The study of uncertainty in palaeoclimate reconstruction

John Haslett, TCD, Ireland

…and many others

Supported by Science Foundation Ireland

IPCC Inter Governmental Panel on Climate Change
Abrupt climate change

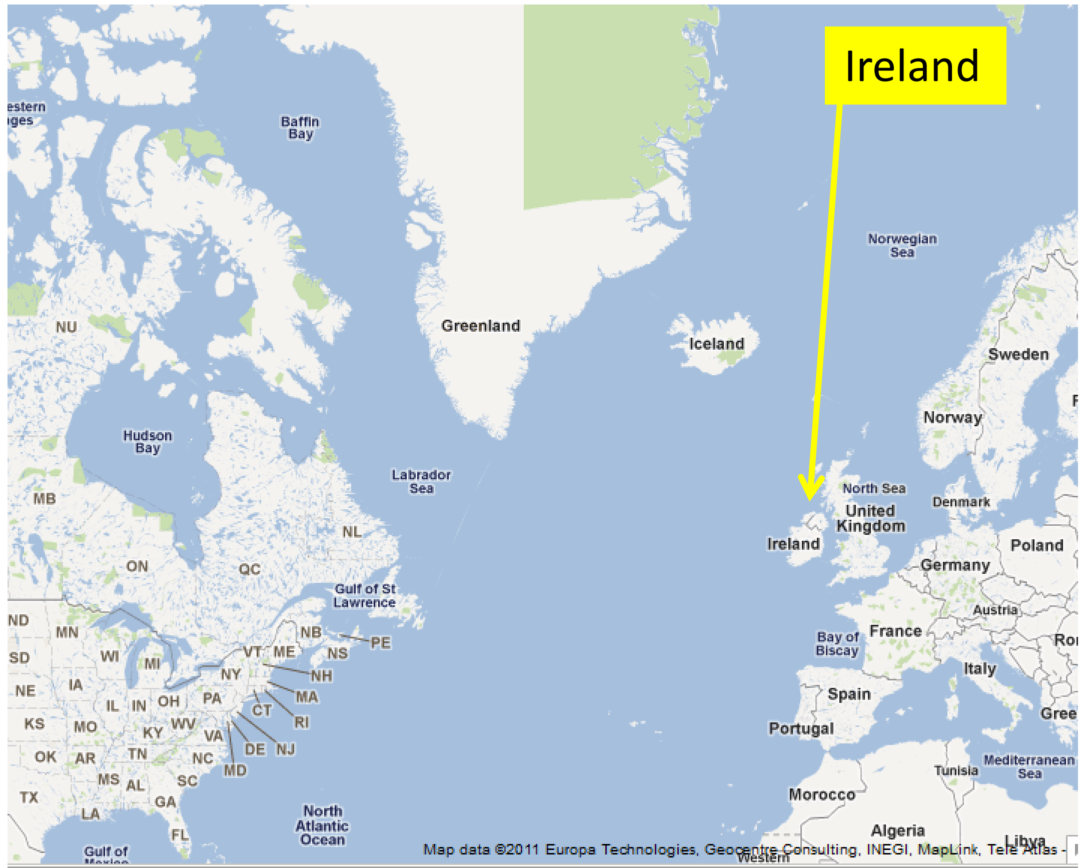
IPCC WG1 2007

“During the last glacial period, abrupt regional warmings (probably up to 16°C within decades over Greenland) occurred repeatedly over the North Atlantic region”

(Fifth Assessment Report due 2014)

What do we know about abrupt changes?

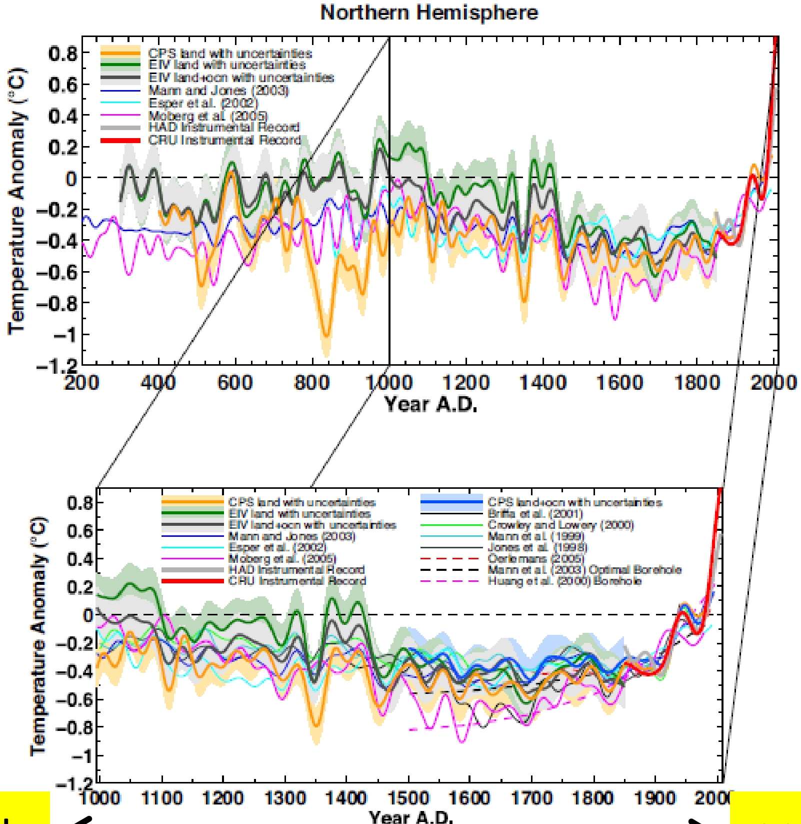
Pitifully little



N Hemisphere Air Temp multi-proxy past 2000 years

What do we know about past millenium?

Quite a lot



past

present

Classic study

Mann et al (2008)

2008 update of 1998

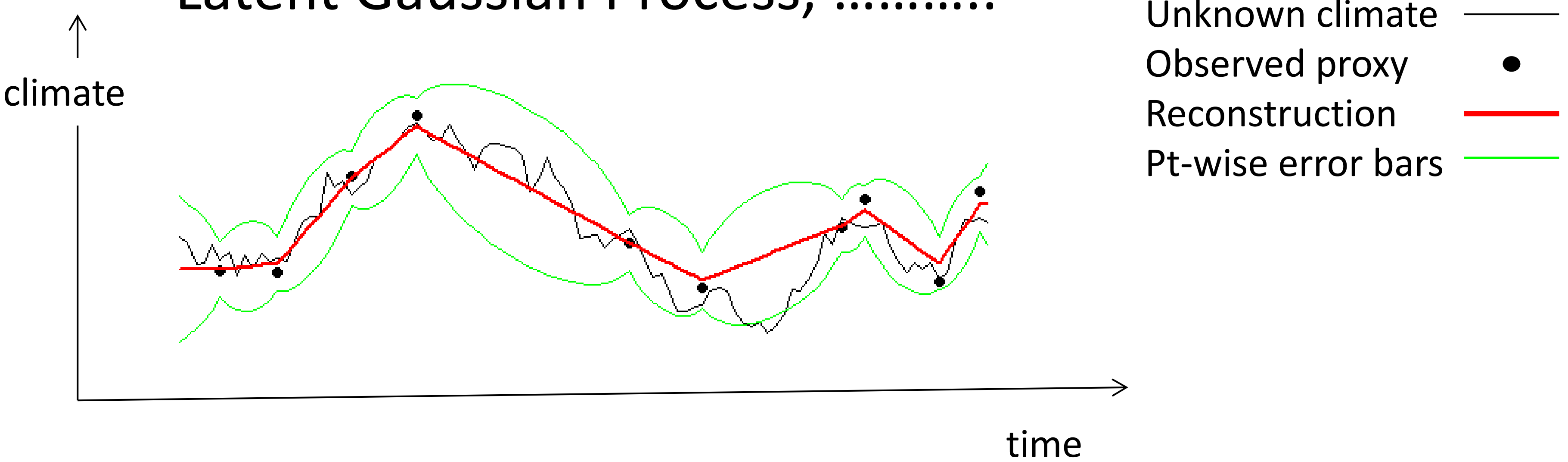
Wiki: Hockey Stick Controversy

Palaeoclimate: **What, How** and Why?

- What? Past climate
  - Temporal scale — 1000 years, 100,000 years
  - Spatial scale — Hemisphere; China; Italian/Irish lakes
  - Physical Process — Ocean, Atmosphere, Ice, Biosphere
  - Statistical summary — *Avg* Atmos temp, *Avg* Ocean *Summer* temp; Atmos moisture *extrema, Abrupt change*
- How
  - Computer Models — *General Circulation Models* (*GCM*)
  - Proxy Data
  - Statistical Methods

Statistical Methods in Palaoeclimate Reconstruction

- Partially observed stochastic processes
  – Simple case
  – BLUP, Kriging, State-space, Latent Gaussian Process, ...



Palaeoclimate: What, How and **Why**?

- Why at YSI2013?
  - Topical, Policy, Challenging, Defined
  - Never-ending source of statistical problems (& grants)
    - Computational
    - Conceptual
    - **Communication**
  - **Uncertainty central**
    - "probably up to 16°C within decades over Greenland"

Statisticians specialists in uncertainty.
How do we communicate it?

Palaeoclimate: What, How and **Why**?

