Implicit parameter estimation

B. Weir, R.N. Miller, and Y.H. Spitz
Oregon State University, CEOAS

Joint work with E. Atkins, A.J. Chorin, and M. Morzfeld
Lawrence Berkeley National Lab/University of California, Berkeley

Supported by the National Science Foundation, Division of Ocean Sciences,
Collaboration in Mathematical Geosciences award #0934956

Parameters of marine ecological models

- Parameters control the growth/death rates of species and their interactions
- Little to no a priori knowledge
- Many are impossible to determine from in situ measurements alone
- Models combine different species into functional groups:
  - Parameters determine dominant species and their behavior
  - Fewer groups = stronger parameter dependence on specific ecosystem
- Assimilate data to find appropriate estimates

State and parameter estimation

Stochastic model with uncertain parameters

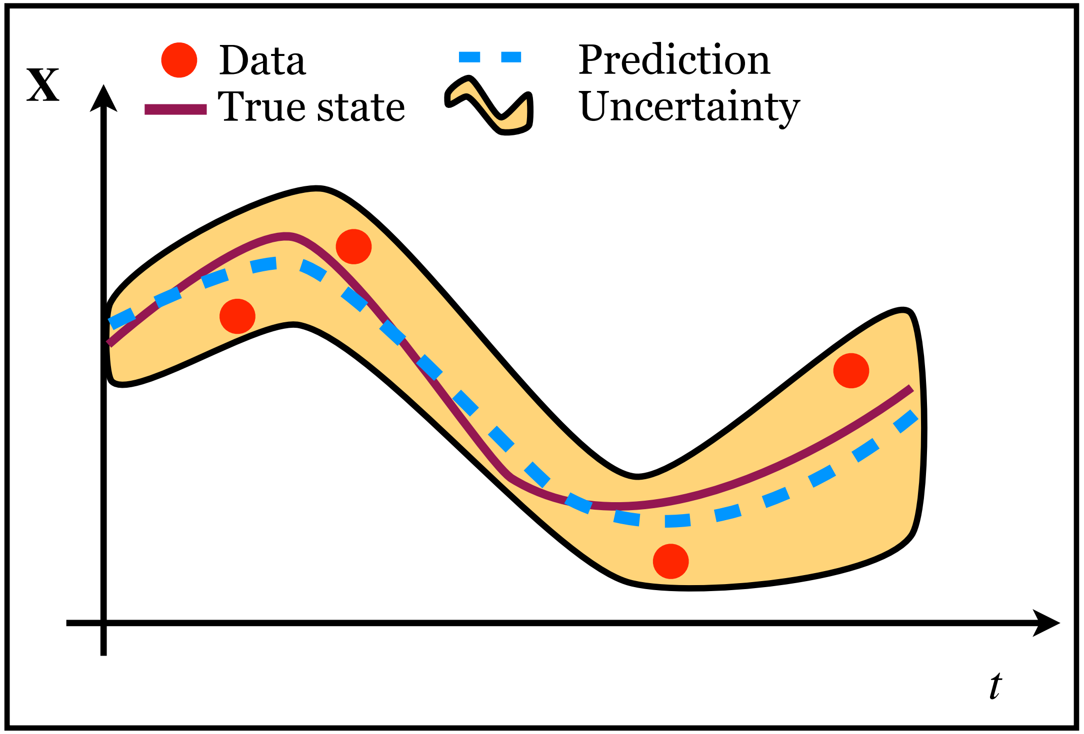
$$\mathbf{X}_m = \mathbf{X}_{m-1} + \tau f(\mathbf{X}_{m-1}, \theta, t_{m-1}) + \sqrt{\tau} G(\mathbf{X}_{m-1}, \theta, t_{m-1}) \mathbf{E}_m$$

Noisy observations of state

$$\mathbf{Y}_n = h(\mathbf{X}_{\mathfrak{m}(n)}, \theta, t_n) + \sqrt{R} \mathbf{D}_n$$

Prediction + uncertainty (target pdf)

$$p(\mathbf{x}_{0:\mathfrak{m}(k)}, \theta \,|\, \mathbf{y}_{1:k})$$



State and parameter estimation

Stochastic model with uncertain parameters

$$\mathbf{X}_m = \mathbf{X}_{m-1} + \tau f(\mathbf{X}_{m-1}, \boldsymbol{\theta}, t_{m-1}) + \sqrt{\tau} G(\mathbf{X}_{m-1}, \boldsymbol{\theta}, t_{m-1}) \mathbf{E}_m$$

Noisy observations of state

$$\mathbf{Y}_n = h(\mathbf{X}_{\mathfrak{m}(n)}, \boldsymbol{\theta}, t_n) + \sqrt{R} \mathbf{D}_n$$

Prediction + uncertainty (target pdf)

$$p(\mathbf{x}_{0:\mathfrak{m}(k)}, \boldsymbol{\theta} \,|\, \mathbf{y}_{1:k})$$

Subscript notation

$$\mathbf{x}_{0:\mathfrak{m}(k)} = \{\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{\mathfrak{m}(k)}\}$$

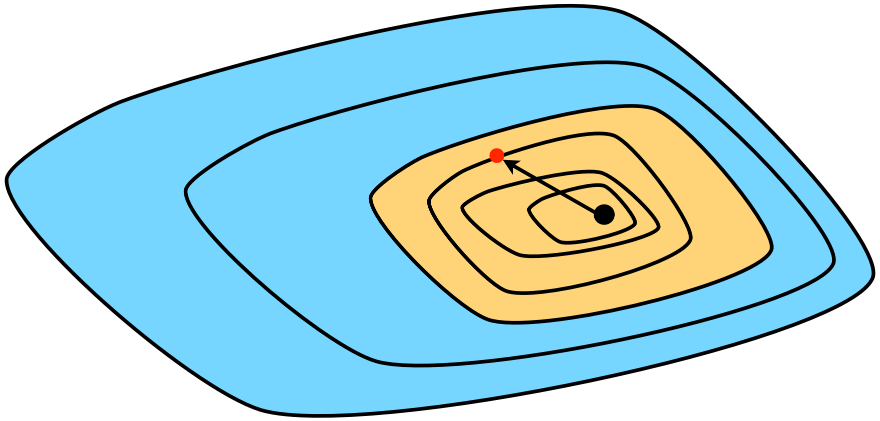
$$\mathbf{y}_{1:k} = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_k\}$$

Can have multiple time steps between observations, e.g.,

$$\mathfrak{m}(k) = 2k$$

The target pdf/density

- Posterior pdf (probability density function)
- Encodes all prior and posterior information about state and parameters
- Variational assimilation finds the target mode
- Monte Carlo methods sample the target



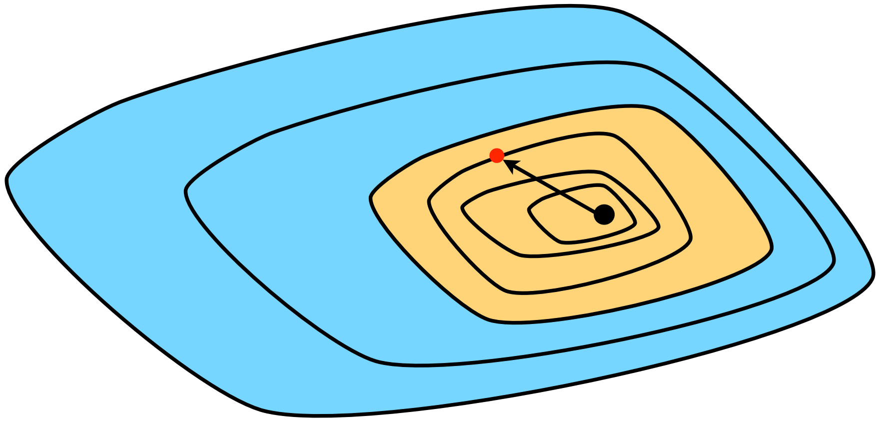
Schematic of target pdf

Black dot = mode
Red dot = sample
Yellow area = probable region
Blue area = improbable region

Implicit sampling

(Chorin, Atkins, Morzfeld, and Tu and BW, RNM, and YHS)

- Monte Carlo method for importance sampling
- No forecast distribution, work directly with target
- Apply particle by particle
- Use numerical optimization to find high probability regions
- Focus sampling within these regions



