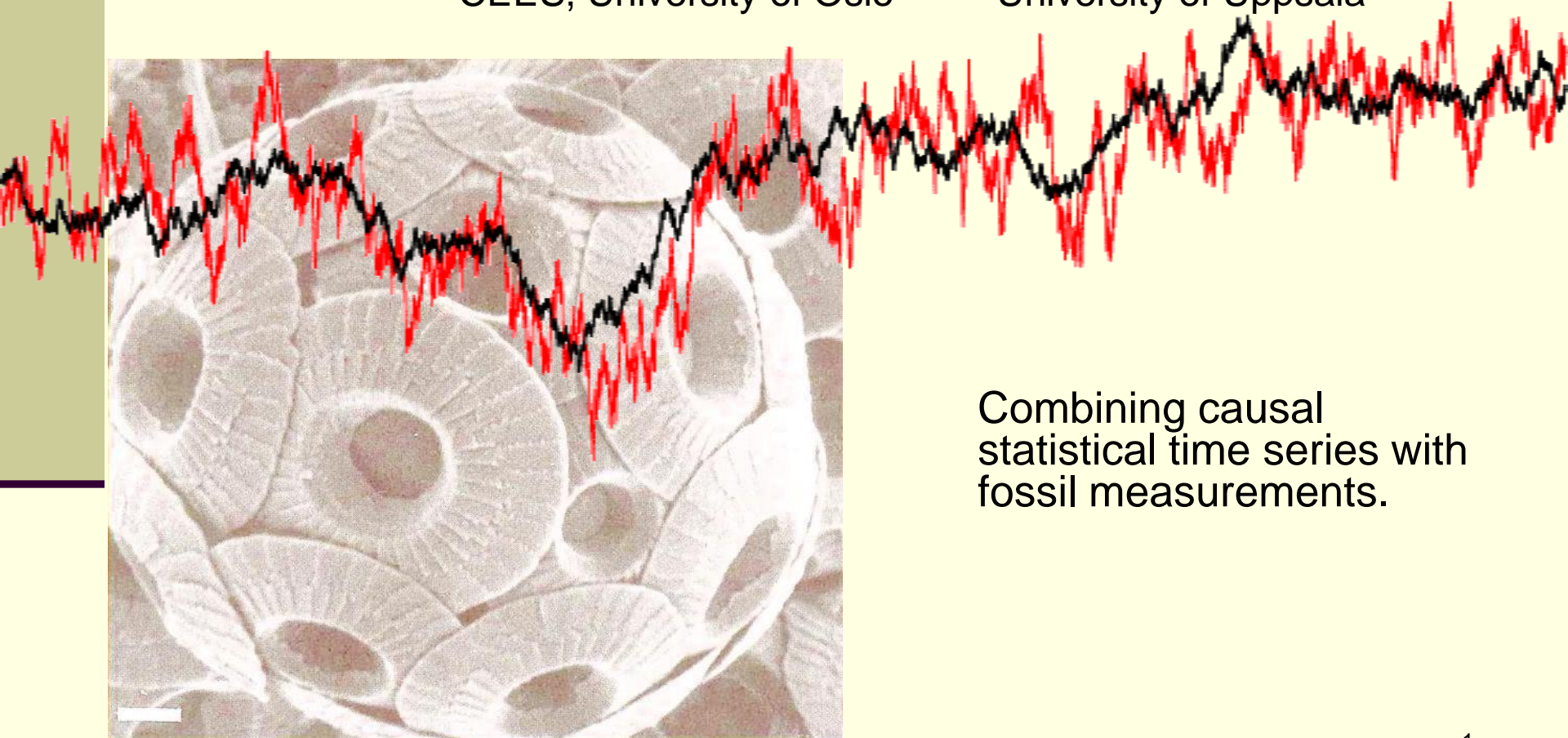


Layered continuous time processes in biology

Tore Schweder and Trond Reitan
CEES, University of Oslo

Jorijntje Henderiks
University of Uppsala



Combining causal
statistical time series with
fossil measurements.

Overview

1. Introduction - Motivating example:
 - Coccolith data (microfossils)
 - Phenotypic evolution: irregular time series related by (possibly common) latent processes
2. Causality in continuous time processes
3. Stochastic differential equation vector processes
 - Ito representation and diagonalization
 - Tracking processes and hidden layers
 - Kalman filtering
 - Model variants
4. Inference and results
 - Bayesian inference on model properties, models, process parameters and process states
 - Results for the coccolith data.
5. Second application: Phenotypic evolution on a phylogenetic tree
 - Primates - preliminary results
6. Conclusion

Irregular time series related by latent processes: evolution of body size in *Coccolithus*