System	**Climate**	**Economy**	**Human Body**	**?**
Subsystem	Atmosphere, Ocean…	Domestic, International…	Kidney, Heart…	
Math Theory, Comp Models				
Proxy Systems				
Data				
Combining Uncertainty				
Communicating Uncertainty				
Policy				

Monte Carlo methods

Experienced Statistician

The *process* of modelling

Adopt	simplified math representation
Solve	to get useful information
Too hard?	stronger and stronger assumptions

Experienced Statistician

The *process* of modelling

Adopt simplified math representation

Solve to get useful information

Too hard? stranger and stranger assumptions

Modelling has got easier

- Linear regression; Normal errors
- glm, gam
- Hierarchical models
- Bayesian methods
- Stochastic processes
- MCMC
- Data-bases, Computers, Algorithms

New challenges

IPS006

Statistical methods in the study of palaeoclimate

09:00 Monday

- *Preliminary multiproxy surface air temperature field reconstruction for China over the past millennium*

  Feng Shi; China

- Paleoclimate reconstruction using statistical nonlinear forward models

  Peter F. Craigmile; United Kingdom

- Bayesian palaeoclimate inference from pollen in Southern Italy Andrew Parnell; Ireland
- Discussion Paul Switzer; USA

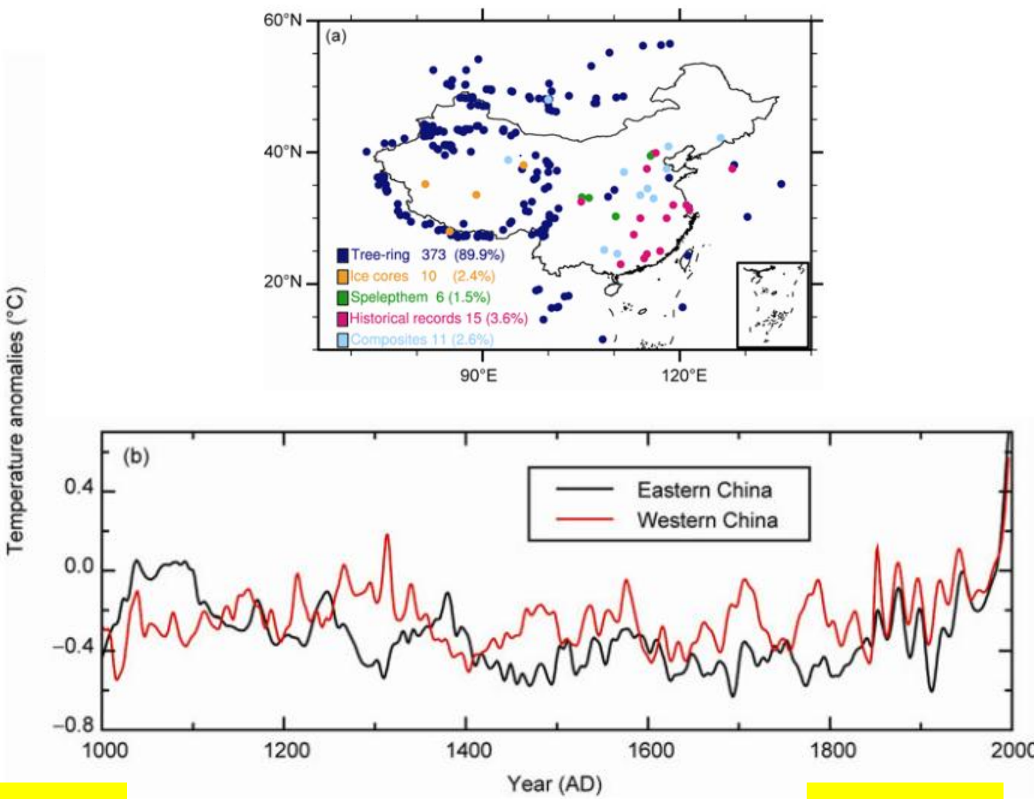
Palaeoclimate Reconstruction

What: Examples

How: Uncertainty as Stat Inference on Latent Climate

Challenges in Communicating Uncertainty

China Air temp "anomalies" wrt present
Multi-proxy – last 1000 years



past ← - - - - - - - - - - - - - - → present

Shi Feng IPS006

Focus: Air Temp (spatial avg, annually)

Proxies:

Tree Rings	373 cores
Ice Core	10 cores
Speleothem	6 cores
Historical	15
Other	11

Heavily pre-processed data

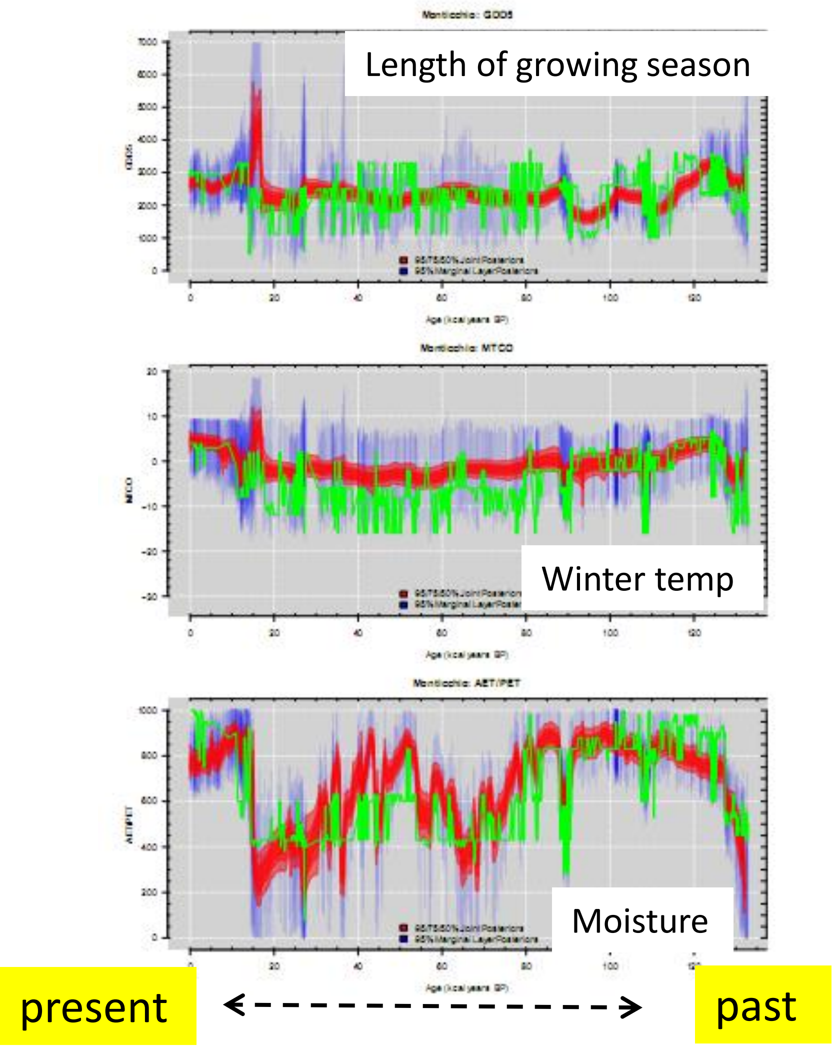
No dating uncertainty issues

Linear methods - 'Regression'

Assumption:

Temp linearly related to proxy

Italy – Multiple aspects of climate
Pollen, last 100,000 years



Parnell IPS006

Focus – Multiple aspects of climate
Uncertainty

Proxy One core
924 samples; 28 taxa

Raw data

No dating uncertainty (unusual for pollen)

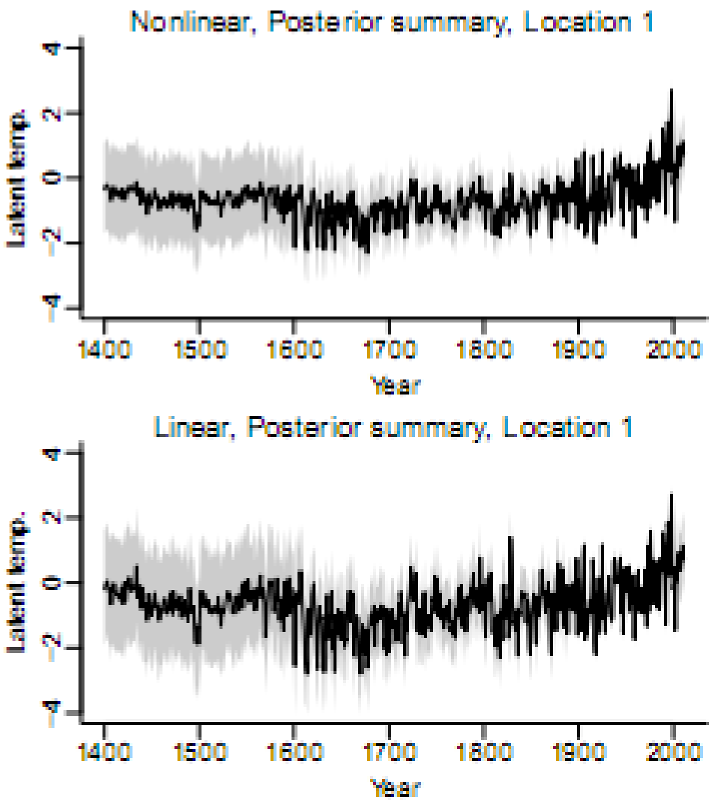
Non-linear, non-Gaussian

Bayesian process modelling

Assumption:

Non-monotone multivariate relationship between proxy and climate

N. America – Air Temp
Tree Rings, past 600 years



past ← - - - - - - - - - → present

Craigmile IPS006

Focus – Air temp

Uncertainty

Proxy Two cores

Tree ring

Heavily processed data

No dating uncertainty issues

Non-Linear, non-Gaussian

Bayesian - process modelling

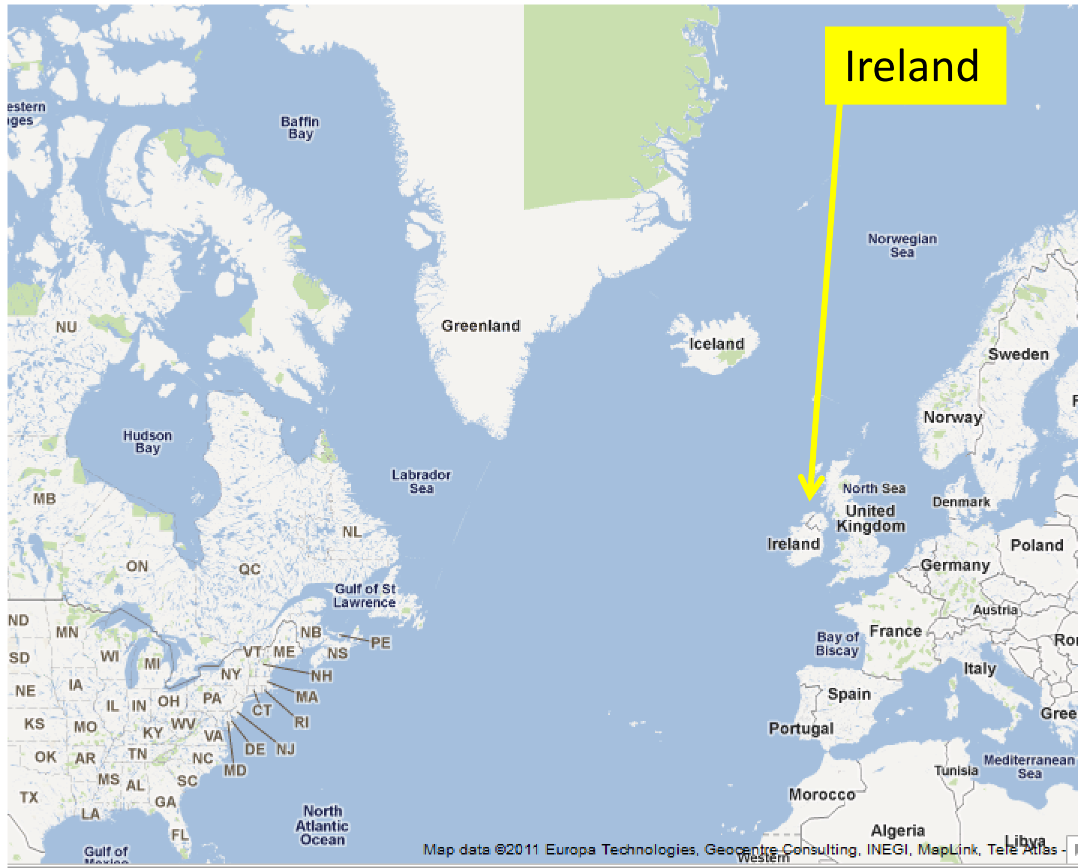
Assumption:

monotone relationship between (processed) proxy and temp

IPCC Inter Governmental Panel on Climate Change
Abrupt climate change

IPCC WG1 2007

"During the last glacial period, abrupt regional warmings (probably up to 16°C within decades over Greenland) occurred repeatedly over the North Atlantic region"

