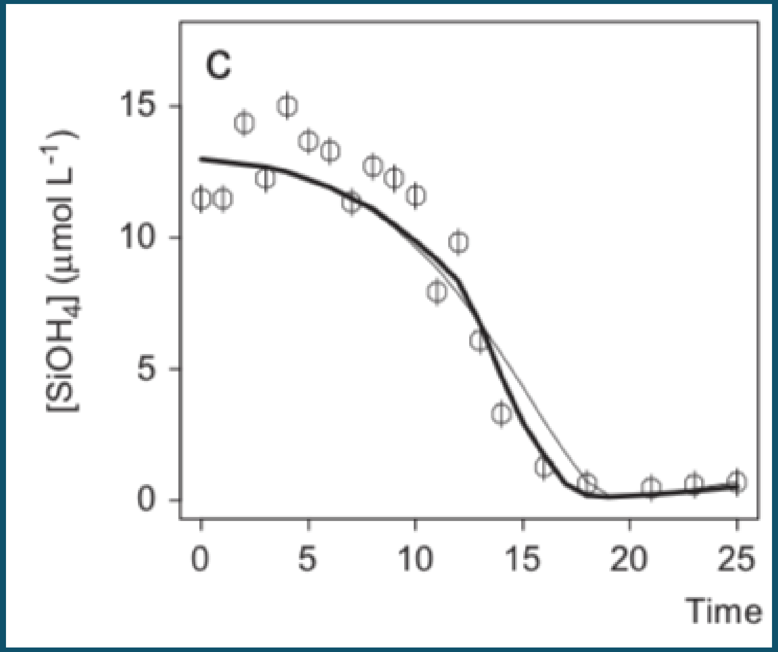
The Adaptive Metropolis algorithm as a tool for model selection given irregular and imperfect time-series data

or

How gambling intellegently pays off!

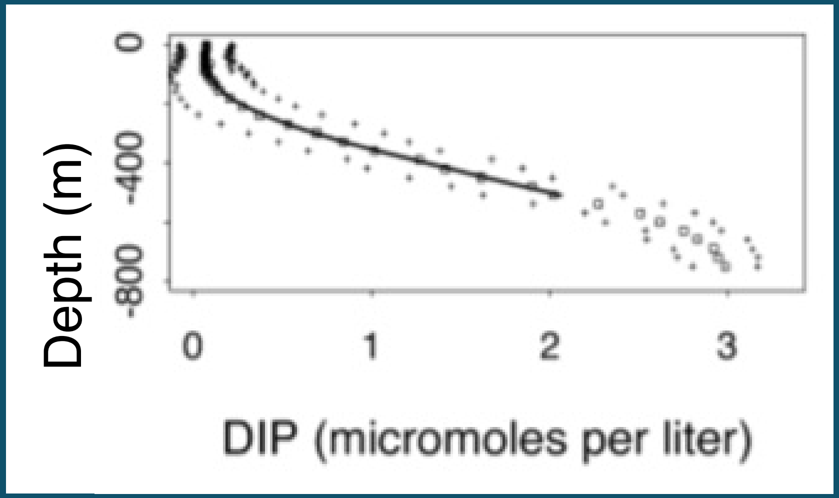
S. Lan Smith, JAMSTEC, Yokohama, Japan

Very Bad



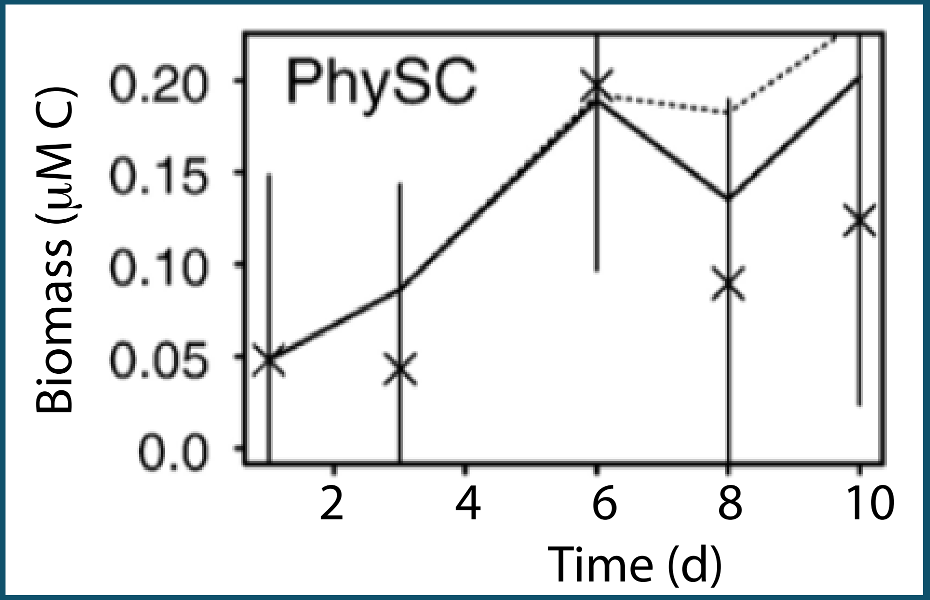
Smith et al. (*Deep Sea Res. II* 2010)

Bad



Smith et al. (*J. Oceanogr.* 2005)

