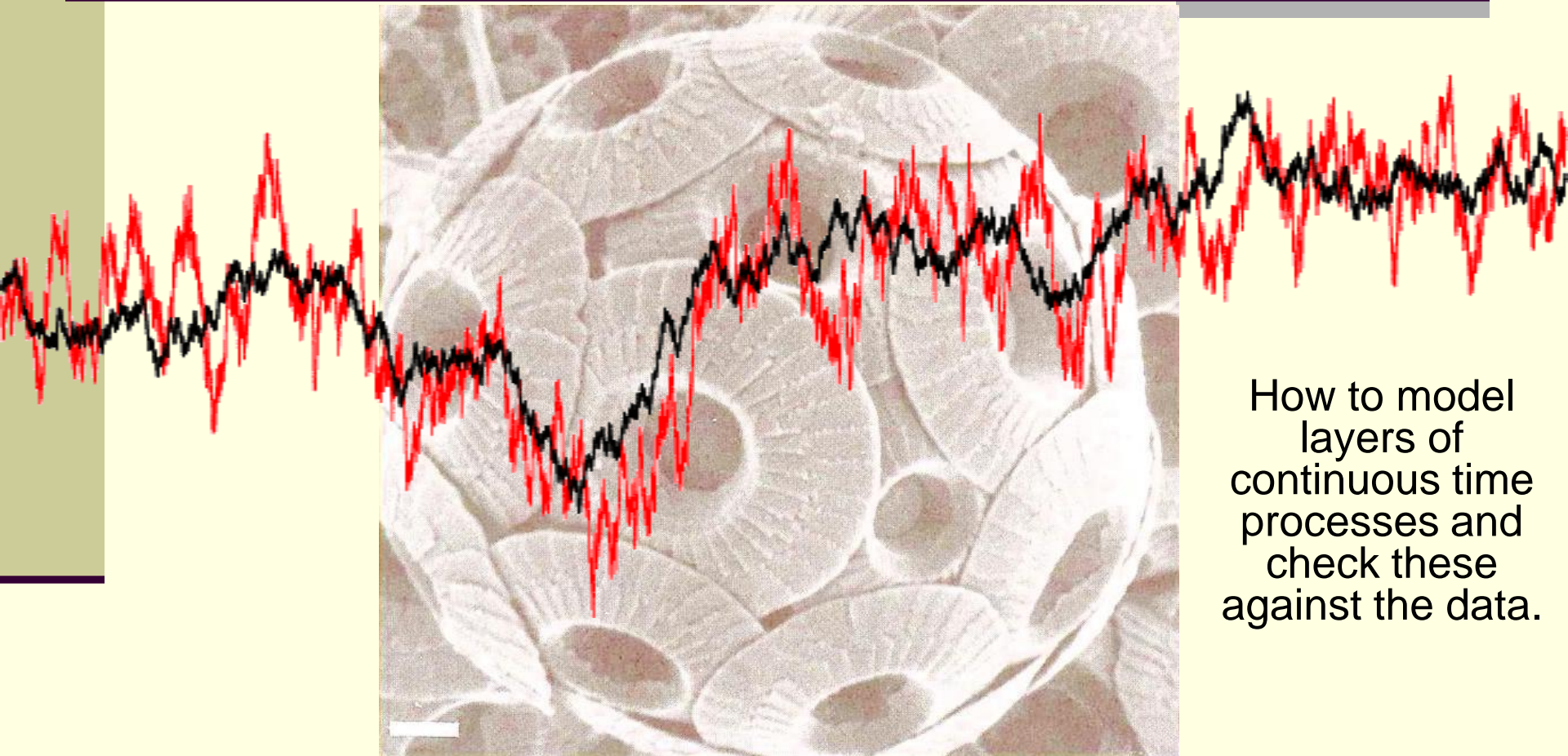


Modelling phenotypic evolution using layered stochastic differential equations (with applications for Coccolith data)



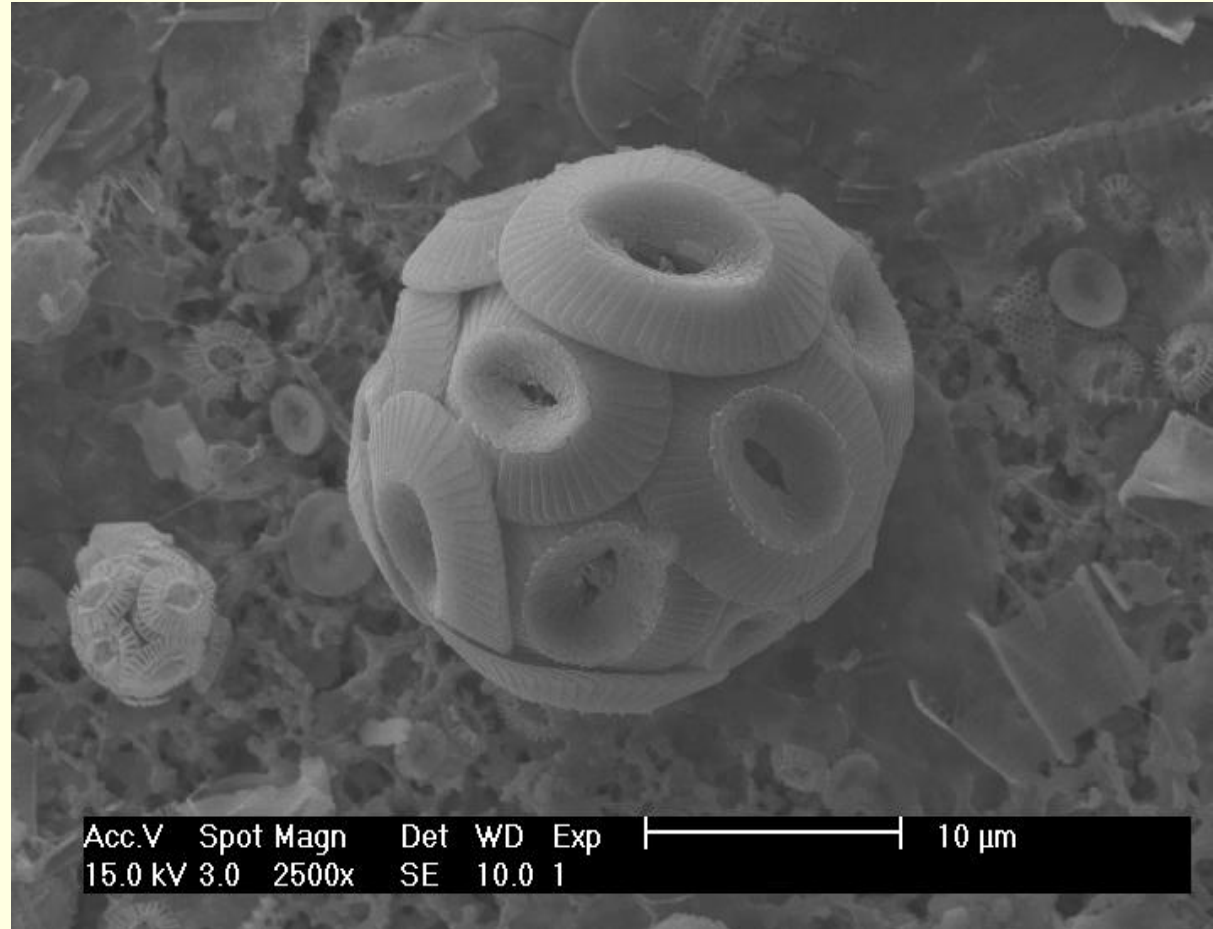
How to model layers of continuous time processes and check these against the data.