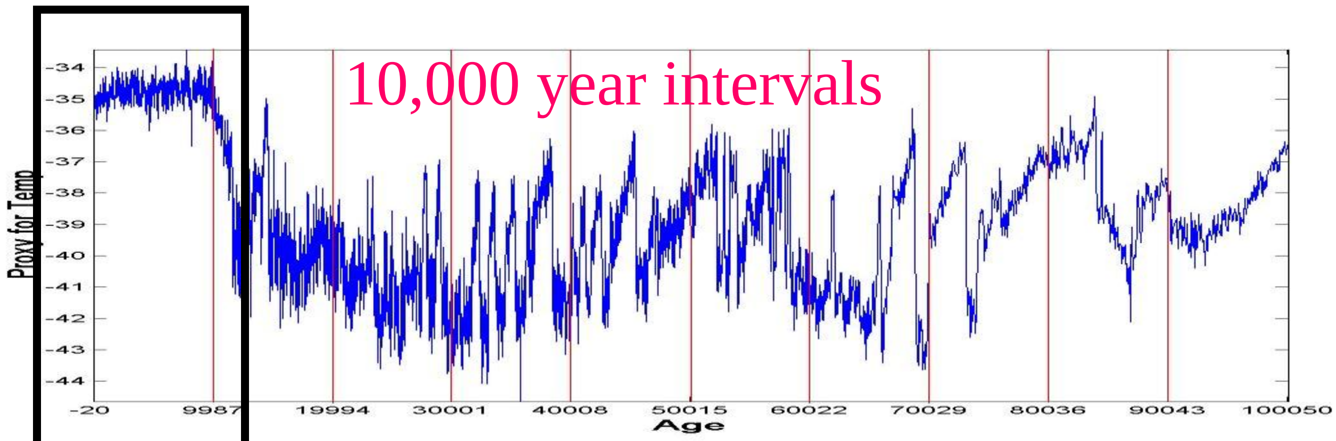


How do we know

anything?

Greenland summit – Air Temp
Oxygen Isotope ice last 100,000 years

Modern ⟵ ⟶ Ancient



Irish study period: next slides

Proxy: Oxygen isotope – proxy for temp

Temporal structure

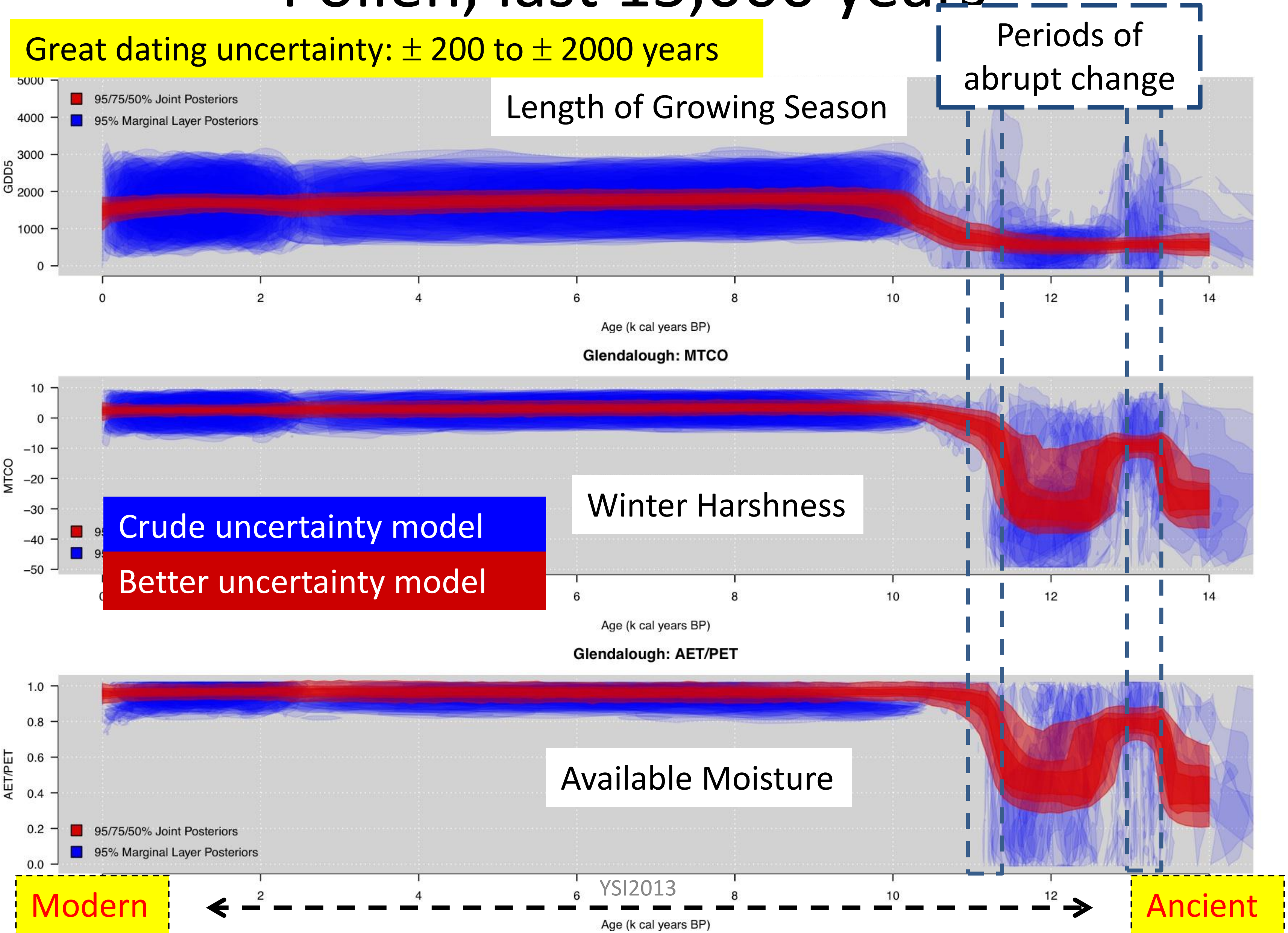
Frequent small changes, occasional large changes

Roughly (*at this scale*) : long tailed random walk

Glendalough Ireland

Courtesy of Sunday Times graphics dept

Ireland – Multiple aspects of climate
Pollen, last 15,000 years



Palaeoclimate Reconstruction

What: Examples

How: Uncertainty as Stat Inference on Latent Climate

Challenges in Communicating Uncertainty

Uncertainty Challenges

- Simple but adequate statistical models of multiple forward stochastic processes
- Data collection and processing
- Inversion – statistical inference, latent processes
- Communicating uncertainty
  - How? To whom?
  - About what?
- Combining uncertainties

"probably up to 16°C within decades over Greenland"

"Uncertainty *Fusion*"