Henderiks – Schweder - Reitan

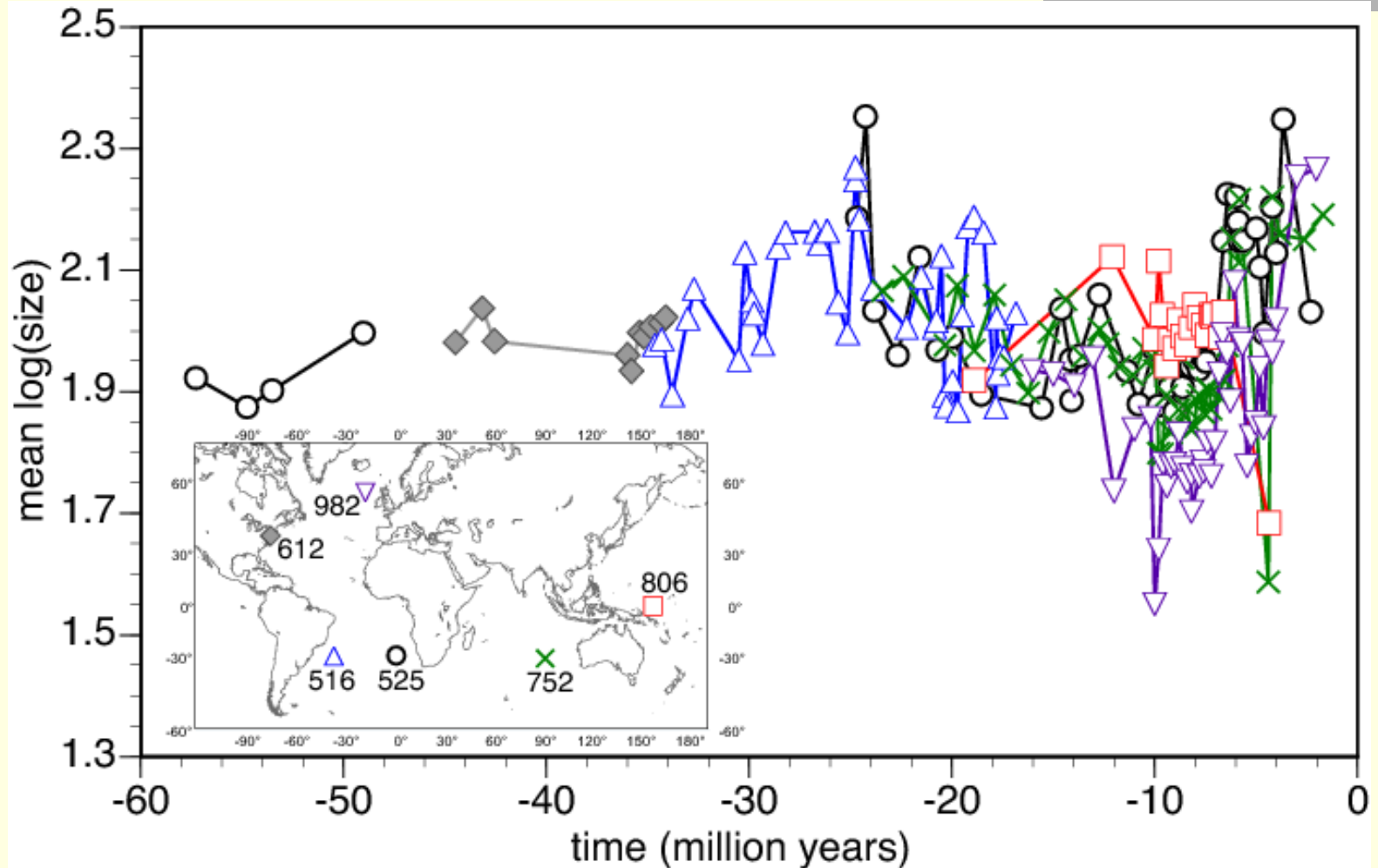
The size of a single cell algae (*Coccolithus*) is measured by the diameter of its fossilized coccoliths (calcite platelets). Want to model the evolution of a lineage found at six sites.

- Phenotype (size) is a process in continuous time and changes continuously.

19,899 coccolith measurements, 205 sediment samples ($1 < n < 400$) of body size by site and time (0 to -60 my).



Our data – Coccolith size measurements



205 Sample mean log coccolith size ($1 < n < 400$) by time and site.

Evolution of size distribution

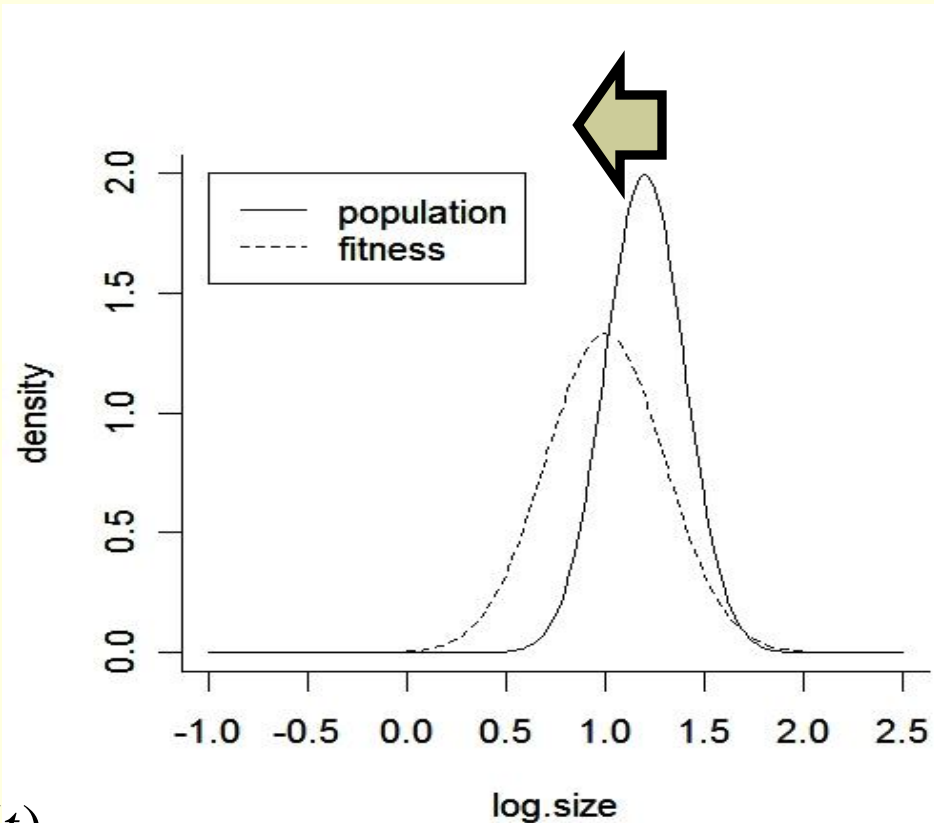
Fitness = expected number of reproducing offspring.

The population tracks the fitness curve (natural selection)
The fitness curve moves about, the population follow.