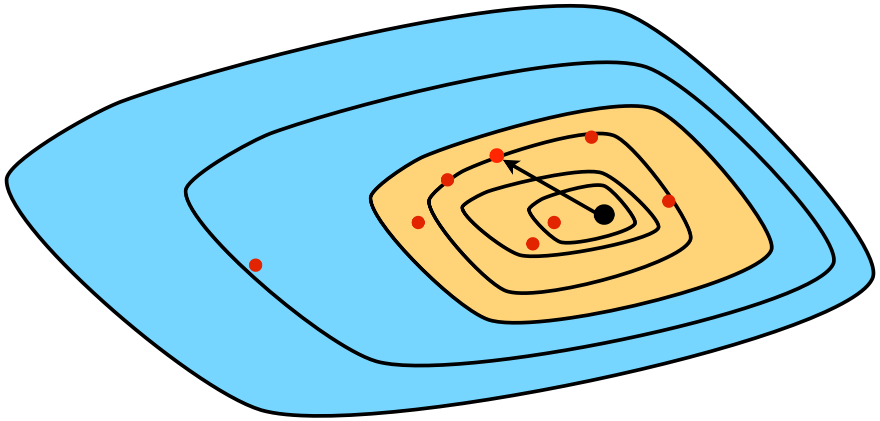
<u>Schematic of target pdf</u>

Black dot = mode
Red dot = sample
Yellow area = probable region
Blue area = improbable region

Implicit sampling

(Chorin, Atkins, Morzfeld, and Tu and BW, RNM, and YHS)

- Monte Carlo method for importance sampling
- No forecast distribution, work directly with target
- Apply particle by particle
- Use numerical optimization to find high probability regions
- Focus sampling within these regions



Schematic of target pdf

Black dot = mode
Red dot = sample
Yellow area = probable region
Blue area = improbable region

What makes this a good idea

- **Nonparametric**: strong theoretical basis for nonlinear/non-Gaussian problems

- **Generally applicable**:
  - smoother and filter forms
  - state and/or parameter estimator
  - applicable to deterministic and stochastic models

- **Optimized for observations**: explores important regions in sample space; does not “blindly” explore space and eliminate improbable samples (like many particle filters and MCMC methods)

- **Many implementations**: allows problem-specific tuning (hint ...)

Notation and definitions

- By construction, target pdf is exponential of a nonlinear sum of squares
- Work with this sum, the **target cost** $J(\zeta;\eta)$, defined such that

$$p(\mathbf{x}_{0:\mathrm{m}(k)}, \theta \,|\, \mathbf{y}_{1:k}) = C(\eta)\exp[-J(\zeta;\eta)]$$

- New variables ζ and η divide state and parameter space:
  ζ is the **estimated variables** and η is the **given/sampled variables**
- Exact definitions of ζ, η, and C depend on problem; C is usually constant
- In a particle filter, e.g., $\mathrm{m}(k) = k$, $\zeta = \mathbf{x}_k$, $\eta = (\mathbf{x}_{0:k\text{-}1}, \theta)$, and $C = 1$ after resampling

An implicit sampling algorithm

1. For a given η, find target mode ζ^*; same as min of target cost, like in 4D-Var

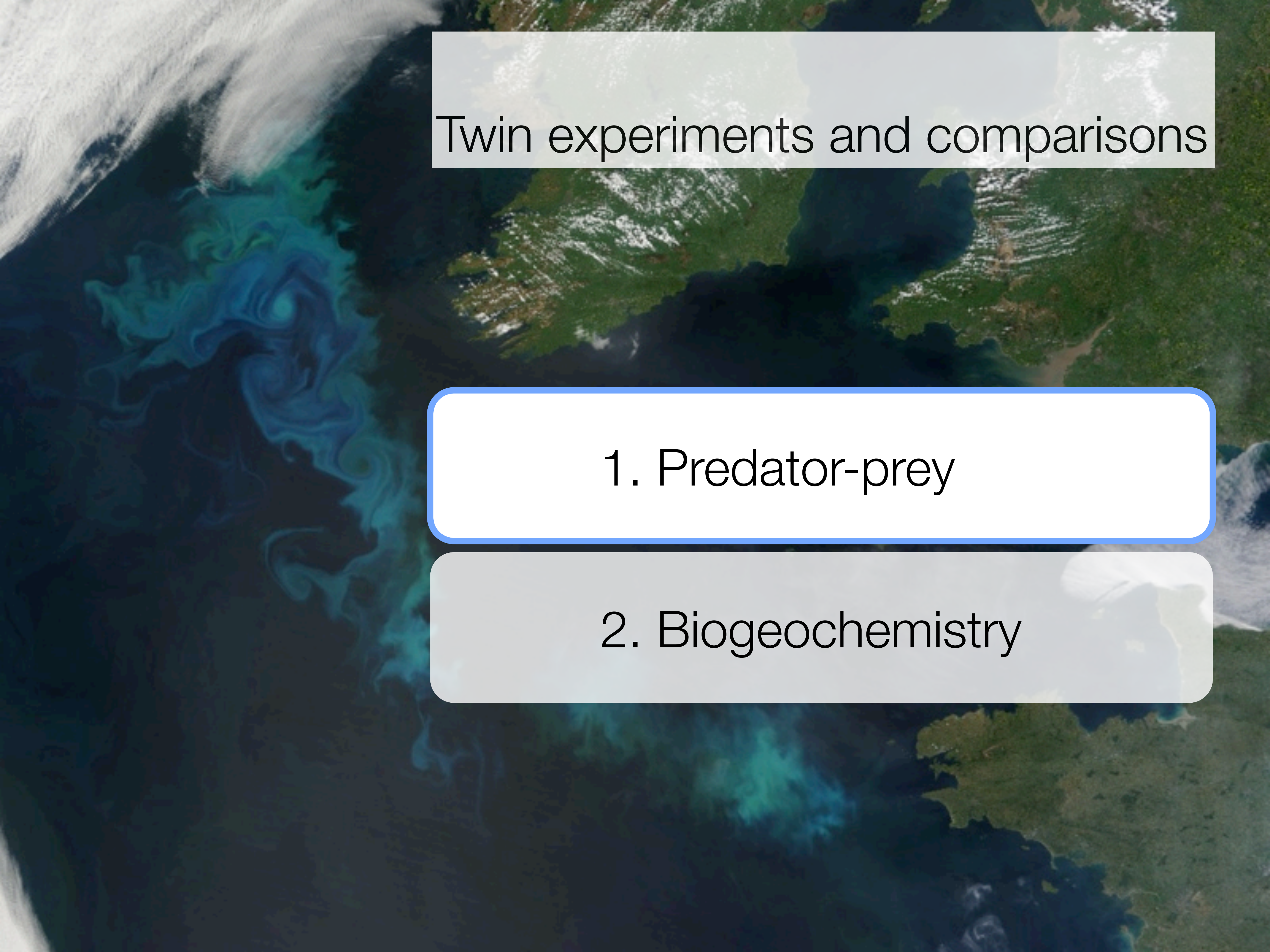
2. Sample a Gaussian with mean ζ^* and covariance H^{-1};
   H is the Hessian of target cost at ζ^*

3. Weigh the sample ζ to account for difference between **proposal cost** K of the Gaussian and true **target cost** J:

$$w = |H|^{-1/2} \exp[-J(\zeta^*;\eta)] \exp[K(\zeta) - J(\zeta;\eta)]$$

- Repeat steps 1-3 N_p times for all choices of η

- Weighted ensemble represents true target pdf



Twin experiments and comparisons

1. Predator-prey
2. Biogeochemistry

Predator-prey

The Lotka-Volterra equations

- Estimate ***2 state variables*** P (prey) and Q (predator) and ***7 unknown parameters*** $\theta = (\theta_1, ..., \theta_7)$ in model equations

$$\frac{dP}{dt} = (\theta_1 - \theta_2 P)P - \theta_3 \frac{PQ}{1+\theta_7 P}$$
$$\frac{dQ}{dt} = (-\theta_4 - \theta_5 Q)Q + \theta_6 \frac{PQ}{1+\theta_7 P}$$

+ noise

color legend: growth, death, and consumption

- State and parameters are positive numbers

- Apply (anamorphosis) transform to variables that are more nearly Gaussian, e.g., $\zeta = (\log P, \log Q, \log \theta)$

Twin experiments (Weir et al. 2013)