Trond Reitan
CEES, institute of
biology, UiO

Our data - Coccoliths

Measurements are on the diameter of Coccoliths from marine, calcifying single cell algae (Haptophyceae).

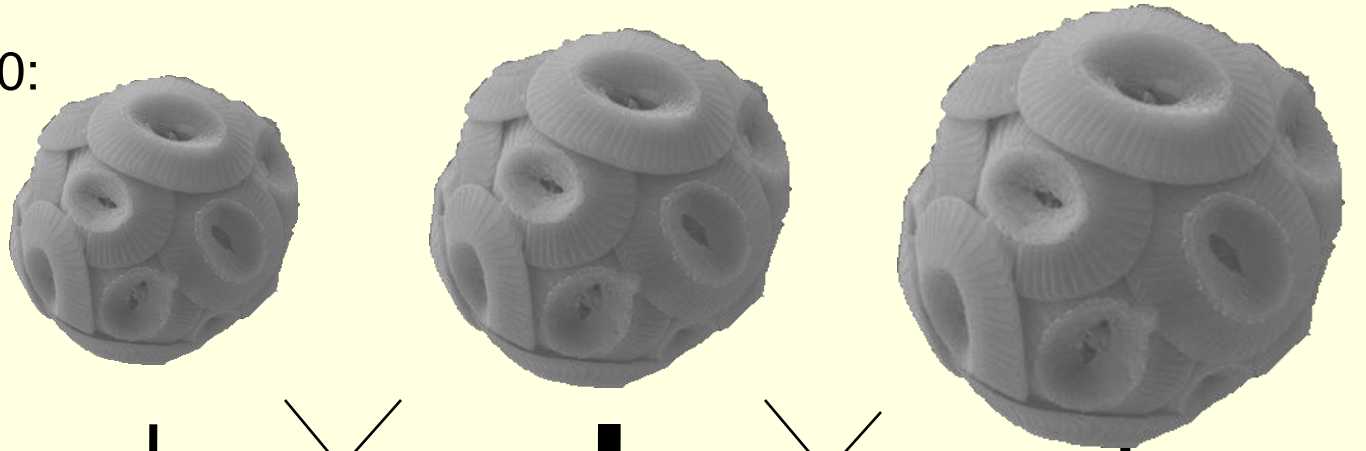
These are covered by calcite platelets ('coccoliths' forming a 'coccosphere')



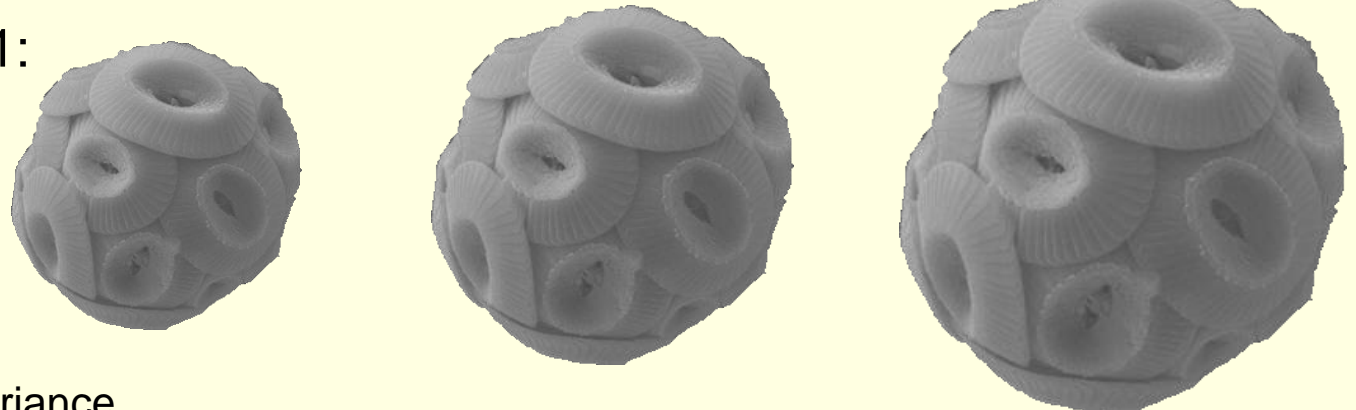
SEM image: J. Henderiks, Stockholm University

Size (phenotype) variation

Generation 0:



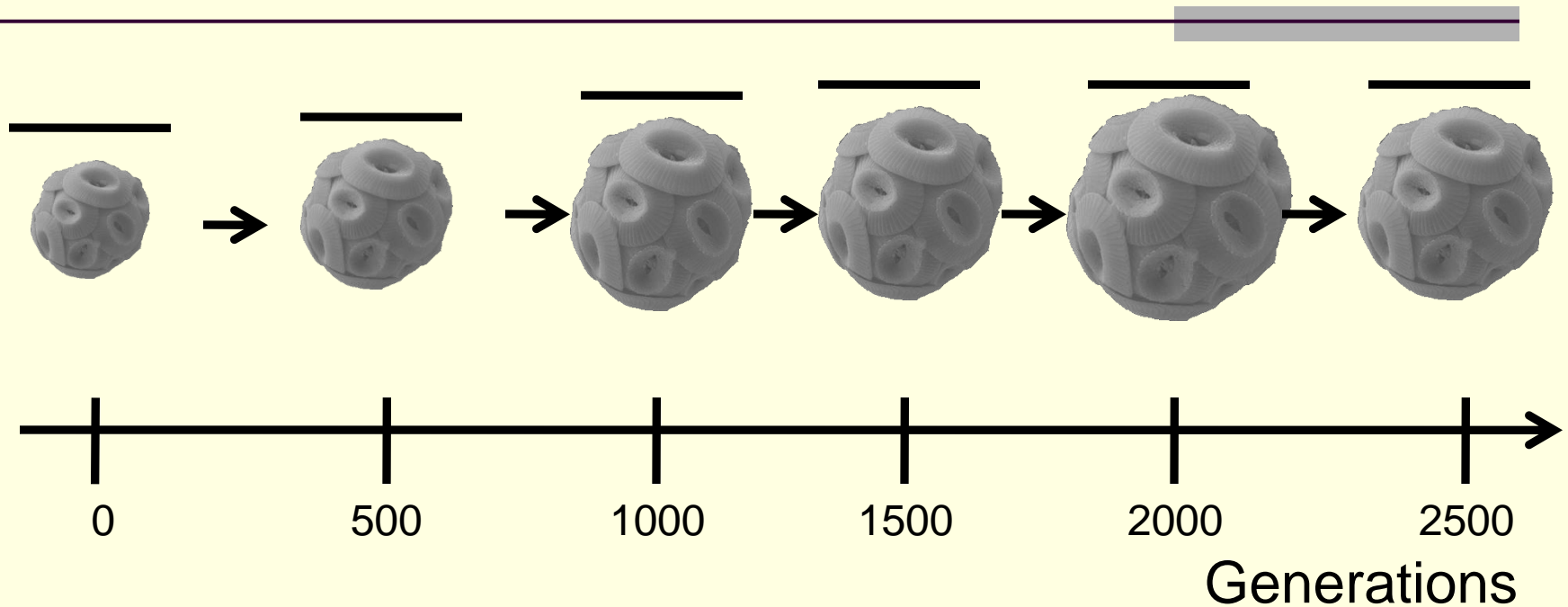
Generation 1:



Assume that the variance is maintained

Optimum

Change of the average in time

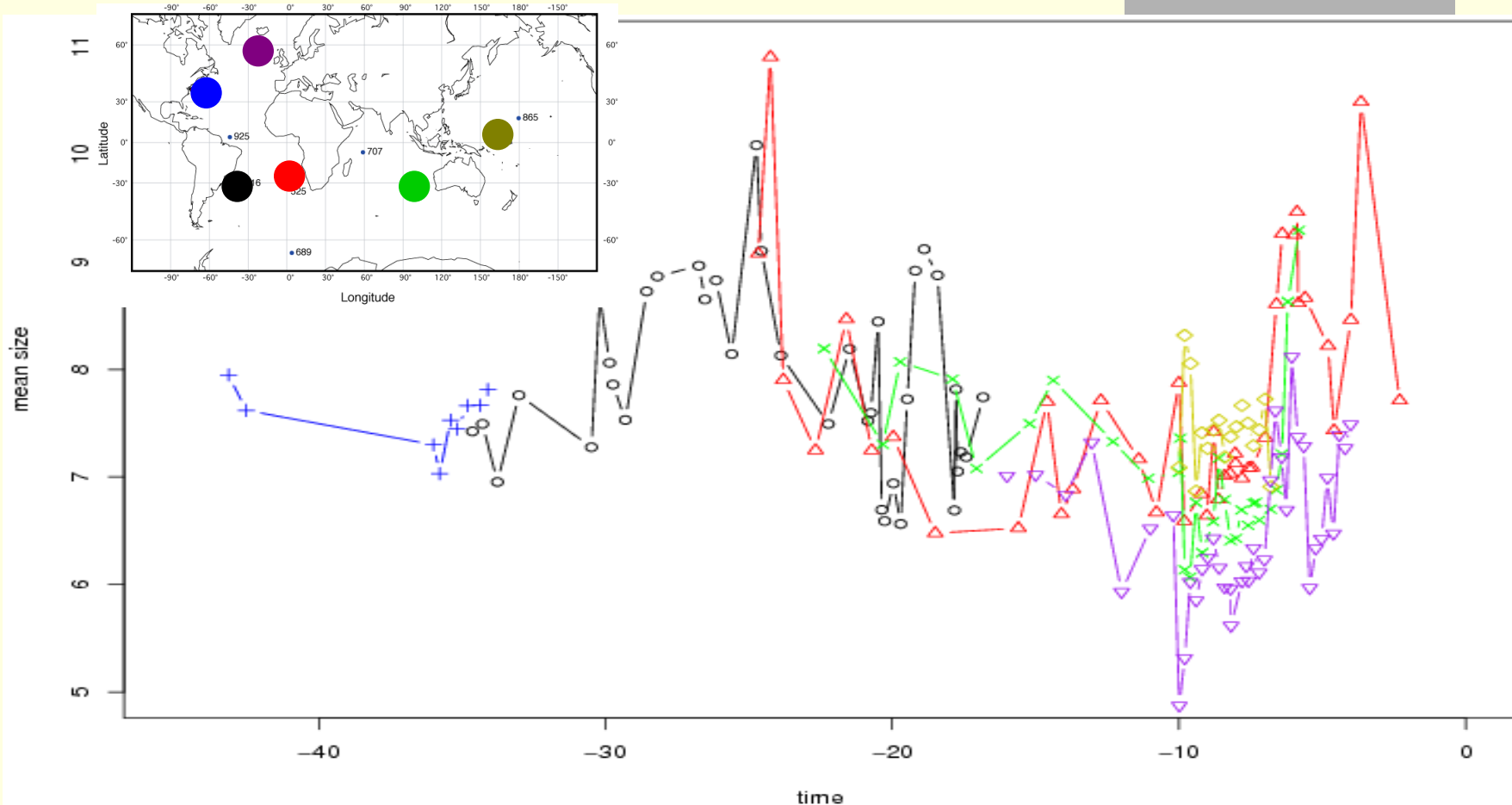


Mean changes. Pulled towards the optimum, but always with some fluctuations. => Stochastic process

Lande, 1976: This is an Ornstein-Uhlenbeck process

But: The optimum may also change!

Our data – Coccolith size measurements



For each site and each observed time: The size of several (1-400) Coccoliths was measured. Mean size shown. (For analysis, the data was log-transformed. After data massage: n=178 different mean sizes.)

Concepts

- Data: Several size measurements for different ages and sites. => average and variance
- Should express something about an underlying set of processes, optima-layers, belonging to the lineage.
- Non-equidistant time series: Continuous in time
- Stochastic
- Can we use the data to say something about the processes?

Background

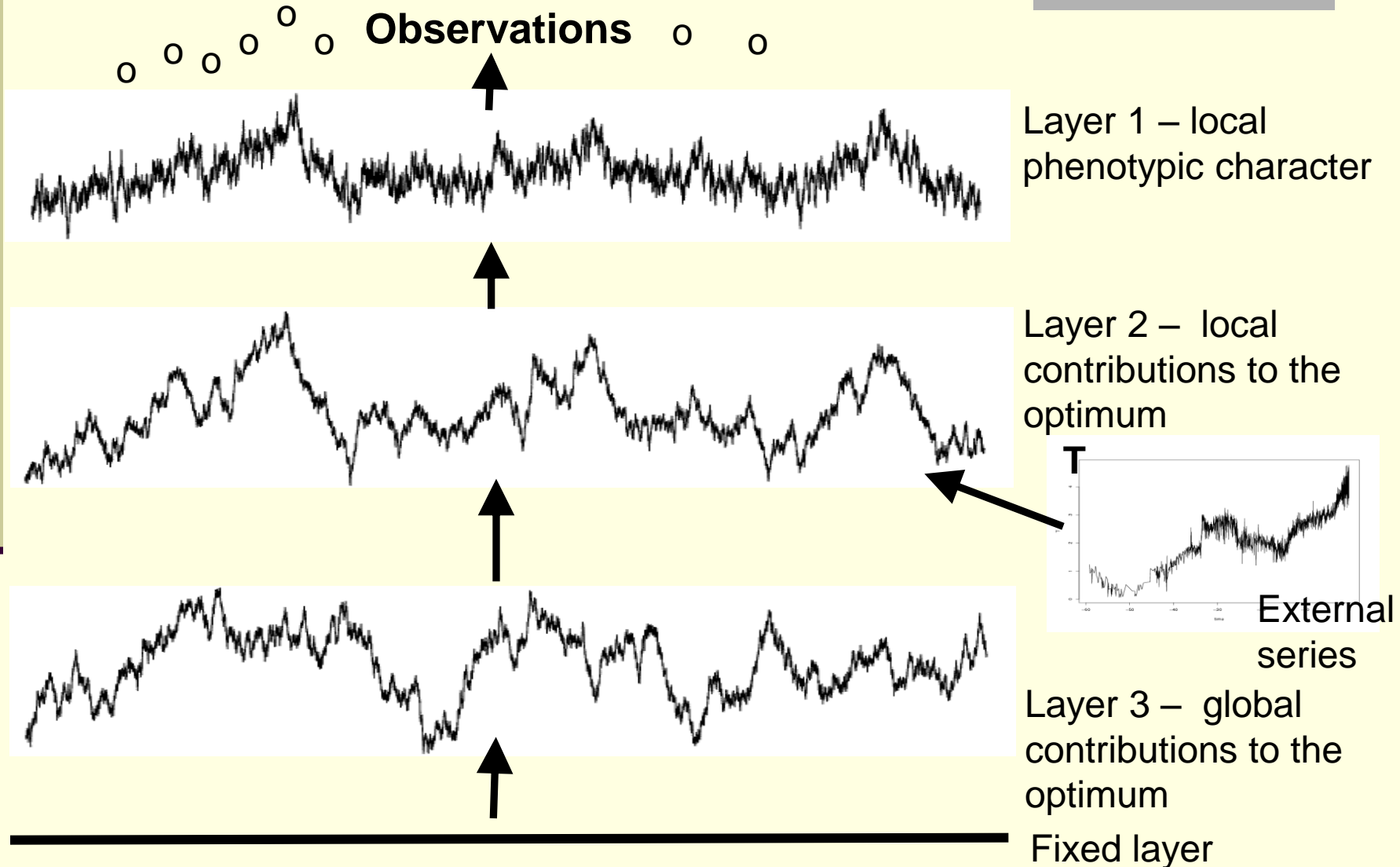
- Data: Jorijntje Henderiks.
- Tore Schweder: Idea and mathematical foundation.
- Me: Inference, programming and control calculations.