With a known fitness, μ , the mean phenotype should be an Ornstein-Uhlenbeck process (Lande 1976).

$$dX(t) = -\alpha [X(t) - \mu] dt + \sigma dW(t)$$

With fitness as a process, $\mu(t)$,
we can make a tracking model: $dX(t) = -\alpha [X(t) - \mu(t)] dt + \sigma dW(t)$ 5

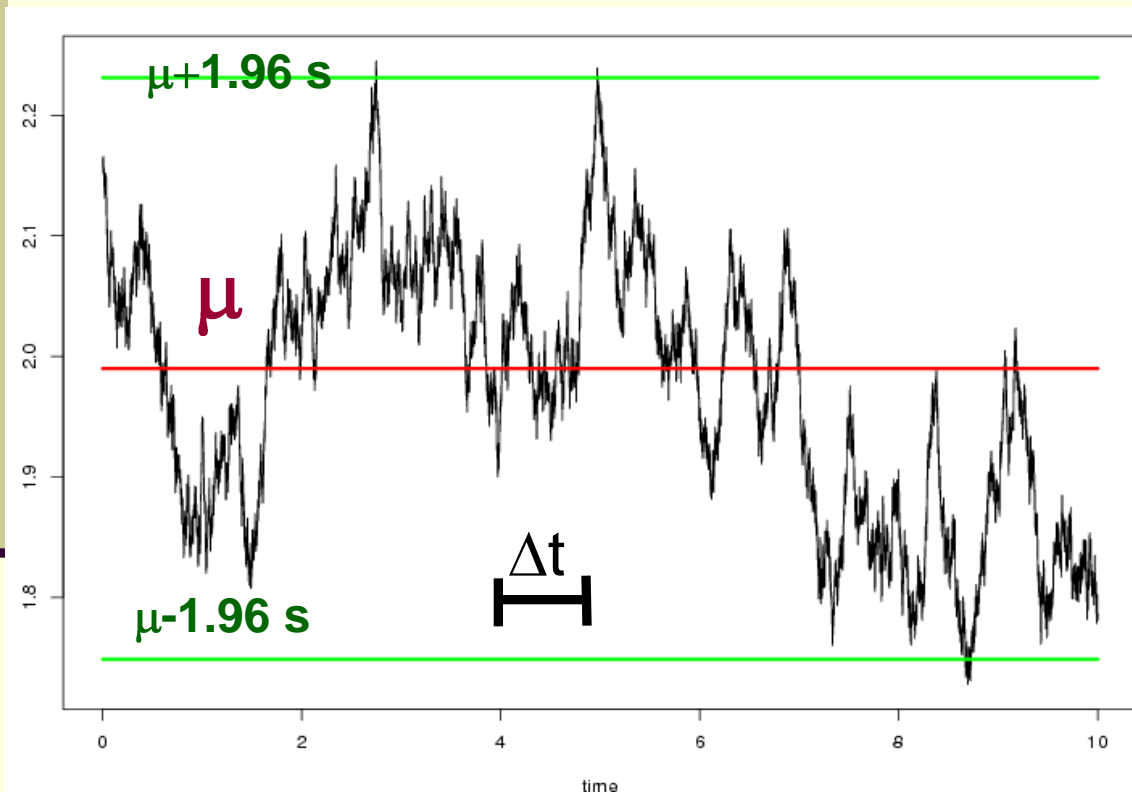


The Ornstein-Uhlenbeck process

$$dX(t) = -\alpha (X(t) - \mu) dt + \sigma dW(t)$$

Attributes:

- Stationary
 - Normally distributed
 - long-term level: μ
 - Standard deviation: $s = \sigma / \sqrt{2\alpha}$
- Markovian
 - α : pull
 - $\text{corr}(x(0), x(t)) = e^{-\alpha t}$
 - Time for the correlation to drop to $1/e$: $\Delta t = 1/\alpha$

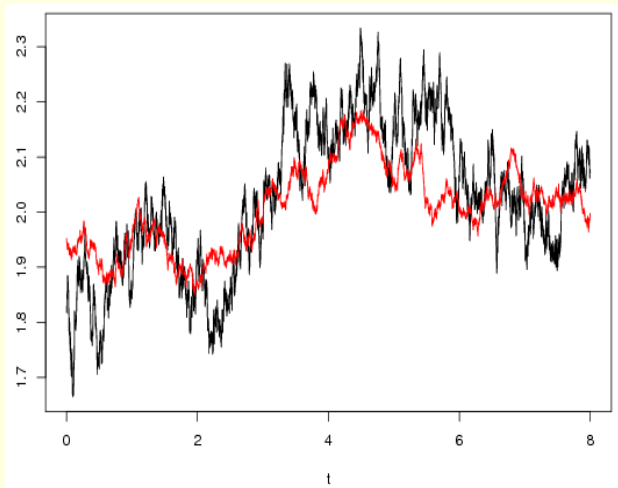


The parameters $(\mu, \Delta t, s)$ can be estimated from the data. In this case: $\mu \approx 1.99$, $\Delta t = 1/\alpha \approx 0.80 \text{ Myr}$, $s \approx 0.12$.

