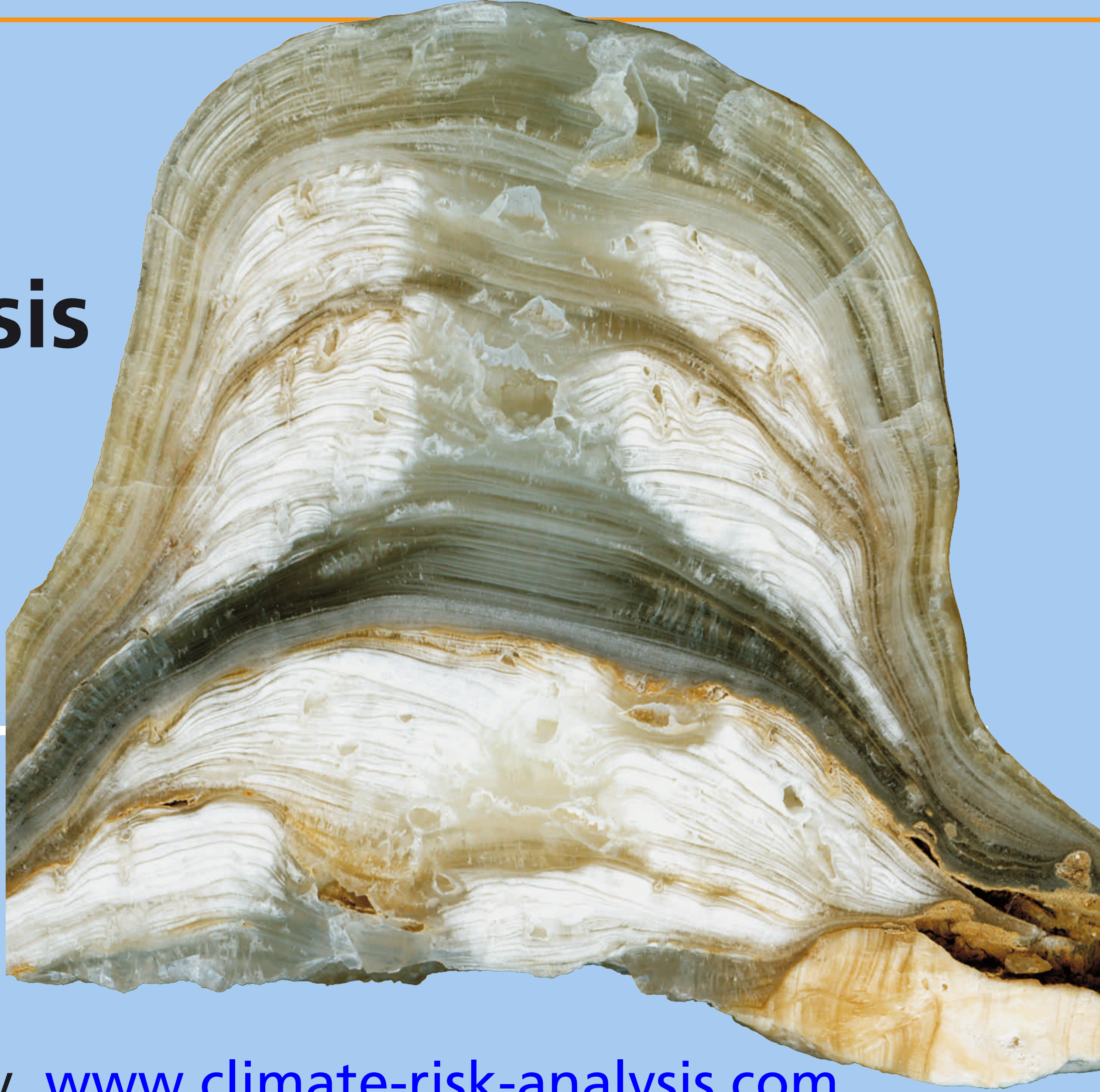
Paleoclimate Time Series Analysis



Manfred Mudelsee

Climate Risk Analysis, Germany, www.climate-risk-analysis.com
Alfred Wegener Institute, Germany

TIES, Hyderabad, 5 January 2012

Courtesy Dominik Fleitmann, Oeschger Centre, Switzerland

Part 1: **Paleoclimate**

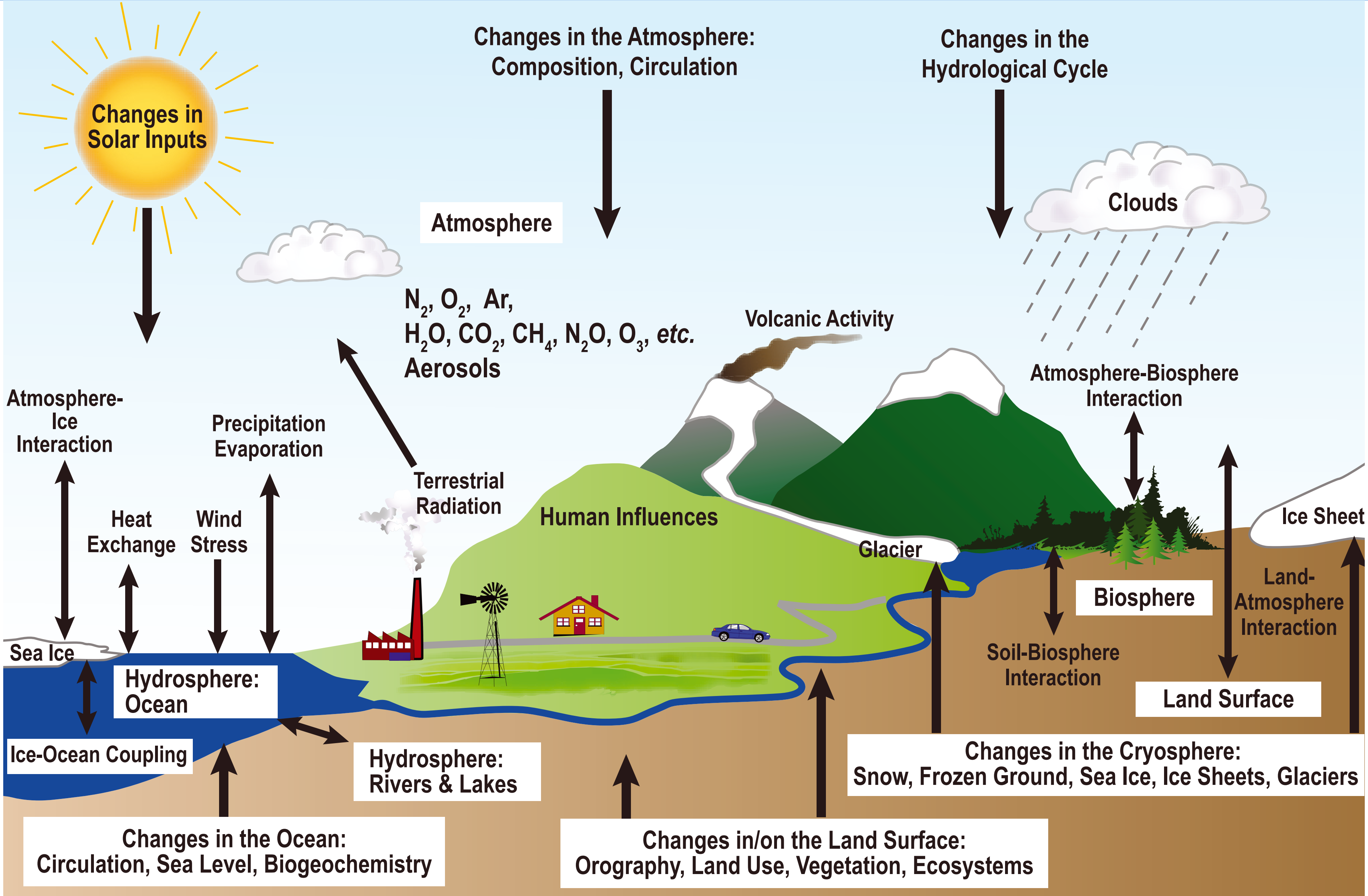
Part 2: **Time Series Analysis**

Part 3: **Examples**

Mathematical formulas: no sweat!

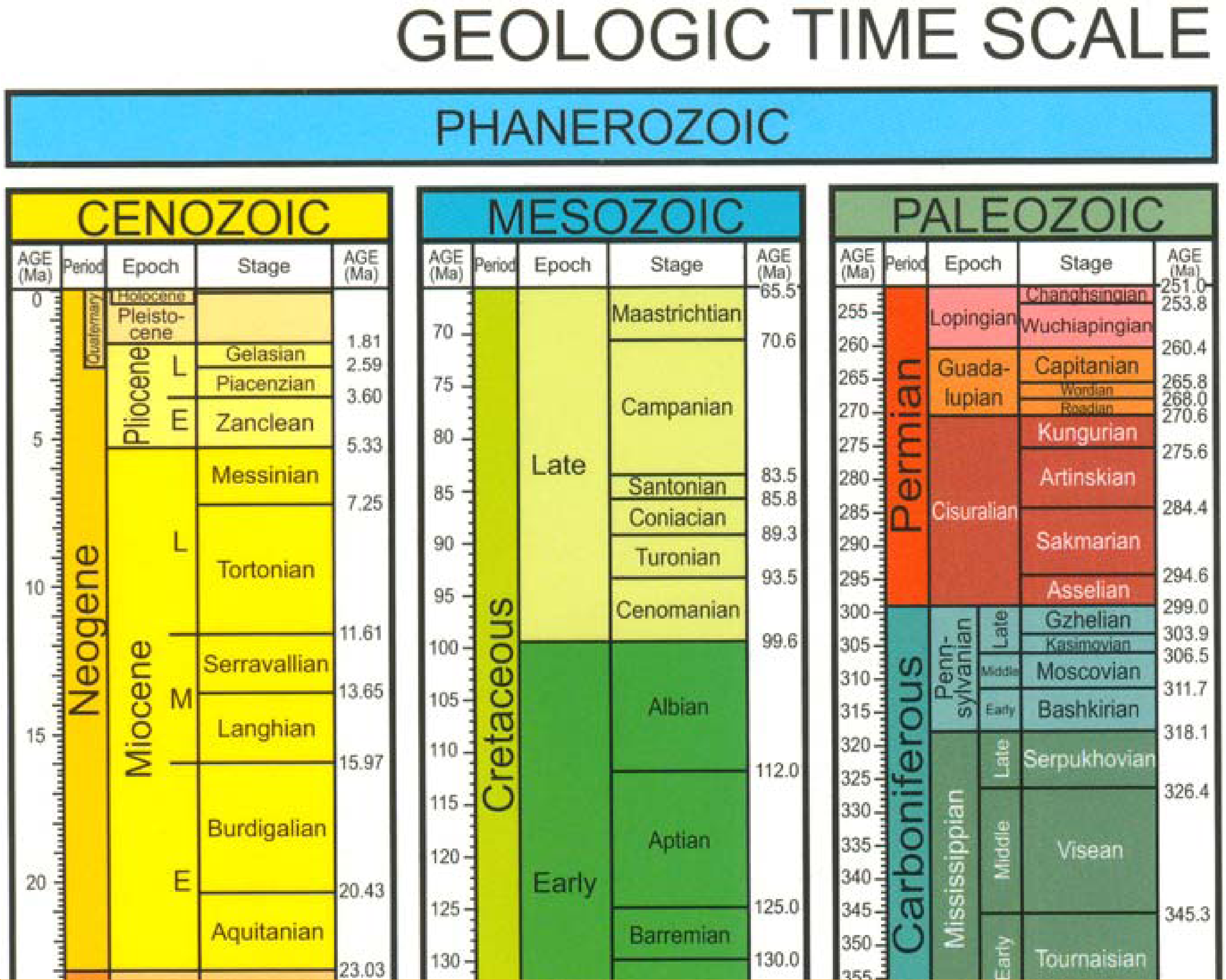
Physics, Geology: no sweat!

Climate



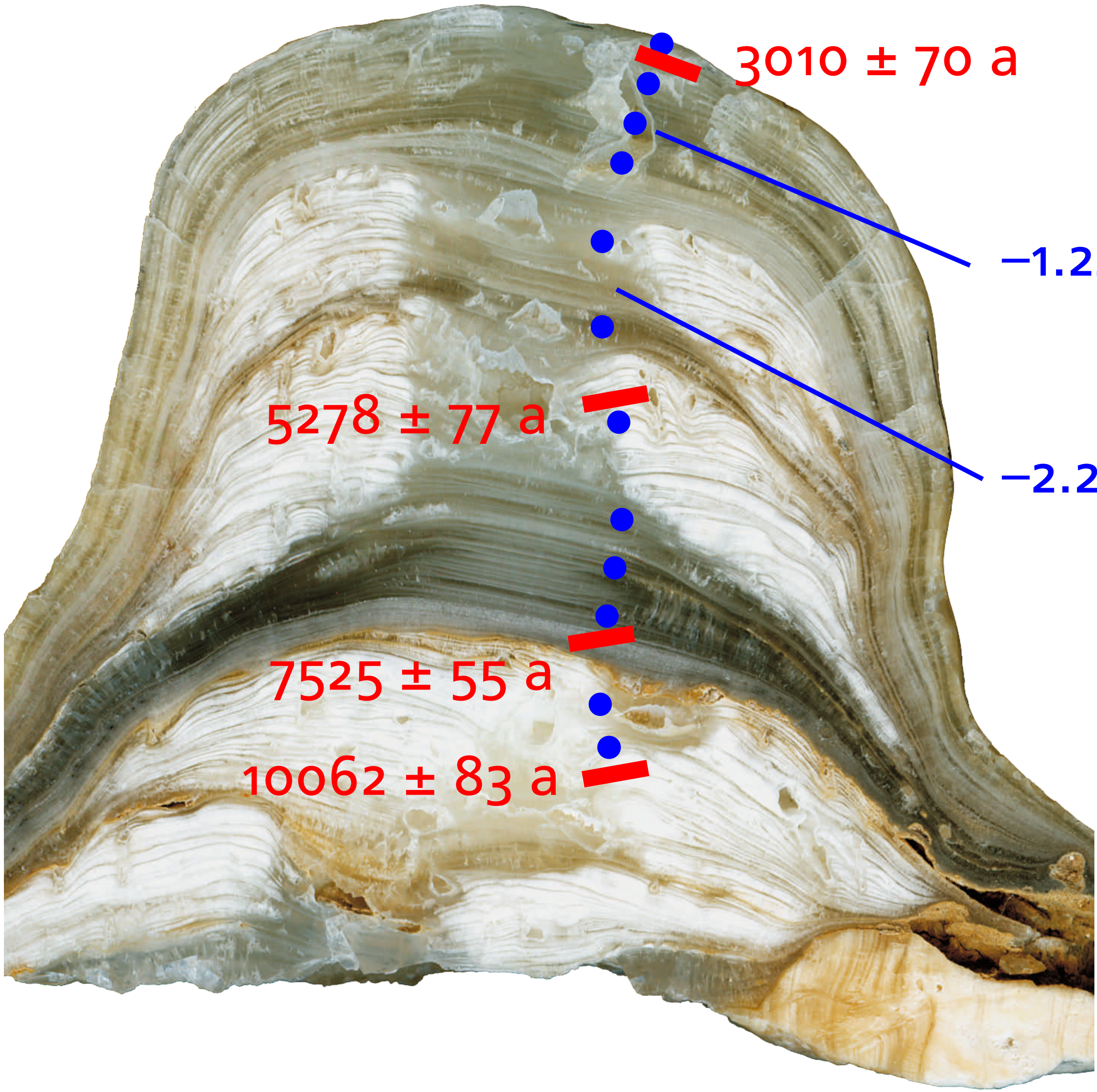
Solomon et al. (2007) *Climate Change 2007: The Physical Science Basis.* Cambridge Univ. Press