More Info, Still Bad



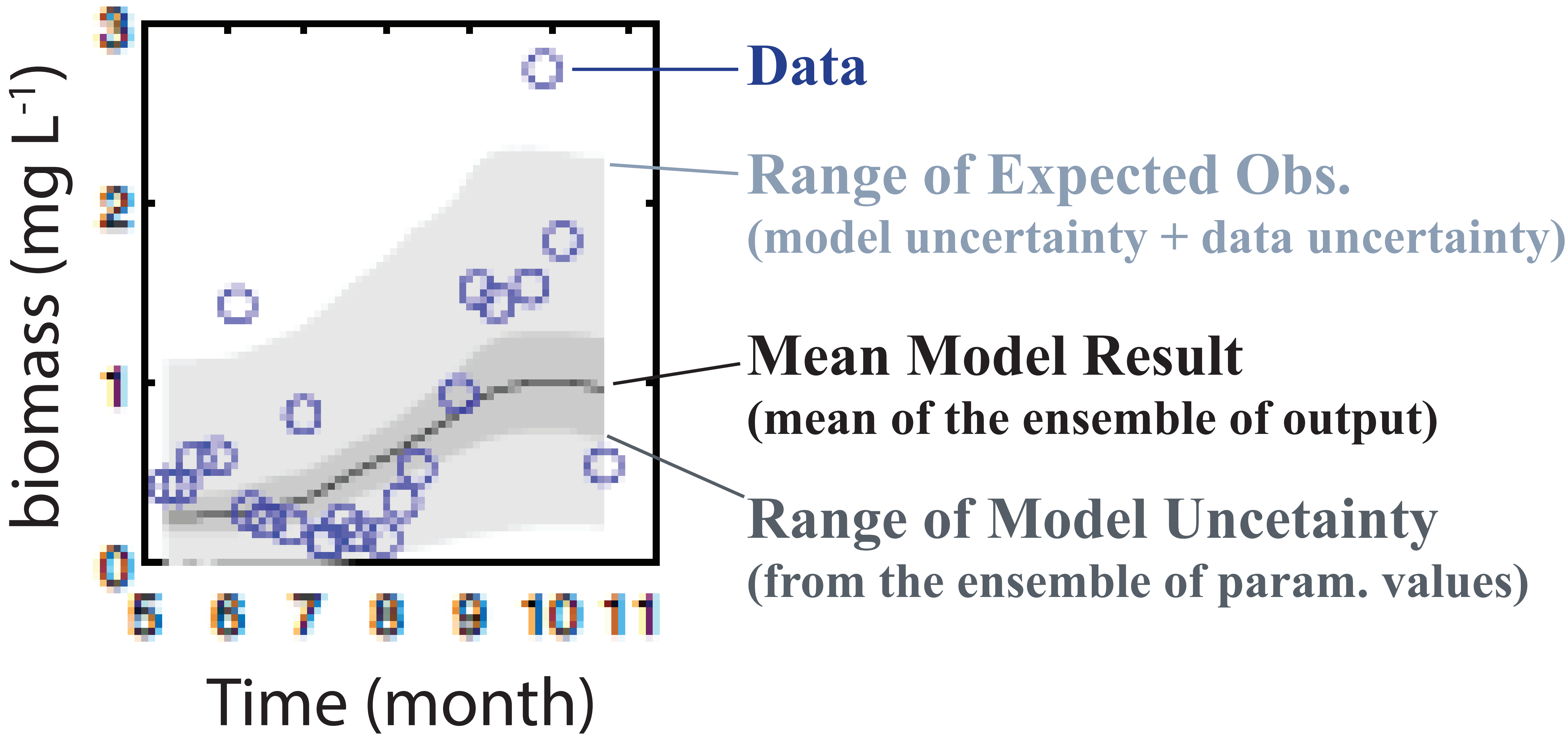
Smith et al. (*J. Mar Sys.* 2007)

No info about model uncertainty!

Can these models be trusted?

How does the range of modelled values compare to the observed range?

Where to expect future obs?



Marko Laine (Fig. 3a, PhD Thesis, Lapeenranta Univ. of Tech., Finland, 2008)

What makes this possible?

Conditional probability, $p(A \mid B)$

'the probability of *A* given *B*'

i.e, if *B* is true

Likelihood, $p(y \mid \Theta)$

'probability of observing *y* given model Θ'

e.g., *p*(wet sidewalks | it's raining)