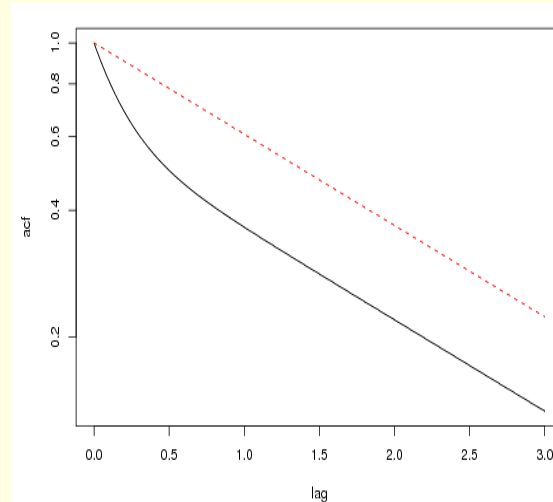
One layer tracking another

$$dX_1(t) = -\alpha_1 (X_1(t) - X_2(t)) dt + \sigma_1 dW_1(t)$$

$$dX_2(t) = -\alpha_2 (X_2(t) - \mu) dt + \sigma_2 dW_2(t)$$



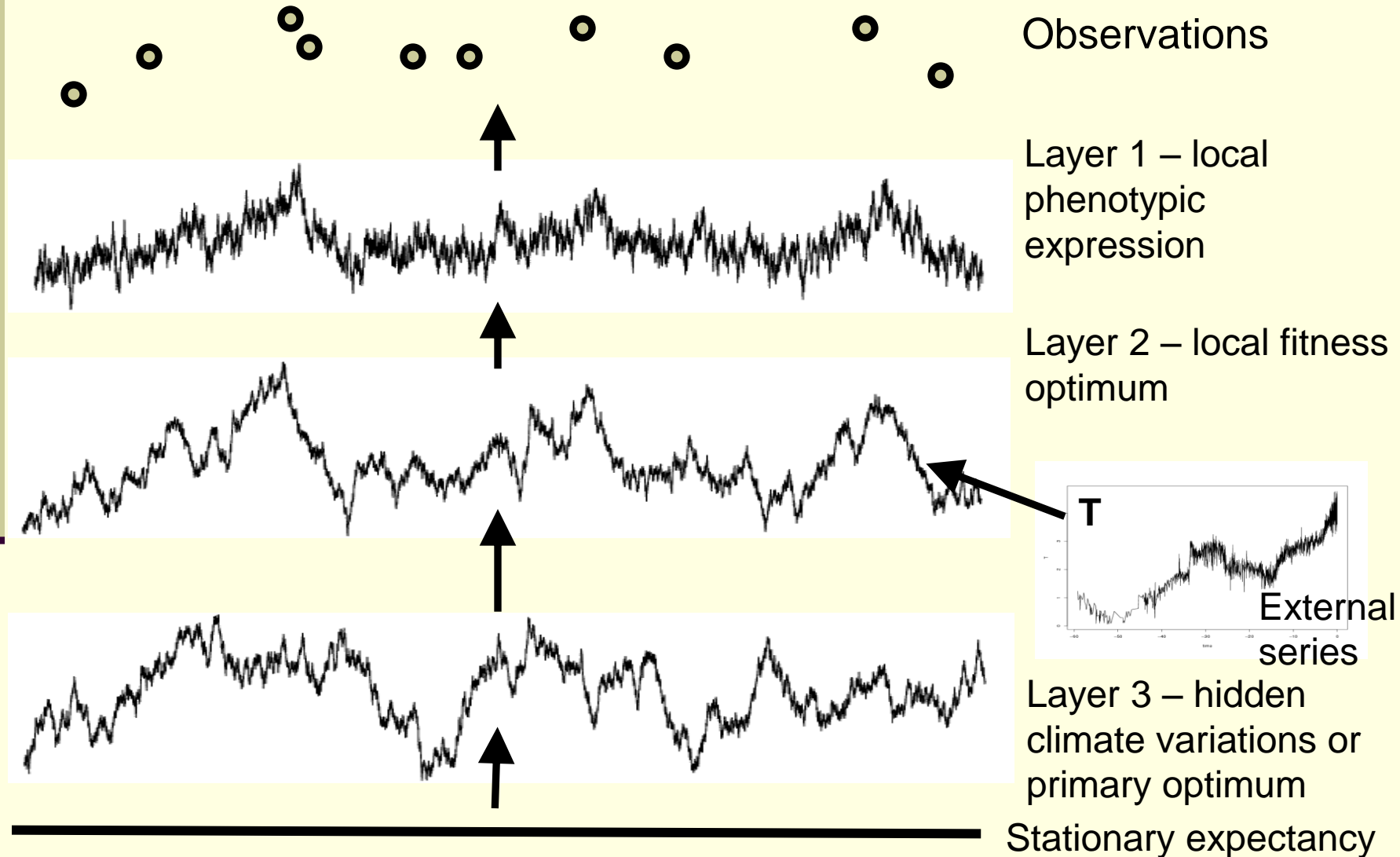
Black process ($\Delta t_1 = 1/\alpha_1 = 0.2$, $s_1 = 2$) tracking red process ($\Delta t_2 = 1/\alpha_2 = 2$, $s_2 = 1$)



Auto-correlation of the upper (black) process, compared to a one-layered SDE model.

A slow-tracking-fast can always be re-scaled to a fast-tracking slow process.
Impose identifying restriction: $\alpha_1 \geq \alpha_2$

Process layers - illustration



Causality in SDE processes

We want to express in process terms the fact that climate affects the phenotypic optimum which again affects the actual phenotype.

Local independence (Schweder 1970):

- Context: Composable continuous time processes.
- Component A is locally independent of B iff transition prob. of A does not depend on B.
- If B is locally dependent on A but not vice versa, we write $A \rightarrow B$.

Local dependence can form a notion of causality in cont. time processes the same way Granger causality does for discrete time processes.

