**	**0.23**	**0.31**	**0.28**	—	±0	Tree-ring width	Moberg et al. (2006)
Jämtland	June–August	13.30	63.10	800–2002	**0.37**	0.16	**0.48**	**0.50**	0.63	±0	Tree-ring width	Linderholm and Gunnarson (2005)
Karakorum Mountains	June and July	74.99	36.37	828–1998	**0.29**	0.08	0.12	**0.29**	0.48	±0	$\delta^{13}C$ tree-ring	Treydte et al. (2009)
Laanila	June–August	27.30	68.50	800–2007	0.09	0.17	**0.53**	**0.56**	0.56	±0	Tree-ring height-increment	Lindholm et al. (2011)
Lake C2	June–August	−77.54	82.47	800–1987	0.14	0.04	0.09	0.06	—	±57	Varved lake sediment	Lamoureux and Bradley (1996)
Lower Murray Lake	July	−69.32	81.21	800–1969	**0.27**	**0.23**	0.11	0.10	0.78	±16	Varved lake sediment	Cook et al. (2009)
Polar Urals	May–September	65.75	66.83	800–1990	**0.52**	**0.38**	**0.40**	**0.30**	—	±0	Tree-ring density	Esper et al. (2002)
Renland	Annual	−26.70	71.30	800–1986	0.09	−0.04	**0.39**	**0.27**	—	±2–20	Ice-core δ^{18}O	Vinther et al. (2008)
ShiHua Cave	May–August	116.23	39.54	800–1985	**0.43**	0.12	0.06	0.05	0.65	±5	Speleothem layer thickness	Tan et al. (2003)
Sol Dav	April–October	98.93	48.30	800–1999	**0.39**	−0.28	−0.06	0.02	0.58	±0	Tree-ring width	D'Arrigo et al. (2001)
Southern Sierra Nevada	June–August	−118.90	36.90	800–1988	**0.20**	**0.20**	**0.21**	**0.21**	—	±0	Tree-ring width	Graumlich (1993)
Southern Colorado Plateau	June–August	−111.40	35.20	800–1996	**0.43**	**0.27**	**0.23**	0.19	0.68	±0	Tree-ring width	Salzer and Kipfmueller (2005)
Teletskoe Lake	Annual	87.61	51.76	800–2002	**0.44**	0.06	**0.31**	0.17	—	±1	Varved lake sediment	Kalugin et al. (2009)
The Alps	June–September	8.00	46.30	800–2004	**0.24**	0.14	**0.81**	**0.80**	0.69	±0	Tree-ring density	Büntgen et al. (2006)
Torneträsk	April–August	19.80	68.31	800–2004	**0.38**	**0.33**	**0.81**	**0.79**	0.79	±0	Tree-ring density	Grudd (2008)
Yamal	June–July	69.17	66.92	800–1996	**0.29**	0.03	**0.61**	**0.59**	0.56	±0	Tree-ring width	Briffa (2000)

Estimating models I

To calibrate our models:

- The response is yearly observations of the NH mean temperature data from the 5x5 gridded HadCRUT3v data set (Brohan et al 2006)
- 15 covariates: the selected proxies with yearly observations
- Observations from 1850 through 1969 (120 observations) - the calibration period.

Note that the NH temperature is centred over the calibration period. The proxies are centred and normalized with mean and standard deviation from the calibration period. No missing observations in the calibration period.
We fit a linear regression model, a nonparametric model and a semiparametric model (with two nonparametric proxies).
Use R and package gam.

Estimating models II

An approximate F test is used to evaluate the significance of the nonlinearity, to determine whether including the nonlinear component of each smooth term in the model resulted in a significantly better fit than a linear relationship. Although the test statistics do not have exact or even asymptotic F distributions, Hastie and Tibshirani (1990) report that simulations show them to be useful approximations.

The model comparison test: approximate F-test. Test statistic

$$F = \frac{(RSS_1 - RSS_2)/RSS_2}{(DF_1 - DF_2)/DF_2},$$

RSS_i is the residual sum of square for model i and DF_i is the approximate degrees of freedom for model i.

Calibration results

TABLE 2. Calibration results, with (first and second columns) the estimated coefficients and corresponding p value of a standard significance test for the linear model, (third column) the p value for the nonlinearity test of the proxies in the nonparametric model, and (fourth column) the p value for the nonlinearity test of the proxies in the semiparametric model.