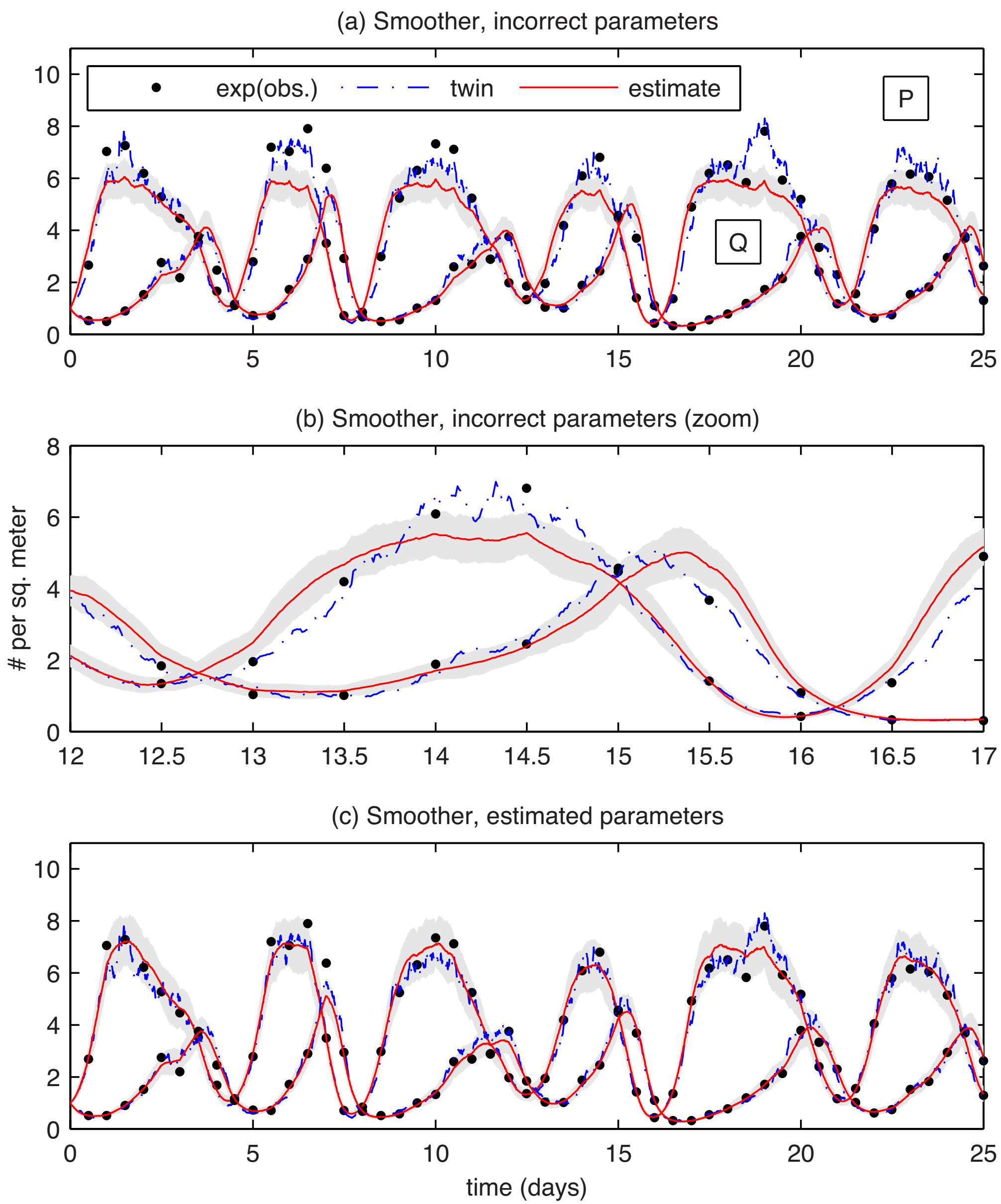
- Observations of P and Q every 50 time steps a total of 50 times; initial condition fixed at $(1,1)$

- Compare two different assimilation techniques:

1. **Smoother**
   - All obs. assimilated at once
   - Target pdf almost Gaussian

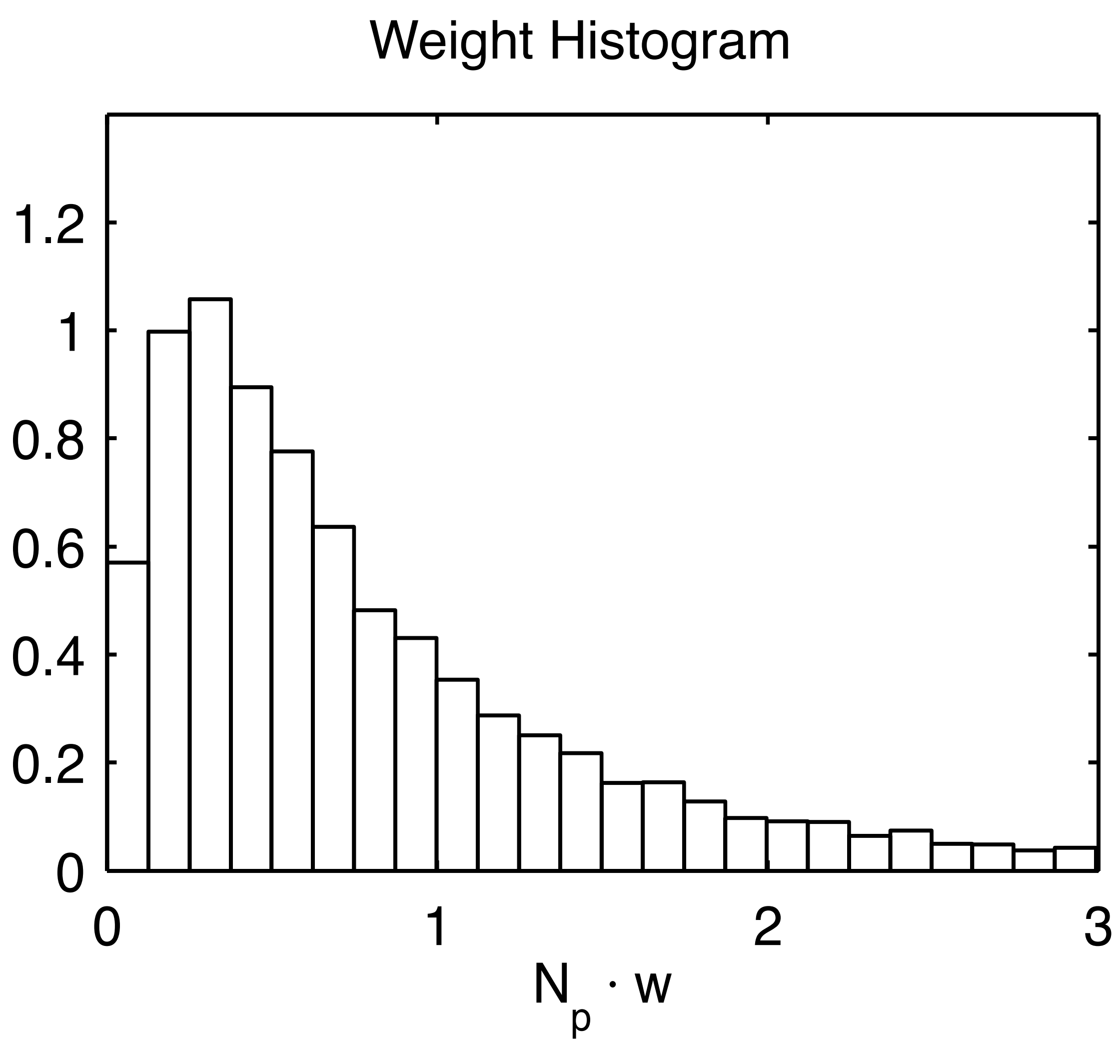
2. **Filter**
   - Each obs. assimilated in sequence (filter)
   - Target pdf non-Gaussian
   - Kernel density/Gaussian mixture used to continue parameter estimate sequentially

Smoother state estimates



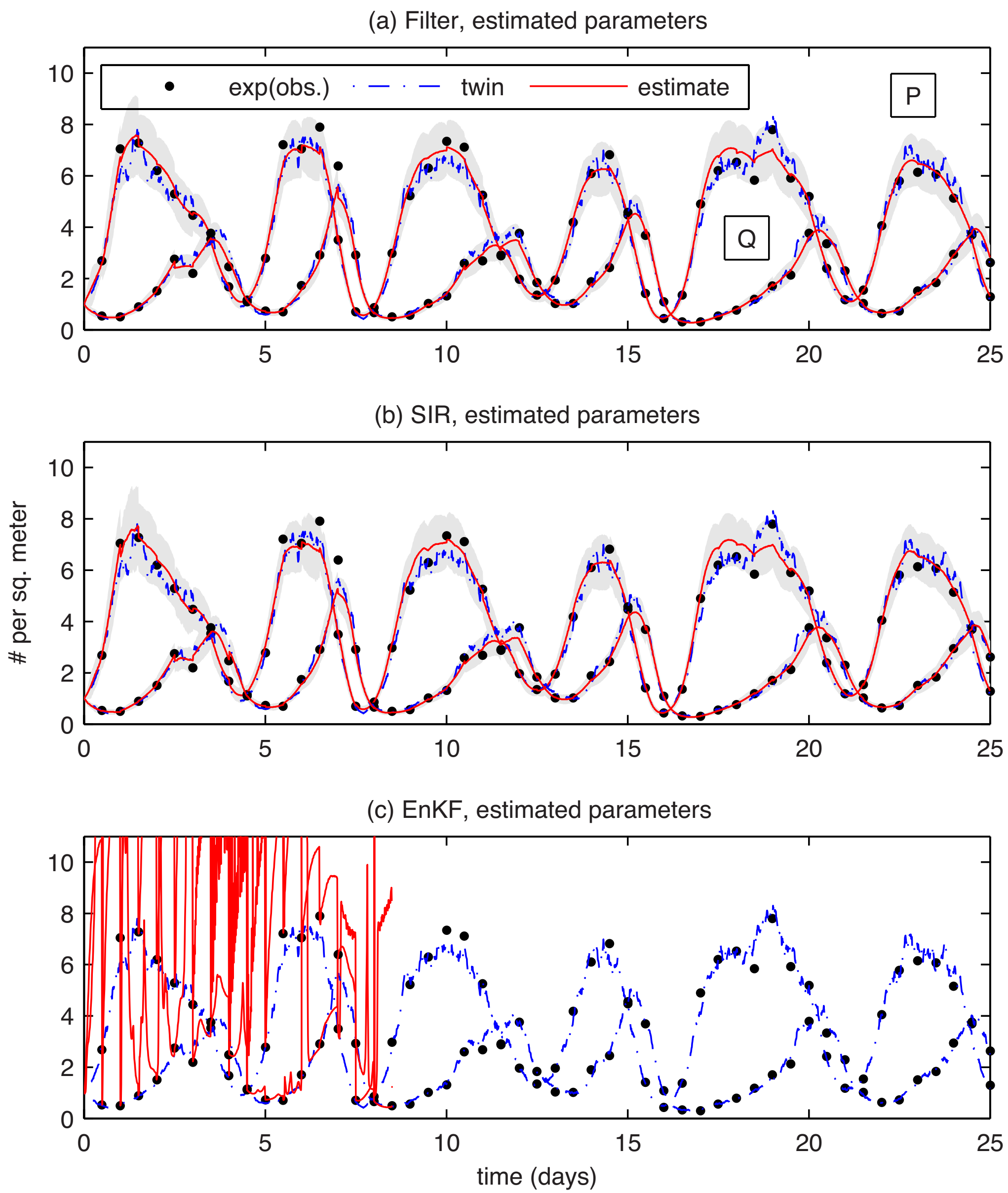
- **Comparison of state estimate in two cases**: (a,b) parameters fixed at incorrect values, (c) estimated parameters
- Substantial noise in model and measurements
- Estimated with 240 particles
- Shaded region = 2 standard deviations