Maximum Likelihood methods

are widely used to estimate param. values

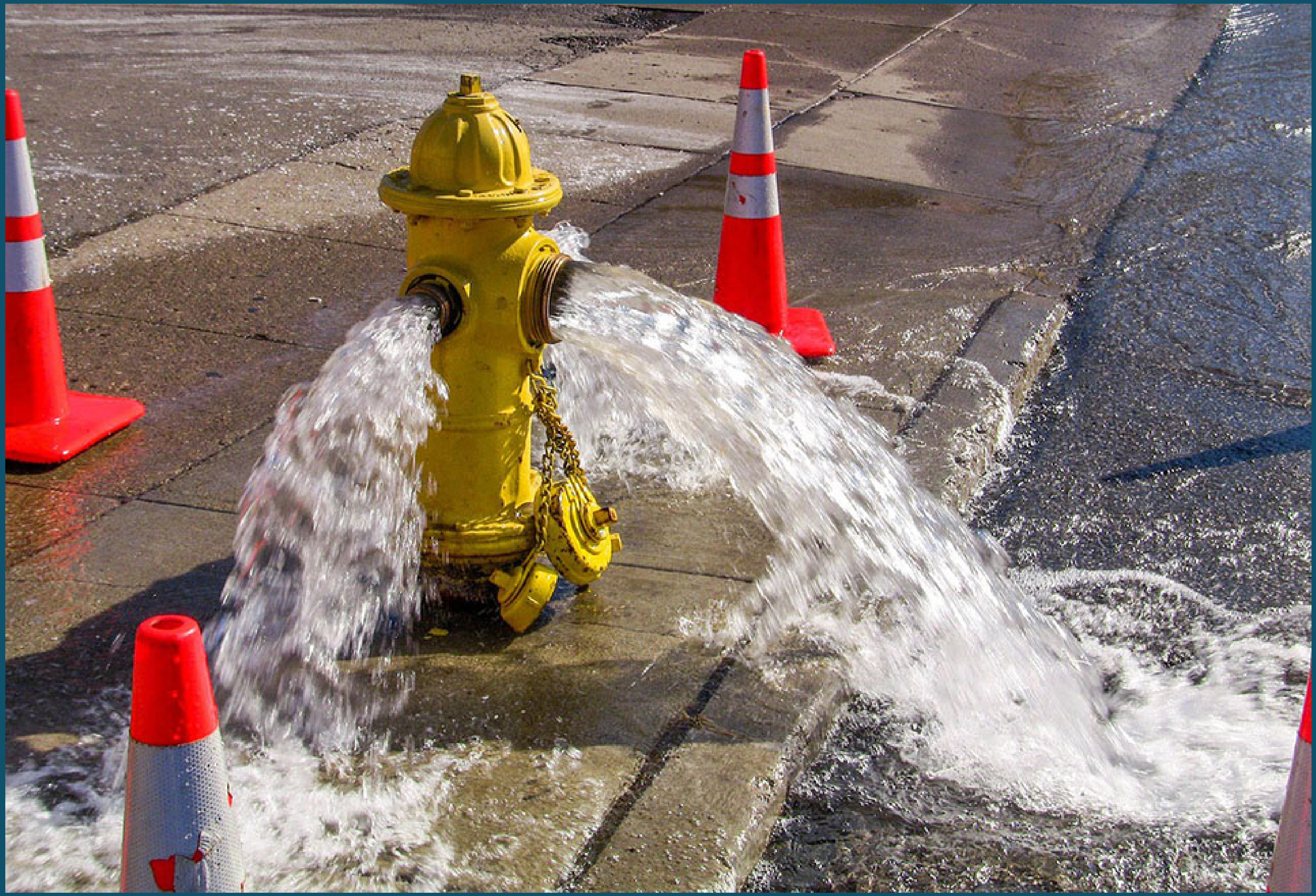
i.e., find param. values that maximize the likelihood of the obs.

This can be useful, <u>but it is NOT sufficient!</u>

Bayes Theorem

$$p(A \mid B)\, p(B) = p(B \mid A)\, p(A)$$

p(wet sidewalks | rain) ≠ p(rain | wet sidewalks)



Bayes Theorem

$$p(\Theta \mid y)\, p(y) = p(y \mid \Theta)\, p(\Theta)$$

$p(\Theta \mid y)$	probability of the model given the data
$p(y)$	probability of the observation(s), y
$p(y \mid \Theta)$	Likelihood of obs. y given model Θ
$p(\Theta)$	probability of the model, a.k.a. the *Prior*

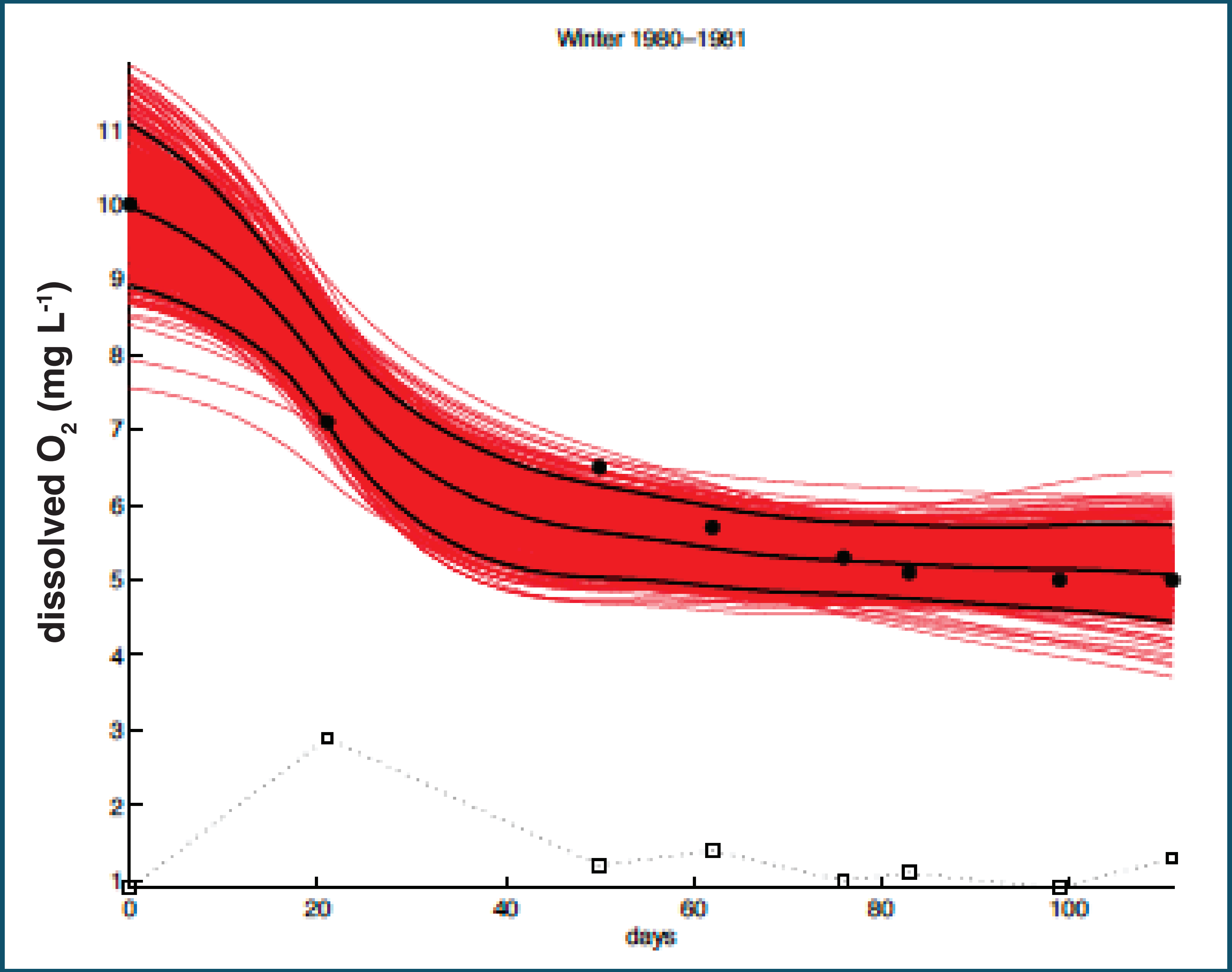
Priors are beliefs or estimates before applying the algorithm