Statistics: Inverting Forward Processes

Science

Forcing
Insolation
Volcanoes
CO_2

Feedback
Ocean/Ice
Bio-sphere
Anthro-CO_2

Climate processes $C(s,t)$

Proxy processes $P(s,t)$

Data collection $Y(s_i,d_i)$

Data processing $\{Z(s,t)\}$

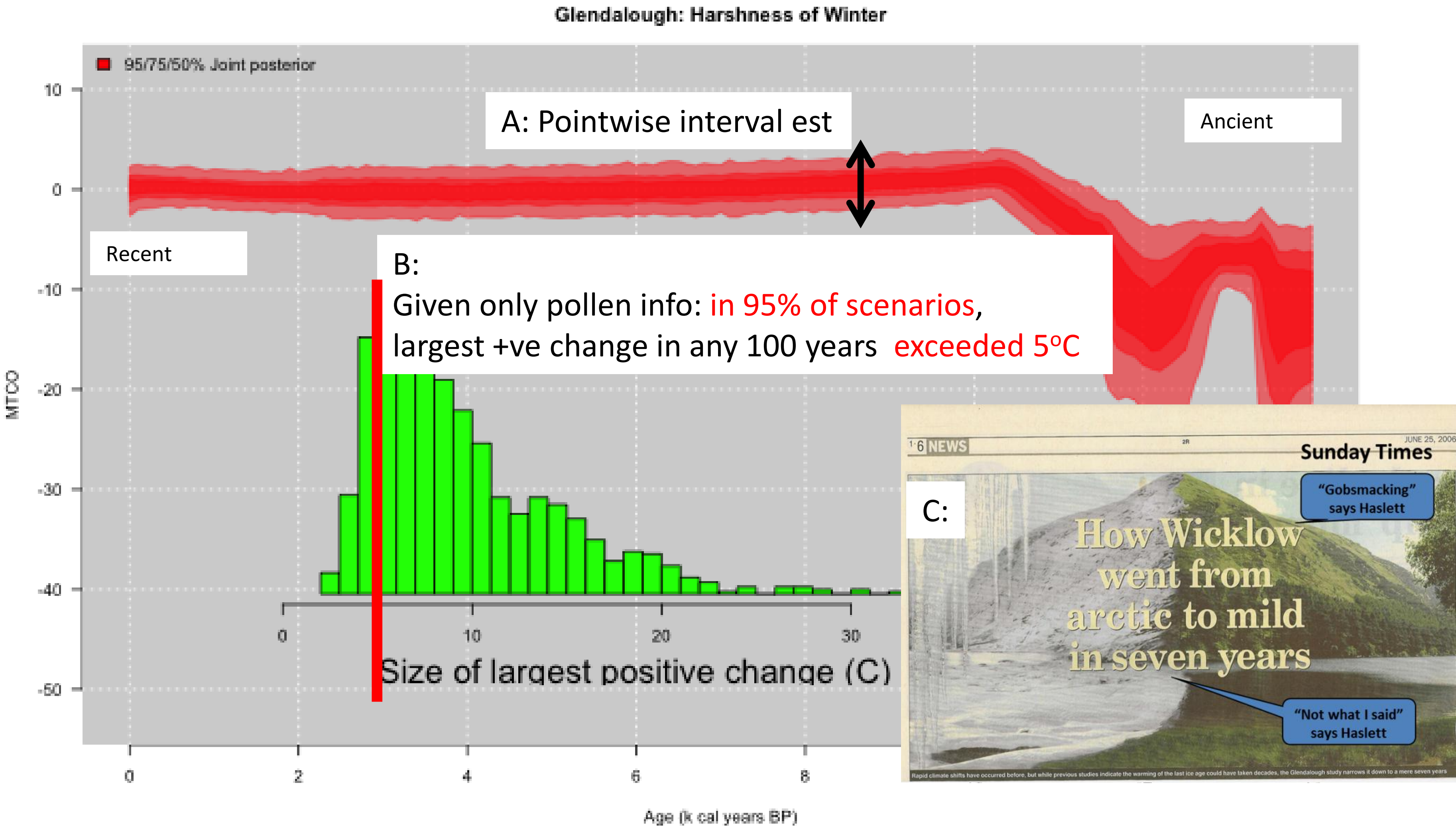
Statistics

Models
Phyiscal/Statistical
- Uncertainty
- Inference

Eg Depth
- Proxy for Age
- Uncertain time

Data
- Raw
- Processed

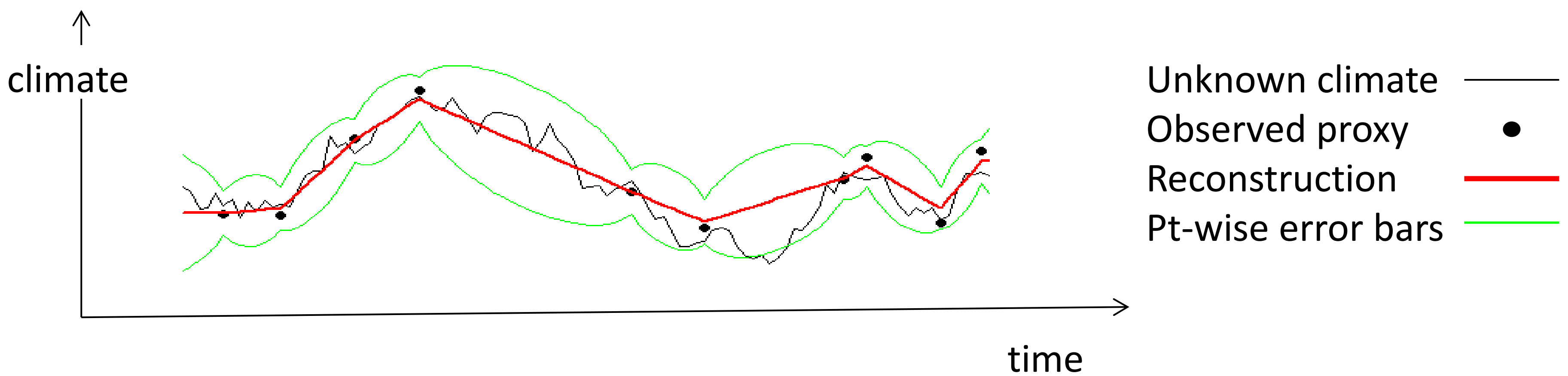
Communicating Uncertainty



Statistical Methods

- Partially observed stochastic processes

Simple case: single core: one location
underlying latent c(t) univariate Gaussian
slices correspond to known irregular times
proxy process ⇒ additive univariate Gaussian errors
with simple variance structure
'Reconstruct' ⇒ make statistical inference

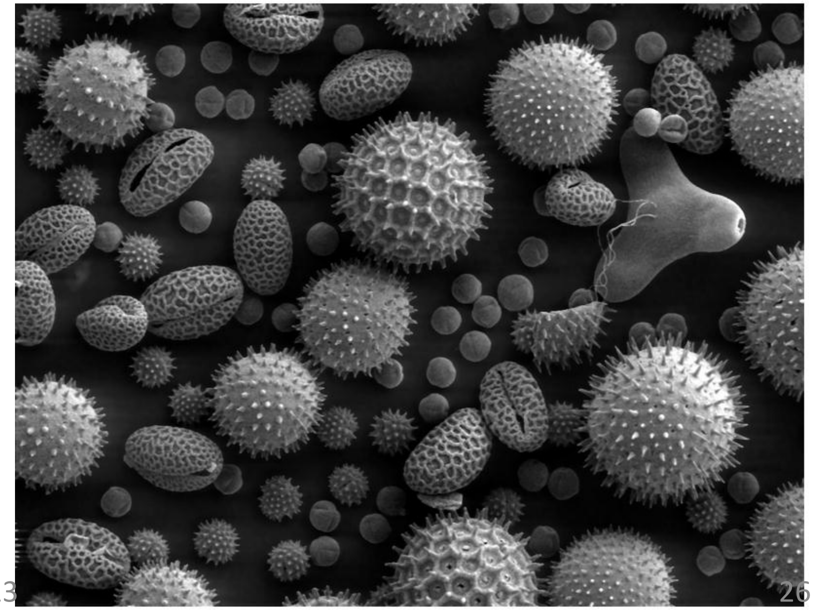


Statistical Methodologies

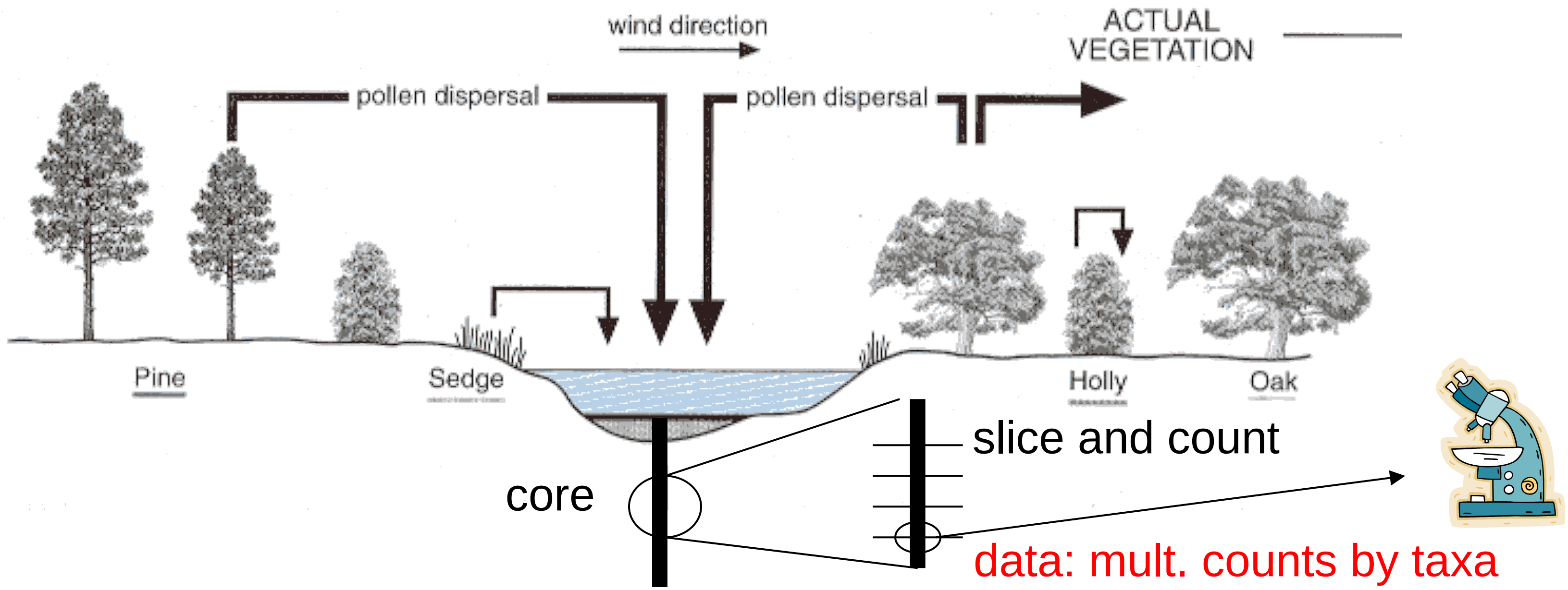
- Partially observed stochastic processes
  - State-space modelling *multivariate* climate
  - Space-time processes *long-tailed* climate inc's
  - Zero-inflated multivariate proxy processes *pollen*
- Multiple irregular time series
  - Time series with uncertain times *sediment*
- Interest in possibly *non-linear functionals*
- Monte Carlo methods in Bayesian Modelling
  - Integrated Nest Laplace Approximations INLA
  - Stochastic partial differential equations SPDE

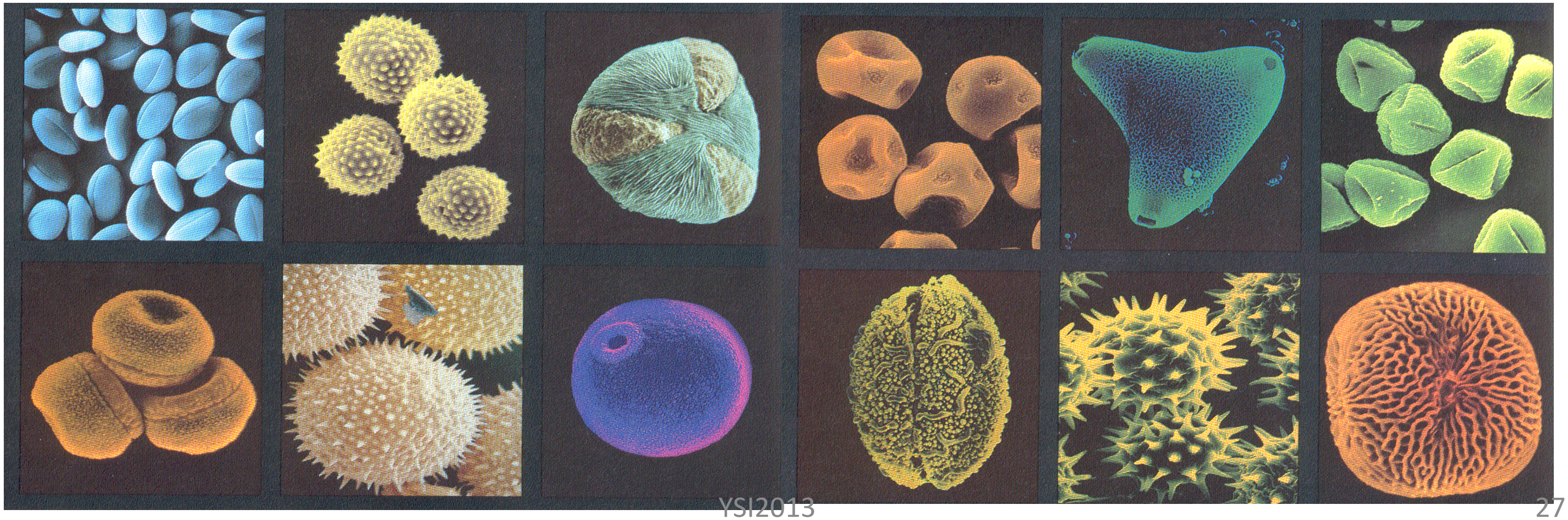
probably up to 16°C within decades over Greenland

Show me the data



Pollen: widely available, but challenging





Asian pollen - palms

From Wikipedia

Roughly 202 genera with around 2600 species are currently known, most of them restricted to tropical, subtropical, and warm temperate climates.