Stochastic differential equation (SDE) vector processes

$X(t)$: p - dimensiona l vector state process

$W(t)$: q - dimensiona l vector Wiener process (Brownian motion)

$\mu(t) = \beta u(t)$: p - dimensiona l level process, β a $p \times r$ regression matrix

A : $p \times p$ dimensiona l pull matrix

Σ $p \times q$ dimensiona l diffusion matrix

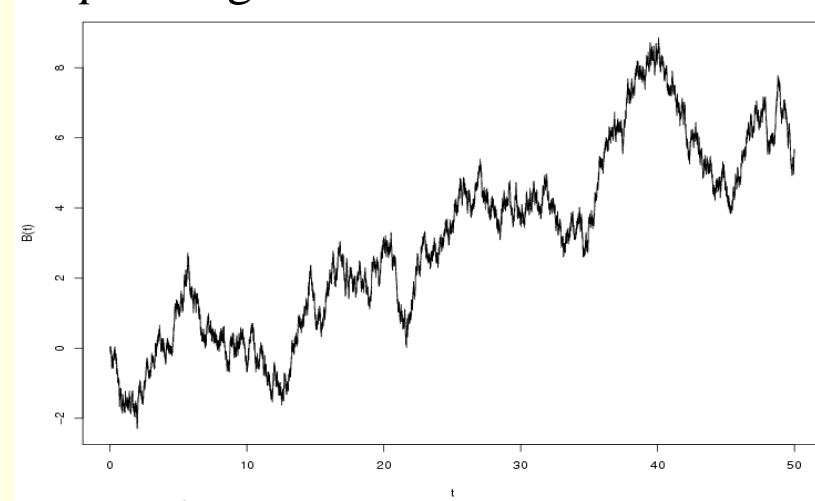
$$dX(t) = (AX(t) + \mu(t))dt + \Sigma dW(t)$$

Ito solution :

$$X(t) = B(t)^{-1} \left\{ X(0) + \int_0^t B(u) \mu(u) du + \int_0^t B(u) \Sigma dW(u) \right\}$$

$$B(t) = e^{-At} = \sum_{n=0}^{\infty} A^n (-t)^n / n! = V^{-1} e^{-\Lambda t} V \text{ (when existing)}$$

Λ is a diagonal eigenvalue matrix of A , V is the left eigenvector matrix.



Stochastic differential equation (SDE) vector processes

Causal structure in linear SDEs :

If we want to have the causal structure $X_L \rightarrow X_{L-1} \rightarrow \dots \rightarrow X_2 \rightarrow X_1$
we would express this as a system of OU - like tracking processes

$$dX_L = -\alpha_L (X_L - \mu)dt + \sigma_L dB_L(t)$$

$$dX_{L-1} = -\alpha_{L-1} (X_{L-1} - X_L)dt + \sigma_{L-1} dB_{L-1}(t)$$

\vdots

$$dX_1 = -\alpha_1 (X_1 - X_2)dt + \sigma_1 dB_1(t)$$

Expressed in vectorial form, this will yield a pull matrix looking like this :

$$A = \begin{pmatrix} -\alpha_1 & \alpha_1 & 0 & \dots & 0 & 0 \\ 0 & -\alpha_2 & \alpha_2 & \dots & 0 & 0 \\ \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ 0 & 0 & 0 & \dots & -\alpha_{L-1} & \alpha_{L-1} \\ 0 & 0 & 0 & \dots & 0 & -\alpha_L \end{pmatrix}$$

The zeros in this matrix determines the local independencies.

Model variants for Coccolith evolution

Model variations:

- 1, 2 or 3 layers (possibly more)
- Inclusion of external time series or extra internal time series in connection to the one we are modelling
- In a single layer:
 - Local or global parameters
 - Correlation between sites (inter-regional correlation)
 - Deterministic response to the lower layer
 - Random walk (no tracking)

Likelihood: Kalman filter

Need a linear, normal Markov chain with independent normal observations:

$$\begin{array}{ccccccc} X(t_1) & \rightarrow & X(t_2) & \rightarrow & \cdots & \rightarrow & X(t_n) & \text{Process} \\ \downarrow & & \downarrow & & & & \downarrow & \\ z_1 & & z_2 & & & & z_n & \text{Observations} \end{array}$$

The Ito solution

$$X(t) = B(t)^{-1} \left\{ X(0) + \int_0^t B(u) \mu(u) du + \int_0^t B(u) \Sigma dW(u) \right\}$$

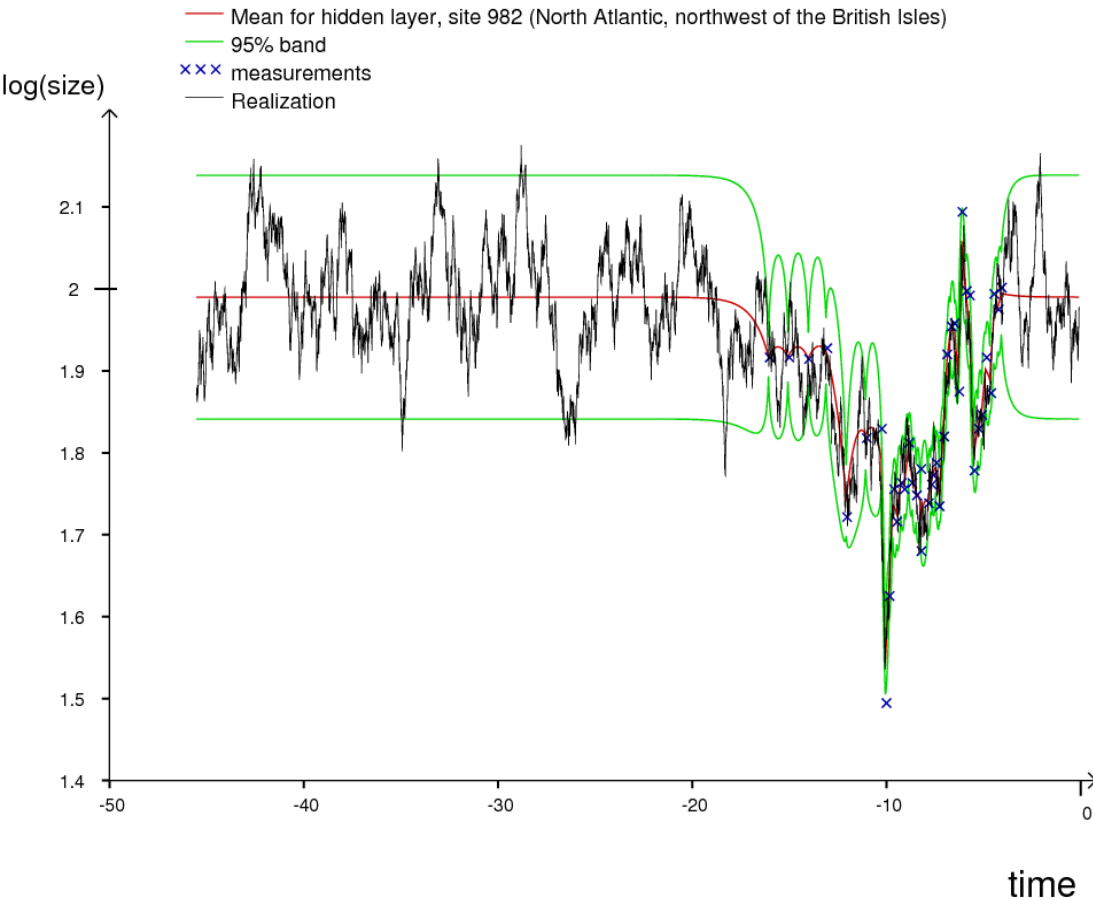
gives, together with measurement variances, what is needed to calculate the likelihood using the Kalman filter:

$$E(z_i | z_1, \dots, z_{i-1}) \equiv m_i \quad \text{and} \quad \text{Var}(z_i | z_1, \dots, z_{i-1}) \equiv v_i$$

$$f(\underline{z} | \theta) = \prod_{i=1}^n f_N(z_i | m_i, v_i, \theta)$$

Kalman smoothing (state inference)

A tracking model with 3 layers and fixed parameters



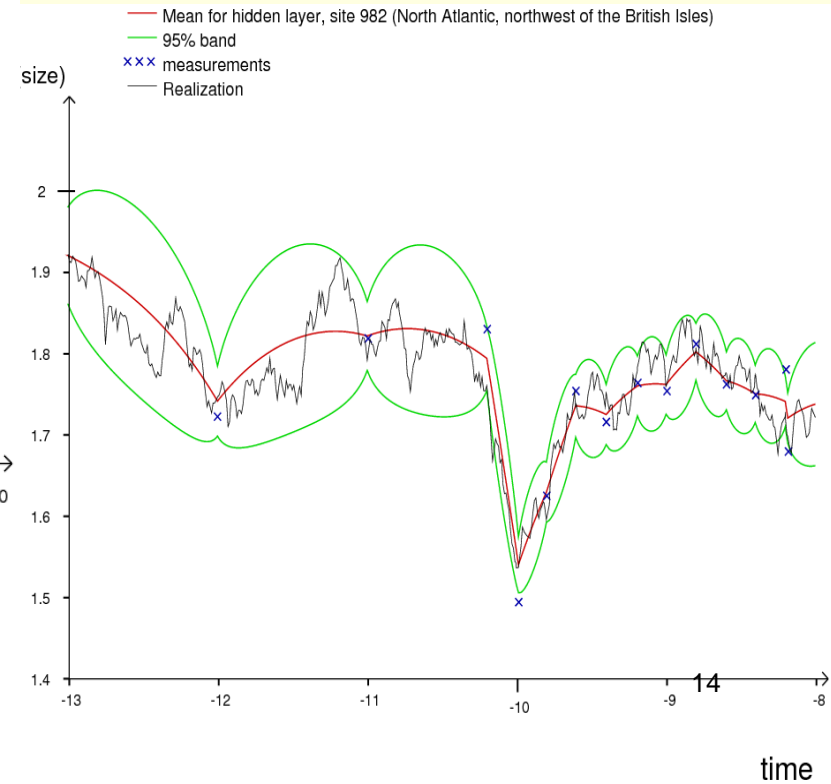
Snapshot:

North Atlantic

Red curve: expectancy

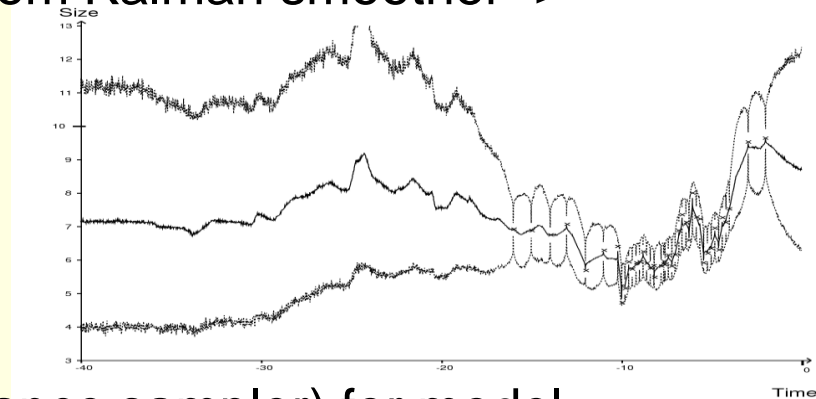
Black curve: realization

Green curve: uncertainty



Parameter and model inference

- Wide but informative prior distributions respecting identifying restrictions
- MCMC for parameter inference on individual models:
- MCMC samples+state samples from Kalman smoother -> Possible to do inference on the process state conditioned only on the data.



- Model likelihood (using an importance sampler) for model comparison (posterior model probabilities).

- Posterior weight of a property C from model likelihoods:

$$W(C) = \frac{\frac{1}{n_C} \sum_{M \in C} f(\text{data} | M)}{\sum_{C'} \frac{1}{n_{C'}} \sum_{M \in C'} f(\text{data} | M)} \quad 15$$

Bayesian inference on the 723 models

Table 2. Posterior weights for different properties. Number of models in parenthesis.

