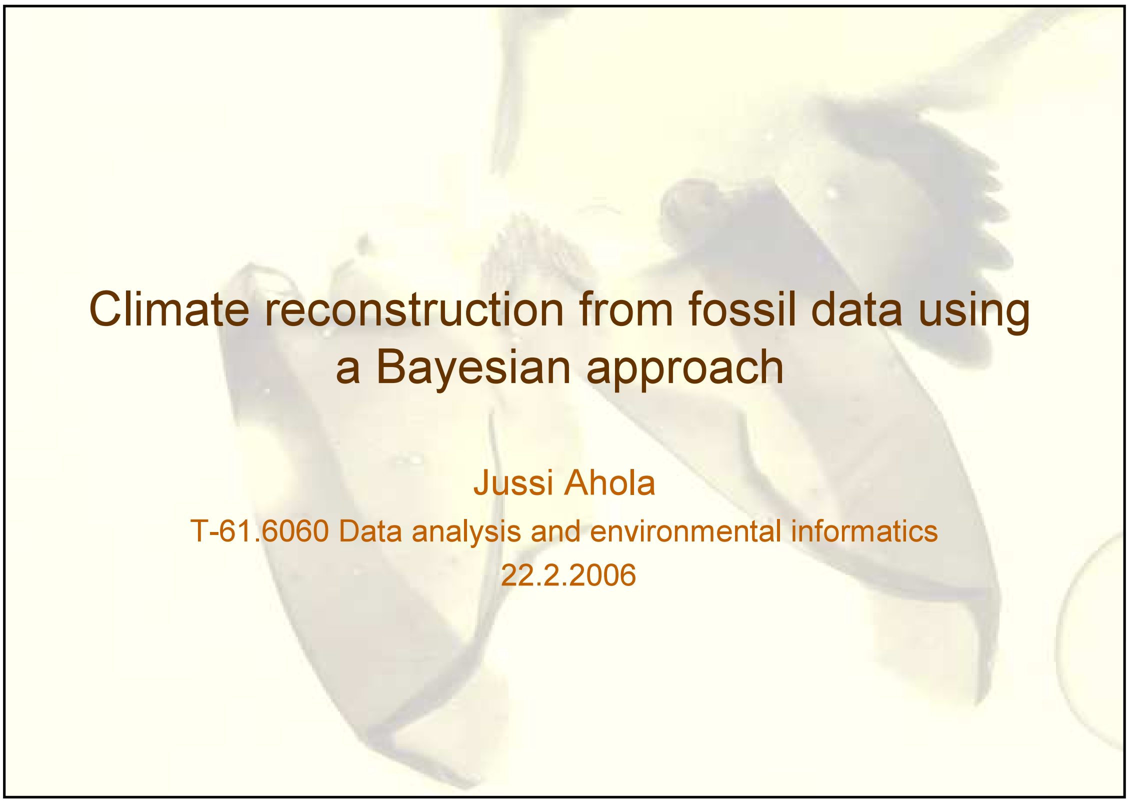


Climate reconstruction from fossil data using a Bayesian approach

Jussi Ahola

T-61.6060 Data analysis and environmental informatics

22.2.2006

Outline

- Introduction
- Data
- Model
- Results
- Conclusions

Starting point

- Forecasting climate change requires information on the past climate behavior
- Meteorological records are too short to cover the full climate variation
  -> History has to be inferred from indirect “proxy” indicators
- Biological indicators preserved in the sedimentary deposits are a major source of proxy-climate data

---

Reconstruction

- Early attempts for climate reconstruction utilized qualitative and descriptive analysis on single indicator species data
- Nowadays numerical techniques and approaches are available allowing quantitative reconstruction
- In general, the methods involve two steps
  - Regression: mapping responses of biological indicators to the contemporary environment
  - Calibration: predicting the environmental variable from the fossil data

Approaches

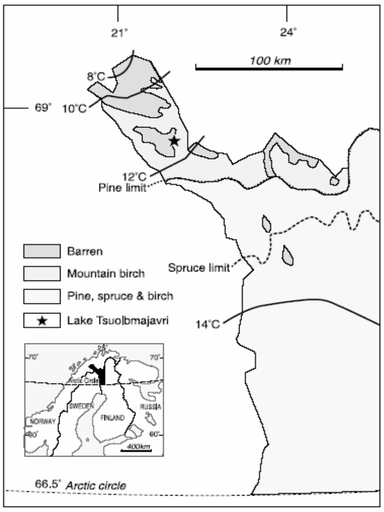
- Ways to perform the quantitative reconstruction can be classified to two categories:
  - Classical: estimate the response function from the climate state to the biological indicator, and reconstruct the climate state by using the inverse of the function
  - Inverse: estimating directly the inverse function from the biological indicators to the climate and applying the function in the reconstruction
- Reconstruction contains many uncertainties to be accounted for
  - Nature in general is not deterministic
  - Measured biological indicator data is “noisy”
  - Applied model only approximates the real world