Smoother weights

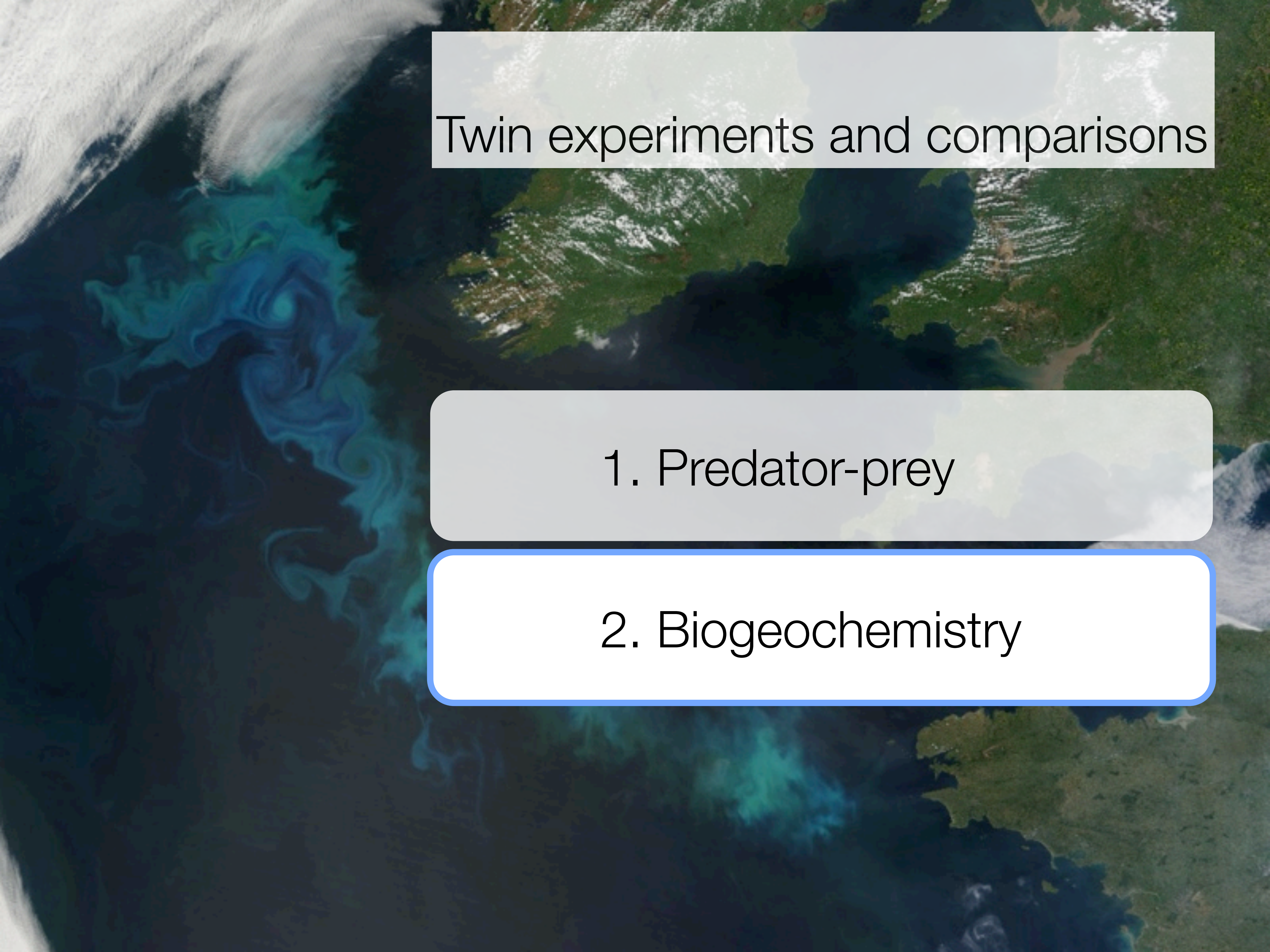


- Histogram computed with 240,000 particles for high resolution
- Delta function at 1 = perfect sampling
- Noticeable drop-off in distribution before zero

Filter/sequential state estimates



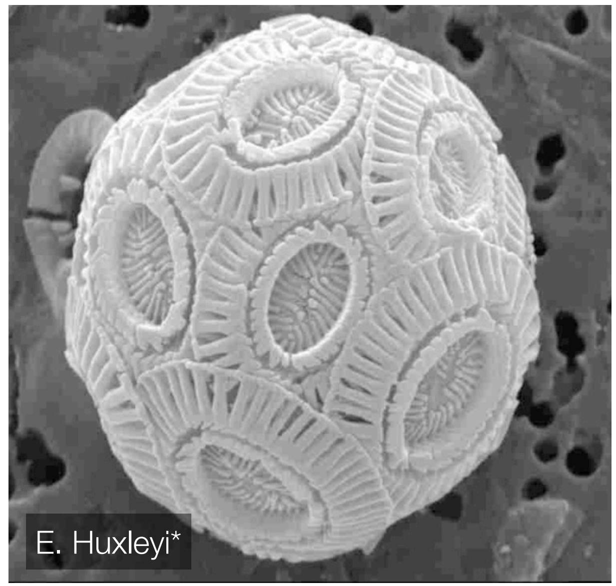
- 2400 particles for (a) implicit filter and (b) SIR filter
- 240,000 particles for (c) EnKF
- EnKF covariance blows up; **works only if observations are denser in time**