e.g., expected parameter values, $\mu_{max} = 1\ d^{-1}$

or, distributions: $\mu_{max} \sim \text{Gaussian}(\text{mean} = 1, \text{var} = 0.25)$

Posterior, the end result after applying the algorithm

Ensemble, a set of {parameter values, simulations}



Marko Laine (Fig. 5, PhD Thesis, Lapeenranta Univ. of Tech., Finland, 2008)

‘Monte Carlo’ Methods

sounds more sophisticated than

but is it really?

random sampling,
as in gambling

Why Gamble?

In order to approximate the integrals needed to calculate probabilities

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}.$$

Bayes formula (Laine 2008, eq. 9)

$$p(y) = \int p(y|\theta)p(\theta)\,d\theta.$$

(Laine 2008, eq. 10)

**We cannot in general calculate these analytically,
but we can use computers to approximate them by
conducting many simulations,
i.e., discretely sampling the solution space**

and much more...

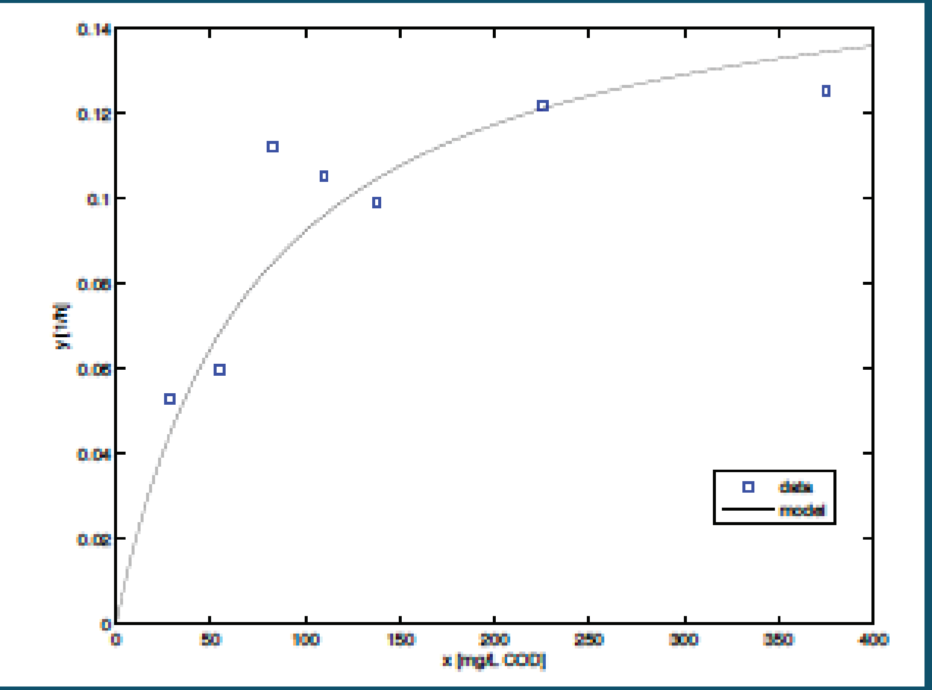
Hastings, WK (*Biometrika* 57, 1970) <- 5, 658 citations (Web of Science Core Collection only)