Paleo



Gradstein et al. (2004) *A Geological Time Scale 2004.* Cambridge Univ. Press

Paleoclimate archive



10 cm

Paleoclimate time series

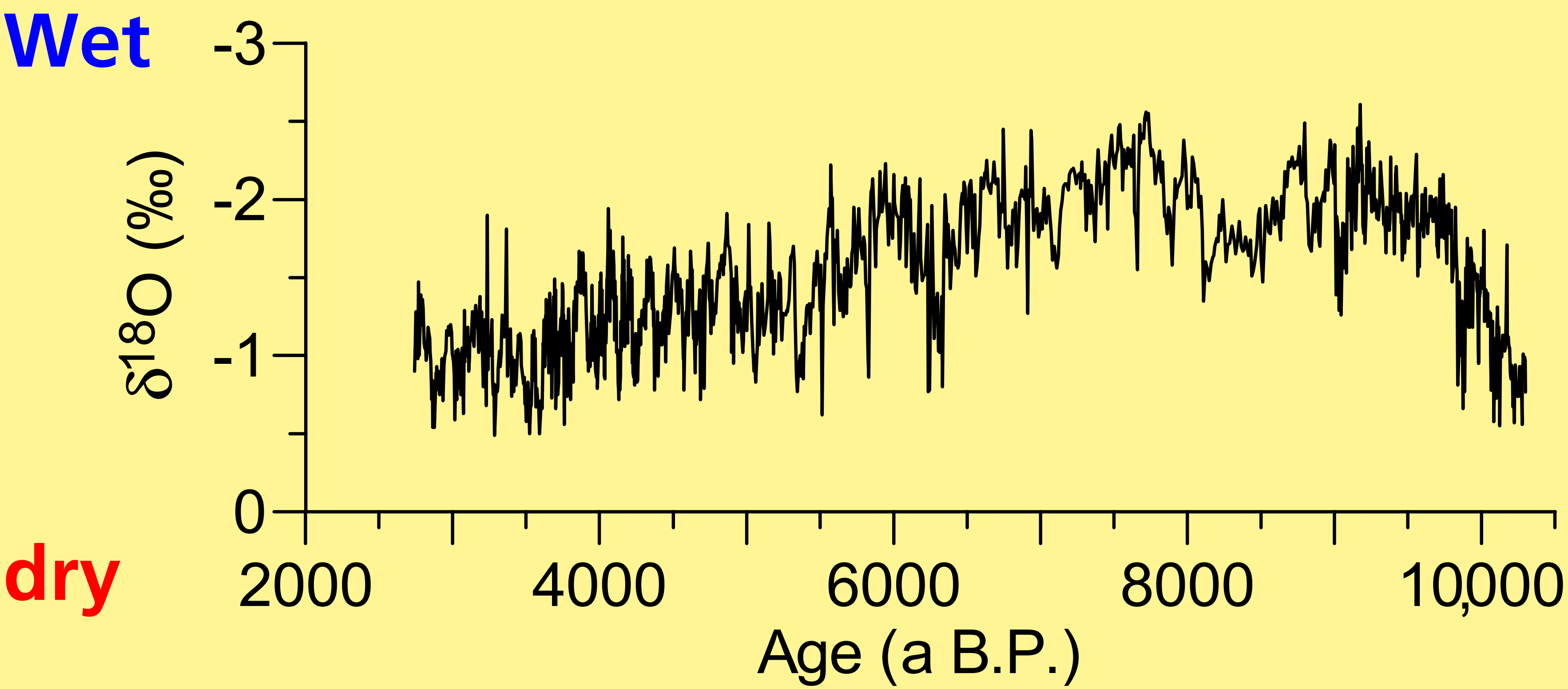
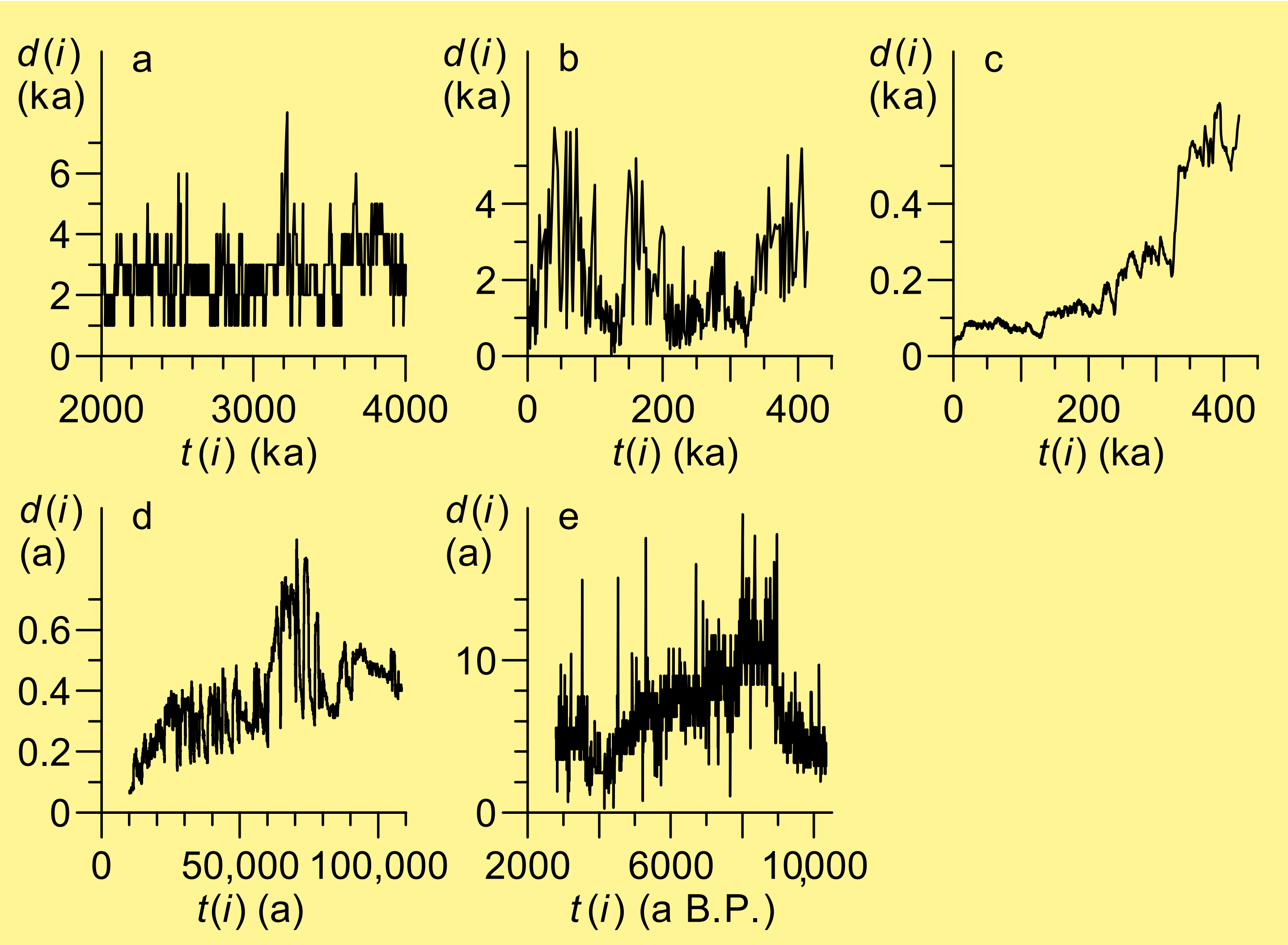


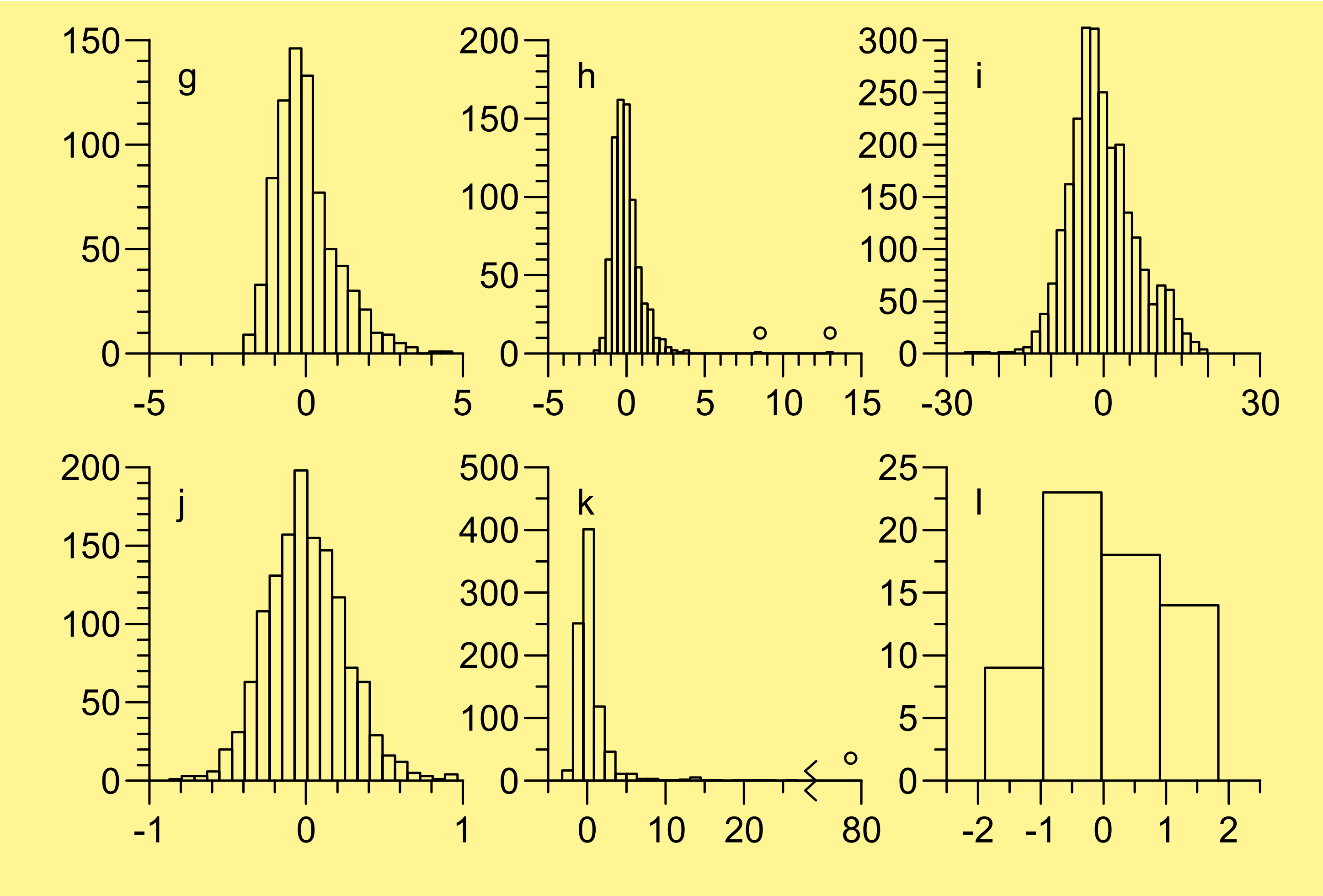
Figure 1.7. Speleothem data: oxygen isotope record from stalagmite Q5 from southern Oman over the past 10,300 years. Along the growth axis of the nearly 1 m long sta-

Spacing and uncertain timescale



Book, Fig. 1.15 (**e,** stalagmite Q5 $\delta^{18}O$)

Noise and statistical distribution

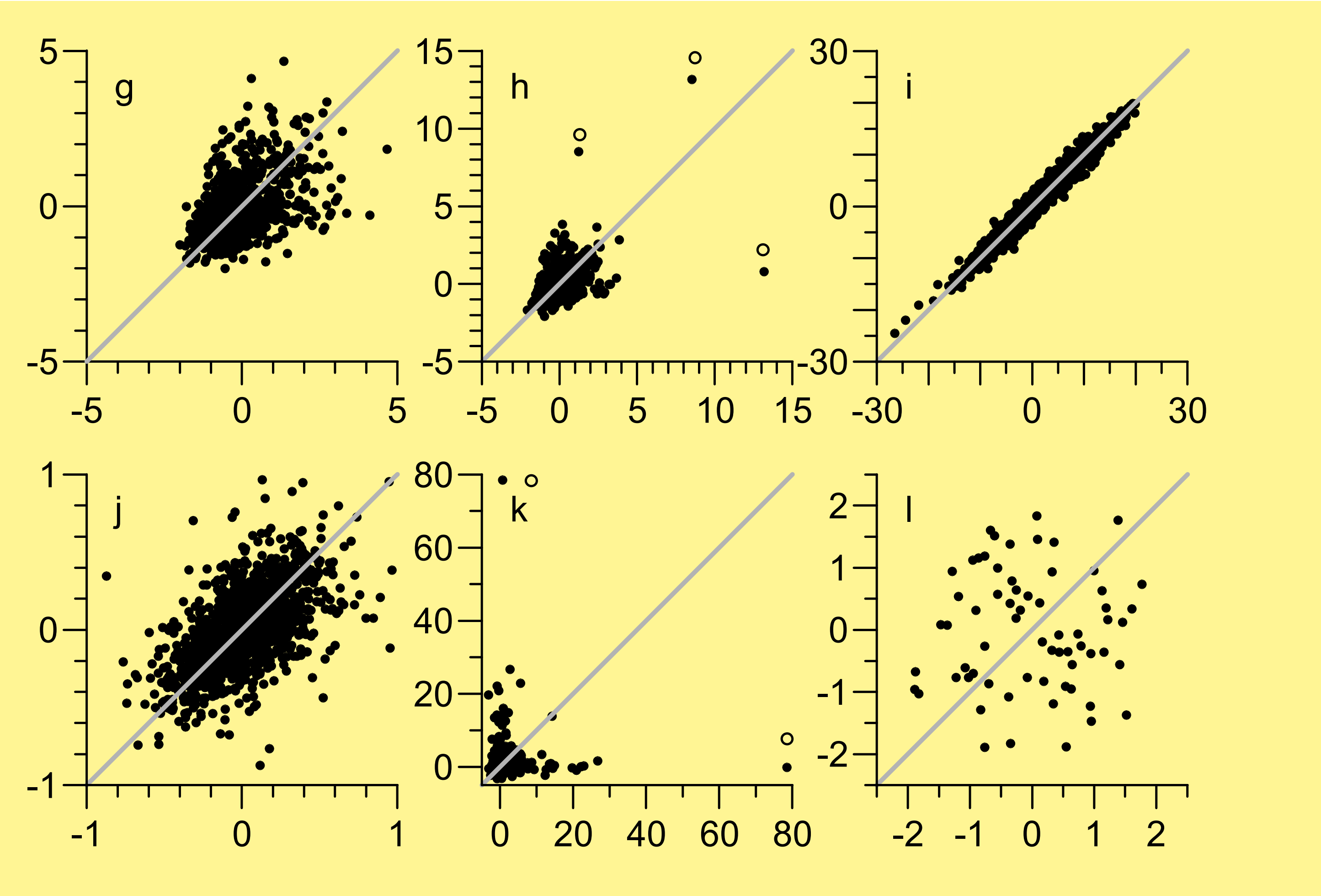


Histogram: inference of PDF

Residuals, "detrended version" of the $x(i)$

Book, Fig. 1.11 (**j,** stalagmite Q5 $\delta^{18}O$)

Persistence



Lag-1 scatterplot $x(i-1)$ vs $x(i)$: inference about persistence

Residuals, "detrended version" of the $x(i)$

Book, Fig. 1.12 (**j**, stalagmite Q5 $\delta^{18}O$)

Climate equation

$$X(i) = X_{\text{trend}}(i) + X_{\text{out}}(i) + S(i) \cdot X_{\text{noise}}(i), \qquad (1.2)$$

Climate equation, discrete time

$T(i)$, time points

$X(i) \equiv X(T(i))$, etc.