Proxy	Linear model: Coefficients	Linear model: *p* value	Nonparametric model: Nonlinear *p* value	Semiparametric model: Nonlinear p value
Avam–Taimyr	0.035	0.026[a]	0.543	—
Columbia Ice Field	−0.002	0.873	0.466	—
Finnish Lapland	0.037	0.014[a]	0.473	—
French Alps	0.021	0.131	0.121	—
Greenland composite	0.002	0.878	0.237	—
Gulf of Alaska	0.021	0.133	0.560	—
Indigirka	0.019	0.127	0.084[b]	0.027[a]
Jämtland	0.013	0.935	0.647	—
Polar Urals	0.060	0.005[c]	0.243	—
Southern Sierra Nevada	0.042	0.002[c]	0.591	—
Southern Colorado Plateau	0.057	0[c]	0.489	—
Teletskoe Lake	0.002	0.914	0.193	—
The Alps	0.012	0.375	0.610	—
Torneträsk	0.006	0.701	0.358	—
Yamal	−0.033	0.081[b]	0.079[b]	0.043[a]

[a] Significance at the 5% level.
[b] Significance at the 10% level.
[c] Significance at the 1% level.

Estimated nonparametric functions

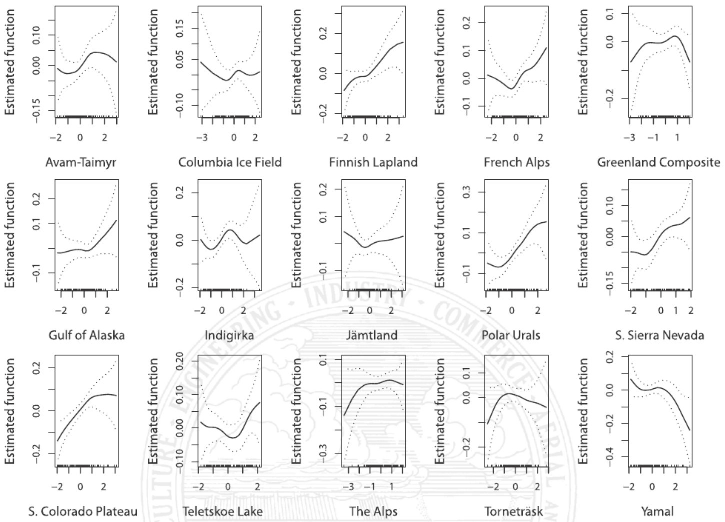


FIG. 1. The estimated functions (i.e., $\hat{m}$) for each proxy for the nonparametric model. The dashed lines in the plots are twice the pointwise standard error bounds. The vertical marks along the bottom illustrate the distribution of the proxies.

Estimated semiparametric functions

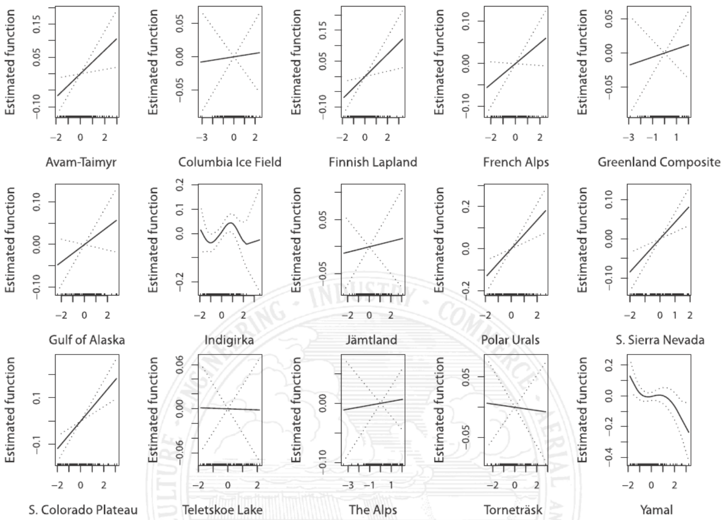


FIG. 2. As in Fig. 1, but for each proxy for the semiparametric model.

Model comparison tests and in-sample correlations

TABLE 3. Model comparison tests with (first column) the Df, (second column) the RSS, and (fourth column) the results from the two tests.

Model	Df	RSS	F statistic	p value
Linear	104	1.634	—	—
Nonparametric	59	0.881	1.120	0.339
Linear	104	1.634	—	—
Semiparametric	98	1.380	3.012	0.009*

* Significance at the 1% level.