Marko Laine (PhD thesis, 2008, Appendix A.3.1

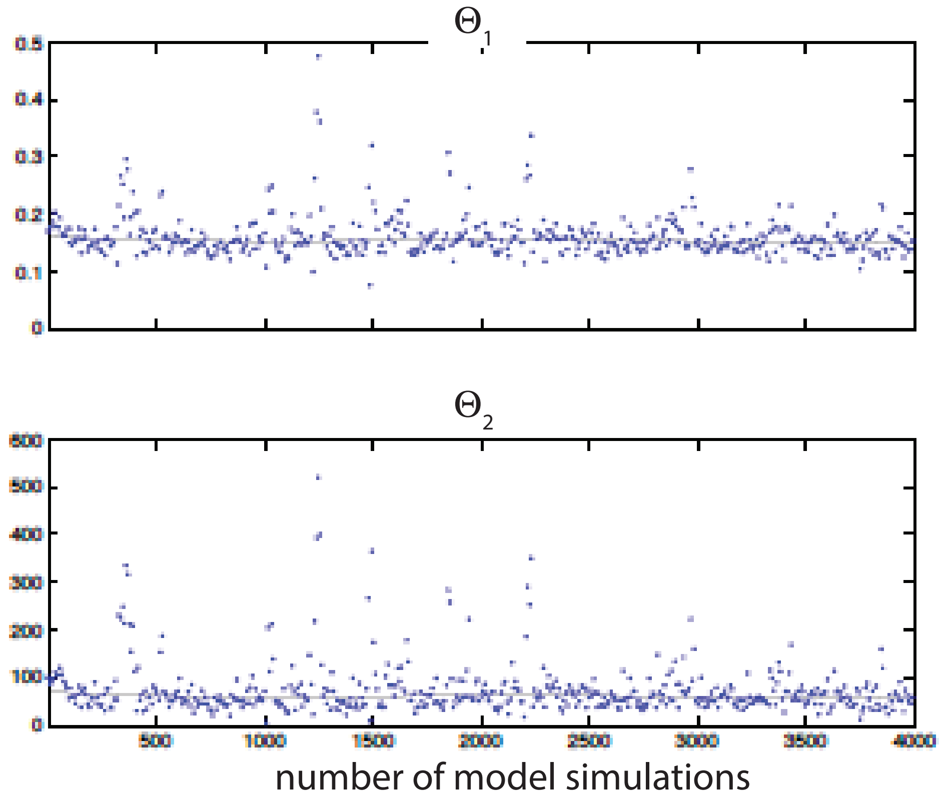
Monod model for growth rate, *y*

$$y = \theta_1 \frac{t}{\theta_2 + t} + \epsilon \quad \epsilon \sim N(0, I\sigma^2)$$