Climate is a paradigm of a complex system. Analysing climate data is an exciting challenge, which is increased by non-normal distributional shape, serial dependence, uneven spacing and timescale uncertainties. This book presents bootstrap resampling as a computing-intensive method able to meet the challenge. It shows the bootstrap to perform reliably in the most important statistical estimation techniques: regression, spectral analysis, extreme values and correlation.

This book is written for climatologists and applied statisticians. It explains step by step the bootstrap algorithms (including novel adaptions) and methods for confidence interval construction. It tests the accuracy of the algorithms by means of Monte Carlo experiments. It analyses a large array of climate time series, giving a detailed account on the data and the associated climatological questions.

Climate Time Series Analysis

Classical Statistical and Bootstrap Methods

Manfred Mudelsee

Manfred Mudelsee received his diploma in Physics from the University of Heidelberg and his doctoral degree in Geology from the University of Kiel. He was then a postdoctoral researcher in Statistics at the University of Kent at Canterbury, research scientist in Meteorology at the University of Leipzig and visiting scholar in Earth Sciences at Boston University; currently he does climate research at the Alfred Wegener Institute for Polar and Marine Research, Bremerhaven. His science focuses on climate extremes, time series analysis and mathematical simulation methods. He has authored over 50 peer-reviewed articles. In his 2003 Nature paper, Mudelsee introduced the bootstrap method to flood risk analysis. In 2005, he founded the company Climate Risk Analysis.

GEOSCIENCES

ISBN 978-90-481-9481-0

9 789048 194810

› springer.com

Mudelsee

Climate Time Series Analysis

Climate Time Series Analysis

Classical Statistical and Bootstrap Methods

Atmospheric and Oceanographic Sciences Library

42

Manfred Mudelsee

Springer

11

Estimation

$$X(i) = X_{\text{trend}}(i) + X_{\text{out}}(i) + S(i) \cdot X_{\text{noise}}(i), \qquad (1.2)$$

Climate equation, stochastic process *X*(*i*)

Time series $\{t(i), x(i)\}_{i=1}^{n}$

Time series analysis uses sample to learn about process: trend parameters, probability of extremes, cycles, etc.

Note relation between data and question(s).

First-order autoregressive model, uneven spacing

$$
\begin{aligned}
X_{\text{noise}}(1) &= \mathcal{E}_{\mathrm{N}(0,\,1)}(1), \\
X_{\text{noise}}(i) &= \exp\left\{-\left[T(i) - T(i-1)\right]/\tau\right\} \cdot X_{\text{noise}}(i-1) \\
&\quad + \mathcal{E}_{\mathrm{N}(0,\,1-\exp\{-2[T(i)-T(i-1)]/\tau\})}(i), \qquad i = 2, \ldots, n.
\end{aligned}
\tag{2.9}
$$

AR(1) process, uneven spacing

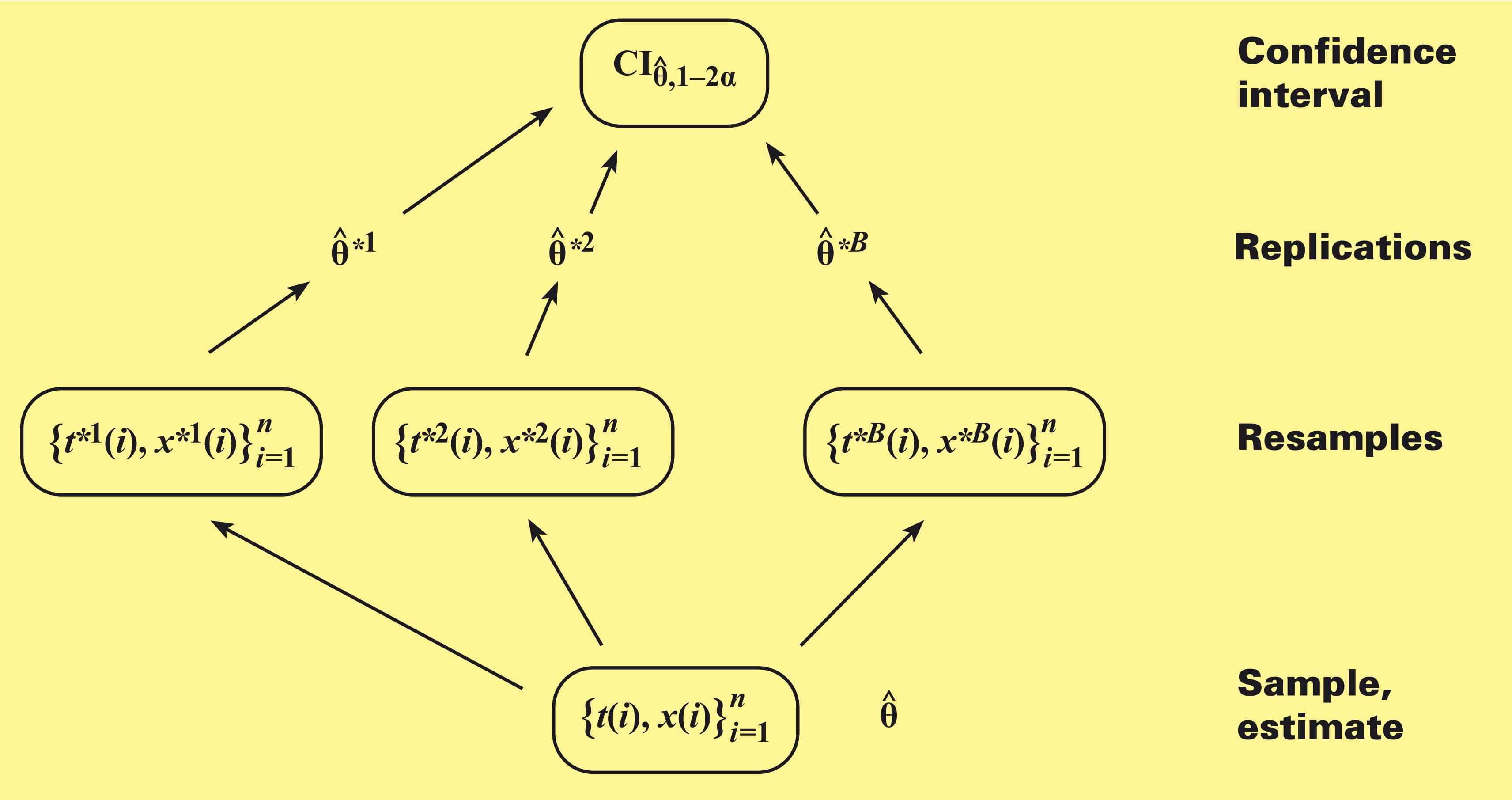
$$
S(\widetilde{\tau}) = \sum_{i=2}^{n} \left[x_{\text{noise}}(i) - \exp\left\{-\left[t(i) - t(i-1)\right]/\widetilde{\tau}\right\} \cdot x_{\text{noise}}(i-1)\right]^2,
$$

Least-squares estimation

Residuals $x_{\text{noise}}(i)$, "detrended version" of the $x(i)$

Book, Eq. (2.11); Robinson (1977) *Stochastic Processes and their Applications* 6:9

Bootstrap principle

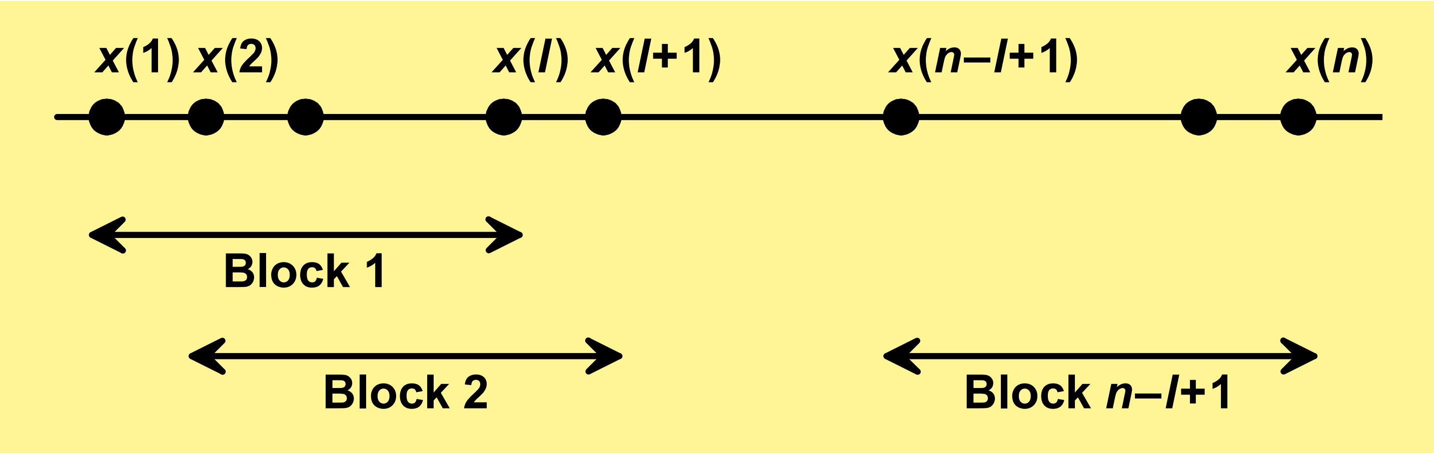


Book, Fig. 3.3

Bootstrap principle

- Bootstrap
  - Resampling
    - Moving block bootstrap (MBB)
    - Autoregressive bootstrap (ARB)
    - Other
  - Confidence interval construction
    - Normal
    - Student's *t*
    - Percentile
    - BCa

Bootstrap resampling



Moving block bootstrap (MBB)

Bootstrap resampling

$$l_{\mathrm{opt}} = NINT\left\{\left[6^{1/2} \cdot \widehat{a} \Big/ \left(1 - \widehat{a}^2\right)\right]^{2/3} \cdot n^{1/3}\right\}, \qquad (3.28)$$

Block length selector (MBB) after Carlstein (1986)

Other block length selectors via

- persistence time
- autocorrelation function

Bootstrap resampling

Step 1	Data	$\{t(i), x(i)\}_{i=1}^{n}$
Step 2	Resampled times unchanged	$\{t^*(i)\}_{i=1}^{n} = \{t(i)\}_{i=1}^{n}$
Step 3	Residuals (Eq. 3.29)	$r(i) = \left[x(i) - \widehat{x}_{\text{trend}}(i) - \widehat{x}_{\text{out}}(i)\right] \Big/ \widehat{S}(i)$
Step 4	Apply MBB (Algorithm 3.1) to residuals	$\{r(i)\}_{i=1}^{n}$
Step 5	Resampled residuals	$\{r^*(i)\}_{i=1}^{n}$
Step 6	Use resampled residuals to produce resamples	$x^*(i) = \widehat{x}_{\text{trend}}(i) + \widehat{x}_{\text{out}}(i) + \widehat{S}(i) \cdot r^*(i)$

Algorithm 3.3. MBB for realistic climate processes, which comprise trend, outlier and variability components.