Morphologically diverse, inhabit nearly every type of habitat within their range, from rainforests to deserts

Distinctive pollen… but

how many of 2600
can be distinguished by eye
from ancient pollen grains?

Changing pollen composition

Data: compositional vector of observed pollen proportions

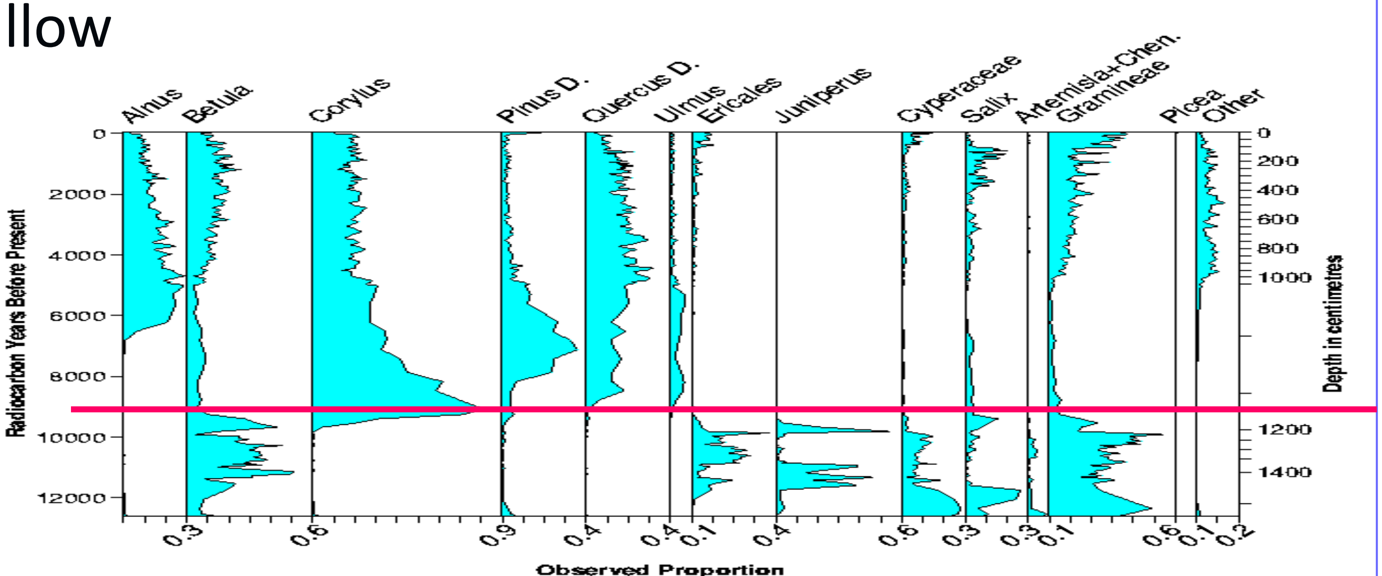
Proxy process: pollen rain,

Statistical issues discrete multivariate,

zero inflated, non Gaussian, non-linear

Pollen composition changes ⇒ Climate changes

Recent Shallow

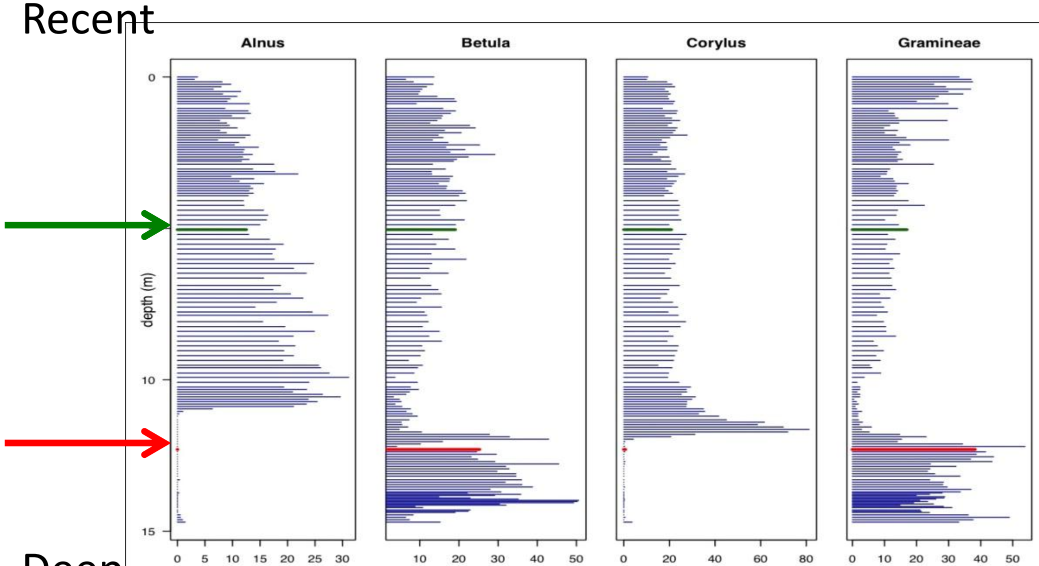


Ancient Deep

Climate Information from Pollen

Shallow
Recent

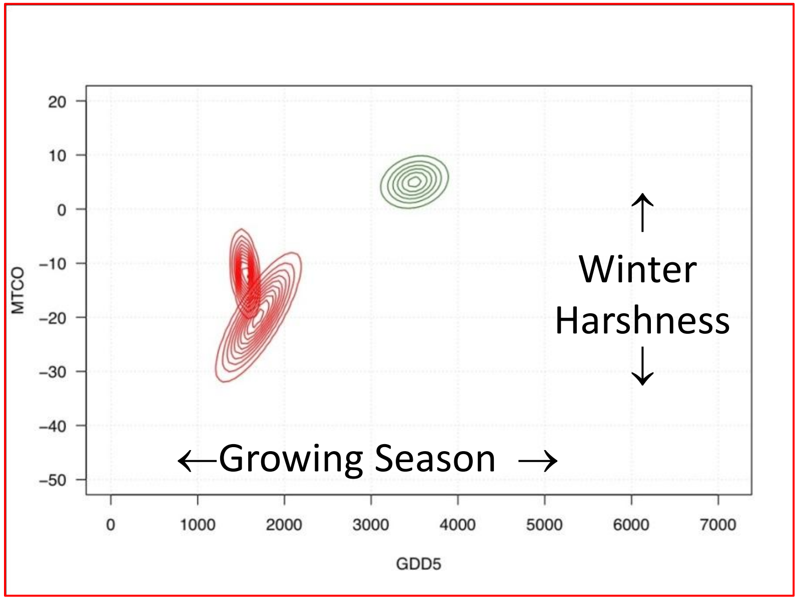
showing 4 of 28 taxa



Deep
Ancient

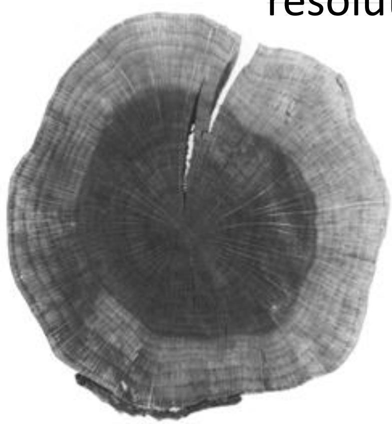
Depths known,
Times uncertain

Pollen expert ⇒
Bivariate (poss bimodal) climate

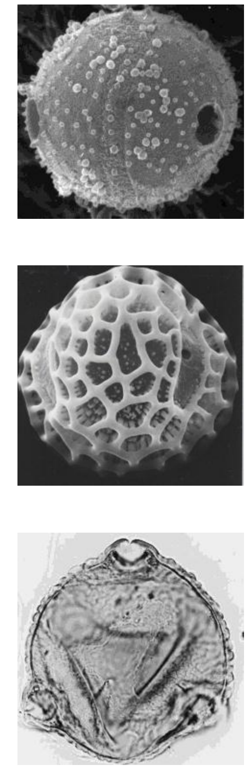


Other Proxies

‘Laminar’ –
annual
high-temporal
resolution



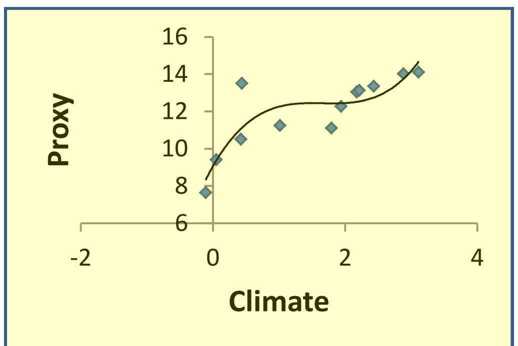
low-
temporal
resolution



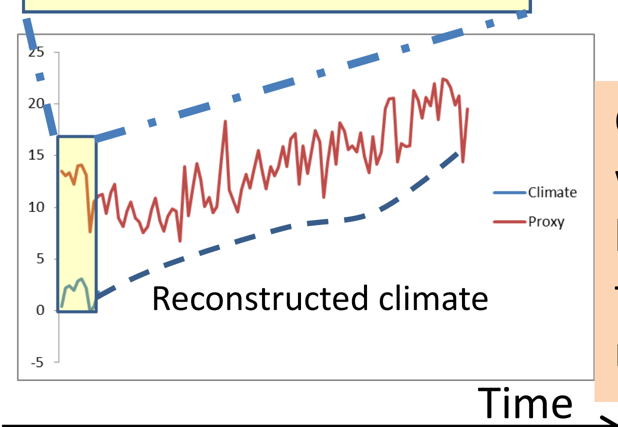
Oak tree GISP ice Sediment Pollen

Calibration data

- Past ~ 150 years
  - Overlapping time series

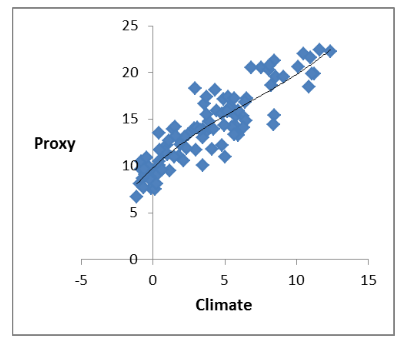


Proxy and known recent climate at different times at *that* location

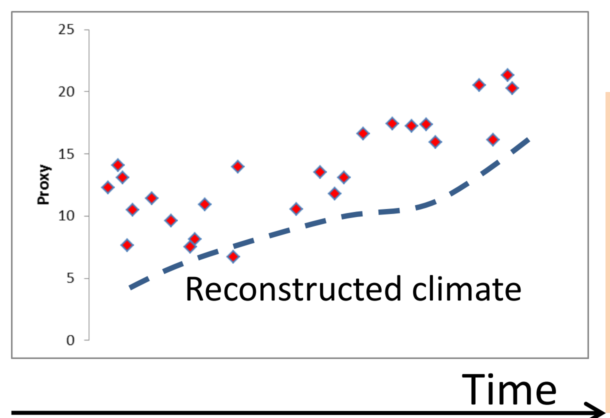


Common when proxy has good temporal resolution

- Modern analogue
  - Space for time



Proxy and known *modern* climate at *many* locations across space