Multiple layers of hidden processes

– why?

- Measured mean size is a noisy indicator of overall mean size at a given moment.
- Even with perfect measurements, what happens between needs inference.
- A phenotype character will track an evolutionary optimum (natural selection).
- The optimum changes also. Can be further divided into layers describing global and local contributions.
- Each layer is responding to what happens in a lower layer.

Process layers - illustration

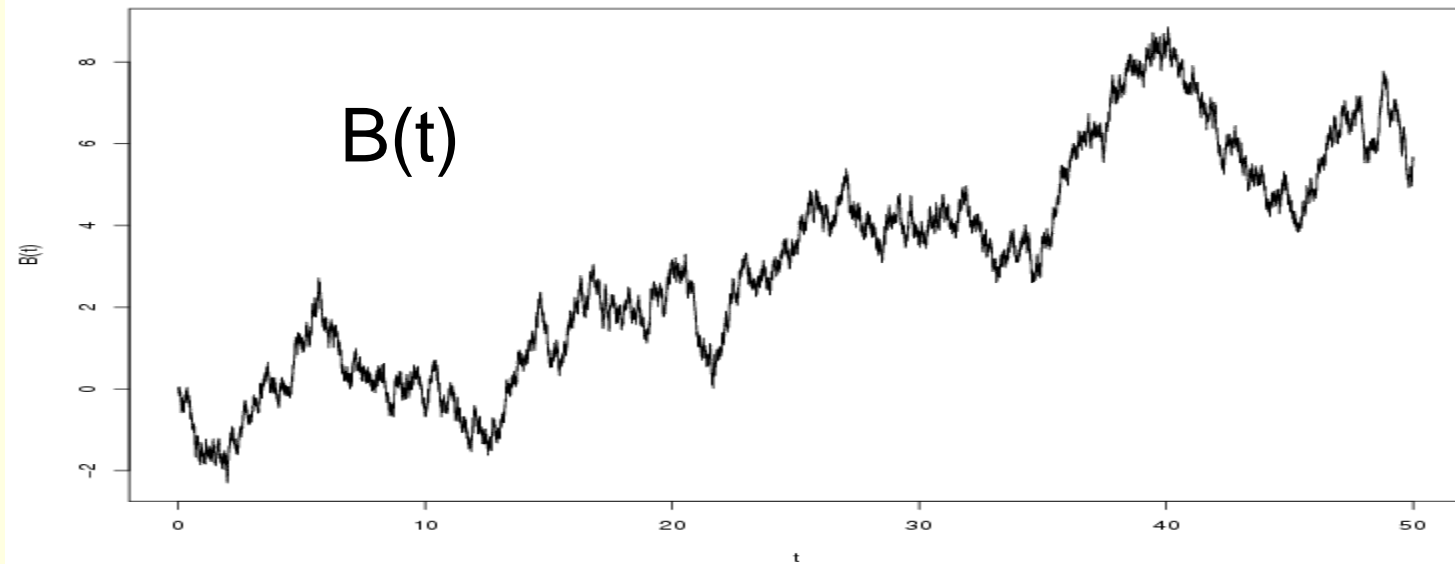


Variants

- Can have different number of layers.
- In a single layer, one has the choice of:
 - Local or global parameters
 - The stochastic contribution can be global or local (site-specific)
 - Correlation between sites (inter-regional correlation)
 - Deterministic response to the lower layer
 - Random walk (not responding to anything else, no tracking)
- In total: 750 models examined

The toolbox – stochastic differential equations, the Wiener process

- Need something that is continuous in time, has a direction and a stochastic element.
- Stochastic differential equations (SDEs):
Combines differential equations with the Wiener process.
- Wiener process: continuous time random walk

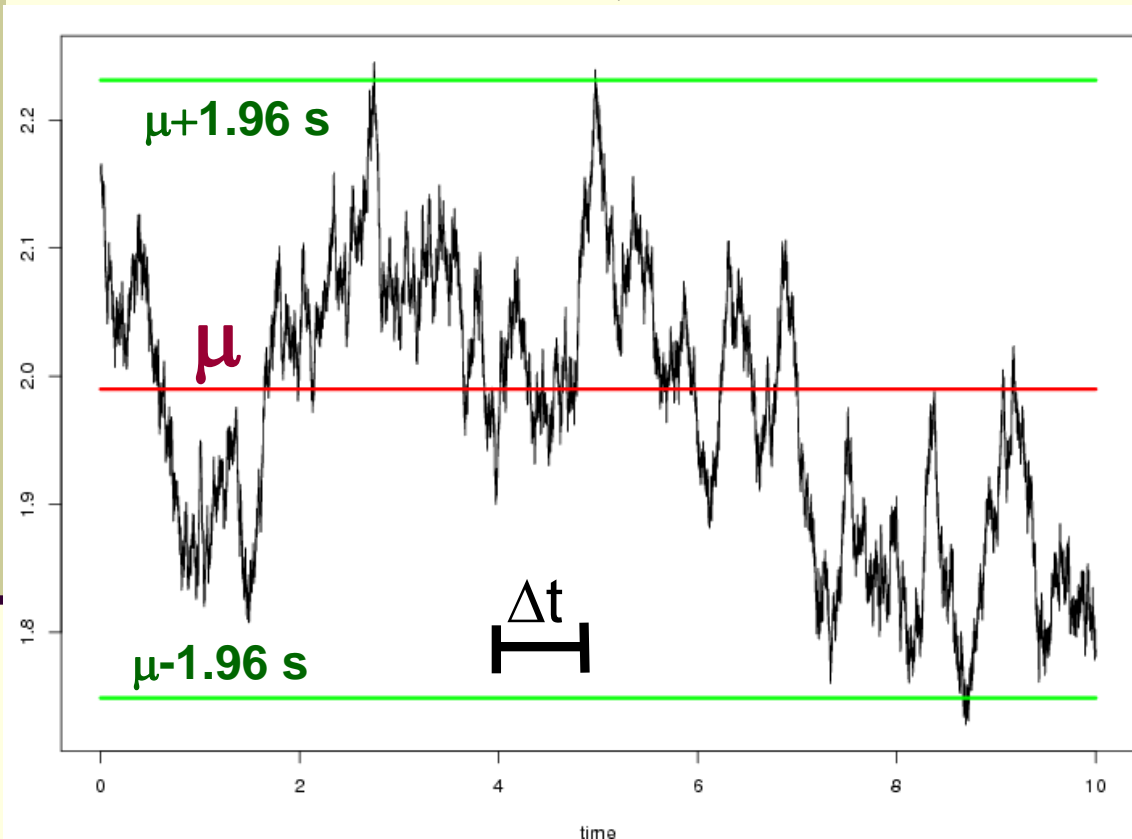


The Ornstein-Uhlenbeck (mean-reverting) process

$$dY(t) = -a(Y(t) - \mu)dt + \sigma dB(t)$$

Attributes:

- Normally distributed
- Markovian
- Expectancy: μ
- Standard deviation: $\mathbf{s} = \sigma / \sqrt{2a}$
- a : pull
- Time for the correlation to drop to $1/e$: $\Delta t = 1/a$

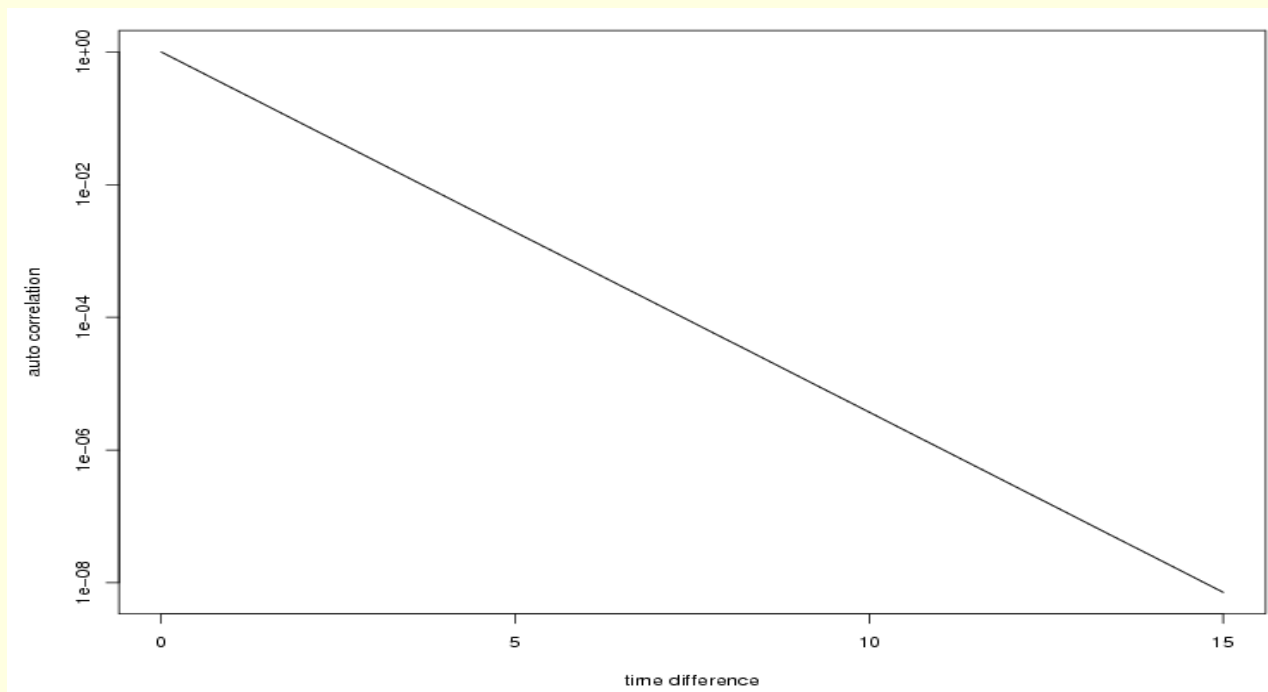


The parameters (μ , Δt , \mathbf{s}) can be estimated from the data. In this case: $\mu \approx 1.99$, $\Delta t \approx 0.80 \text{ Myr}$, $\mathbf{s} \approx 0.12$.

Autocorrelation

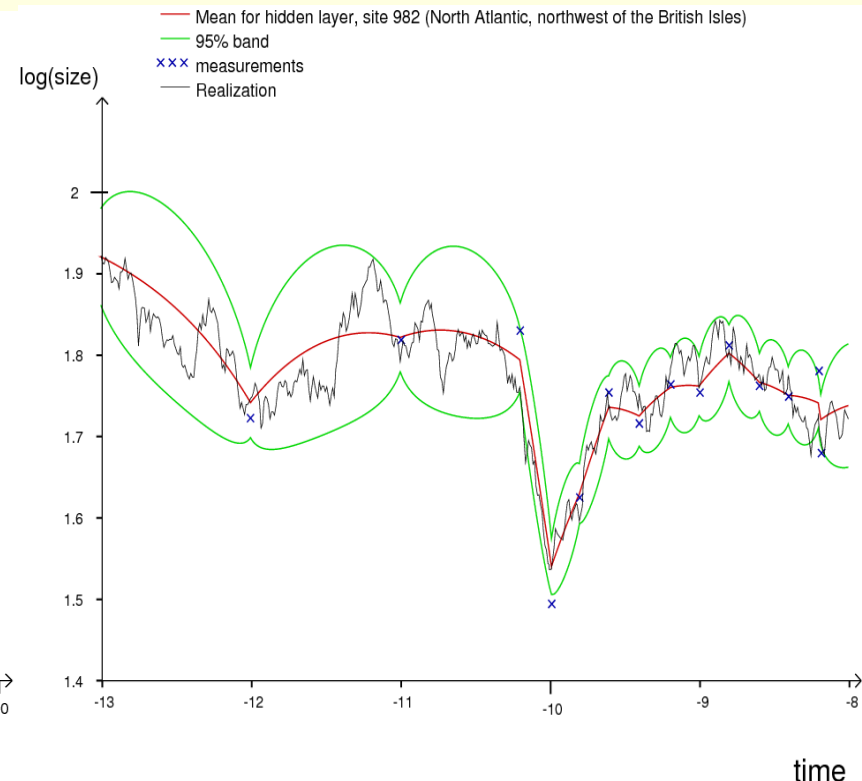
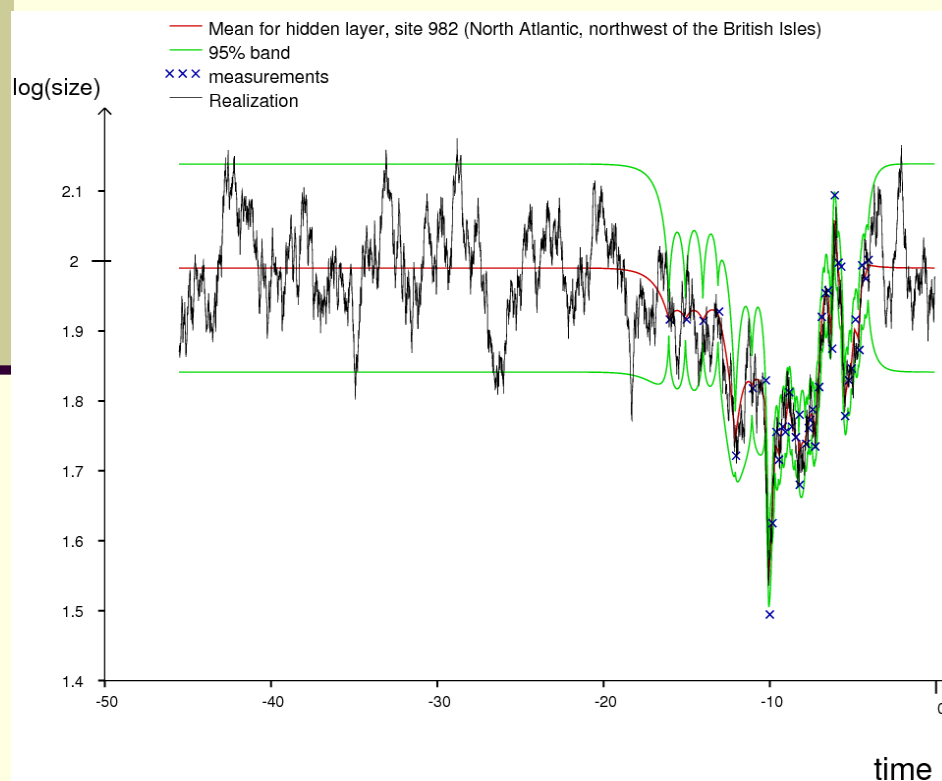
- The autocorrelation function is the correlation between the process state at one time, t_1 , and the process state a later time, t_2 .
- A function of the time difference, $t_2 - t_1$.
- For a single layer OU process,

$$c(t_2 - t_1) \equiv \text{corr}(x(t_1), x(t_2)) = e^{-(t_2 - t_1) / \Delta t}$$