Bootstrap confidence intervals

$$\widehat{\mathrm{se}}_{\widehat{\theta}^*} = \left\{ \sum_{b=1}^{B} \left[\widehat{\theta}^{*b} - \left\langle \widehat{\theta}^{*b} \right\rangle \right]^2 \Big/ (B-1) \right\}^{1/2}, \qquad (3.30)$$

Bootstrap standard error

Bootstrap confidence intervals

$$\mathrm{CI}_{\widehat{\theta},1-2\alpha} = \left[\widehat{\theta}^*(\alpha); \widehat{\theta}^*(1-\alpha)\right], \tag{3.33}$$

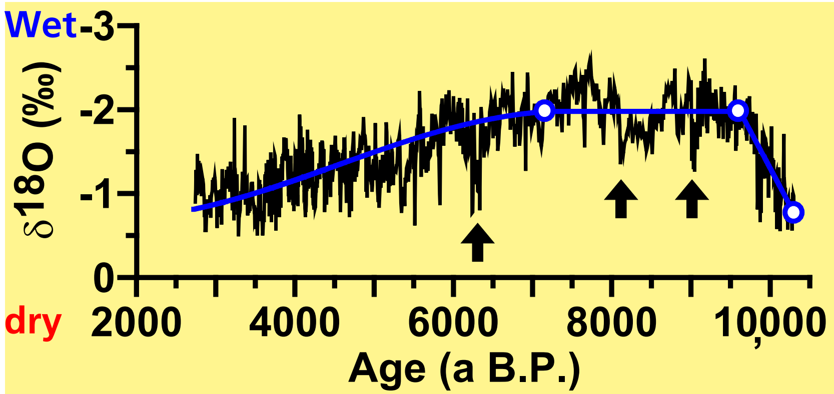
Percentile confidence interval

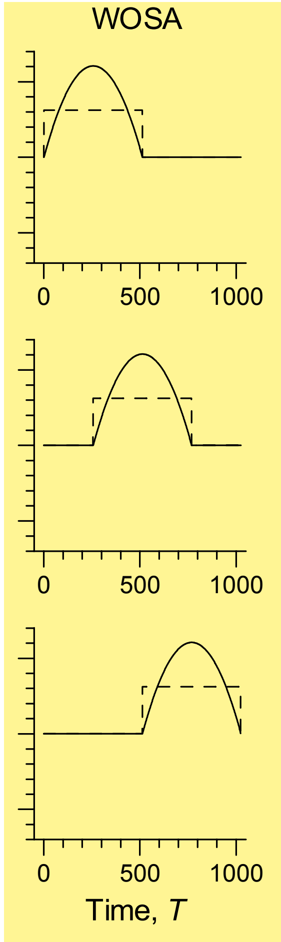
α small $\Rightarrow$ B large

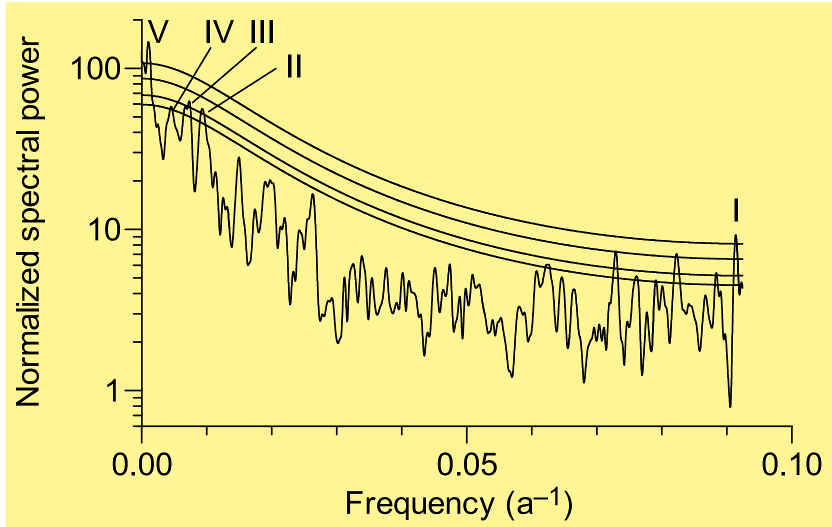
typically

$\alpha = 0.025$ (95% CI) or $\alpha = 0.05$ (90% CI), $B \approx 2000$

Spectrum estimation, stalagmite Q5, Holocene







Oxygen isotope record (Q5)

- Monsoon rainfall proxy
- Trend: ramp + sinusoid
- Interval [2.7 ka; 8.0 ka]
- $\bar{d}$ = 5.4 a, n = 973

Estimated spectrum

- Lomb–Scargle
- Oversampling factor 64
- Highest $f = 1.0\ f_{Ny}$

Welch I taper, 6 segments, 50% overlap

Book, Figs. 4.18 (modified), 5.5 (modified), 5.8

Spectrum estimation, stalagmite Q5, Holocene

Spectrum

6-dB bandwidth

$B_s \approx 0.001 \text{ a}^{-1}$

Peaks

$T_{\text{period}} = 10.9$ a (I)

$T_{\text{period}} = 107$ a (II)

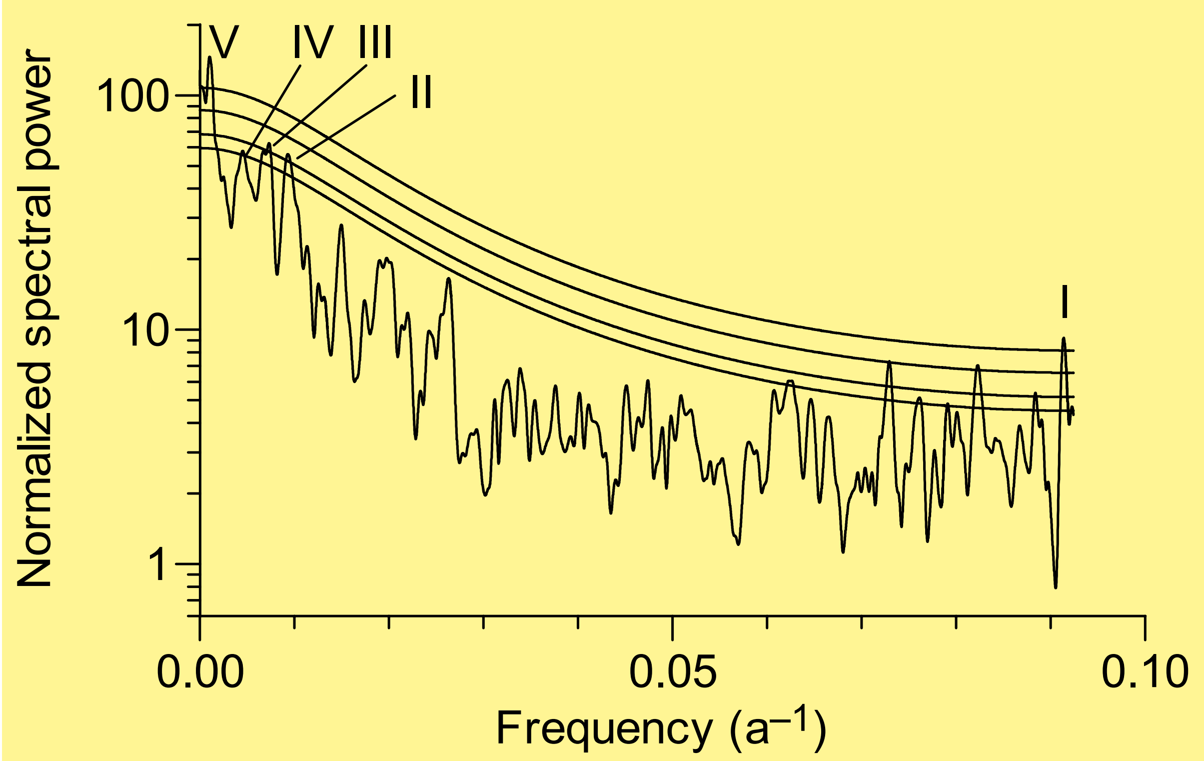
$T_{\text{period}} = 137$ a (III)

$T_{\text{period}} = 221$ a (IV)

$T_{\text{period}} = 963$ a (V)

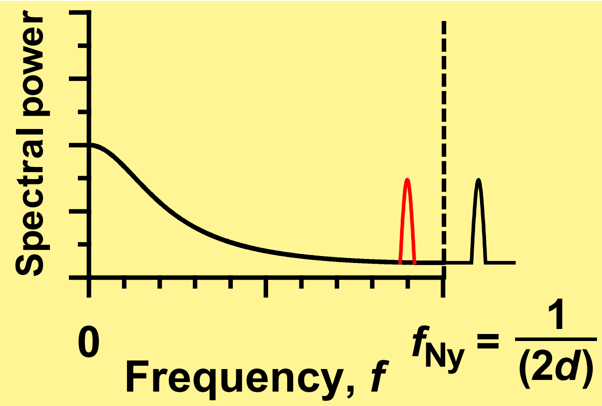
AR(1) alternatives

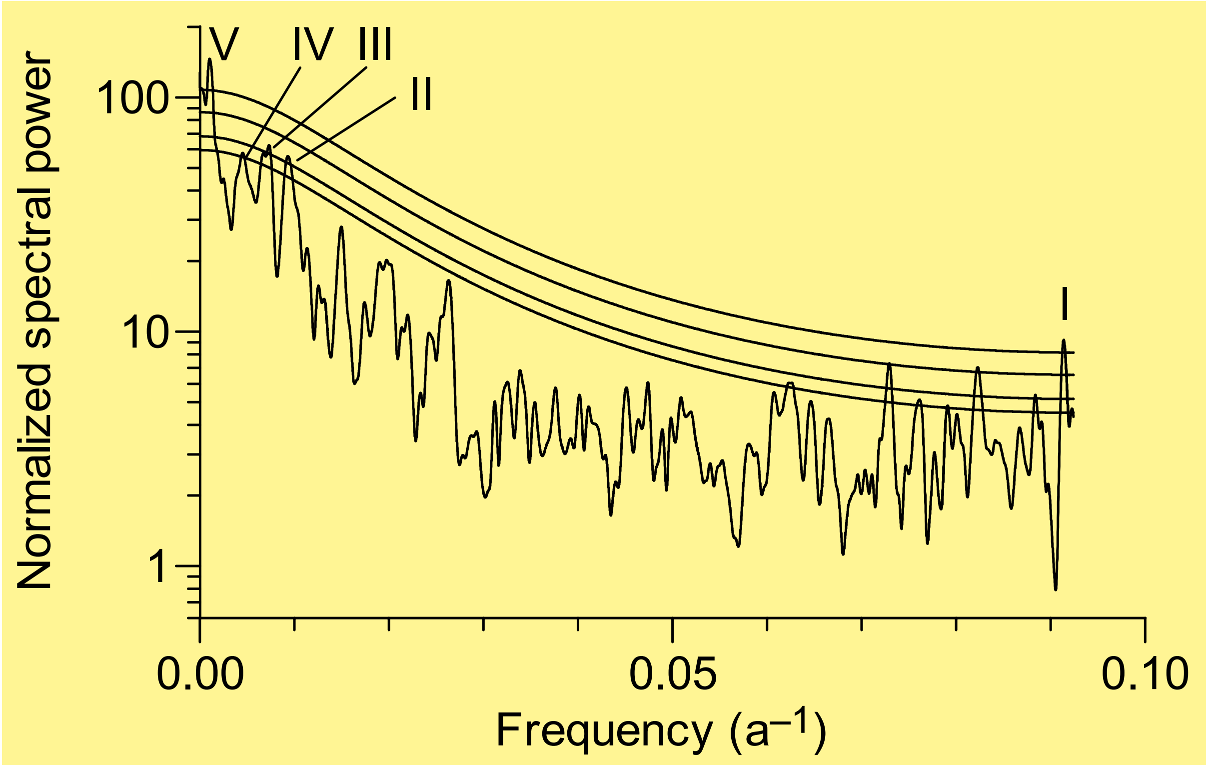
90%, 95%, 99%, 99.9%



Book, Fig. 5.8; 6-dB bandwidth: decay on flanks of a peak to $10^{-6/10} \approx 0.251 \times$ peak maximum

Spectrum estimation, stalagmite Q5, Holocene



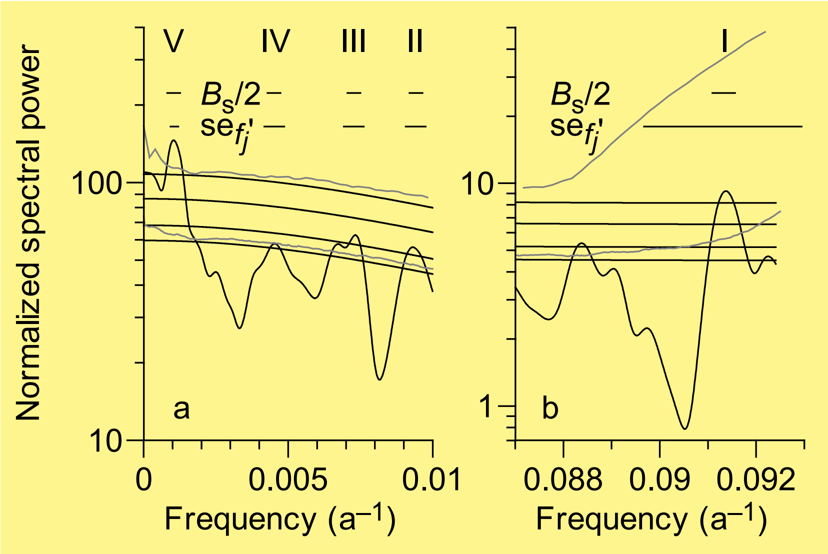


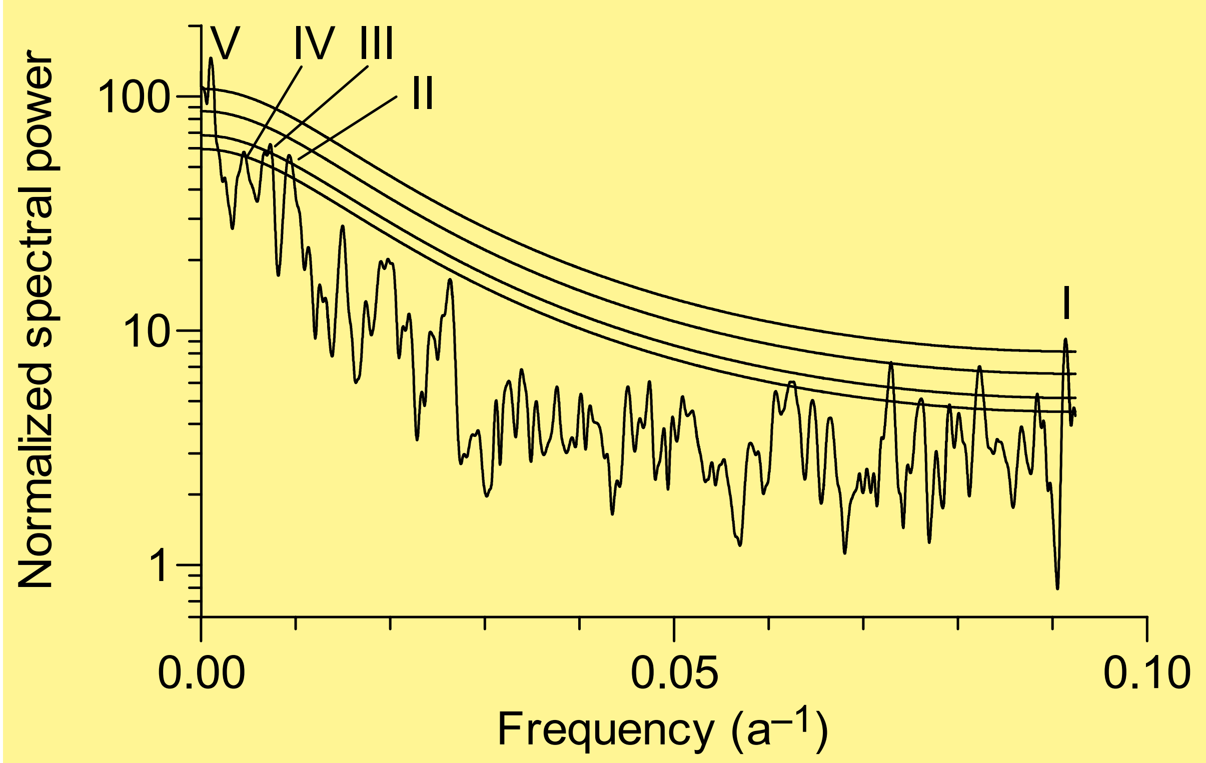
Aliasing

Likely not a problem here

- Sampling length 3.9 years
- Monsoon rainfall mostly during JAS, not full year
- Statistical test: AR(1) + annual cycle ⇒ no peaks (T_{period}: I–V)

Spectrum estimation, stalagmite Q5, Holocene





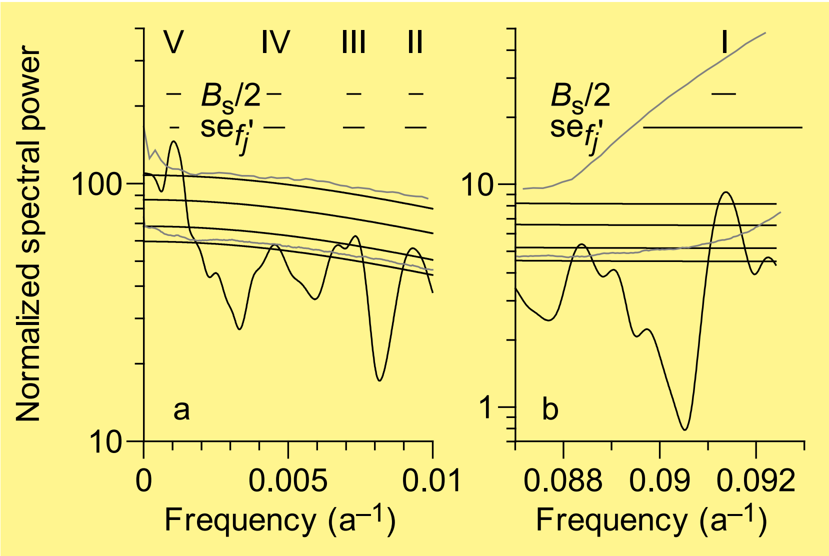
Timescale error effects?