Twin experiments and comparisons

1. Predator-prey
2. Biogeochemistry

Simplified model concept


E. Huxleyi*

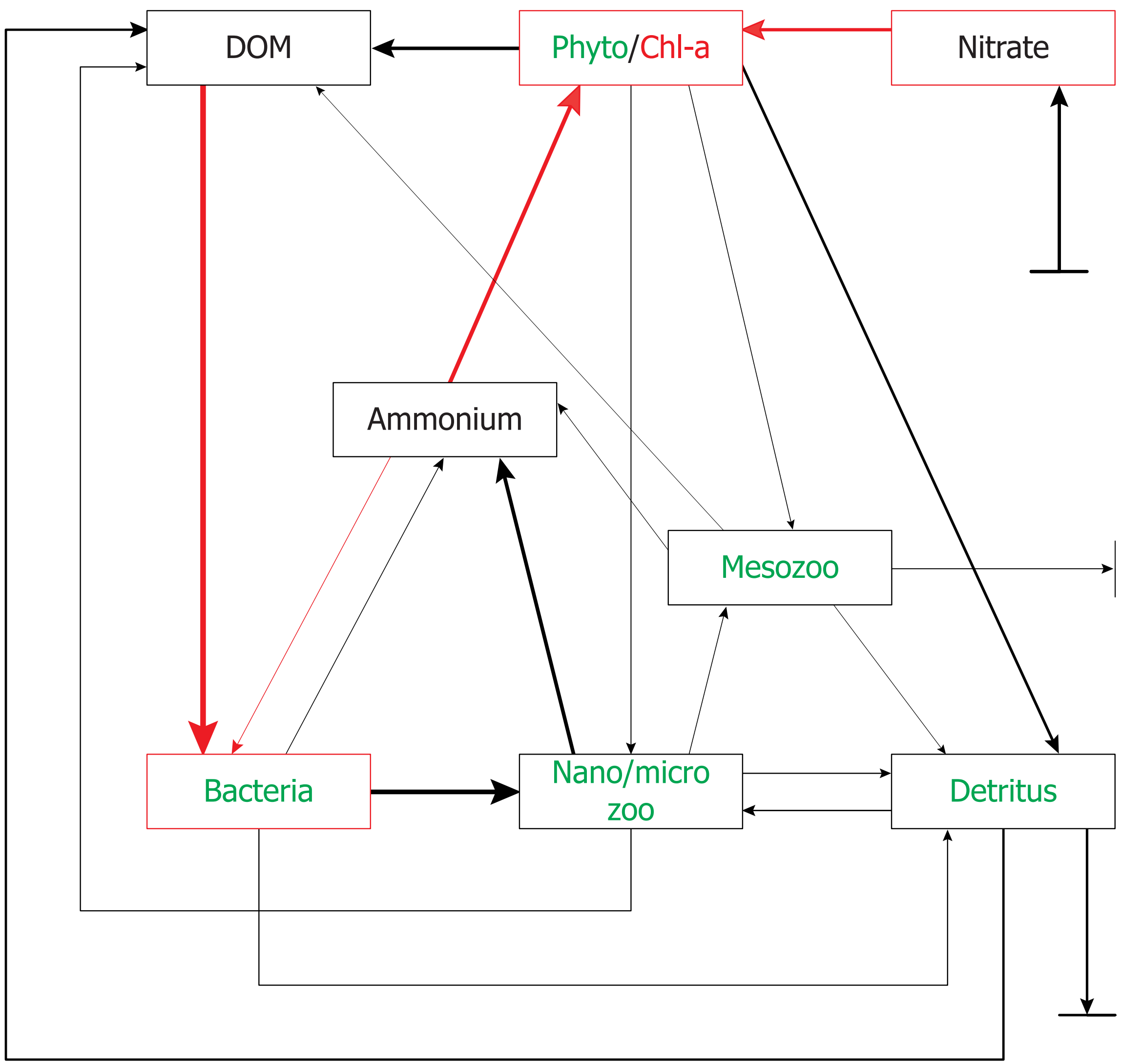

Tunicate/salp*

Nutrients → Phytoplankton → Zooplankton

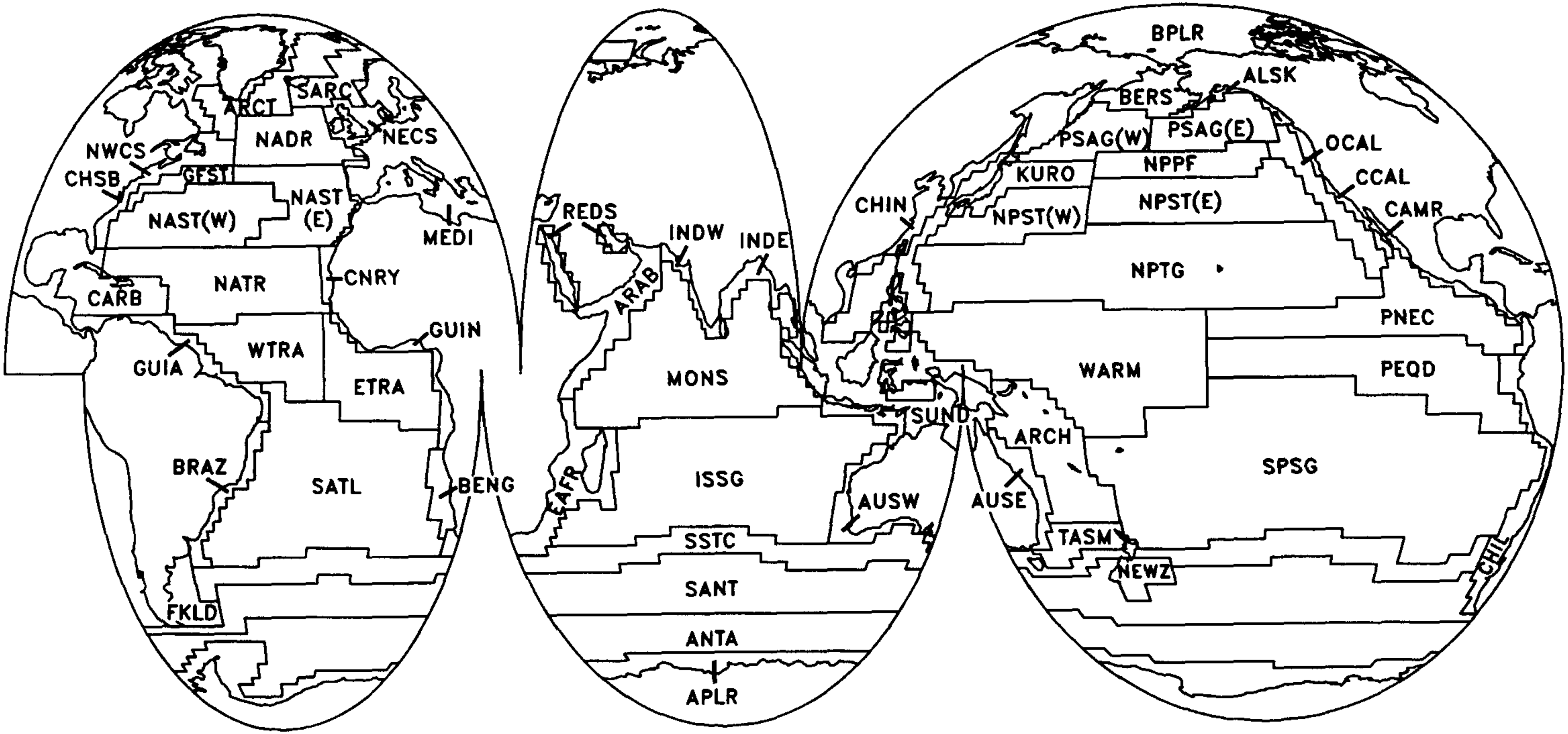
(remineralization)

* http://angelicquewhite.com

Nitrogen cycle model (Spitz et al. 2001)



- Depth-averaged, nitrogen-based **ODE** for mixed layer concentrations
- 49 parameters
- Entering arrow = entrainment due to mixed layer deepening
- Exiting arrows = sinking
- (DOM, dissolved organic matter)



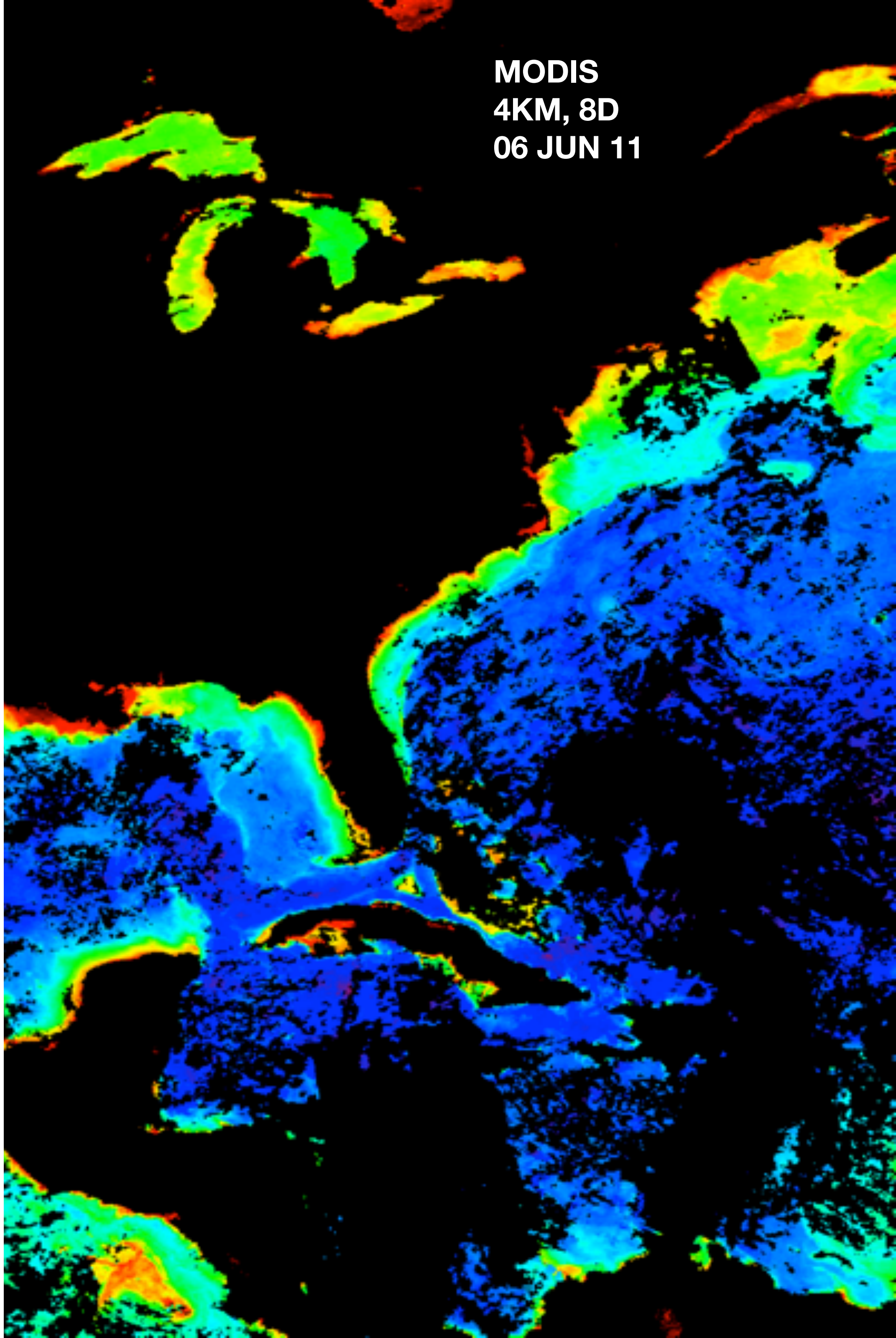
Can we use parameter estimates to divide the ocean into ecosystems?