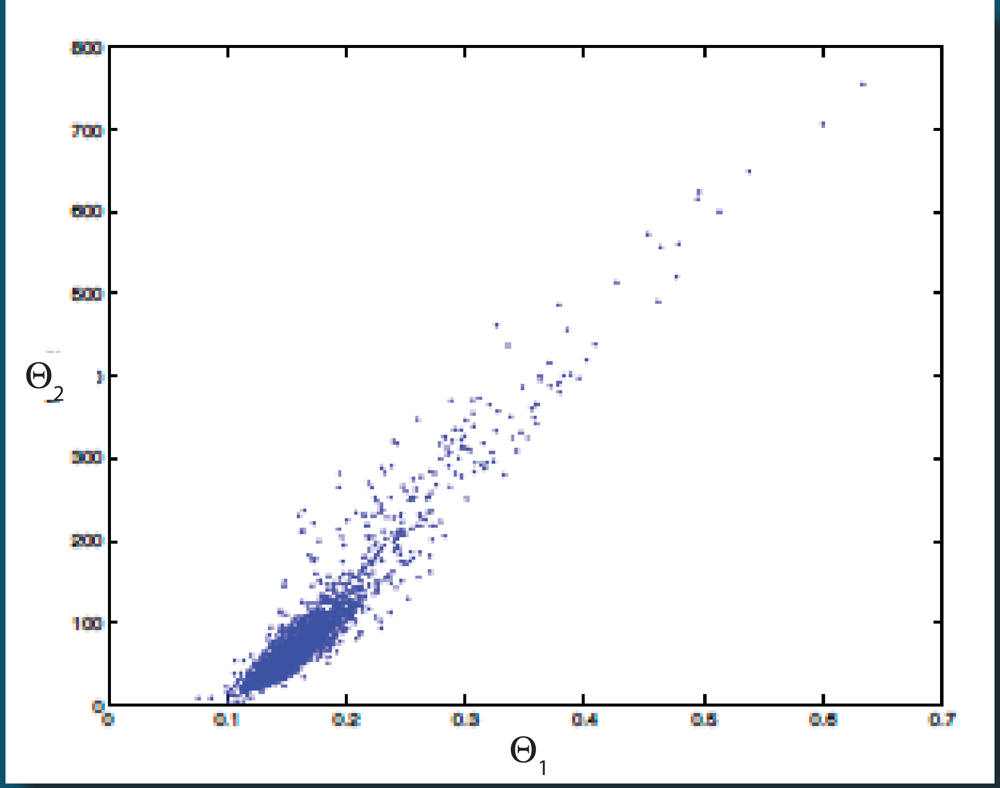
Data



Posterior Ensemble of Parameter Values

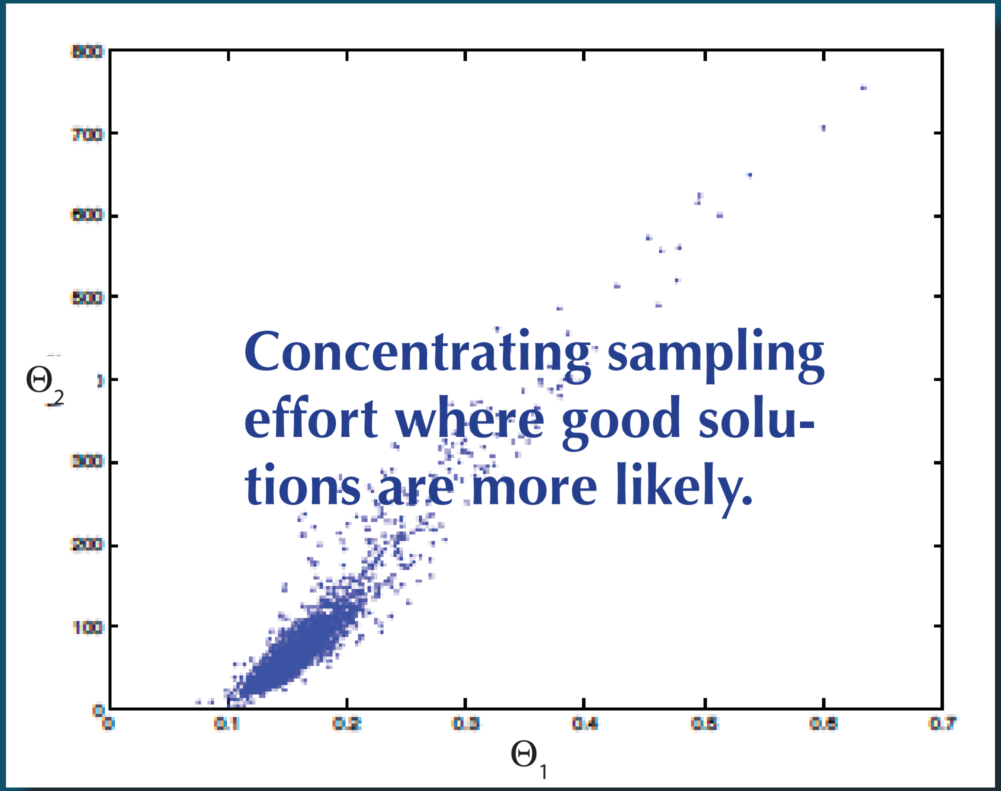


The posterior estimates are correlated!



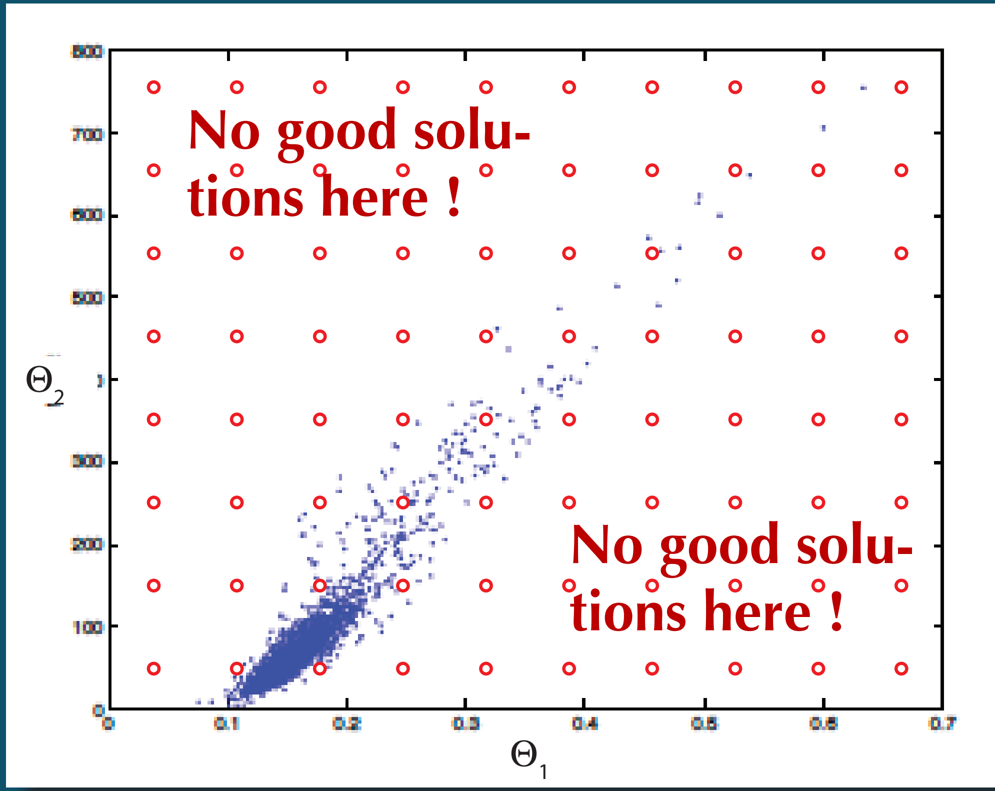
Smart Monte Carlo

AM samples Efficiently by exploiting the Shape of the Param. Distribution



Dumb Monte Carlo

Naive or "Brute Force" samplig wastes effort.



This becomes much more important for higher dimensional problems.

Imagine fitting 10 parameters!

Smart Gambling
patient, strategic

Less Risk.
Consistent payoffs.

Reckless Gambling

Pedro Grendene Bartelle bet big and won big (US$ 3.5 millon) at the roulette table.

But could he repeat that?

The Adaptive Metropolis (AM) Algorithm

Marko Laine (PhD Thesis, Lapeenranta U. Tech., Finland, 2008)

Haario et al. (*Bernoulli* 7, 2001) <- 876 citations (Web of Science Core Collection only)

Based on the Metropolis-Hastings algorithm, but
modified to adapt its Proposal Function based on its past history

=> 1. AM is not Markovian.
But it's more efficient, and it does converge.
2. AM adapts how far & in which direction to "jump" in parameter space.

Automatically samples the standard errors (Gibbs Sampling),
which are used to calculate the Sum of Squares & Likelihood,
yielding an ensemble of σ_d separately for each data type, *d*

$$SSQE_d = \sum_n \left(\frac{x_{mod,n} - x_{obs,n}}{\sigma_d}\right)^2$$

=> Automatic weigthing for data of different kinds, with different units.
Widths of ensembles do indeed cover the range of data.