Objective

- Applying Bayesian multinomial Gaussian response model to fossil chironomid assemblages in a tree-line lake in Finnish Lapland to reconstruct Holocene palaeotemperatures
  - All unknown quantities – parameters as well as the data – are treated as random variables
  - The gained knowledge and reconstructions are explicitly conditioned on the statistical model, prior knowledge and data in hand

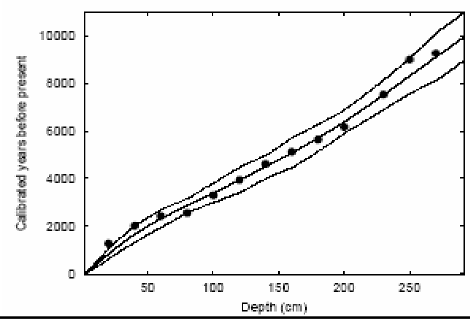
Study site

- Lake Tsuolbmajavri is located on the north-western Finland, 526m above sea level at the boundary between boreal forest and tundra
  - It is small (A=13.9 ha), shallow (Z_{max}=5.35 m), clear and oligotrophic
  - Measured mid-summer water temperature is 10.9°C and site-specific altitude-corrected mean July air temperature is 11.0°C.
  - The mean annual air temperature is about 2°C; annual precipitation is about 350mm (50% falls between June and September)



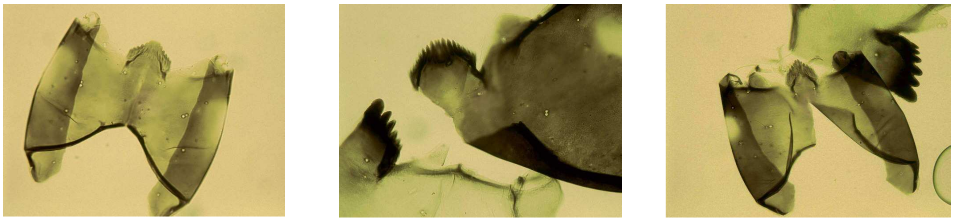
Fossil data

- A 291-cm-long sediment core was sampled in 1997 at the central part of the lake from the frozen lake surface
- For chronology 14 radiocarbon (^{14}C) datings were determined from the core and age-depth models were developed by non-parametric weighted regression
- The dates suggest a fairly linear sediment accumulation rate



Chironomid analysis

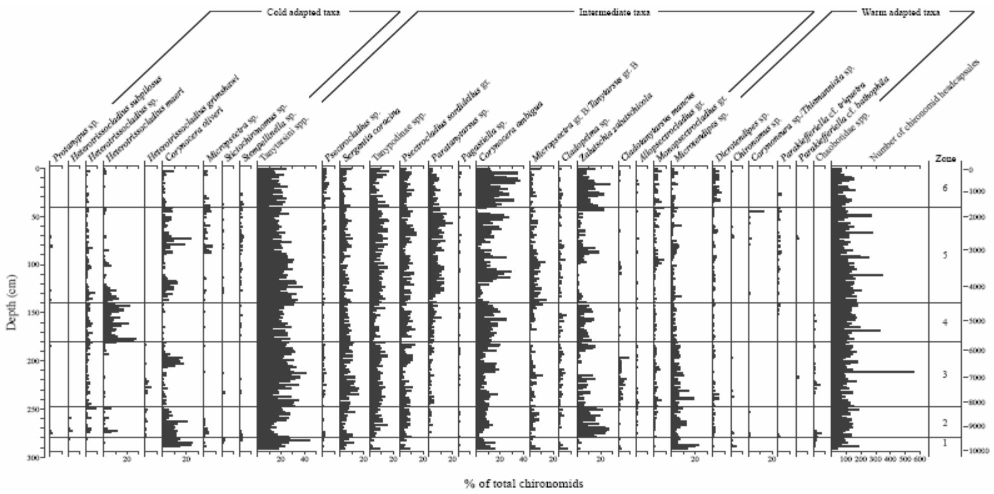
- The sediment core was sampled for subfossil chironomids at intervals of 2 cm, which is equivalent to a temporal resolution of 50–70 years
- Subsequently, a few samples were taken at 1 cm interval from sequences with rapid changes in chironomid composition
- Chironomid head capsules were counted from the samples using complex, manual procerdure (including chemical processing, sieving, microscoping etc.)
  - 50 chironomid taxa were recorded from the total of 148 samples