Longhurst (1995) ecological provinces

Satellite chlorophyll observations

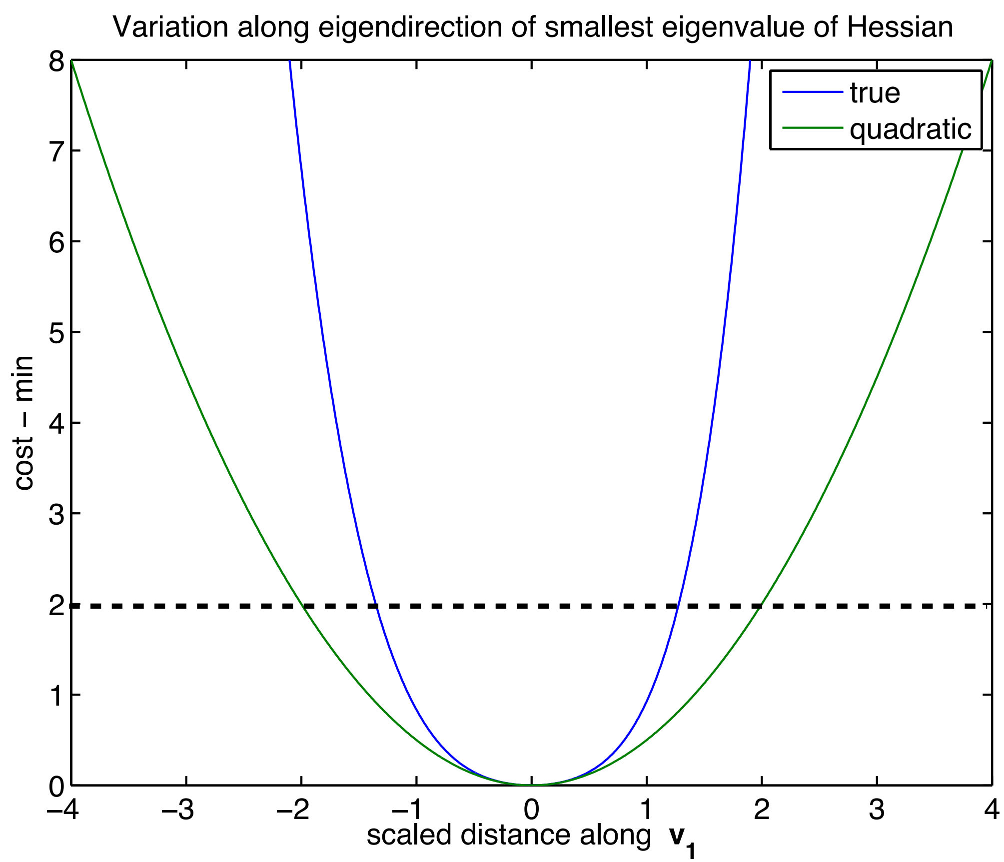
- Far away from a few time-series studies, satellite chlorophyll is only available data
- Best case scenario: data is available every day
- Gaussian importance sampling fails
- To see why, consider target cost (deterministic model):

$$J(\theta) = -\log\left[p(\theta \,|\, \mathbf{y}_{1:k})\right]$$



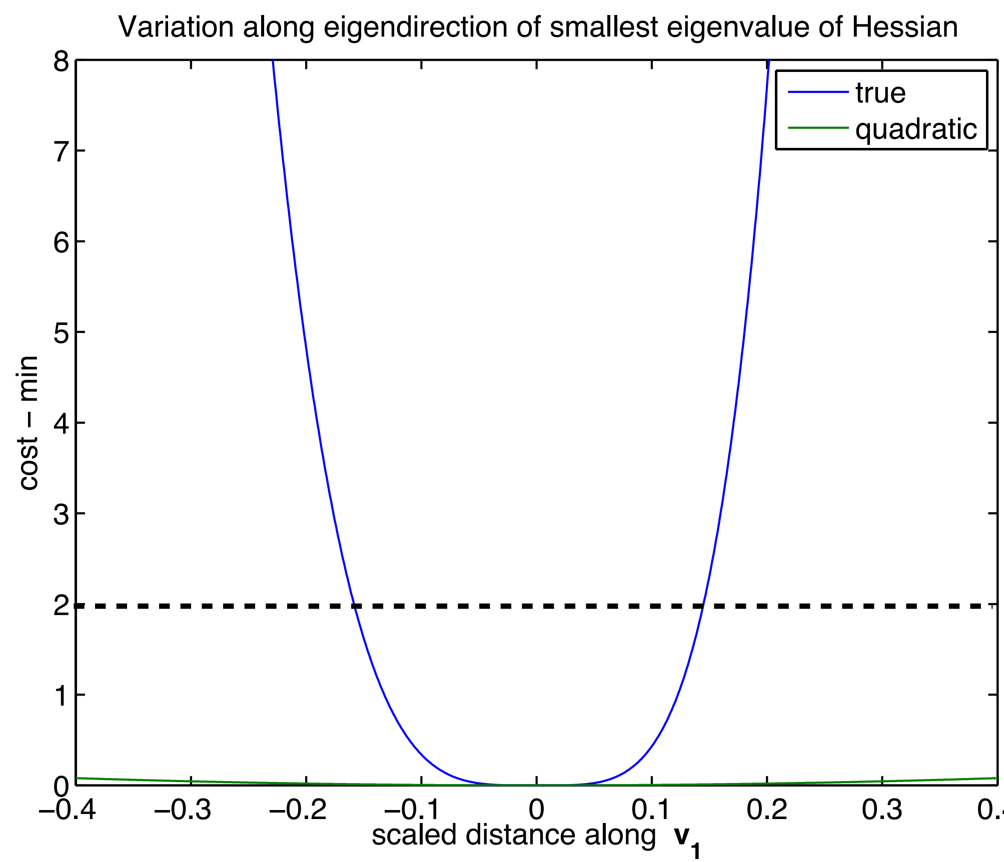
Cost function transect

- Hessian over-predicts uncertainty of estimate
- Higher-order modes are important
- Inefficient: spends a lot of time sampling the tail
- Problem gets worse as background covariance of parameters increases



Cost function transect

- Hessian over-predicts uncertainty of estimate
- Higher-order modes are important
- Inefficient: spends a lot of time sampling the tail
- Problem gets worse as background covariance of parameters increases



Adaptive sampling

- Each sample gives us information about the target cost, use to build a better **global** quadratic approximation
- Find a Hessian such that $\mathbb{E}_K[K(\theta; H) - J(\theta)] = 0$
- In certain cases, equivalent to minimizing the variance of the weights
- Solution is underdetermined in more than 1 dimension, apply rank-1 update similar to BFGS
- No need to discard any samples or restart estimation?