Not sensitive to initial estimates (starting values) of fitted params.
Allows fits of coupled equations with strong non-linearities

Haario et al. (*Bernoulli* 7, 2000)

Marko Laine (PhD Thesis, Lapeenranta Univ. of Tech., Finland, 2008)

Metropolis algorithms

a broad class of statistical methods for sampling distributions

Monte Carlo Markov Chain (MCMC), Simulated Annealing, etc

usually 'Markovian', i.e., 'jumps' depend only on present state

Adaptive

here 'jumps' do depend on past history

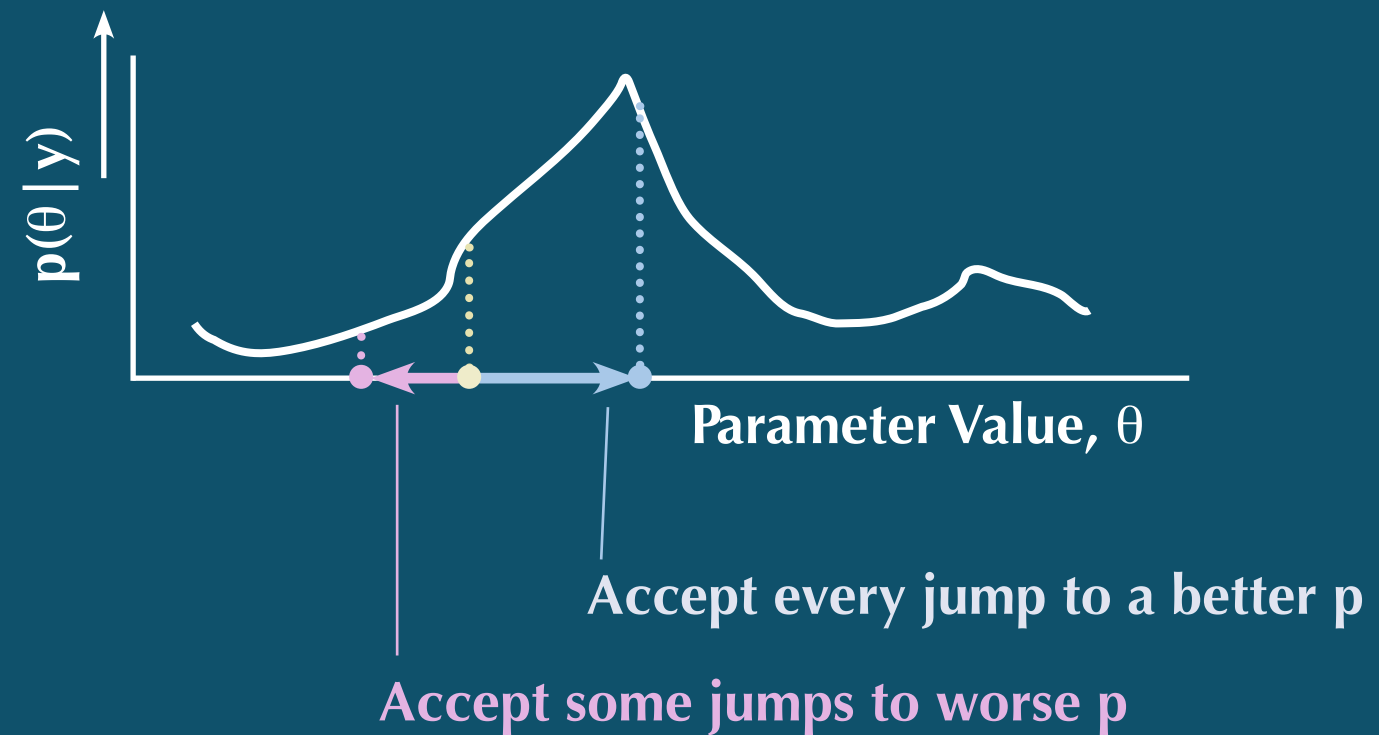
The 'Proposal function' decides the direction & magnitude of 'jumps'

Here it is a mult-variate Gaussian distribution,
based on the past 'chain' of parameter values already sampled

Metropolis algorithms

y observations

θ parameters



Bayes Theorem:

$$p(\theta\,|\,y) = \frac{p(y\,|\,\theta)\; p(\theta)}{p(y)}$$

$p(y\,|\,\theta)$ is the Likelihood of observing y given the model (e.g., assuming Gaussian errors)

$p(\theta)$ is the 'prior estimate' of θ

$p(y)$ is the probability of the observations, which we do not know ... but it cancels out!

'accepting' a jump means 'moving' to the new parameter value, θ^*

$$\text{acceptance probability} = \min\left(1, \frac{p(\theta^*\,|\,y)\; q(\theta^*, \theta)}{p(\theta\,|\,y)\; q(\theta, \theta^*)}\right)$$

$q(\theta^*, \theta)$ is the 'transition density', i.e., decides the probability of jumping from θ to θ^*

Gibbs Sampling for model-data mismatch

If the model-data errors (residuals) are Gaussian

and if we assume a prior such that σ^{-2} is Gamma distributed

$$p(\sigma^{-2}) \sim \Gamma(n_0/2,\ n_0 S_0^2/2)$$

then the conditional distribution of σ^{-2}, given model and data is

$$p(\sigma^{-2} \mid y, \theta) \sim \Gamma(\ (n_0+n)/2,\ (n_0 S_0^2+\Sigma_R)/2\ )$$

where Σ_R is the sum of squared residuals (un-weighted),
and S_0 is the prior mean estimate for σ

At each step in the chain, we can then sample the posterior σ
based on its prior estimate and the sum of squared residuals

This gives an automatic way to assign weights to the data,
so that the posterior distribution (ensemble of simulated values)
will have the same width as, i.e., span or cover, the data

It's Automatic!