Contemporary data

- The modern training data used for calibration models include surface-sediment chironomid head capsules and air temperature data from 63 lakes in north-western Finland.
  - The study lakes are generally small, clear and oligotrophic
- Chironomid data are expressed as the head capsule count
  - Rare taxa (less than 2% relative abundance in more than 4 lakes) were excluded
  - The remaining data included 52 chironomid taxa, 45 of which are common with the fossil data
- Mean July temperatures were estimated for each lake using data of nearby climate stations from years 1961-90
  - One measurement was removed as an outlier
  - The remaining data included 62 measurements ranging from 7.9°C to 13.8°C

Chironomid stratigraphy



Problem statement

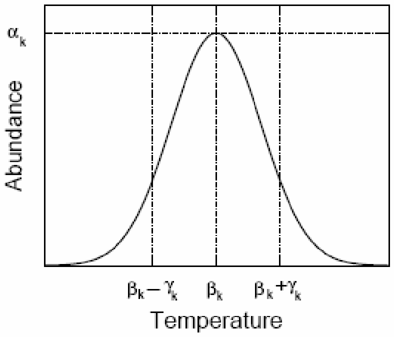
- Let y_{ik} be the count of taxon k at site i and $y_i = (y_{i1}, \ldots, y_{im})^T$ be the actual counts at site i
  -> $Y = [y_1, \ldots, y_n]$ is the $m{\times}n$ matrix of abundances (m=52, n=63)
- Furthermore let this be denoted as $\tilde{y}_{ik} = y_{ik}/y_{i+}$, where $y_{i+} = \sum_h y_{ih}$ (relative abundance)
- Let x_i be the temperature at site I
  -> $X = (x_1, \ldots, x_n)^T$ is the vector of temperatures at sites
- Let $y_0 = (y_{01}, \ldots, y_{0m})^T$ be a fossil assemblage and x_0 the air temperature at the time the fossil assemblage was formed
- Given Y, X and y_0, the task is to reconstruct the temperature x_0.
  - In other words the idea is to estimate the posterior distribution $f(x_0|X,Y,y_0)$

Bayesian approach

- Let θ be the vector of parameters of the statistical model
  -> $f(x_0|X,Y,y_0) = \int f(x_0,\theta|X,Y,y_0)d\theta$
  $\propto \int f(x_0,\theta)f(Y,y_0|X,x_0,\theta)d\theta$
- In other words, the model specifies the response functions that consist of two parts
  - Likelihood term $f(Y,y_0|X,x_0,\theta)$
  - Prior probability distribution $f(x_0,\theta)$

Bummer model

- Bummer uses a multinomial response model
  -> $E(\tilde{y}_{ik}|x_i,\theta_k) = \lambda_{ik} / \sum_j \lambda_{ij}$, where λ_{ik} is the response of taxon k to temperature at site i
- It is assumed that the response curve has a unimodal Gaussian shape
  -> $\lambda_{ik} = \alpha_k \exp\{-[(\beta_k - x_i)/\gamma_k]^2\}$
- Hence $\theta_k = (\alpha_k, \beta_k, \gamma_k)$ represents hyperparameters that control scaling factor, optimum and tolerance of taxon k to the temperature



Likelihood term

- It is assumed that sites are mutually independent given the temperatures and taxon-specific model parameters

  -> $f(Y,y_0|X,x_0,\theta) = \prod_i f(y_i|X,x_0,\theta) = \prod_i f(y_i|x_i,\theta)$
- It is further assumed that the taxon abundances y_i are multinomially distributed

  -> $(y_{i1},y_{i2},\ldots,y_{im}|y_{i+},p_i) \sim \mathrm{Mult}(y_{i+},p_i)$, where $p_i = (p_{i1},\ldots,p_{im})$ and p_{ik} is the probability of an individual from site *i* to be of taxon *k*
- Random variables p_i are assumed to be Dirichlet distributed

  -> $(p_{i1},\ldots,p_{im}|x_i,\theta) \sim \mathrm{Dirichlet}(\lambda_{i1},\ldots,\lambda_{im})$, given the Gaussian responses λ_{ij}

Prior distributions

- It is assumed that each temperature x_0 to be reconstructed as well as all species-specific parameters α_k, β_k, *and* γ_k are mutually independent in the prior