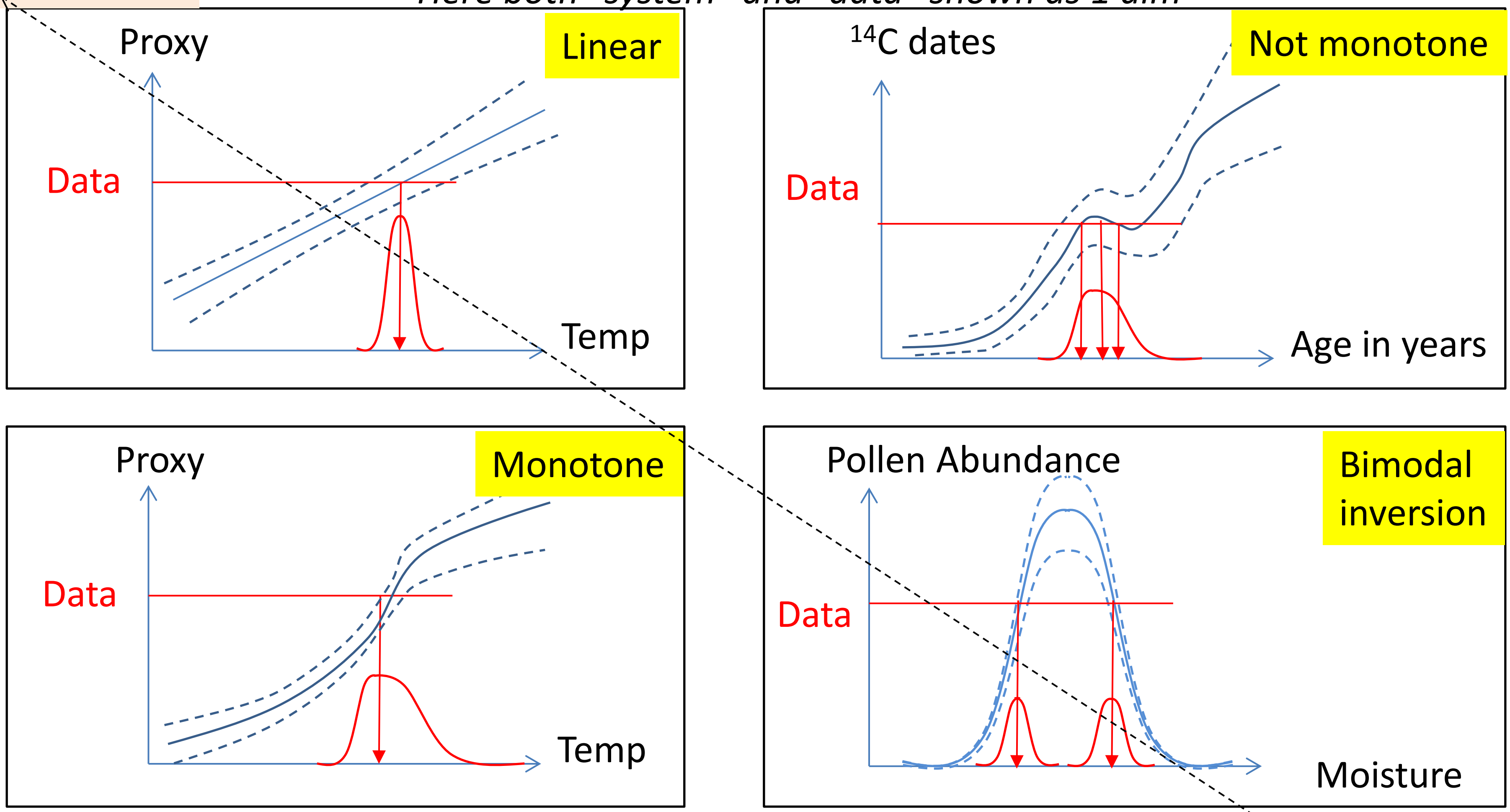


Common when proxy has poor temporal resolution

Inverting Simple Proxy Relationships

Easy proxies

Here both "system" and "data" shown as 1 dim



Hard proxies

Forward: Climate → Proxy

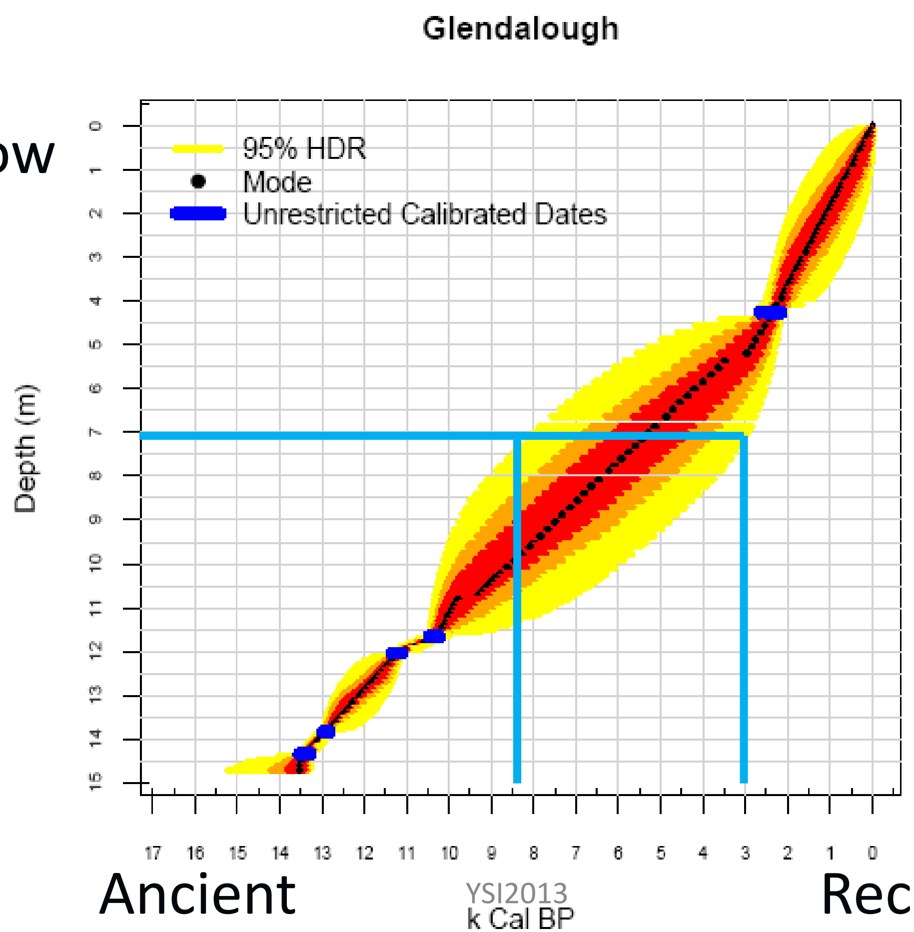
Inverse: Proxy → Climate

Temporal Uncertainty: Depth ⇒Age?

Glendalough: 150 depths; 5 ^{14}C dates

Shallow

Deep



Ancient

Recent

Random

Chronologies

Superimposed piece wise linear monotone functions

$d(t)$

…

consistent with ^{14}C data

Working with uncertain times

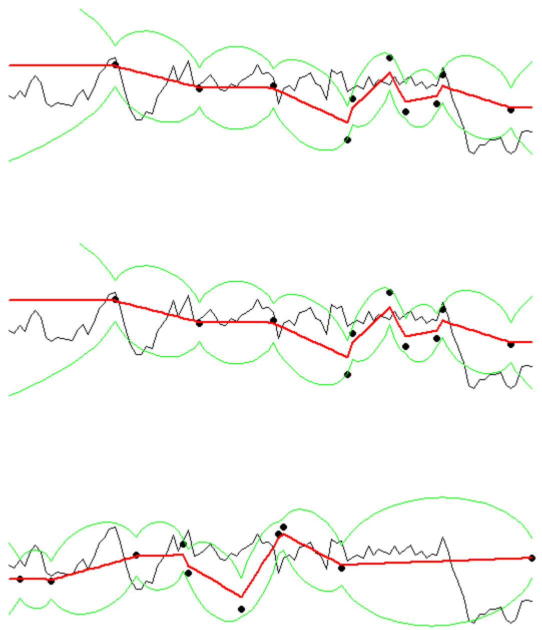
Proxy values known $y=\{y_1, y_2, y_3, \ldots\}$

Depths known: $d=\{d_1, d_2, d_3, \ldots\}$

Chronology unknown $t= \{t(d_1), t(d_2), t(d_3) \ldots \}$

Seek distribution of $c= \{c(t_1), c(t_2), c(t_3) \ldots \}$

Same *y* values, different times



Algorithm

1. Sample chronology from $p(t \mid d)$
(*depth order constrained*)
2. Compute $p(c \mid y, t)$
3. Repeat

and then Combine… $p(c \mid y, d)$

Combining Uncertainties

- (proxy, depth) pairs $\Rightarrow$ random (climate, age)
- Ensemble of (climate, age) pairs $\Rightarrow$
  random history

- Sample histories $c(t)$ for each set of times
  - Combine random histories
  - *Monte Carlo marginalising*

Reconstructing Palaeoclimate

- Data:
  - (climate proxy, age proxy) pairs
- What was the most likely palaeoclimate?
  - What are the associated uncertainties?