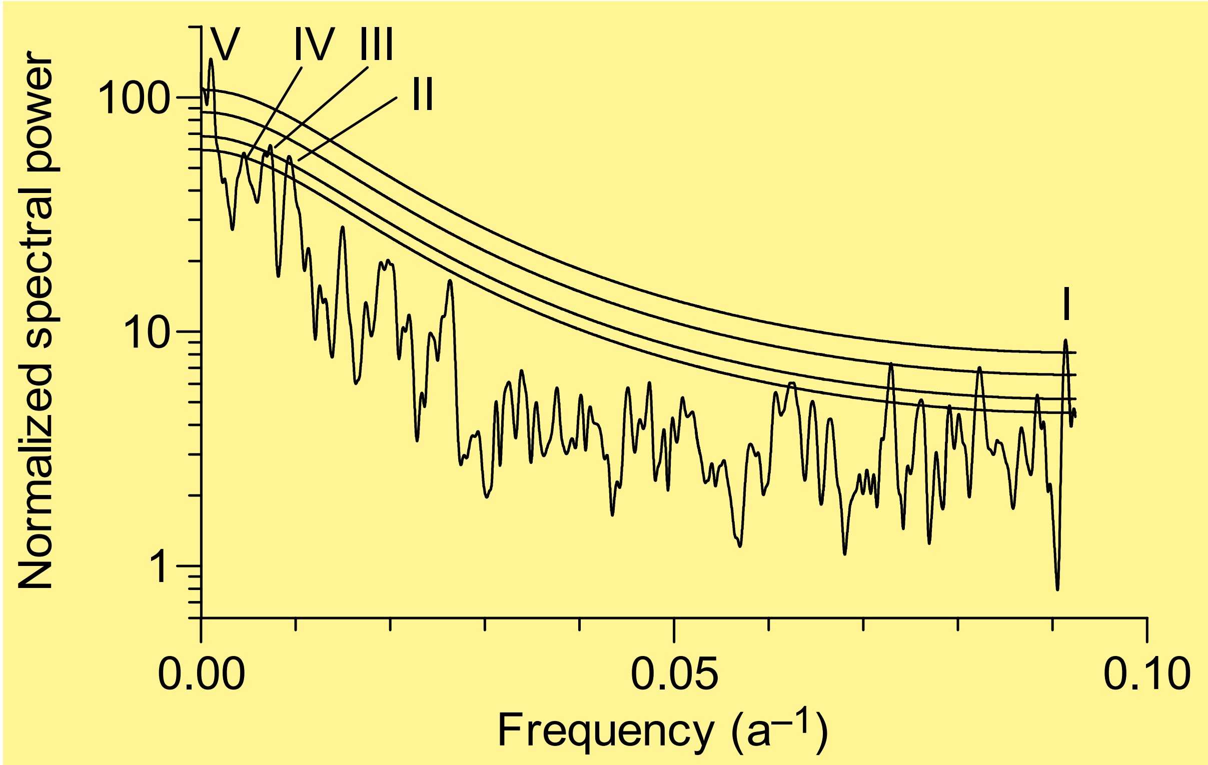
Test using $t(i)^*$ simulation

Spectrum estimation, stalagmite Q5, Holocene

Step 1	Time series	$\{t(i), x(i)\}_{i=1}^{n}$
Step 2	Bias-corrected Lomb–Scargle	
	spectrum (Algorithm 5.3)	$\widehat{h}'(f_j)$
Step 3	Estimated, bias-corrected	
	persistence time	$\widehat{\tau}'$
Step 4	Determine area	$A_{\widehat{h}'}$
	under spectrum within $[0; (2\bar{d})^{-1}]$	
Step 5	Generate AR(1) data (Eq. 2.9)	$\{t(i), x^*(i)\}_{i=1}^{n}$
Step 6	Use timescale model	
	to resample times	$\{t^*(i)\}_{i=1}^{n}$
Step 7	Bias-corrected Lomb–Scargle	
	spectrum estimate	$\widehat{h}'^{*b}(f_j)$
	for $\{t^*(i), x^*(i)\}_{i=1}^{n}$	
	(b, counter),	
	scaled to area	$A_{\widehat{h}'}$
Step 8	Go to Step 5 until $b = B$	
	replications exist	
Step 9	Test at each f_j whether	
	$\widehat{h}'(f_j)$ exceeds a pre-defined	
	upper percentile of $\left\{\widehat{h}'^{*b}(f_j)\right\}_{b=1}^{B}$	

Spectrum estimation, stalagmite Q5, Holocene



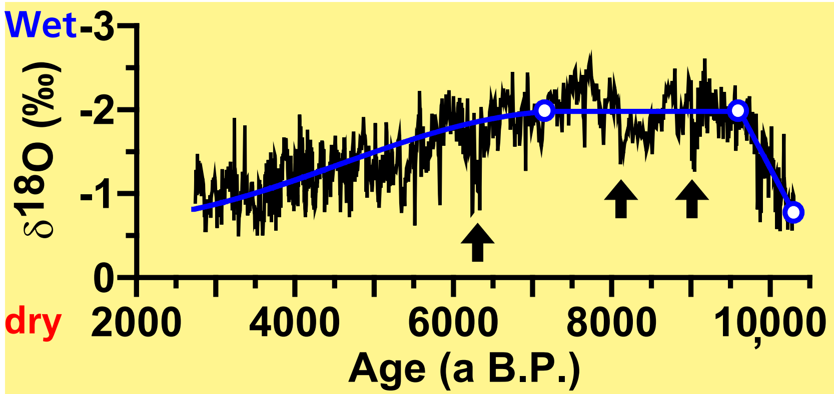


Timescale error effects?

- Test using $t(i)^*$ simulation
- Peak I likely affected, possibly also peaks II to IV
- Peak V (T_{period} = 963 a) robust

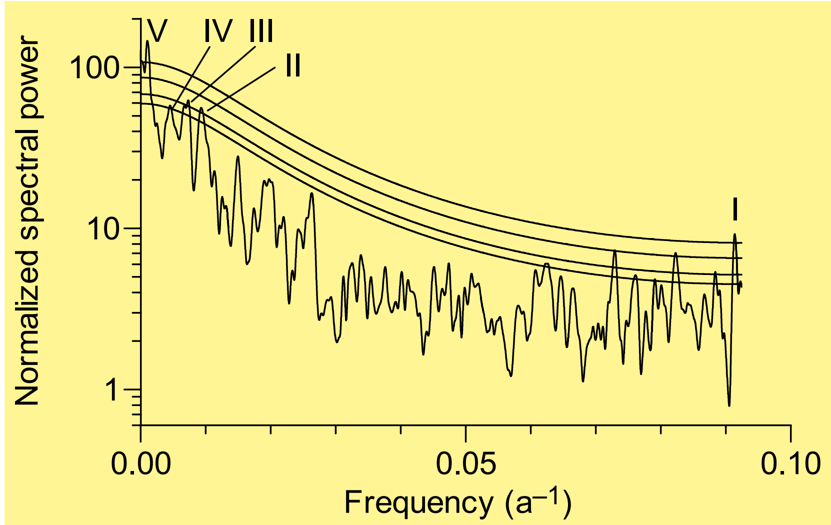
Book, Algorithms 5.5, 5.6, Figs. 5.8, 5.11, Section 5.2.8

Spectrum estimation, stalagmite Q5, Holocene



Oxygen isotope record (Q5)

Monsoon rainfall proxy



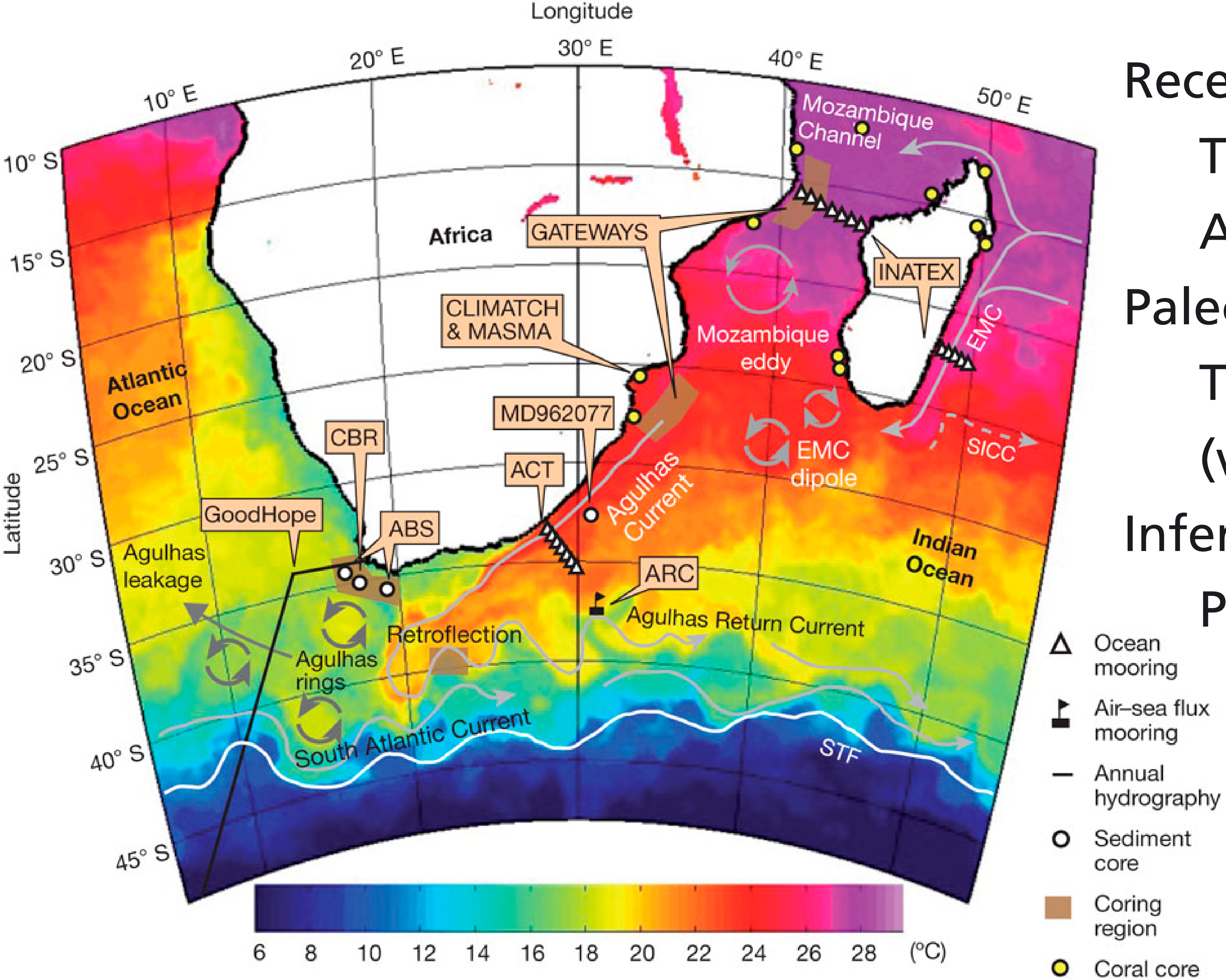
Estimated spectrum

Use spectrum to filter out frequency ranges of interest.

Study physics of Sun–monsoon system.

Take into account all relevant records (multiple test).

Correlation estimation, Agulhas current, Pleistocene



Recent:
- Temperature
- Agulhas flow

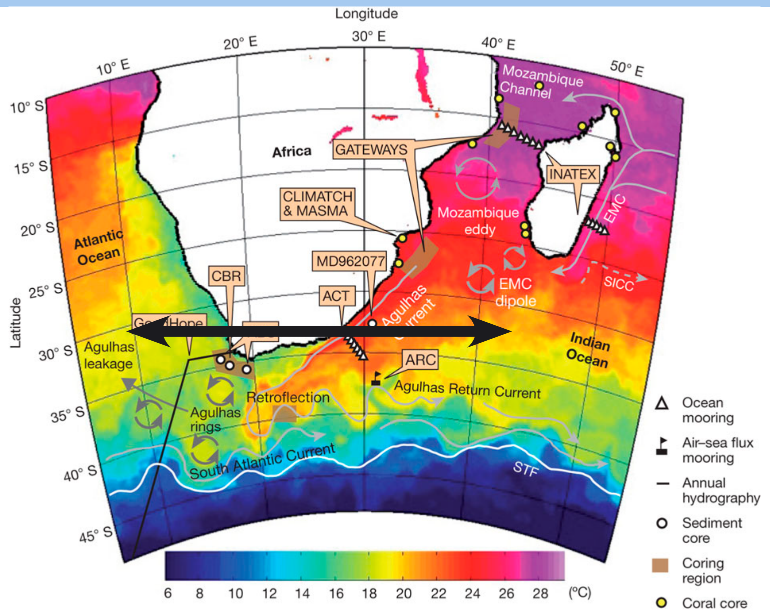
Paleo:
- Temperature
- (via sediment core)

Inference target:
- Paleo Agulhas flow

Beal et al. (2011) *Nature* 472:429

Part 3: Examples

Correlation estimation, Agulhas current, Pleistocene



Recent:

- Temperature (X)
- Agulhas flow (Y)

Paleo:

- Temperature (via sediment core)

Inference target:

- Paleo Agulhas flow

Method:

- Look for max(r_{XY}) with bootstrap CI

Beal et al. (2011) *Nature* 472:429

Correlation estimation, Agulhas current, Pleistocene

$$X(i) = X_{\text{trend}}(i) + X_{\text{out}}(i) + S(i) \cdot X_{\text{noise}}(i)$$
$$= \mu_X + S_X \cdot X_{\text{noise}}(i).$$