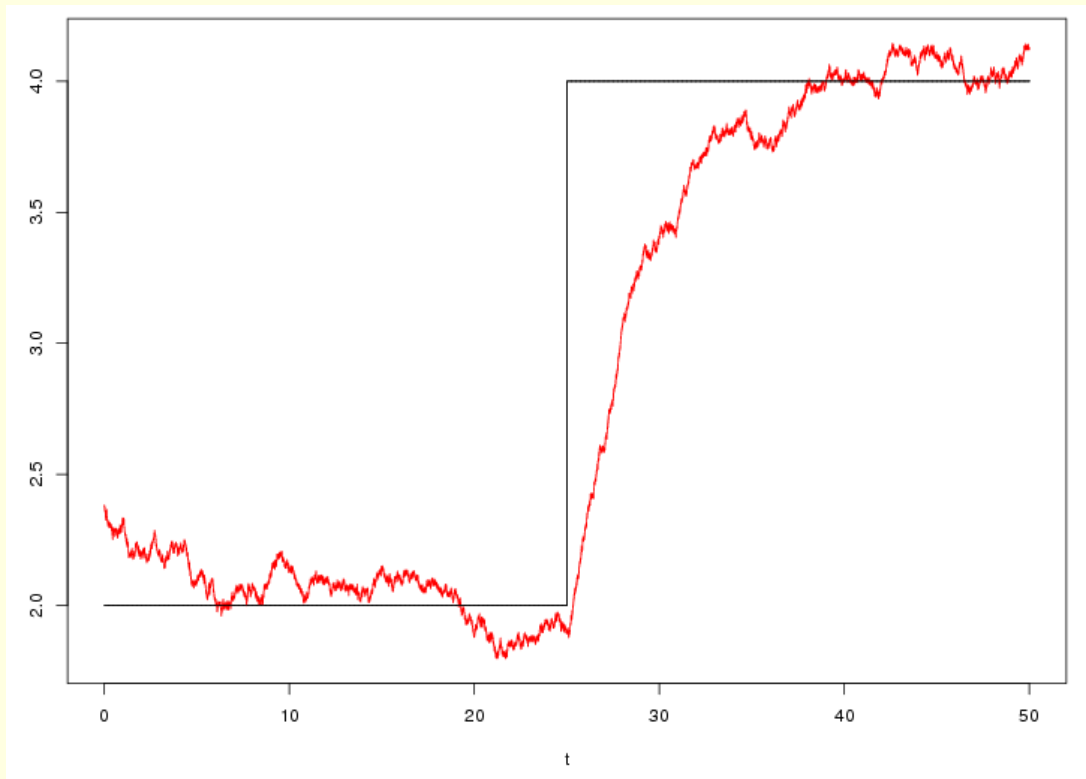
Curiosity: OU process conditioned on data

- It's possible to realize an OU process conditioned on the data, also (using the Kalman framework, which will be described later).



Ou process tracking another process

$$dY(t) = -a(Y(t) - \mu(t))dt + \sigma dB(t)$$



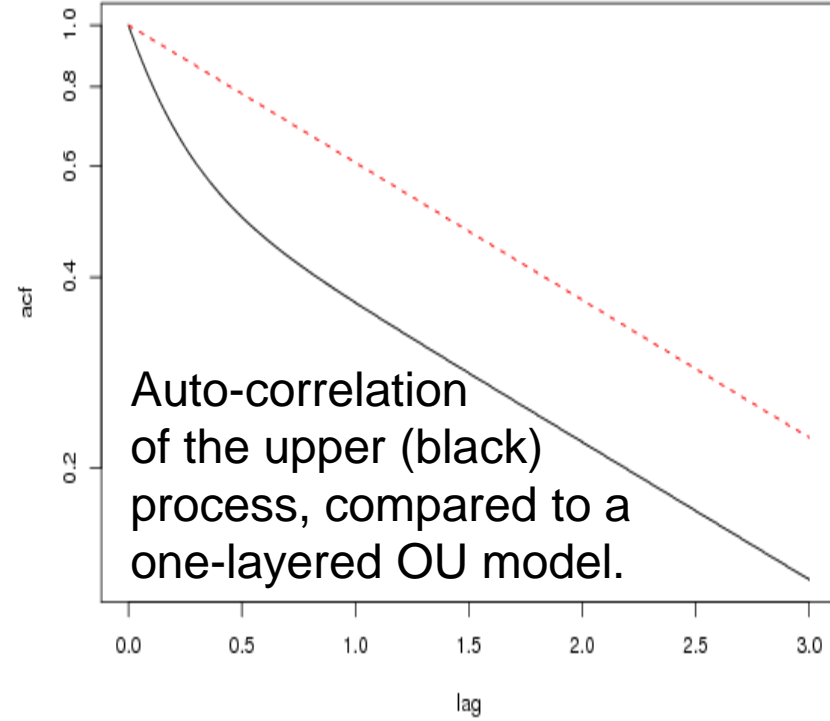
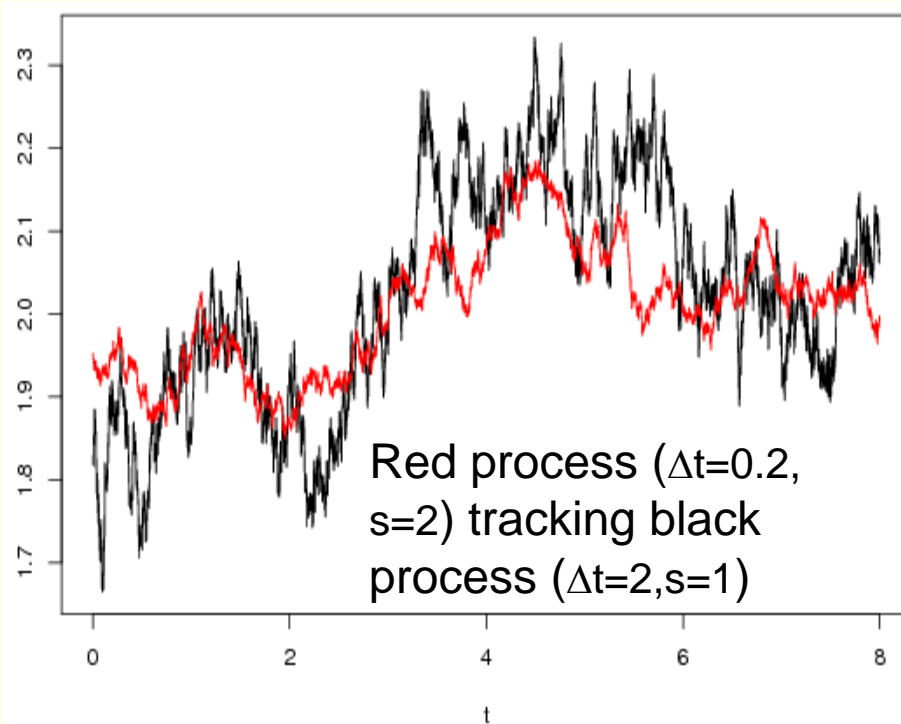
Tracks the underlying function/process, $\mu(t)$, with response time equal to the characteristic time, $\Delta t = 1/a$

Idea: Let the underlying process be expressed the same way.