- Conceptual algorithm
  - Find palaeoclimate histories that are consistent with the data
  - Summarise

Palaeoclimate Reconstruction

What: Examples

How: Uncertainty as Stat Inference on Latent Climate

Challenges in Communicating Uncertainty

Communicating Uncertainty *scientist to scientist*

Scientist 1 Chronology expert

Generates random chronologies

Publishes random chronologies

(or algorithm)

Scientist 2 Pollen expert

Downloads random chronologies

Reconstructs random climate for each

Publishes random climates

(or algorithm)

Marginal and Joint Reconstruction
given single core

- Marginal Data: (less info)
  - (climate proxy, age proxy) pair, each slice
  - What was the most likely palaeoclimate?
  - What are the associated uncertainties?

For each slice?

- Joint Data: (more info)
  - (climate proxy, age proxy) pair, all slices
  - What was the most likely palaeoclimate?
  - What are the associated uncertainties?

For all slices taken jointly?

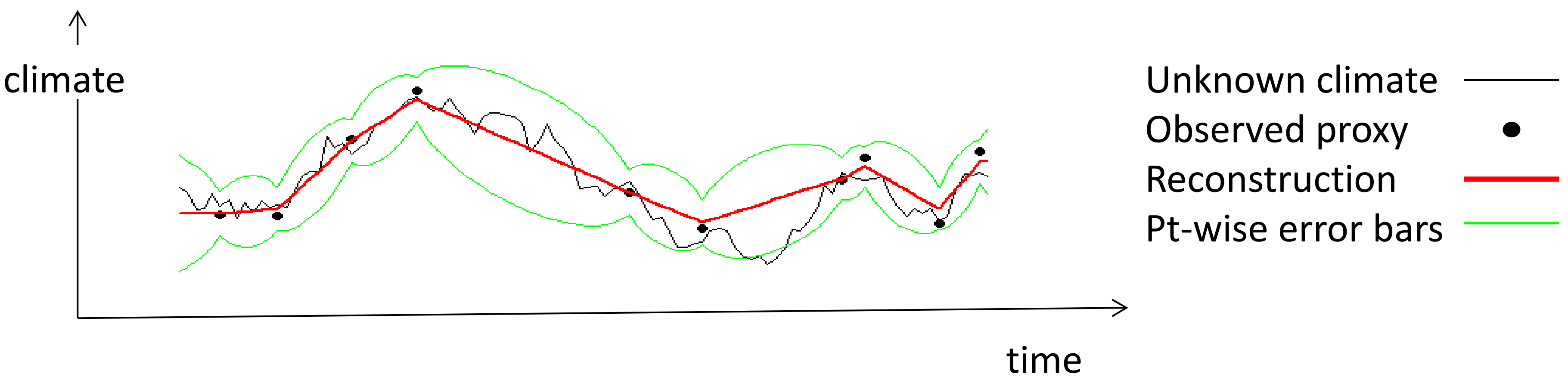
Palaeoclimate Reconstruction
given multiple cores, multiple proxies

- Data:
  - (climate proxy, age proxy) pairs
    - Multiple cores, proxies, slices
  - What was the most likely palaeoclimate?
    - Aspects of climate
    - Space-time histories
  - Associated uncertainties
    - Define, compute, present

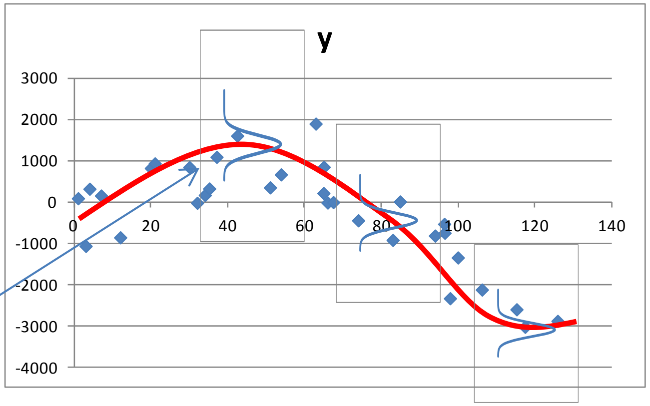
Statistical Methods

- Partially observed stochastic processes

Simple case: single core: one location
underlying latent c(t) univariate Gaussian
slices correspond to known irregular times
proxy process $\Rightarrow$ additive univariate Gaussian errors
with simple variance structure
'Reconstruct' $\Rightarrow$ make statistical inference



Statistics by Ski Slalom



- Marginal data ⇒gates
  - Different "widths"
- Physics defines the smoothness
- *Challenge*
  - 'Location' of gates uncertain

Collect sample paths *only* for those that successfully pass thru *all* the gates?

Joint inference

Inference by rejection sampling

- Model $y = f(x, z, \theta)$ statistical/physical
  - Observe data y, covariates x
  - Infer latent processes z, fixed parameters θ

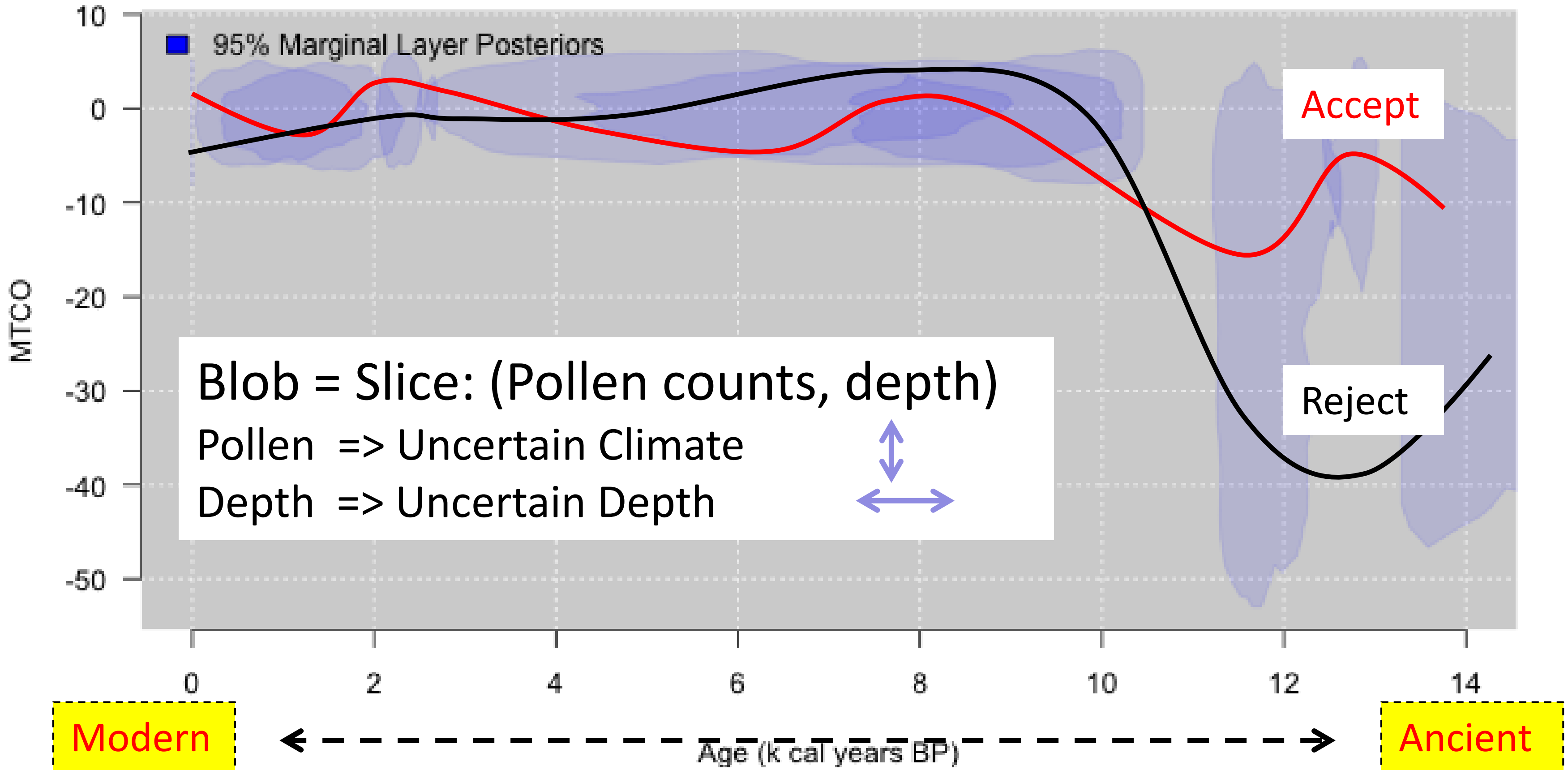
- Conceptual algorithm
  - Propose z, θ
  - Statistically consistent with y, x?
  - Yes No