Climate equation, discrete time

$$\{t(i), x(i)\}_{i=1}^{n}$$

Time series

$$Y(i) = \mu_Y + S_Y \cdot Y_{\text{noise}}(i).$$

Climate equation, discrete time

$$\{t(i), y(i)\}_{i=1}^{n}$$

Time series

Bivariate setting, equal time points

Absent trends, absent outliers, constant variability

Correlation estimation, Agulhas current, Pleistocene

$$r_{XY} = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{X(i) - \bar{X}}{S_{n,X}} \right) \cdot \left(\frac{Y(i) - \bar{Y}}{S_{n,Y}} \right), \qquad (7.5)$$

Pearson's correlation coefficient (r_{XY} estimates ρ_{XY})

$-1 \leq r_{XY} \leq 1$

$$\bar{X} = \sum_{i=1}^{n} X(i) \,/\, n \qquad \bar{Y} = \sum_{i=1}^{n} Y(i) \,/\, n$$

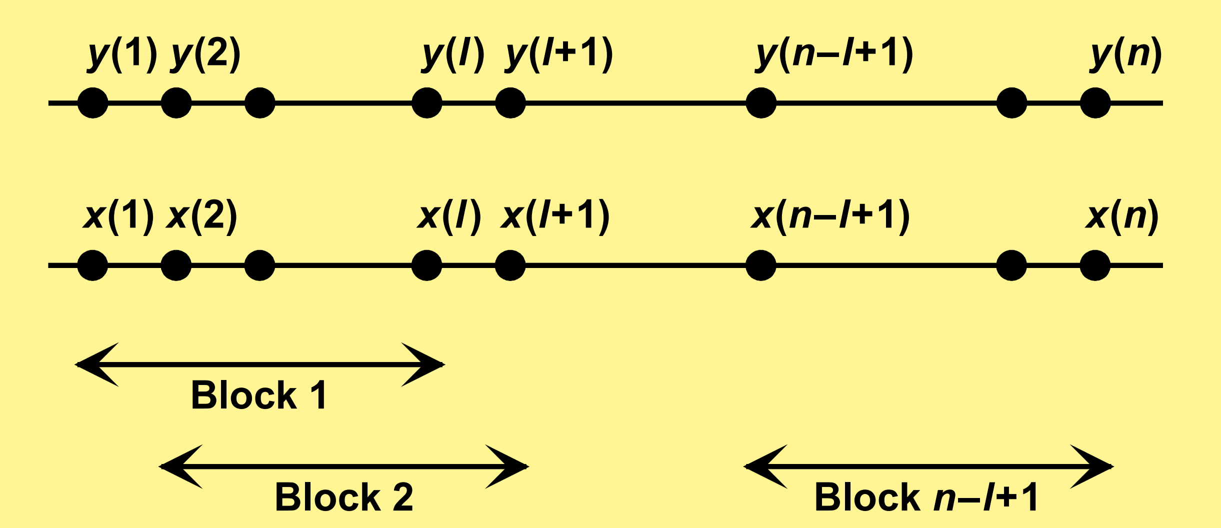
Sample means

$$S_{n,X} = \left\{ \sum_{i=1}^{n} \left[X(i) - \bar{X} \right]^2 / \, n \right\}^{1/2}$$

$$S_{n,Y} = \left\{ \sum_{i=1}^{n} \left[Y(i) - \bar{Y} \right]^2 / \, n \right\}^{1/2}$$

Sample standard deviations (denominator n)

Correlation estimation, Agulhas current, Pleistocene



Pairwise-moving block bootstrap (pairwise-MBB)

$$l_{\mathrm{opt}} = NINT\left\{\left[6^{1/2}\cdot \widehat{\bar{a}}'_{XY}\Big/\left(1-\widehat{\bar{a}}'^{\,2}_{XY}\right)\right]^{2/3}\cdot n^{1/3}\right\}.$$

Block length selector

$$\widehat{\bar{a}}'_{XY} = \left[\widehat{\bar{a}}'_{X}\cdot\widehat{\bar{a}}'_{Y}\right]^{1/2}$$

Correlation estimation, Agulhas current, Pleistocene

Table 7.4. Monte Carlo experiment, Pearson's correlation coefficient with Fisher's z-transformation for bivariate lognormal AR(1) processes. The number of Monte Carlo simulations and the properties of $\{T(i), X(i), Y(i)\}_{i=1}^{n}$ are identical to those in the first experiment (Table 7.1), with the exception that $\rho_{\mathcal{E}}$ is here given via Eq. (7.24). The construction of CIs followed Algorithms 7.1 and 7.2.

n	$\gamma^{a}_{r_{XY}}$						Nominal
	True correlation, ρ_{XY}						
	0.3			0.8			
	CI type			CI type			
	Bootstrap		Classical	Bootstrap		Classical	
	Student's t	BCa		Student's t	BCa		
10	0.820	0.701	0.748	0.864	0.778	0.875	0.950
20	0.876	0.808	0.805	0.904	0.859	0.807	0.950
50	0.932	0.875	0.848	0.898	0.864	0.737	0.950
100	0.939	0.866	0.836	0.895	0.856	0.684	0.950
200	0.941	0.879	0.781	0.897	0.853	0.633	0.950
500	0.907	0.876	0.767	0.899	0.846	0.554	0.950
1000	0.911	0.885	0.730	0.913	0.866	0.551	0.950

Alert! MC runs suffer from programming bug (which affects small r_{XY}).

Correlation estimation, Agulhas current, Pleistocene

Table 7.6. Monte Carlo experiment, Pearson's and Spearman's correlation coefficients with Fisher's z-transformation for bivariate lognormal AR(1) processes: calibrated CI coverage performance. The number of Monte Carlo simulations and the properties of $\{T(i), X(i), Y(i)\}_{i=1}^{n}$ are identical to those in the first experiment (Table 7.1), with $\rho_{\mathcal{E}}$ given by Eq. (7.24) and Table 7.8, respectively. Calibrated Student's t CIs were constructed after Eq. (3.47) using two loops of pairwise-MBB resampling with block length selection after Eqs. (7.31) and (7.32). The first loop (bootstrap of samples) used $B = 2000$ resamplings, the second loop (bootstrap of resamples) used 1000 resamplings. In the second loop, the block length was not re-estimated but overtaken from the first loop. The spacing of the λ values for the calibration (Eq. 3.45) is 0.001.

n	$\gamma^{a}_{r_{XY}}$		$\gamma^{a}_{r_S}$		*Nominal*
	True correlation, ρ_{XY}		*True rank correlation*, ρ_S		
	0.3	0.8	0.3	0.8	
10	0.917	0.836			0.950
20	0.959	0.937			0.950
50	0.964	0.947			0.950
100	0.969	0.947			0.950
200	0.966	0.946			0.950

Alert! MC runs suffer from programming bug (which affects small r_{XY}).

Correlation estimation, Agulhas current, Pleistocene

Monte Carlo experiments, in terms of CI coverage accuracy

1. Classical CIs fail completely for non-Gaussian shapes.

2. Usage of Spearman's r_S instead of Pearson's r_{XY} advised:
   If distributional shapes Gaussian, then
      both perform similar
   else if shapes are non-Gaussian (e.g., skewed), then
      Pearson's r_{XY} performs badly.

3. Calibration (expensive) increases accuracy (small n) strongly.

Conclusion

Paleoclimate time series analysis is exciting!

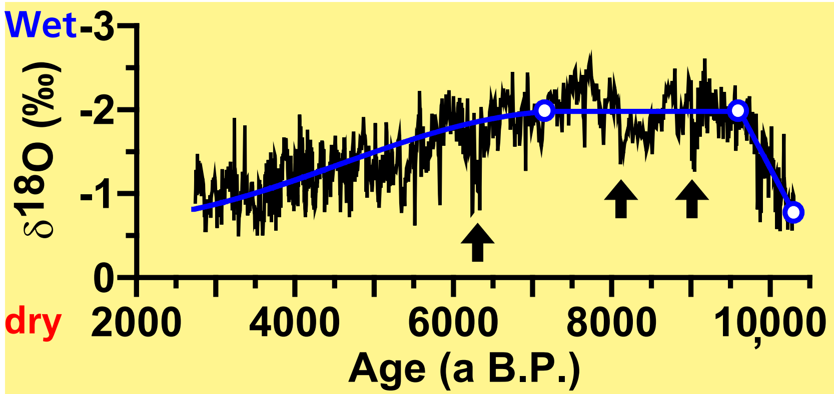
Conclusion

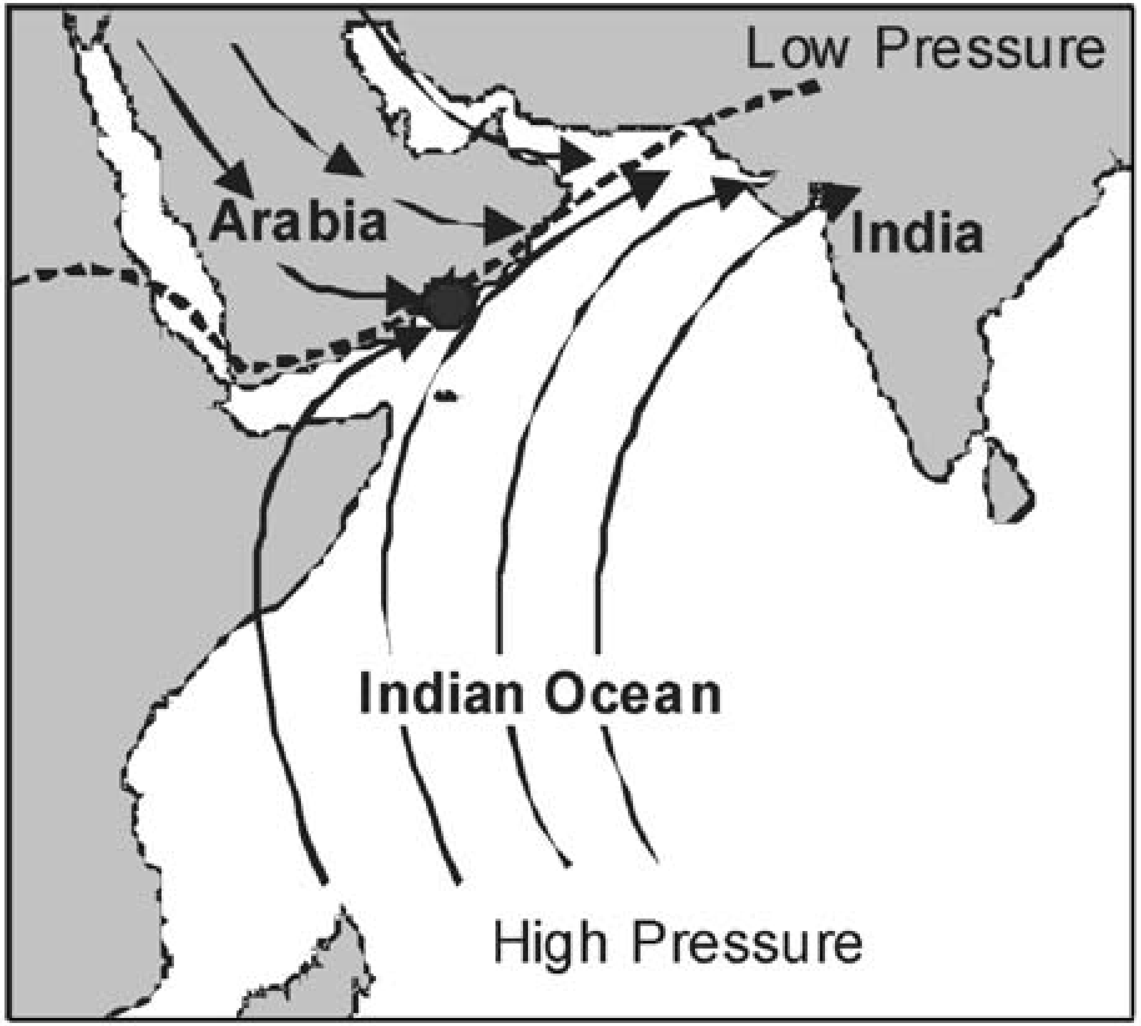
Paleoclimate time series analysis is exciting!

Thanks!

Postdoc job:
www.climate-risk-analysis.com

Paleoclimate time series

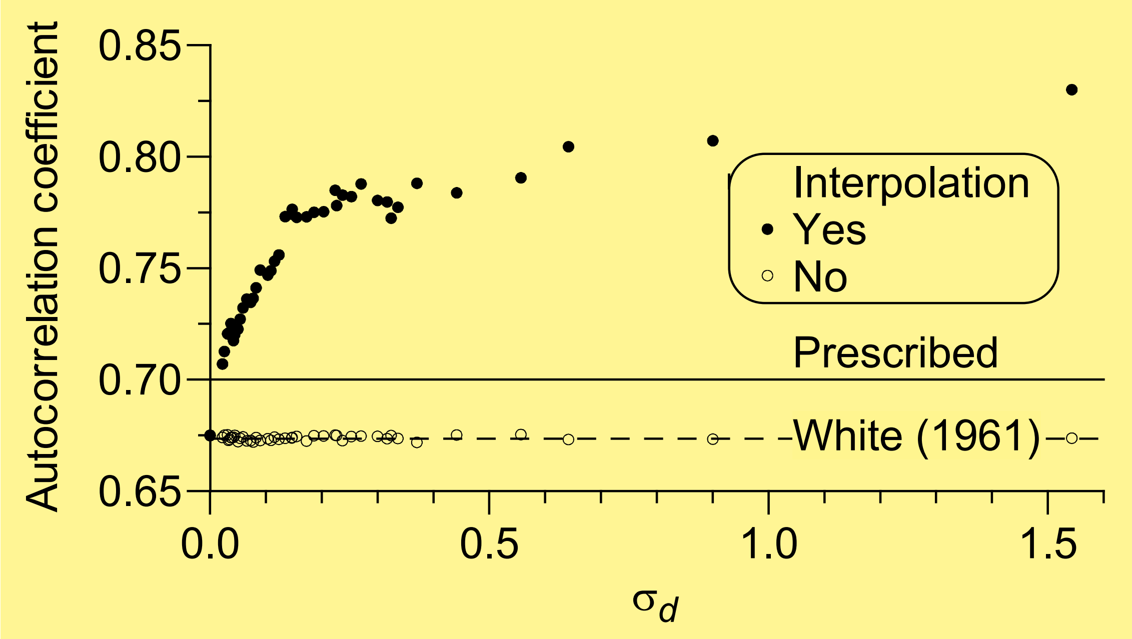




Book, Figs. 4.18 (modified); Fleitmann et al. (2004: Fig. 1a) *Quaternary Science Reviews* 23:935

First-order autoregressive model

Why not interpolate?