OU-like process tracking another OU process

$$dY(t) = -a(Y(t) - \mu(t))dt + \sigma dB(t)$$

$$d\mu(t) = -b(\mu(t) - \mu_0)dt + \omega dW(t)$$



Vectorial linear stochastic differential equation

The two coupled SDEs on the previous slide can be written as:

$$d\underline{X}(t) = [A\underline{X}(t) + \underline{m}(t)]dt + \Sigma d\underline{W}(t) \quad \text{when}$$

$$\underline{X}(t) = \begin{pmatrix} y(t) \\ \mu(t) \end{pmatrix} \quad A = \begin{pmatrix} -a & a \\ 0 & -b \end{pmatrix} \quad \underline{m}(t) = \begin{pmatrix} 0 \\ b\mu_0 \end{pmatrix} \quad \Sigma^2 = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \omega^2 \end{pmatrix}$$

Generalization to 3 layers and several sites per layer:

$$\underline{X}(t) = \begin{pmatrix} \underline{X}_1(t) \\ \underline{X}_2(t) \\ \underline{X}_3(t) \end{pmatrix} \quad A = \begin{pmatrix} -A_1 & A_1 & 0 \\ 0 & -A_2 & A_2 \\ 0 & 0 & A_3 \end{pmatrix} \quad \underline{m}(t) = \begin{pmatrix} 0 \\ 0 \\ A_3 \underline{\mu}_0 \end{pmatrix} \quad \Sigma^2 = \begin{pmatrix} \Sigma_1^2 & 0 & 0 \\ 0 & \Sigma_2^2 & 0 \\ 0 & 0 & \Sigma_3^2 \end{pmatrix}$$

Solving vectorial linear SDEs

Solve by eigen-representation: $VA = \Lambda V$

Eigenvectors: V

Eigenvalues, $\Lambda = \text{diag}(\underline{\lambda})$

Formal solution:

$$\underline{X}(t) = V^{-1} e^{\Lambda t} V \underline{X}(0) + V^{-1} \int_0^t e^{\Lambda(t-u)} V \underline{m}(u) du + V^{-1} \int_0^t e^{\Lambda(t-u)} V \Sigma d\underline{B}(u)$$

Gaussian process, only expectation and covariance needed!

$$E \underline{X}(t) = V^{-1} e^{\Lambda t} V \underline{X}(0) + V^{-1} \int_0^t e^{\Lambda(t-u)} V \underline{m}(u) du$$

$$\text{cov}(\underline{X}(v), \underline{X}(t)) = V^{-1} \left[\int_0^v e^{\Lambda(v-u)} V \Sigma^2 V' e^{\Lambda(t-u)} du \right] (V^{-1})'$$

Why linear SDE processes?

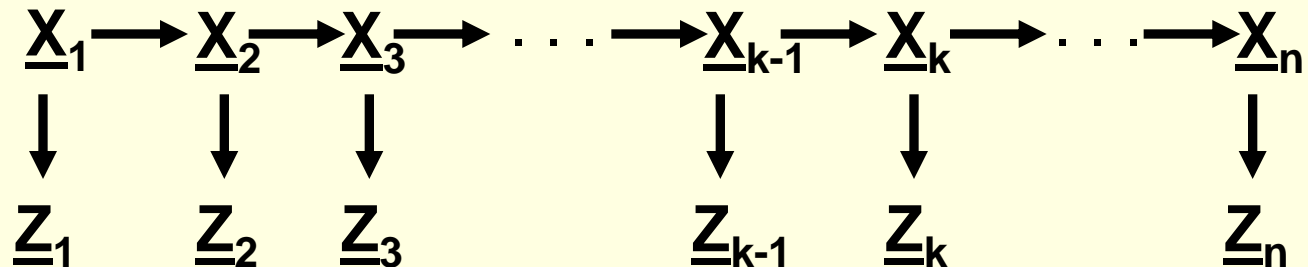
- Parsimonious: Simplest way of having a stochastic continuous time process that can track something else.
- Tractable: The likelihood, $L(\theta) \equiv f(\text{Data} | \theta)$, can be analytically calculated. (θ = model parameter set)
- Some justification from biology, see Lande (1976), Estes and Arnold (2007), Hansen (1997), Hansen et. al (2008).
- Great flexibility...

Inference

- Don't know the details of the model or the model parameters, θ . Need to do inference.
- Classic: Search for $\hat{\theta} \equiv \arg \max_{\theta} f(Data | \theta)$
 - Use BIC for model comparison.
- Bayesian: $f(\theta | Data) = \frac{f(Data | \theta) f(\theta)}{\int f(Data | \theta) f(\theta) d\theta}$
 - Need a prior distribution, $f(\theta)$, on the model parameters.
 - Use $f(Data|M)$ for model comparison.
- Technical: Numeric methods for both kinds of analysis.
 - ML: Multiple hill-climbing
 - Bayesian: MCMC + Importance sampling

Calculating the likelihood - Kalman filtering

- Basis:
 - Hidden linear normal Markov process,
 - Observations centered normally around the hidden process,



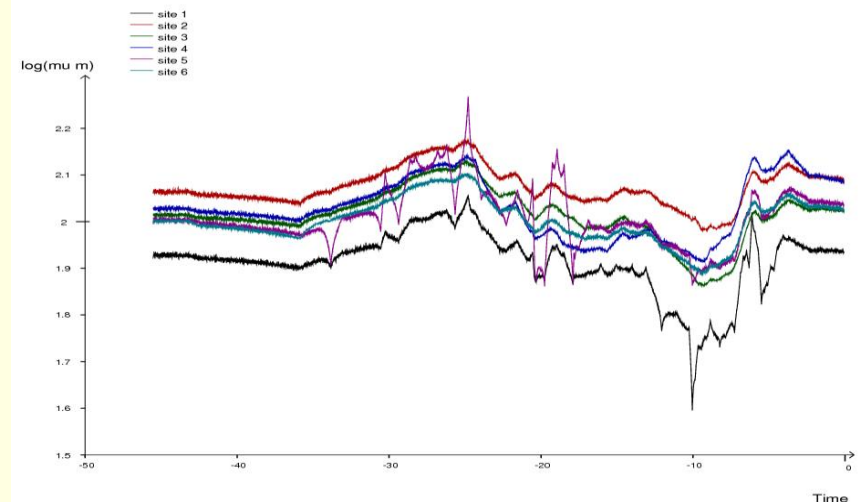
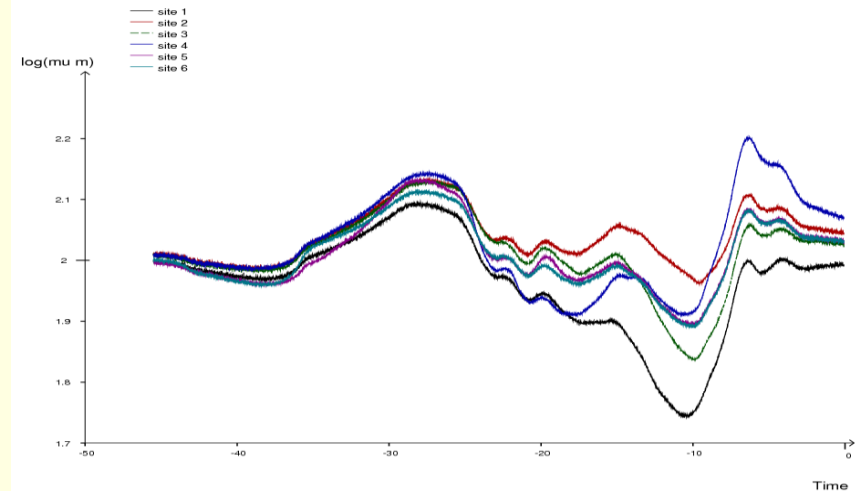
- We can find the transition and covariance matrices for the processes.
- Analytical results for doing inference on a given state and observation given the previous observations.
- Can also do inference on the state given all the observations (Kalman smoothing)

Do we have enough data for full model selection?