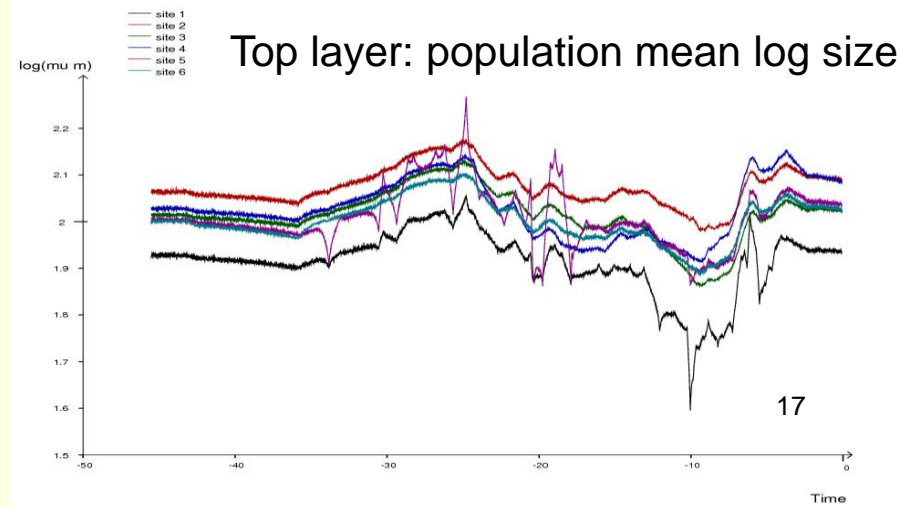
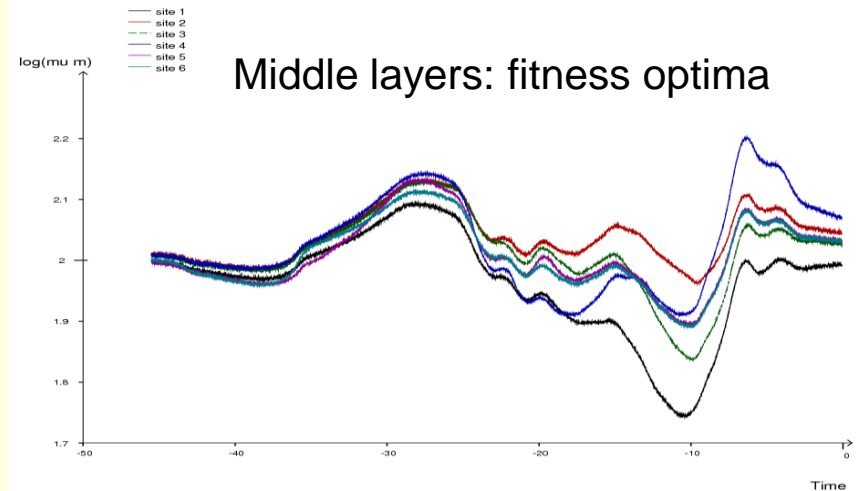
Property	Options			
Number of Layers:	1	2	3	
	4.3% (19)	45.8% (112)	49.9% (592)	
Regionality in:	none or μ_0	pull	diffusion	
	0.15% (252)	79.5% (300)	20.4% (171)	
Regional parameters in:	no layer	upper layer	middle layer	lower layer
	0.09% (252)	20.7% (144)	23.1% (180)	56.1% (147)
Inter-regional correlations:	none	intermediate (6D)	perfect (1D)	
	19.0% (39)	62.3% (169)	18.7% (515)	
Random walk in lowest layer	no	yes		
	90.4% (393)	9.6% (330)		

$$W(C) = \frac{\frac{1}{n_C} \sum_{M \in C} f(data | M)}{\sum_{C'} \frac{1}{n_{C'}} \sum_{M \in C'} f(data | M)}$$

Results

- Best 5 models in good agreement. (together, 19.7% of summed integrated likelihood):

- Three layers.
- Common expectancy in bottom layer .
- No impact of exogenous temperature series.
- Lowest layer: Inter-regional correlations, $\rho \approx 0.5$. Site-specific pull.
- Middle layer: Intermediate tracking.
- Upper layer: Very fast tracking.



Phenotypic evolution on a phylogenetic tree: Body size of primates

Measurements $i \in \{1, \dots, 209\}$ (92 extant measurements, 117 fossils)

$X_{i,1}(t)$ = mean phenotype (log mass)

$X_{i,2}(t)$ = fitness optimum

$X_{i,3}(t)$ = underlying environment, primary optimum

$X_{i,3} \rightarrow X_{i,2} \rightarrow X_{i,1}$ same tracking model as for coccoliths

Phylogenetic tree :

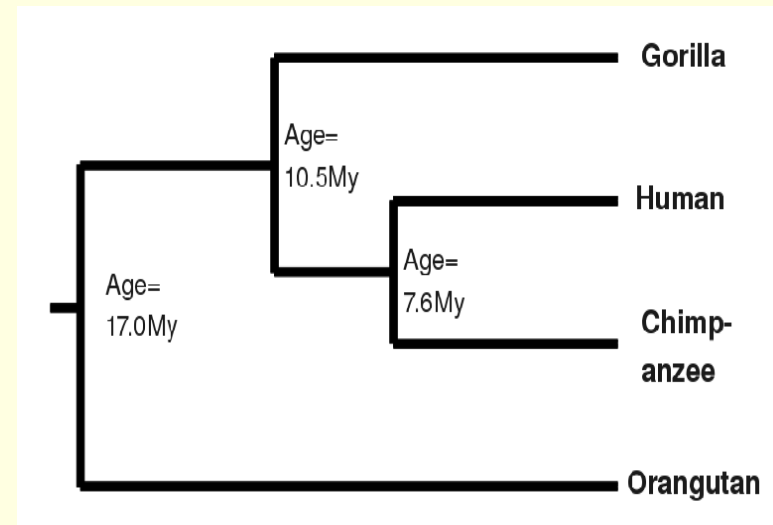
$T_{i,j}$ = Time when i split from j .

$X_{i,l}(t) = X_{j,l}(t)$ when $t \leq T_{i,j}$.