Monte Carlo experiment

$n = 50$,

uneven spacing (σ_d),

$\tau = -1/\ln(0.7) \approx 2.804$,

$\bar{a} = \exp(-\bar{d}/\tau)$ = 0.7

Equivalent autocorrelation coefficient

$\widehat{\bar{a}} = \exp(-\bar{d}/\widehat{\tau})$

Estimated equivalent autocorrelation coefficient

averaged over n_{sim} = 10,000 simulations

Book, Fig. 2.3 (modified)

Climate theory

Climate, $X(T)$
Weather, $Y(T)$

$$\frac{dX(T)}{dT} = F(X(T), Y(T)), \qquad \text{timescale } \tau_X,$$
$$\frac{dY(T)}{dT} = G(X(T), Y(T)), \qquad \text{timescale } \tau_Y,$$

Weather components "leave a trace" in climate.

$$\frac{dX(T)}{dT} \simeq F(X(0), Y(T)),$$
$$= W(T),$$

$$X(T+1) = X(T) + \mathcal{E}_{\mathrm{N}(0,\,\sigma^2)}(T).$$

Climate does not run away: negative feedback (~ X)

$$X(T+1) = a \cdot X(T) + \mathcal{E}_{\mathrm{N}(0,\,\sigma^2)}(T),$$

$W(T)$, Wiener process ("continuous-time noise increment")
Hasselmann (1976) *Tellus* 28:473; Book, Eqs. (2.25), (2.26), (2.28), (2.29), (2.30), Section 2.5.1

Climate theory

Support of simple AR(1) climate noise model

1. Weather components (Hasselmann 1976)
2. Climate equation $$X(i) = X_{\mathrm{trend}}(i) + X_{\mathrm{out}}(i) + S(i) \cdot X_{\mathrm{noise}}(i),$$
3. We have also trend, outliers/extremes and variability to describe climate.

Book, Eq. (1.2)

Bootstrap resampling

Autoregressive bootstrap (ARB)

- Take noise residuals $r(i)$
- Fit AR(1) model
- Take white-noise residuals $\epsilon(i) = [r(i) - \widehat{a}' \cdot r(i-1)]$
- Draw randomly from $e(i)$ with replacement (ordinary bootstrap)
- Plug in resampled $e(i)$
- Plug in resampled $r(i)$

Bootstrap confidence intervals

$$\widehat{\mathrm{se}}_{\widehat{\theta}^*} = \left\{ \sum_{b=1}^{B} \left[\widehat{\theta}^{*b} - \left\langle \widehat{\theta}^{*b} \right\rangle \right]^2 \Big/ (B-1) \right\}^{1/2}, \qquad (3.30)$$

Bootstrap standard error

Bootstrap confidence intervals

$$\mathrm{CI}_{\widehat{\theta},1-2\alpha} = \left[\widehat{\theta} + z(\alpha) \cdot \widehat{\mathrm{se}}_{\widehat{\theta}^*}; \widehat{\theta} - z(\alpha) \cdot \widehat{\mathrm{se}}_{\widehat{\theta}^*}\right], \qquad (3.31)$$

Normal confidence interval

$z(\alpha)$ Percentage point of normal distribution

Example

$z(1 - 0.025) \approx 1.959964$

"±2-sigma interval is 95% CI."

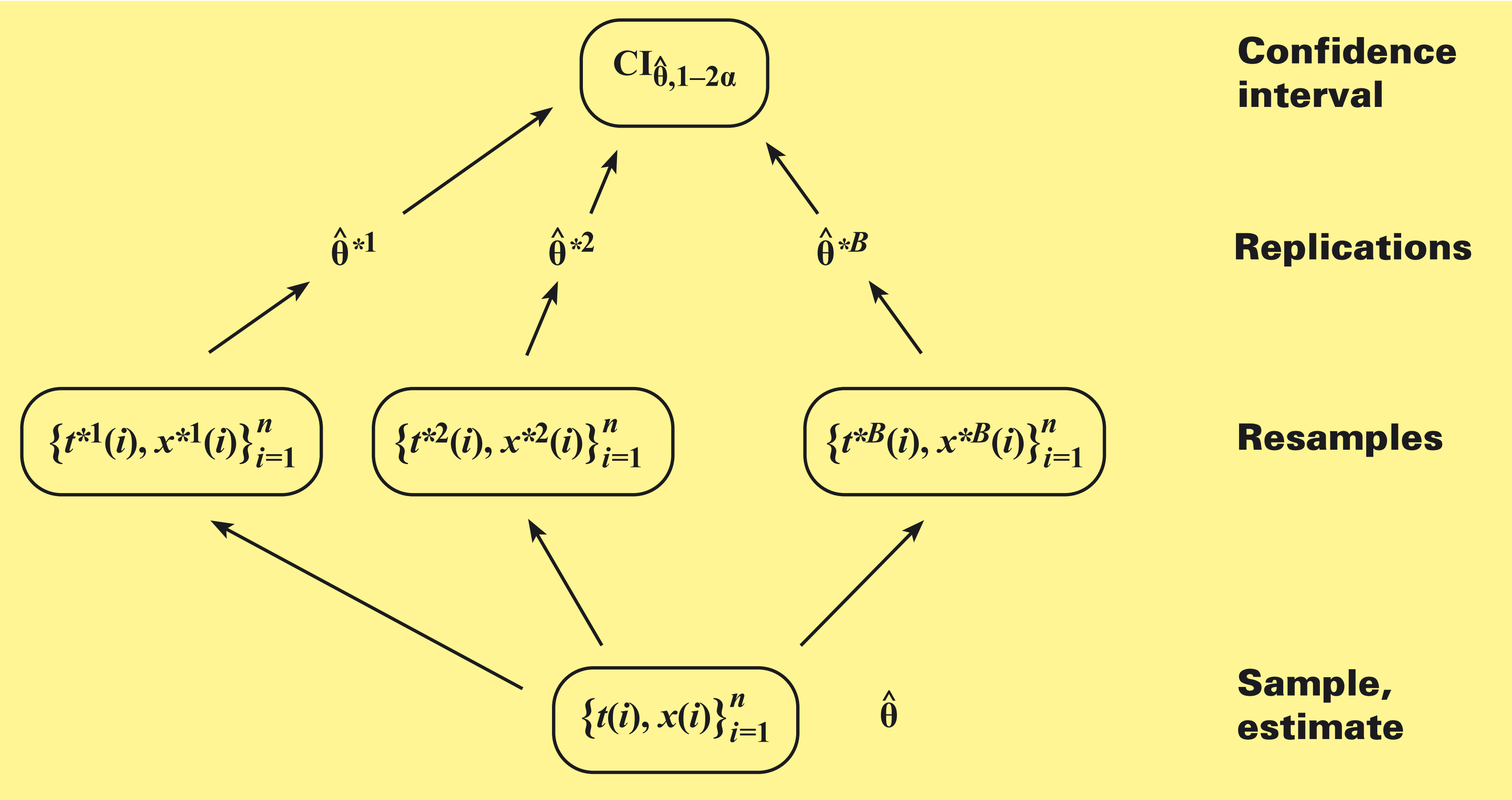
Bootstrap confidence intervals

$$\mathrm{CI}_{\widehat{\theta},1-2\alpha} = \left[\widehat{\theta} + t_{\nu}(\alpha) \cdot \widehat{\mathrm{se}}_{\widehat{\theta}^*} ; \widehat{\theta} - t_{\nu}(\alpha) \cdot \widehat{\mathrm{se}}_{\widehat{\theta}^*}\right], \qquad (3.32)$$

Student's *t* confidence interval

$t_v(\alpha)$ Percentage point of Student's *t*-distribution
with *v* degrees of freedom (*v* = *n* – number of parameters)
Takes into account that not only standard error,
but also fit value estimated.
Negligible difference to normal CI if *v* above ~ 30

Bootstrap confidence intervals



Book, Fig. 3.3

Bootstrap confidence intervals

$$\mathrm{CI}_{\widehat{\theta},1-2\alpha} = \left[\widehat{\theta}^*(\alpha); \widehat{\theta}^*(1-\alpha)\right], \tag{3.33}$$

Percentile confidence interval

α small $\Rightarrow$ B large

typically

$\alpha = 0.025$ (95% CI) or $\alpha = 0.05$ (90% CI), $B \approx 2000$

Bootstrap confidence intervals

$$\mathrm{CI}_{\widehat{\theta},1-2\alpha} = \left[\widehat{\theta}^*(\alpha); \widehat{\theta}^*(1-\alpha)\right], \tag{3.33}$$

Percentile confidence interval

Assume that our estimator underestimates, $E[\hat{\theta}] < \theta$.

Then likely also the bootstrap replications are "shifted".

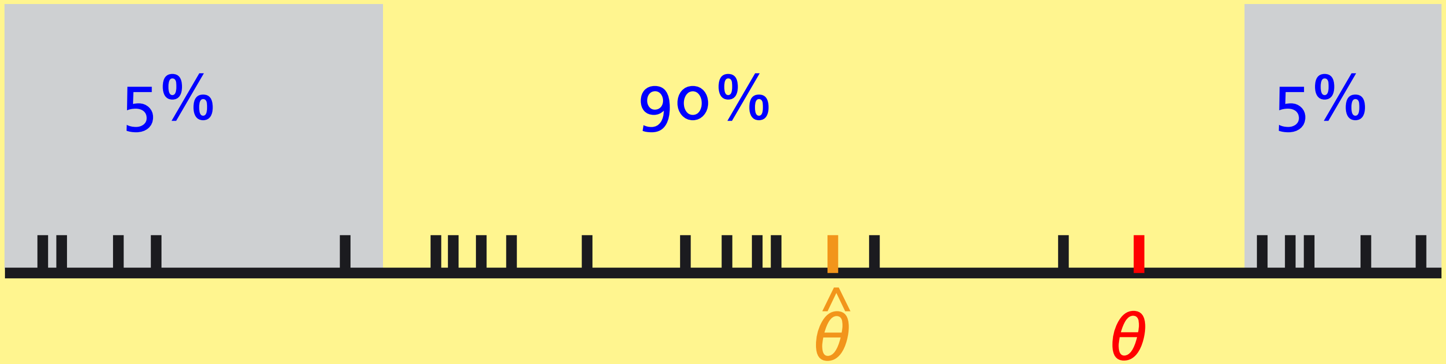
There is an option for bias correction.

Bootstrap confidence intervals

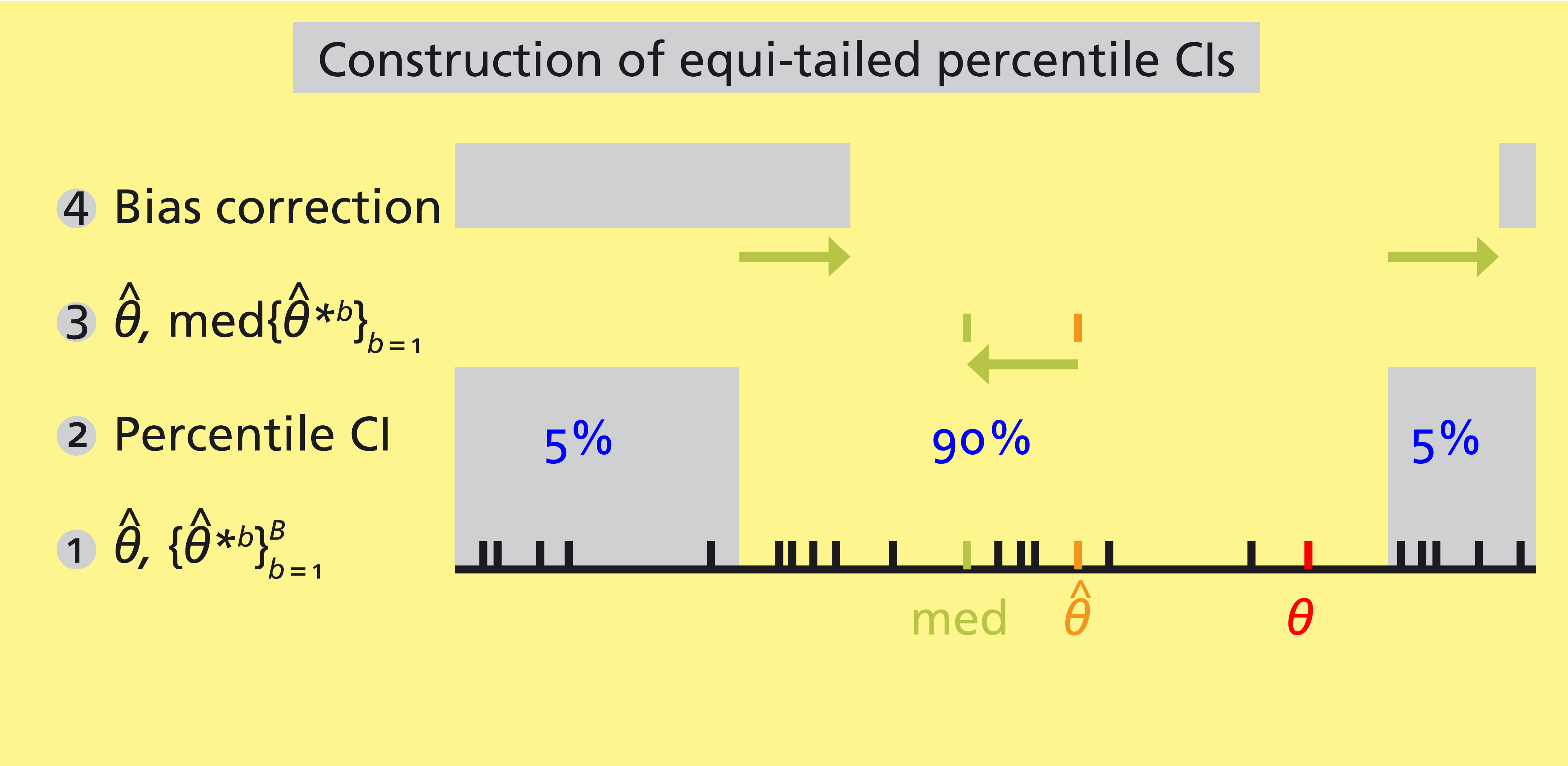
Construction of equi-tailed percentile CIs

2 Percentile CI

1 $\hat{\theta}$, $\{\hat{\theta}^{*b}\}_{b=1}^{B}$



Bootstrap confidence intervals



Bootstrap confidence intervals

$$\mathrm{CI}_{\widehat{\theta},1-2\alpha} = \left[\widehat{\theta}^*(\alpha 1); \widehat{\theta}^*(\alpha 2)\right], \qquad (3.34)$$

Bias-corrected and accelerated (BCa) confidence interval

$$\alpha 1 = F\left(\widehat{z}_0 + \frac{\widehat{z}_0 + z(\alpha)}{1 - \widehat{a}\left[\widehat{z}_0 + z(\alpha)\right]}\right)$$

$$\alpha 2 = F\left(\widehat{z}_0 + \frac{\widehat{z}_0 + z(1-\alpha)}{1 - \widehat{a}\left[\widehat{z}_0 + z(1-\alpha)\right]}\right)$$

$$\widehat{z}_0 = F^{-1}\left(\frac{\#\left\{\widehat{\theta}^{*b} < \widehat{\theta}\right\}}{B}\right)$$