- Assume the data has been produced by a given model in this framework.
 - Can we detect it with the given amount of data?
 - How much data is needed in order to reliably detect this by classic and Bayesian means?
- Check artificial data against the original model plus 25 likely suspects.
- So far: Slight tendency to find the correct number of layers with the Bayesian approach. BIC seems generally too stingy on the number of layers.

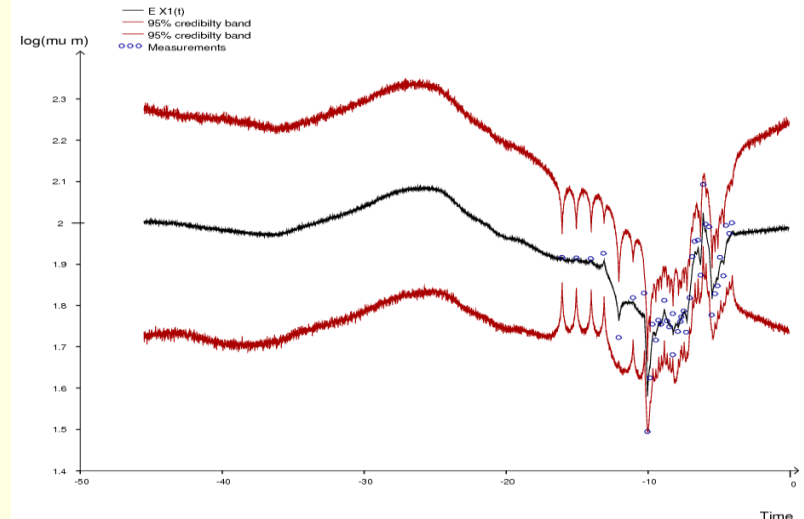
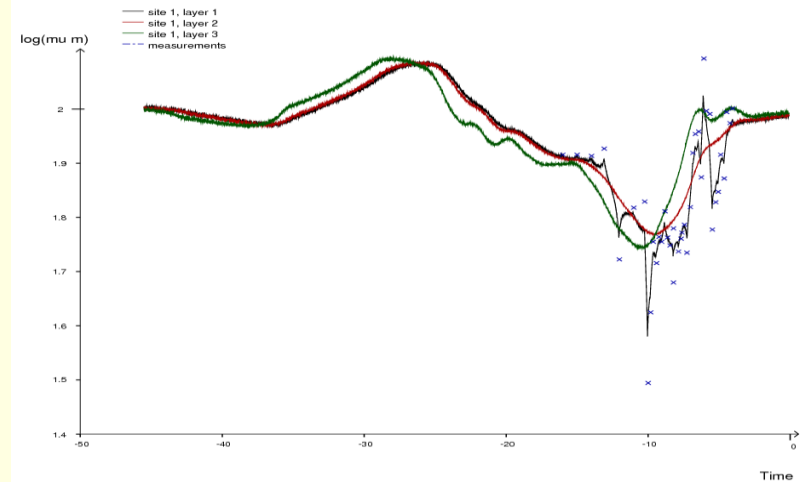
Bayesian model comparison result

- Best: 3 layer model ($\text{Pr}(M|D)=9.9\%$)
 - Lowest layer: Inter-regional correlations, $\rho=0.63$. $\Delta t \approx 6.1$ Myr.
 - Middle layer: Deterministic, $\Delta t \approx 1.4$ Myr.
 - Upper layer: Wide credibility band for the pull, which is local, $\Delta t \in (1\text{yr}, 1\text{Myr})$



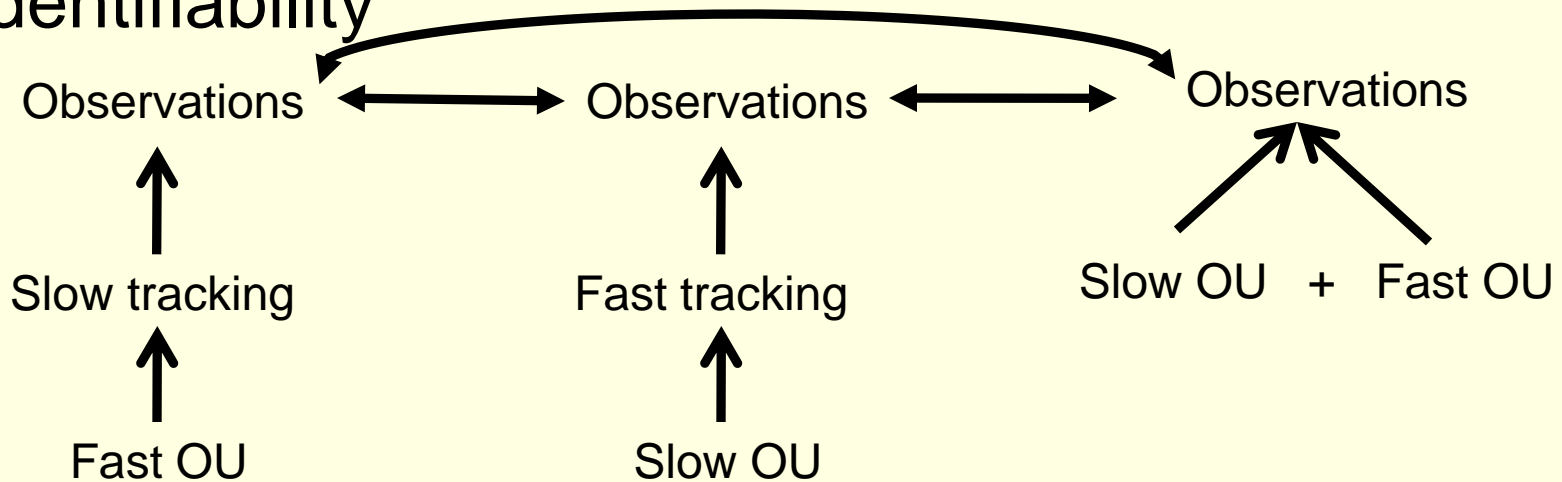
Bayesian model comparison result

- Best: 3 layer model ($\Pr(M|D)=9.9\%$)
 - Lowest layer: Inter-regional correlations, $\rho=0.63$. $\Delta t \approx 6.1$ Myr.
 - Middle layer: Deterministic, $\Delta t \approx 1.4$ Myr.
 - Upper layer: Wide credibility band for the pull, which is local, $\Delta t \in (1\text{yr}, 1\text{Myr})$



Problems / Future developments

- Identifiability



- Multimodal observations - speciation
- Handling several lineages and several phenotypes of each lineage: phylogeny, hierarchical models

Conclusions

- Possible to do inference on a model with multiple layers.
- There are methods for doing model comparison.
- Not enough data to get conclusive results regarding the model choice.
- Possible to find sites that “don’t follow the norm”.
- Productive framework which may be used in other settings also.
- Gives biological insight into processes in spite of sparse data.
- Bayesian priors wanted!

Links and bibliography

- Presentation:

http://folk.uio.no/trondr/stoch_layers7.ppt

http://folk.uio.no/trondr/stoch_layers7.pdf

- Bibliography:

- 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