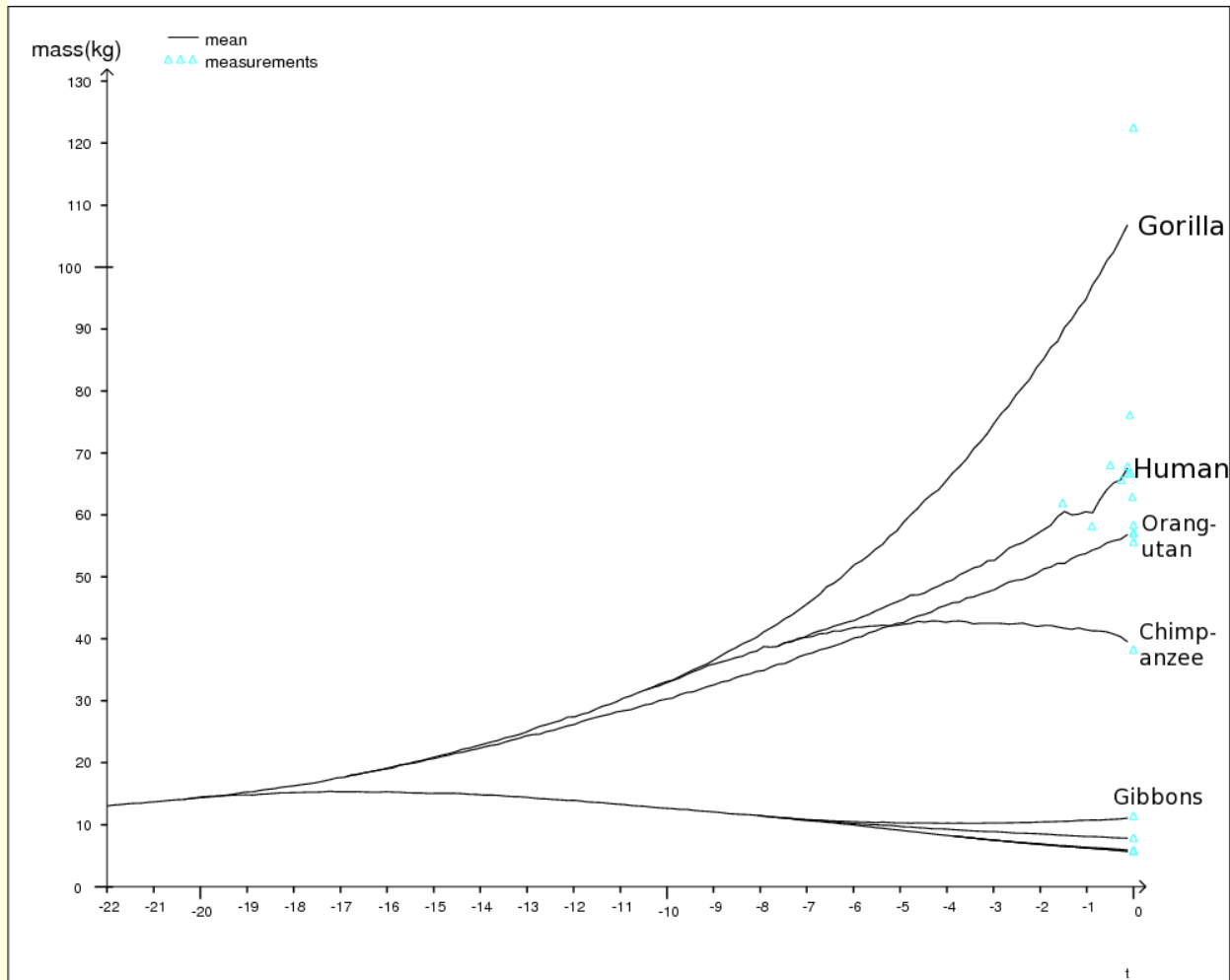
Condition on these equalities for the given tree .

Observations $Z_{i,k} = X_{i,k}(t_k) + N(0, \sigma_k^2)$ will

be multi - normally distributed with mean vector and covariance matrix as parametric expressions from the SDE model.



First results for some primates



Why linear SDE processes?

- Parsimonious: Simplest way of having a stochastic continuous time process that can track something else.
- Tractable: The likelihood, $L(\theta) \equiv f(\text{Data} \mid \theta)$, can be analytically calculated by the Kalman filter or directly by the parameterized multi-normal model for the observations. (θ = model parameter set)
- Can code for causal structure (local dependence/independence).
- Some justification in biology, see Lande (1976), Estes and Arnold (2007), Hansen (1997), Hansen et. al (2008).
- Great flexibility, widely applicable...
- Thinking and modelling might then be more natural in continuous time.
- Allows for varying observation frequency, missing data and observations arbitrary spaced in time.
- Extensions:
 - Non-linear SDE models...
 - Non-Gaussian instantaneous stochasticity (jump processes).

Bibliography

- Commenges D and Gégout-Petit A (2009), A general dynamical statistical model with causal interpretation. *J.R. Statist. Soc. B*, 71, 719-736
- Lande R (1976), Natural Selection and Random Genetic Drift in Phenotypic Evolution, *Evolution* 30, 314-334
- Hansen TF (1997), Stabilizing Selection and the Comparative Analysis of Adaptation, *Evolution*, 51-5, 1341-1351
- Estes S, Arnold SJ (2007), Resolving the Paradox of Stasis: Models with Stabilizing Selection Explain Evolutionary Divergence on All Timescales, *The American Naturalist*, 169-2, 227-244
- Hansen TF, Pienaar J, Orzack SH (2008), A Comparative Method for Studying Adaptation to a Randomly Evolving Environment, *Evolution* 62-8, 1965-1977
- Schuss Z (1980). *Theory and Applications of Stochastic Differential Equations*. John Wiley and Sons, Inc., New York.
- Schweder T (1970). Decomposable Markov Processes. *J. Applied Prob.* 7, 400–410

Source codes, examples files and supplementary information can be found at <http://folk.uio.no/trondr/layered/>.