The Robbins-Monro (RM) iteration

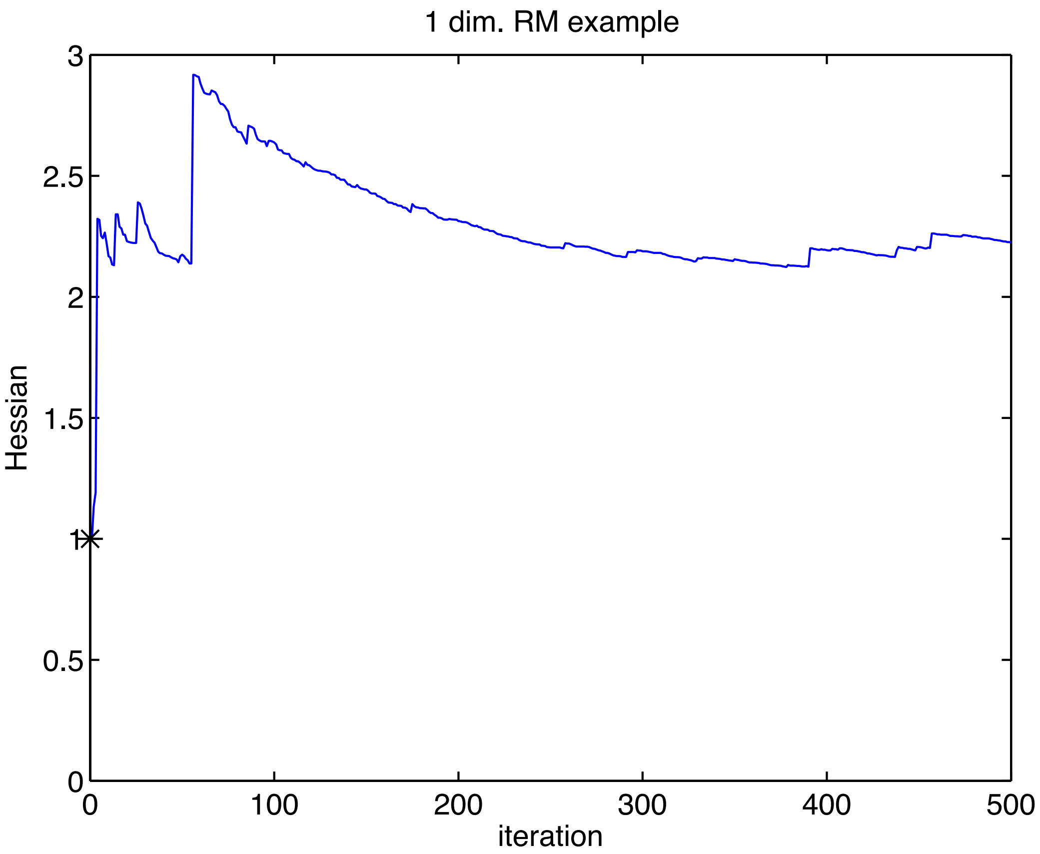
H_0 is the Hessian of the cost function at its minimum. Suppose $\theta^* = 0$. Define the "eigenvalue"

$$\begin{aligned} \nu_n &= \theta^t H_n \theta / \theta^t \theta, \\ &= \exp(\mu_n) \end{aligned}$$

since H_n is positive definite. The RM iteration is

$$\begin{aligned} \mu_{n+1} &= \mu_n + \epsilon_n [K(\theta; H_n) - J(\theta)], \\ H_{n+1} &= H_n + (\nu_{n+1} - \nu_n)\theta\theta^t / \theta^t \theta, \end{aligned}$$

where $\epsilon_n = C n^{-\alpha}$.



The Robbins-Monro (RM) iteration

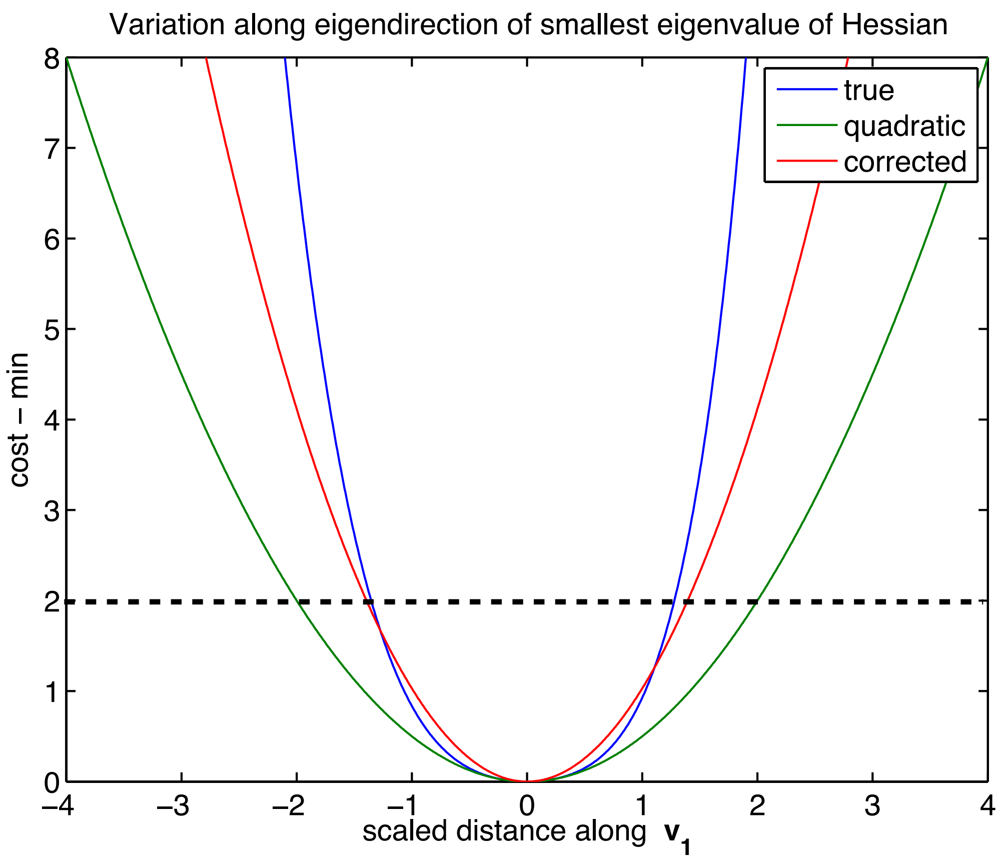
H_0 is the Hessian of the cost function at its minimum. Suppose $\theta^* = 0$. Define the "eigenvalue"

$$\begin{aligned}\nu_n &= \theta^t H_n \theta / \theta^t \theta, \\ &= \exp(\mu_n)\end{aligned}$$

since H_n is positive definite. The RM iteration is

$$\begin{aligned}\mu_{n+1} &= \mu_n + \epsilon_n [K(\theta; H_n) - J(\theta)], \\ H_{n+1} &= H_n + (\nu_{n+1} - \nu_n)\theta\theta^t / \theta^t \theta,\end{aligned}$$

where $\epsilon_n = Cn^{-\alpha}$.



The Robbins-Monro (RM) iteration

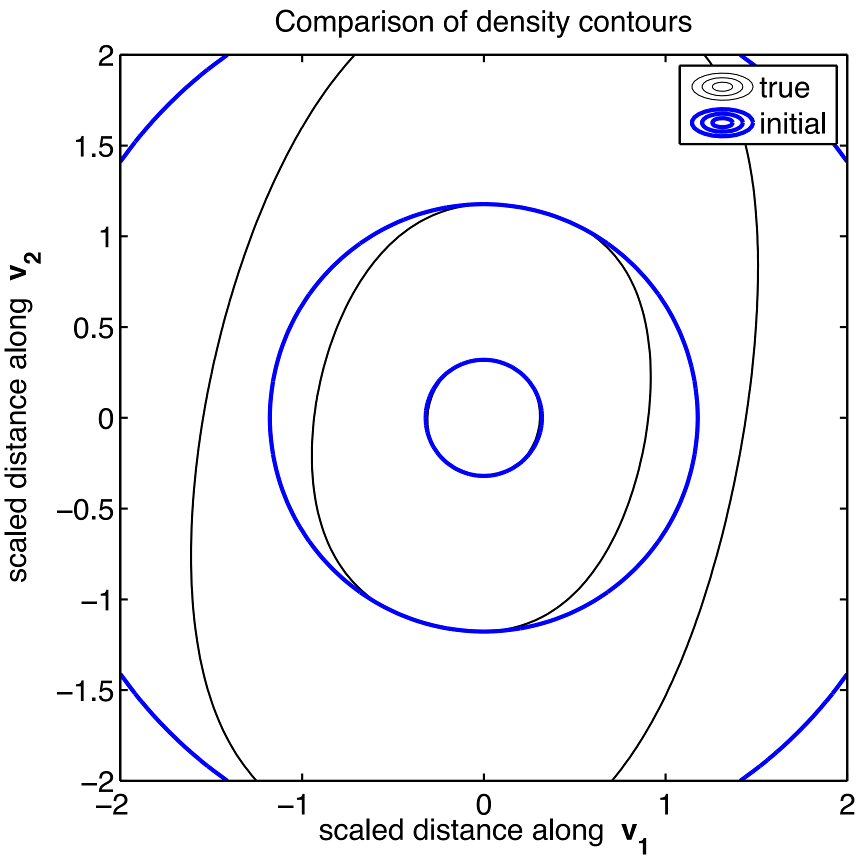
H_0 is the Hessian of the cost function at its minimum. Suppose $\theta^* = 0$. Define the "eigenvalue"

$$\nu_n = \theta^t H_n \theta / \theta^t \theta, \\ = \exp(\mu_n)$$

since H_n is positive definite. The RM iteration is

$$\mu_{n+1} = \mu_n + \epsilon_n [K(\theta; H_n) - J(\theta)], \\ H_{n+1} = H_n + (\nu_{n+1} - \nu_n)\theta\theta^t / \theta^t \theta,$$

where $\epsilon_n = C n^{-\alpha}$.