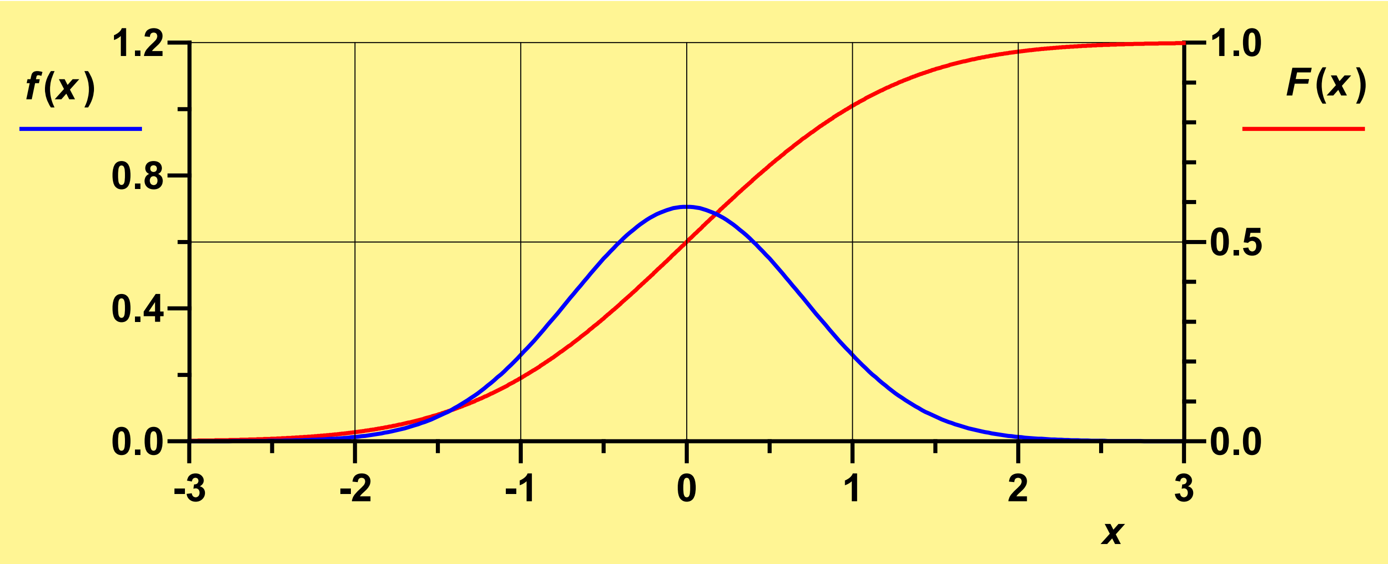
$$\widehat{a} = \frac{\sum_{j=1}^{n}\left[\left\langle\widehat{\theta}_{(j)}\right\rangle - \widehat{\theta}_{(j)}\right]^3}{6\left\{\sum_{j=1}^{n}\left[\left\langle\widehat{\theta}_{(j)}\right\rangle - \widehat{\theta}_{(j)}\right]^2\right\}^{3/2}},$$

Bias correction

Book, Eqs. (3.35), (3.36), (3.37),(3.38), Section 3.9
F^{-1}, inverse normal distribution function; e.g., $F^{-1}(0.5) = 0$

$\hat{a}$, acceleration; $\langle\rangle$, average;
$\hat{\theta}_{(j)}$, jackknife value (leave one point, j, out)

Bootstrap confidence intervals



$$f(x) = (2\pi)^{-1/2} \exp\left(-x^2/2\right)$$

Standard normal probability density function (PDF)

$$F(x) \;=\; \int_{-\infty}^{x} f(x')dx',$$

Standard normal probability distribution function

Correlation estimation, Agulhas current, Pleistocene

Step 1	Bivariate time series	$\{t(i), x(i), y(i)\}_{i=1}^{n}$
Step 2	Pearson's r_{XY} (Eq. 7.5)	
Step 3	Fisher's z-transformation	$z = \tanh^{-1}(r_{XY})$
Step 4	Estimated, bias-corrected	
	persistence time, process $X(i)$,	$\widehat{\tau}'_X$
	using mean-detrended	
	time series, $\{x(i) - \bar{x}\}_{i=1}^{n}$	
Step 5	Analogously, process $Y(i)$	$\widehat{\tau}'_Y$
Step 6	Estimated, bias-corrected	
	equivalent autocorrelation	
	coefficient, process $X(i)$	$\widehat{\bar{a}}'_X = \exp\left(-\bar{d}/\widehat{\tau}'_X\right)$
Step 7	Analogously, process $Y(i)$	$\widehat{\bar{a}}'_Y = \exp\left(-\bar{d}/\widehat{\tau}'_Y\right)$

Correlation estimation, Agulhas current, Pleistocene

Step 8	Effective data size, obtained by plugging in $\widehat{\bar{a}}'_X$ for a_X and $\widehat{\bar{a}}'_Y$ for a_Y in Eq. (2.38)	n'_ρ
Step 9	Approximate, classical normal CI for r_{XY}, obtained from re-transforming z	$\mathrm{CI}_{r_{XY},1-2\alpha} = \left[\tanh\left[z + z(\alpha)\cdot(n'_\rho - 3)^{-1/2}\right]; \tanh\left[z - z(\alpha)\cdot(n'_\rho - 3)^{-1/2}\right]\right]$

$$n'_\rho = n\left\{1 + \frac{2}{n}\frac{1}{1-a_Xa_Y}\left[a_Xa_Y\left(n - \frac{1}{1-a_Xa_Y}\right) - (a_Xa_Y)^n\left(1 - \frac{1}{1-a_Xa_Y}\right)\right]\right\}^{-1}. \quad (2.38)$$

Correlation estimation, Agulhas current, Pleistocene

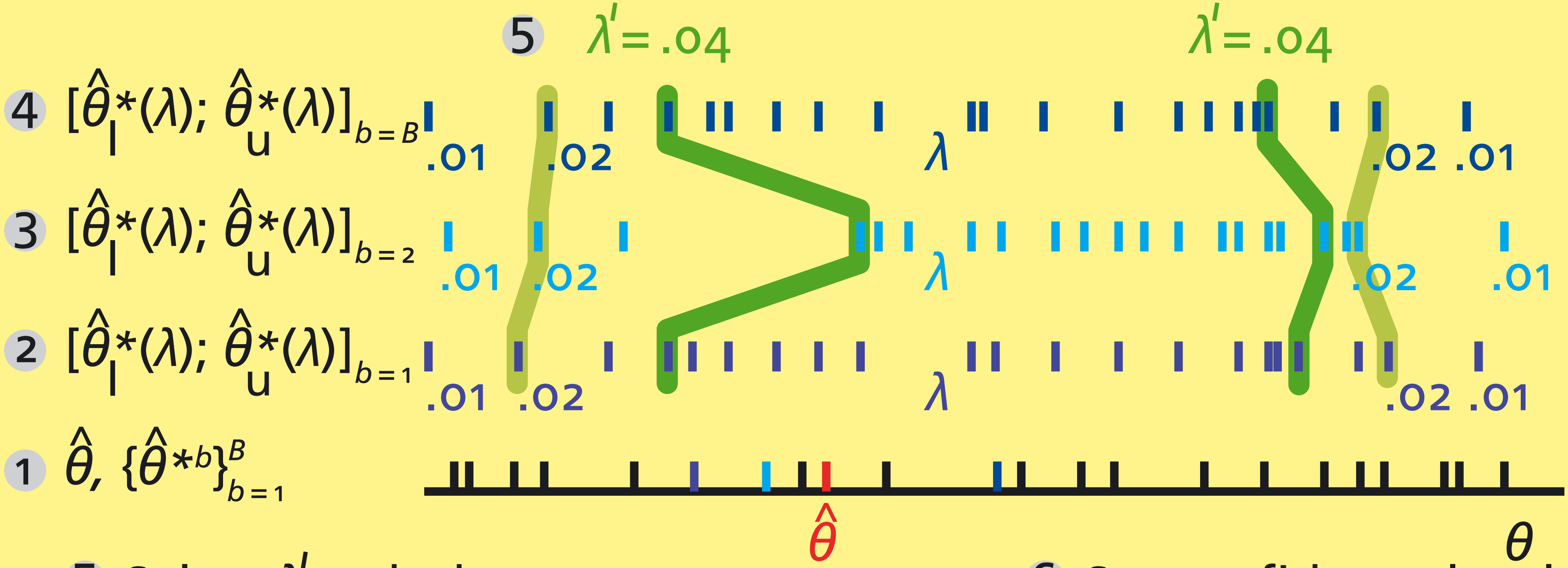
Step 1	Bivariate time series	$\{t(i), x(i), y(i)\}_{i=1}^{n}$
Step 2	Pearson's r_{XY} (Eq. 7.5)	
Step 3	Fisher's z-transformation	$z = \tanh^{-1}(r_{XY})$
Step 4	Estimated, bias-corrected	
	persistence time, process $X(i)$,	$\widehat{\tau}'_X$
	using mean-detrended	
	time series, $\{x(i) - \bar{x}\}_{i=1}^{n}$	
Step 5	Analogously, process $Y(i)$	$\widehat{\tau}'_Y$
Step 6	Select block length	l
Step 7	Apply MBB with l	$\{x^{*b}(i)\}_{i=1}^{n} = \{x(f(i))\}_{i=1}^{n}$
	(Algorithm 3.1) to x values	(b, counter)
Step 8	Overtake bootstrap index	$f(i)$
	for resampled y values	$\{y^{*b}(i)\}_{i=1}^{n} = \{y(f(i))\}_{i=1}^{n}$
Step 9	Resample	$\{x^{*b}(i), y^{*b}(i)\}_{i=1}^{n}$

Correlation estimation, Agulhas current, Pleistocene

Step 9	Resample	$\left\{x^{*b}(i), y^{*b}(i)\right\}_{i=1}^{n}$
Step 10	Bootstrap replications,	
	Pearson's r_{XY} and Fisher's z	$r_{XY}^{*b}, \quad z^{*b} = \tanh^{-1}\left(r_{XY}^{*b}\right)$
Step 11	Go to Step 7 until $b = B$	
	(usually $B = 2000$)	
	replications exist	$\left\{z^{*b}\right\}_{b=1}^{B}$
Step 12	Calculate CI (Section 3.4)	
	for Fisher's z	$\mathrm{CI}_{z,1-2\alpha} = \left[z_{\mathrm{l}}; z_{\mathrm{u}}\right]$
Step 13	Re-transform lower and upper	
	endpoints to obtain	
	pairwise-MBB CI for r_{XY}	$\mathrm{CI}_{r_{XY},1-2\alpha} = \left[\tanh\left(z_{\mathrm{l}}\right); \tanh\left(z_{\mathrm{u}}\right)\right]$

Correlation estimation, Agulhas current, Pleistocene

Calibration of equi-tailed bootstrap confidence intervals



5 Select λ' such that estimate $\hat{\theta}$ is in $(1-2\alpha) \times B$ cases in { }

6 Set confidence level to $(1-2\lambda')$

Correlation estimation, Agulhas current, Pleistocene

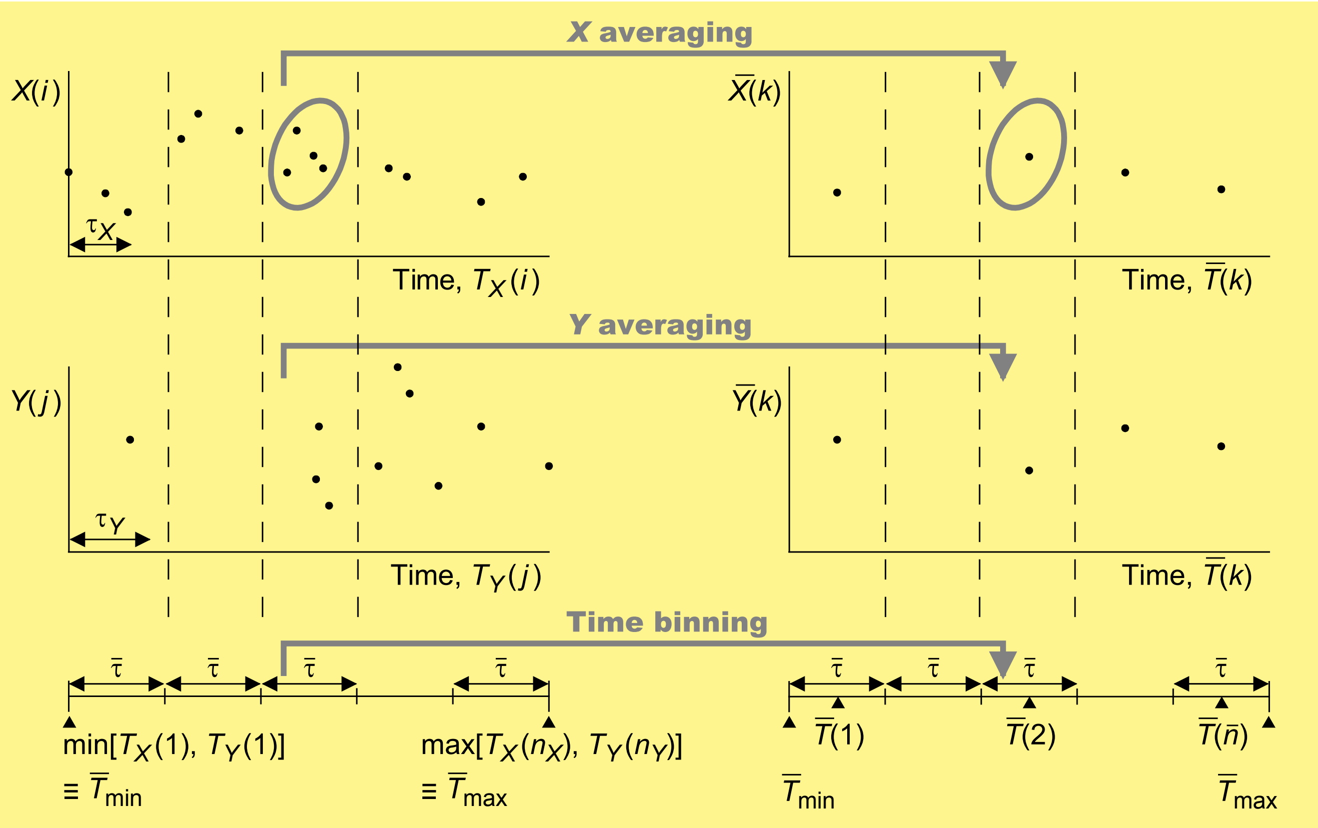
$$\widehat{\theta}_{\mathrm{l}}^{*b}(\lambda), \qquad \lambda = 0.01, \ldots, 0.99.$$

Grid of confidence levels

$$\widehat{p}(\lambda) = \frac{\# \left\{ \widehat{\theta}_{\mathrm{l}}^{*b}(\lambda) < \widehat{\theta} < \widehat{\theta}_{\mathrm{u}}^{*b}(\lambda) \right\}}{B},$$

Calibration curve

Correlation estimation, Agulhas current, Pleistocene



1. If persistence exists, then we can recover correlation information.
2. Binning is better than interpolation.