TABLE 4. In-sample correlation between the reconstructions and measured NH temperature (1850–1969).

Model	Correlation with NH temperature
Linear model	0.80
Semiparametric model	0.86
Nonparametric model	0.92

Robustness of the nonlinearity test

TABLE 5. Results of testing the robustness of the nonlinearity test, based on leave-one-out sampling. As there are 15 proxies, we can choose 15 different sets of 14 proxies each and test for non-linearity and see whether a particular proxy tests positive for nonlinearity. We count the number of times a proxy is found significantly nonlinear (at 10% level) in the 15 possible calibrated nonparametric models. For instance, Yamal was found to be non-linear 10 out of 15 times, while Teletskoe Lake was only found nonlinear twice out of the 15 tests.

Proxy	Significantly nonlinear
French Alps	2
Greenland composite	1
Indigirka	13
Polar Urals	2
Telteskoe Lake	2
Yamal	10
Others	0

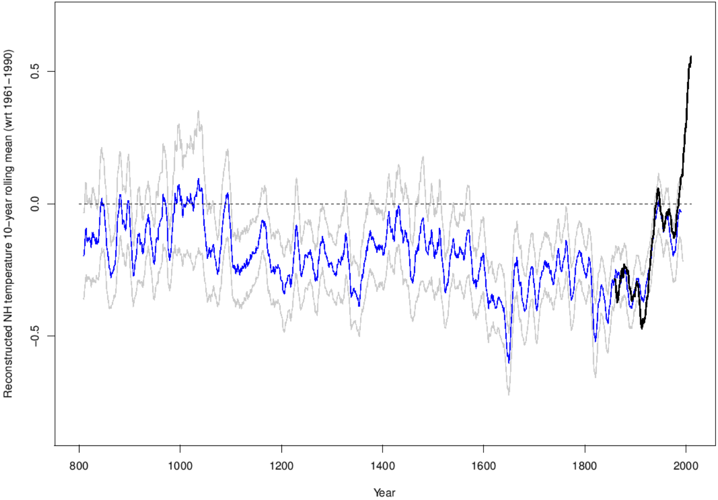
Concluding so far

- Our analysis shows that relative to the present group of proxies two are non-linear relative to the NH mean temperature. The series are the tree-ring width records Indigirka and Yamal.
- That only two out of 15 proxies tested positive for non-linearity is support for the general assumption that climate proxies can be used in linear reconstruction attempts.

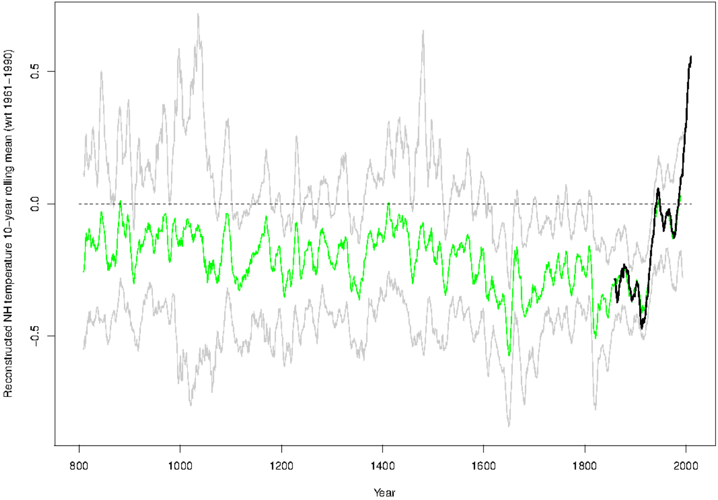
Temperature reconstructions

- Use the three models to reconstruct temperature from 800-1990 (minimize problems with missing proxy data)
- Estimate uncertainties in the reconstructions by bootstraping the observations (temperature and proxies from the calibration period 1850-1969). 1000 bootstrap resamples, calibrate models and perform reconstructions on all resamples. Find the lower 2.5 % and upper 97.5 % percentiles
- Note that the reconstructions presented are for comparative purposes only