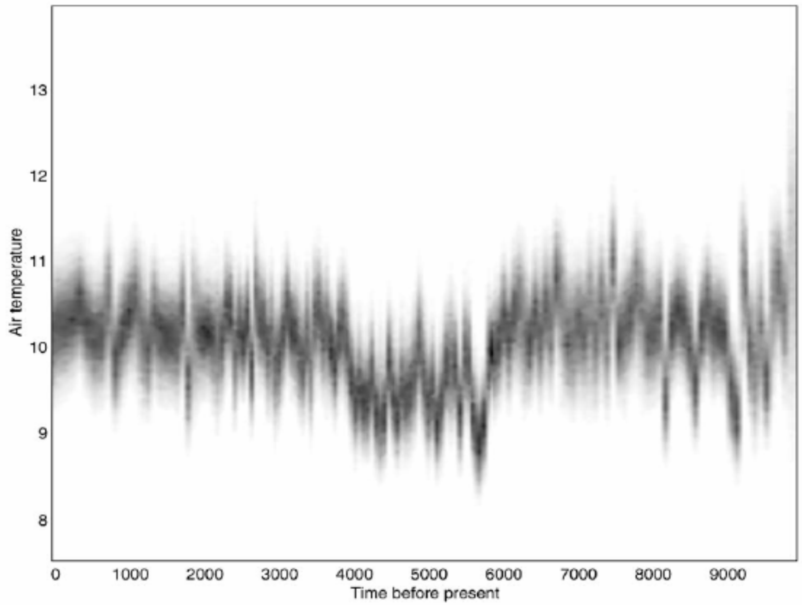
  -> The joint prior distribution $f(x_0,\theta)$ can be determined by individual marginal priors
- The following parameter *apriori* distributions were chosen
  - x_0 is normal distributed with mean 11 and variance 1
  - α_k is uniformly distributed in [0.1, 50]
  - β_k is normal distributed with mean 11 and variance 3
  - γ_k is gamma distributed with mean 3 and variance 1

Modeling

- The integration of the model was implemented using stochastic Markov chain Monte Carlo (MCMC) methods
- Statistical accuracy of the reconstructed temperature was assessed using the contemporary data
  - 0.8°C root mean square error of the prediction calculated by leave-one-out cross-validation
  - Maximum bias along the temperature gradient 0.98°C
- Bummer has high predictive and low inherent systematic and maximum bias

  -> Model should be sufficiently precise for tracking temperature changes, given the assumed sensitivity of the proxy source

Inferred temperature



Reliability of the reconstruction

- There are many uncertainties and potential problems to be taken into account in the interpretation
  - Many of finer scale variations are smaller that the average accuracy of the model
  - Chronological control and chironomid analysis involve uncertainties
  - Stochastic nature of the environment
- Large-scale patterns were assessed by comparing with independent proxy records from the same region

  -> Results seems accurate enough, apart from the late-Holocene (due to shallowing of the lake by progressive sedimentation)
- The uncertainty in the fine-scale changes was minimized by
  - Using taxa whose ecological indicator value is well-established
  - Ensuring that the taxa were well represented in fossil and contemporary data
  - Clarifying the mechanisms and causations behind the changes

Interpretation of the reconstruction

- Climate during the early Holocene was particularly unstable
  - However, concerning the very first few hundreds of years of the Holocene the prediction accuracy (credibility) is quite low
- Fairly low but steadily rising summer temperatures during the early Holocene.
  - There are, however, three successive cooling events with amplitudes of 1–1.5°C at around 9200, 8600 and 8300 cal yr BP
- The interval between ca 8000 and 5800 cal yr BP appears to have been warm and stable
- Relatively long-lasting temperature minimum between ca 5800 and 4000 cal yr BP (ca 1°C drop in the inferred mean)
  - Within the cold interval, particularly cool summers appear to have prevailed at about 5800, 5000 and 4200 cal yr BP
- The inferred late-Holocene climate is surprisingly featureless, fluctuating between 10-10.5°C during the past 4000 years.
  - A brief cold oscillation can, however, be distinguished with some confidence at about 1800 cal BP

Summary

- A Bayesian model was implemented for reconstructing long-term temperature changes from subfossil chironomid remains
  - Representing the temperatures as posterior distributions involves explicit information on the model statistical uncertainty
  - Embedded ecological knowledge results in more plausible model
  - Classical approach enables improved extrapolation capability
- Obtained summer temperature reconstruction is in general consistent with other independent proxy sources from the region
  - Unstable early Holocene, cooling events at 9200, 8600 and 8300 cal yr BP, thermal maximum ca 8000-5800 cal yr BP, and distinct cooling around 5800mcal yr BP
  - Late-Holocene temperature is apparently not reconstructed accurately due to the shallowing of the lake by progressive sedimentation

References

Vasko, K., Toivonen, H.T.T., Korhola, A., 2000. *A Bayesian multinomial Gaussian response model for organism-based environmental reconstruction*. Journal of Paleolimnology 24, 243–250.

Toivonen, H.T.T., Mannila, H., Korhola, A., Olander, H., 2001. *Applying Bayesian statistics to organism-based environmental reconstruction*. Ecological Applications 11, 618–630.

Korhola, A., Vasko, K., Toivonen, H.T.T., Olander, H., 2001. *Holocene temperature changes in northern Fennoscandia reconstructed from chironomids using Bayesian modelling*. Quaternary Science Reviews 21, 1841–1860