The Robbins-Monro (RM) iteration

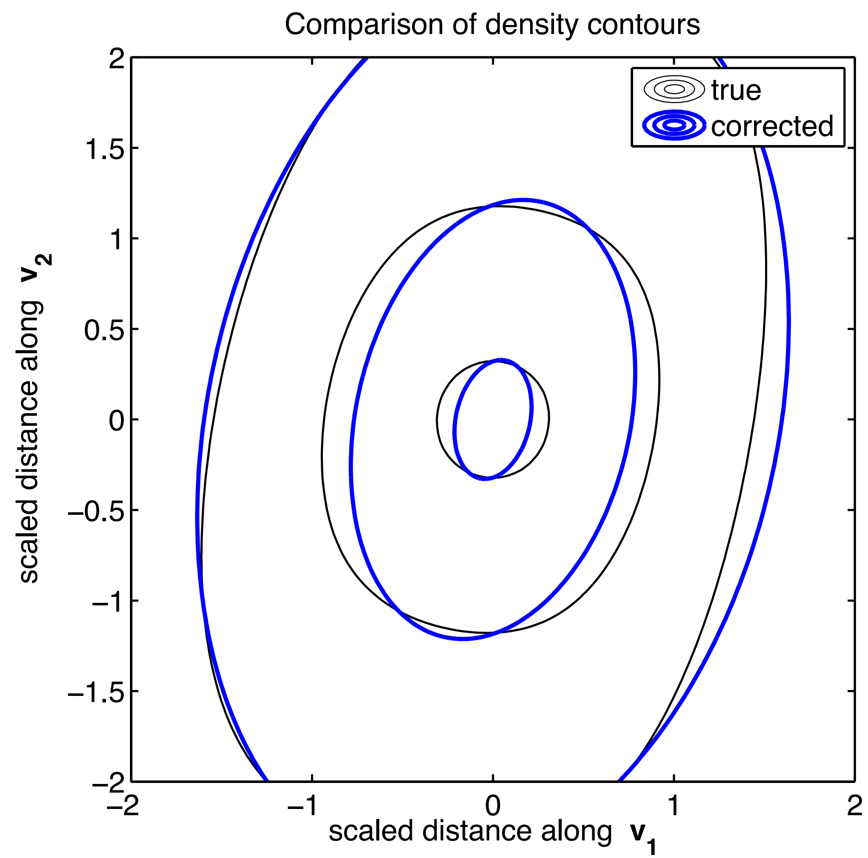
H_0 is the Hessian of the cost function at its minimum. Suppose $\theta^* = 0$. Define the "eigenvalue"

$$\nu_n = \theta^t H_n \theta / \theta^t \theta, \\ = \exp(\mu_n)$$

since H_n is positive definite. The RM iteration is

$$\mu_{n+1} = \mu_n + \epsilon_n [K(\theta; H_n) - J(\theta)], \\ H_{n+1} = H_n + (\nu_{n+1} - \nu_n)\theta\theta^t / \theta^t \theta,$$

where $\epsilon_n = Cn^{-\alpha}$.



The Robbins-Monro (RM) iteration

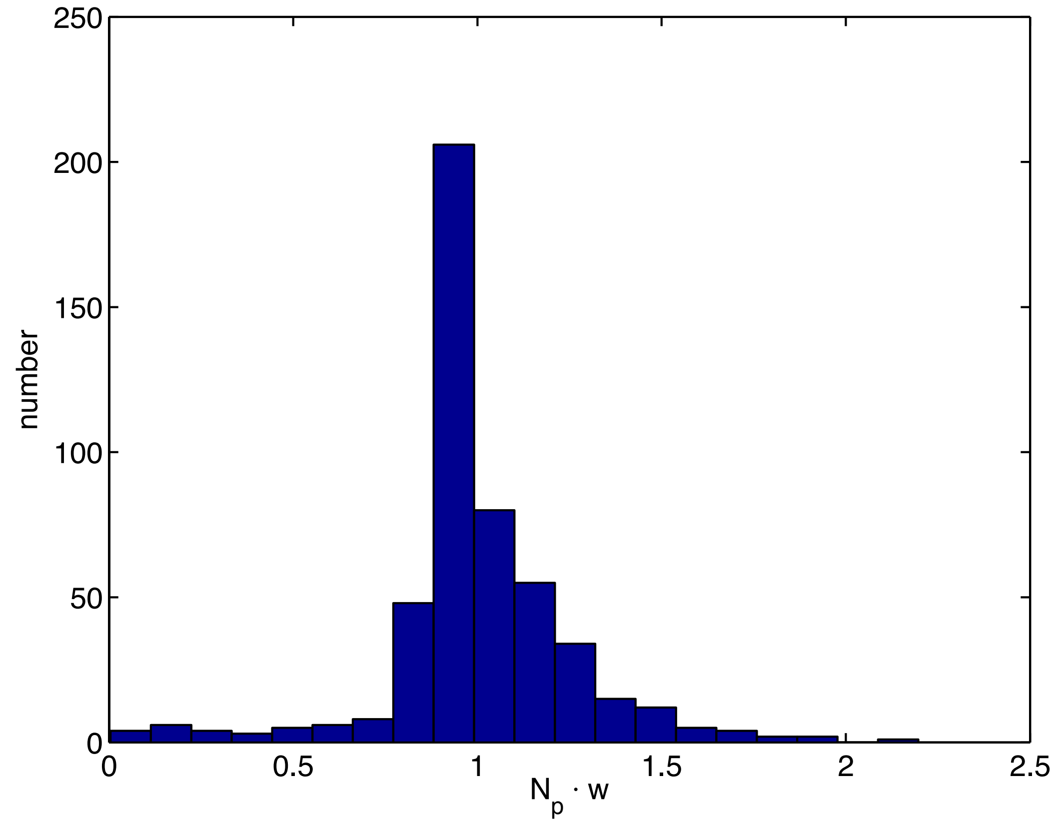
H_0 is the Hessian of the cost function at its minimum. Suppose $\theta^* = 0$. Define the "eigenvalue"

$$\nu_n = \theta^t H_n \theta / \theta^t \theta, \\ = \exp(\mu_n)$$

since H_n is positive definite. The RM iteration is

$$\mu_{n+1} = \mu_n + \epsilon_n [K(\theta; H_n) - J(\theta)], \\ H_{n+1} = H_n + (\nu_{n+1} - \nu_n)\theta\theta^t / \theta^t \theta,$$

where $\epsilon_n = C n^{-\alpha}$.



Conclusions

- Implicit sampling is theoretically applicable to state and parameter estimation in a very general setting
- In strongly non-Gaussian problems, can use a Robbins-Monro iteration to refine the Hessian and generate samples with acceptable weights
- Refinement and sampling significantly improves the confidence limits from those given by local Gaussian assumption
- If chlorophyll is the only information about parameters, can find more accurate estimates than quadratic/Gaussian interpretation suggests
- This lets us define ecological regions with greater precision

References

[1] Atkins, E., Morzfeld, M., & Chorin, A. J. (2012). Implicit particle methods and their connection with variational data assimilation. *Mon. Weather Rev.*, *accepted*.

[2] Chorin, A. J., & Tu, X. (2009). Implicit sampling for particle filters. *Proc. Natl. Acad. Sci.*, *106*, 17,249–17,254.

[3] Chorin, A. J., Morzfeld, M., & Tu, X. (2010). Implicit particle filters for data assimilation. *Comm. App. Math. and Comput. Sci.*, *5*, 221–240.

[4] Morzfeld, M., & Chorin, A. J. (2012). Implicit particle filtering for models with partial noise, and an application to geomagnetic data assimilation. *Nonlin. Processes Geophys.*, *19*, 365–382.

[5] Morzfeld, M., Tu, X., Atkins, E., & Chorin, A. J. (2012). A random map implementation of implicit filters. *J. Comput. Phys.*, *231*, 2049–2066.

[6] Spitz, Y. H., Moisan, J. R., & Abbott, M. R. (2001). Configuring an ecosystem model using data from the Bermuda Atlantic Time Series (BATS). *Deep-Sea Res. II*, *48*, 1733–1768.

[7] Weir, B., Miller, R. N., & Spitz, Y. H. (2013). Implicit estimation of ecological model parameters. *Bull. Math. Biol.*, doi: 10.1007/s11538-012-9801-6

Thank you!