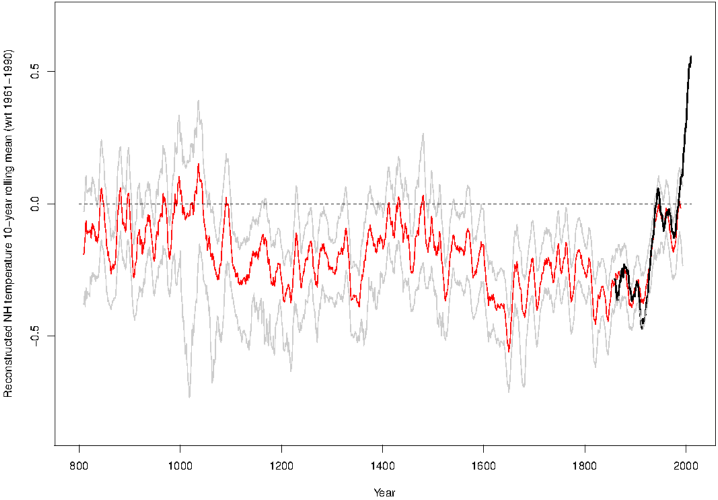
Temperature reconstructions: linear



Temperature reconstructions: nonparametric



Temperature reconstructions: semiparametric



Artificial example

Specify a nonlinear model for three of the proxies, generate artificial temperatures (Y) 1850-1969, and then check whether the method detects this model.

$$Y = \sum_{j=1}^{3} m_j(X_j) + e,$$

X_1, X_2 and X_3: Southern Colorado Plateau (as its coefficient is found significant in the linear model), Indigirka and Yamal (as these are found significantly nonlinear).
$m_1 = 0.06 \cdot X_1$, $m_2 = 0.1 \cdot \cos^2(X_2))$ and $m_3 = 0.02 \cdot X_3^2 - 0.02 \cdot X_3^3$. The functions are choosen to resemblance the calibrated functions from the semiparametric model.
e: noise from a normal distribution (zero mean, sd = 0.15)

Artificial example II

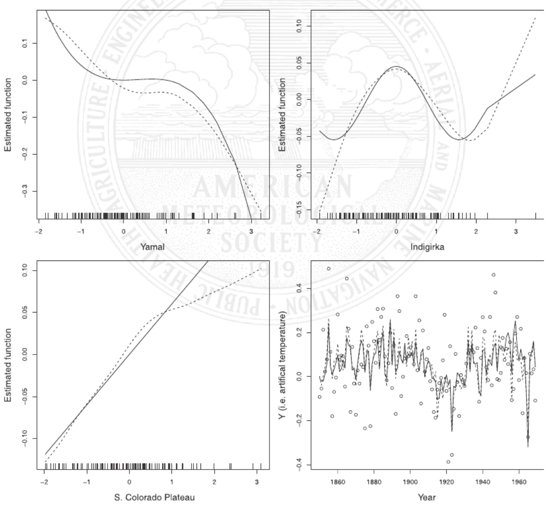


FIG. A1. (from top left to bottom left) The estimated functions (dashed curves) for the fitted nonparametric model and the true functions (solid curves) for the artificial case. (bottom right) Points are the artificial Y with noise, the solid curve is the artificial Y without noise, and the dashed curve is the reconstruction obtained from the nonparametric model.

Artificial example III

TABLE A1. Results from the approximate nonlinear test for the fitted nonparametric model for the artificial case.

Proxy	Nonlinear p value
Southern Colorado Plateau	0.599
Indigirka	0.001*
Yamal	0.051**

* = Significance at the 1% level.
** = Significance at the 5% level.

Conclude that the method indeed is capable of detecting the underlying functions and the artificial temperature.
The nonlinear effect test detects correctly Indigirka and Yamal.

Concluding remarks I

- Our findings partly support the general assumptions that temperature proxies can be used in linear reconstructions attempts
- An alternative test; generalized likeliood ratio (GLR) test of Fan and Jiang (2005) (need conditional bootstrap)
- Should compare nonparametric/semiparametric reconstructions to more commonly used methods (as e.g. RegEM)

Concluding remarks II

- Temperature reconstruction is important for understanding climate variability
- The temperature reconstruction problem offers a difficult statistical modeling challenge
- Note that some studies (e.g. Christiansen et al (2009), McShane and Wyner (2011)) claim that current methods may underestimate uncertainties in temperature reconstructions
- Bayesian models on the way

$\Rightarrow$ more work for statisticians!

Some references

Christiansen et al (2009). A surrogate ensemble study of climate reconstruction methods: Stochasticity and robustenss. *Journal of Climate*

McShane and Wyner (2011). A statistical analysis of temperature proxies: Are reconstructions of surface temperatures over the last 1000 years reliable? *Annals of Applied Statistics*

Li et al (2010). The value of multiproxy reconstruction of past climate. *JASA*.