  $\Rightarrow$Reject with low prob $\Rightarrow$Reject with high prob

Algorithmically inefficient

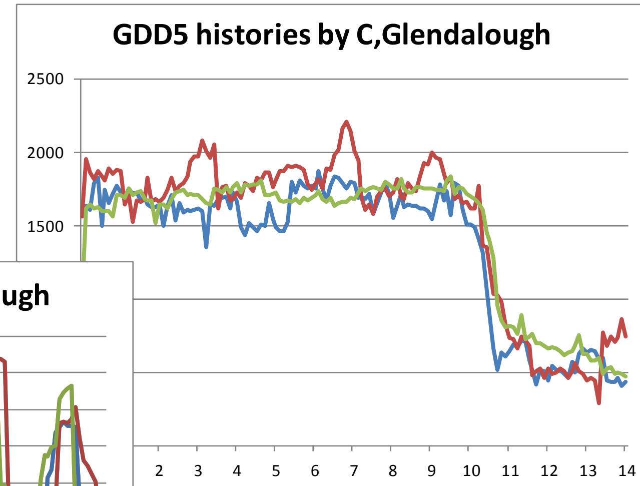
Data as gates with uncertain time

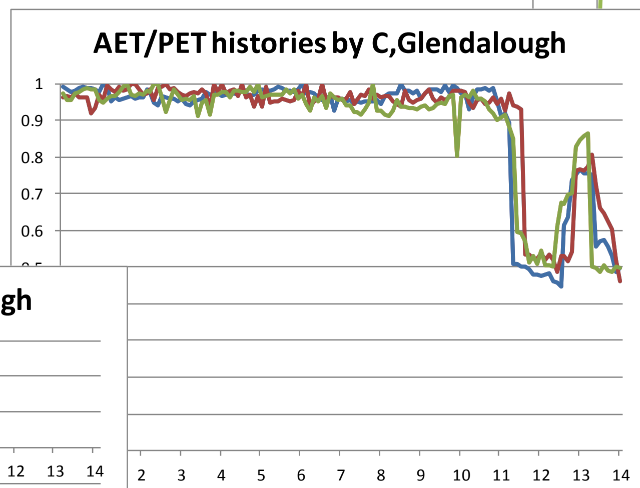
Winter harshness

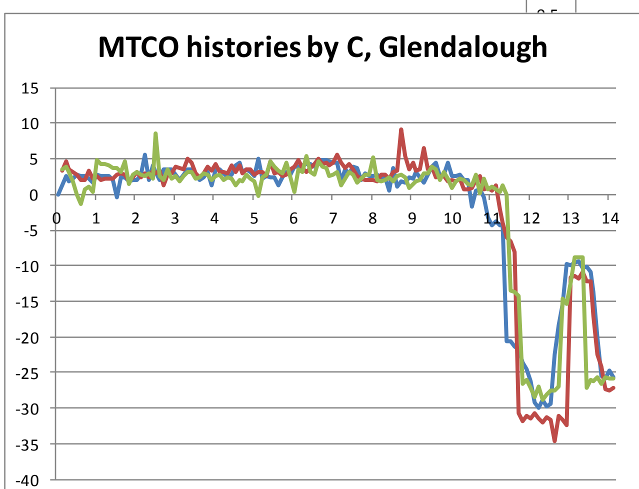


Sampled histories at one location

Three aspects of climate
One three-dim history

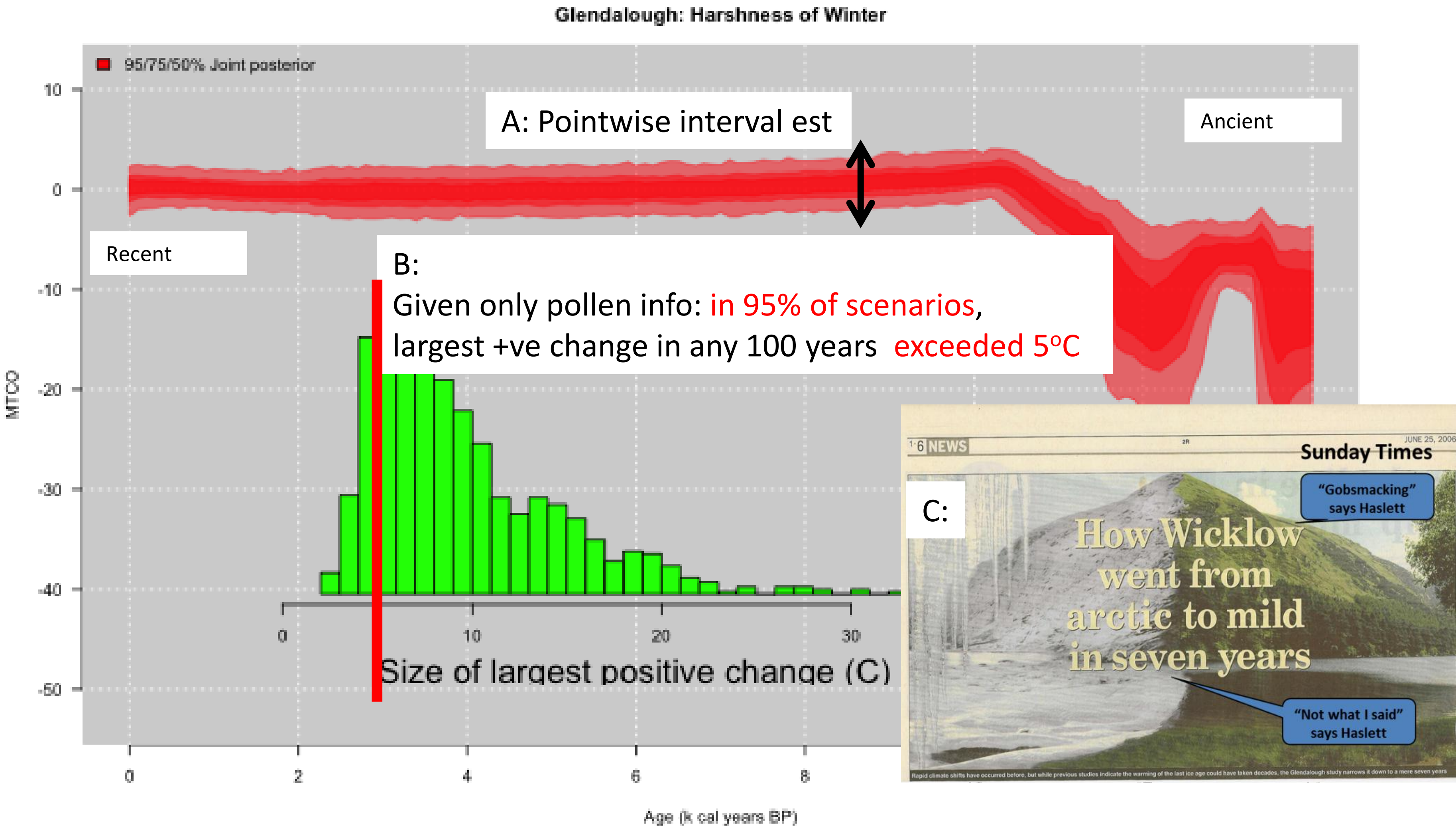






One site in Ireland

Communicating Uncertainty



Palaeoclimate Reconstruction

Examples

Statistical Methods:

Uncertainty in Stat Inference on Climate

Challenges in Communicating Uncertainty

Uncertainty Challenges

- Communicating uncertainty
  - Scientist to scientist
  - Scientist to public
- Modularising uncertainty

Modularising Uncertainty

- Scientist 1 Pollen expert
  - publishes (samples from) $p^{pollen}(c_i \mid y_i)$ as $\{ c_i^{pollen} \}$
- Scientist 2 Tree Ring /Dendro expert
  - publishes (samples from) $p^{dendro}(c_i \mid y_i)$ as $\{ c_i^{dendro} \}$
- Scientist 3 SpaceTime Series expert
  - accepts (samples from dists) $\{ c_i^{pollen} \}$ $\{ c_i^{dendro} \}$
  - accepts random chonologies $\{t_i\}$
  - rejects ensembles of pairs $\{ c_i^{pollen}, c_i^{dendro} \}$

    not jointly consistent -with each other

    -with 'smoothness' – in space

    – in time
  - publishes jointly *acceptable* space/time *histories* $\{ c^{pollen}, c^{dendro} \}$

Modularising Uncertainty

Not (quite?) regular Bayesian paradigm

Regular Bayes

Posteriors based on partial data

Prior for $\left[c(t,s)\,|\,\theta_c\right]$

Cond Indep proxy processes

Likeli-hoods

	Likelihoods	Posteriors based on partial data
depth/time	$\left[d_i\,|\,t_i,\theta_{chron}\right]$	$\left[t_i\,|\,d_i\right]$
pollen	$\left[y_i^{pollen}\,|\,c(t_i,s_j),\theta_{pollen}\right]$	$\left[c(t_i,s_j)\,|\,y_i^{pollen}\right]$
dendro	$\left[y_i^{dendro}\,|\,c(t_i,s_j),\theta_{dendro}\right]$	$\left[c(t_i,s_j)\,|\,y_i^{dendro}\,|\right]$

$$\Rightarrow\left[\left\{c(t_i,s_j)\right\},\theta\,|\,\left\{d_i,y_i^{pollen},y_i^{dendro}\right\}\right]$$

$$\Rightarrow\left[c\,|\,\left\{d_i,y_i^{pollen},y_i^{dendro}\right\}\right]$$

Uncertainty in palaeoclimate reconstruction

- Science
  - Climate multivariate (not just temperature)
  - Driven by physics, forcing (and computer models)
  - Multiple processes
    - Including multi-modal reconstruction
    - Uncertain times
- Challenges in Uncertainty
  - Modularising
  - Communicating

Conceptual
Algorithmic

Challenges

- Communicating uncertainty
  - Scientist to scientist
  - Scientist to public
- Modularising uncertainty
- Fast algorithms
  - Avoid MCMC (as much as possible)
- Inference for computer model parameters

Monte Carlo methods

Resources

Past 1000 yrs http://en.wikipedia.org/wiki/Temperature_record_of_the_past_1000_years

Past ~11000 years, Holocene http://www.ipcc.ch/ipccreports/tar/wg1/073.htm

Past 400ka http://en.wikipedia.org/wiki/Ice_core#Ice_core_data

Abrupt Change http://www.ncdc.noaa.gov/paleo/abrupt/index.html

Haslett, John, et al. "Bayesian palaeoclimate reconstruction." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 169.3 (2006): 395-438.

Li, Bo, Douglas W. Nychka, and Caspar M. Ammann. "The value of multiproxy reconstruction of past climate." *Journal of the American Statistical Association* 105.491 (2010).

Tingley, Martin P., et al. "Piecing together the past: Statistical insights into paleoclimatic reconstructions." *Quaternary Science Reviews* 35 (2012): 1-22.

See also IPS006

General

- The long summer : how climate changed civilization Fagan, Brian (2004), Granta
- Collapse : how societies choose to fail or survive, Diamond, Jared (2006) Penguin
- Was Agriculture Impossible during the Pleistocene but Mandatory during the Holocene? A Climate Change Hypothesis Richerson, Boyd, Bettinger *American Antiquity*, 2001
- Climate stability and the development of agricultural societies Feynman, Ruzmaikin, *Climatic Change* (2007)