Parameters specified for the algorithm:

1. prior estimates of parameter values (mean, co-variance)
2. prior estimates of σ (one for each data type)
   and prior estimates of their accuracy, i.e., compared to # of obs.

In most Metropolis algorithms, e.g., MCMC,
the length scale for jumps in parameters must be specified
but not for AM.

This ratio quantifies relative model skill.

$$\frac{p(M_1|y)}{p(M_2|y)} = \frac{p(y|M_1)}{p(y|M_2)} \frac{p(M_1)}{p(M_2)},$$

Marko Laine (2008)), equation 6

see Smith (*J. Geophys. Res.* 2011) for details of how to apply this

Akaike Information Criterion,

$$AIC = -2\log L + 2P + \frac{2P(P+1)}{(N - P - 1)}$$

where $\log L$ = log likelihood (ensemble mean),
N = no. of observations, P = no. of parameters fitted.

Difference in AIC for model m,

$$\Delta_m = AIC_m - \min\{AIC_i\}$$

Akaike weight for each model: $$w_m = \frac{\exp\{-\Delta_m/2\}}{\sum_i \exp\{-\Delta_i/2\}}$$

relative normalized (0,1) weight that each model is the best of the set of models

Anderson et al. (*J. Wildlife Mngmt.* 64, 2000)

Example: Inferring Combined Effects of T & N Concentrations

Growth rates increase exponentially with T (Eppley. *Fish. Bull.* 1972; Bissinger et al. *L&O* 2008).

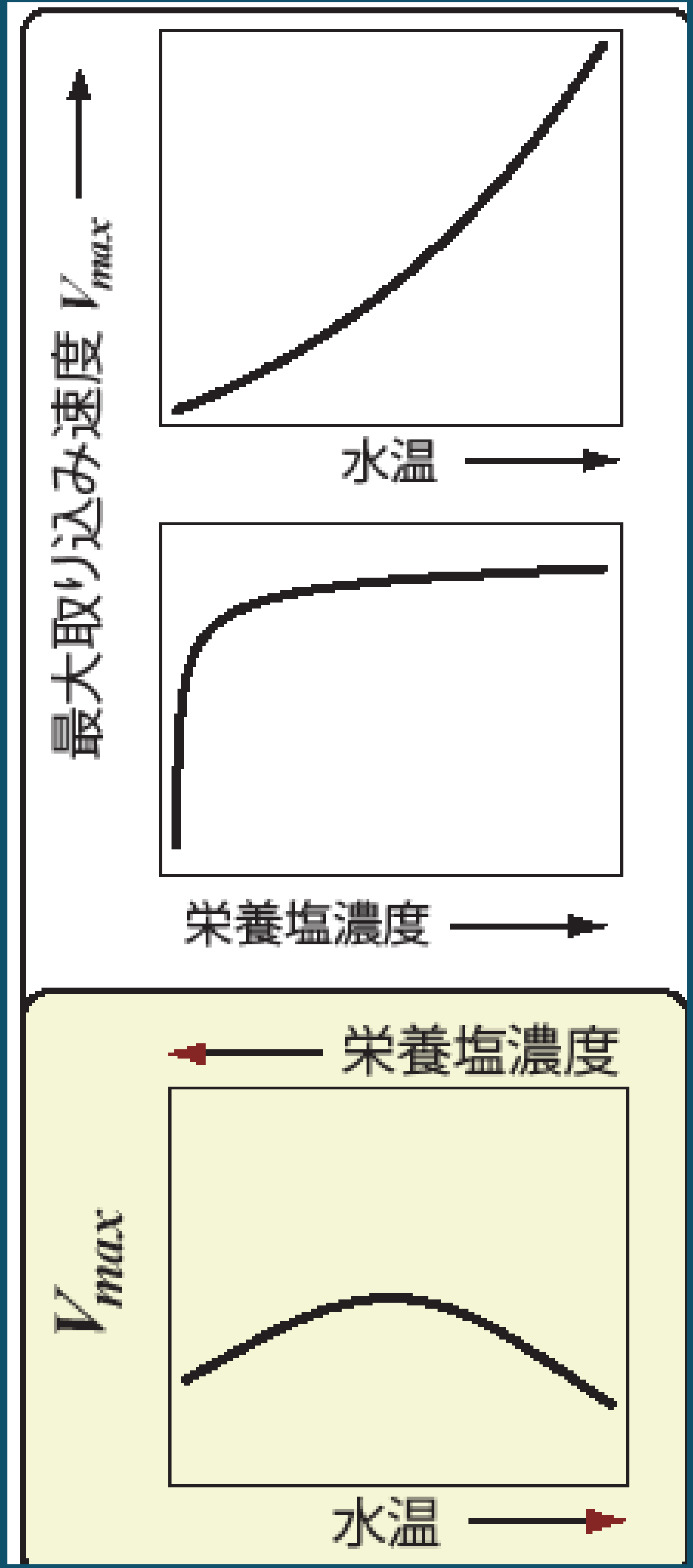
For uptake or growth, V_{max} is usually assumed to be independent of nutrient concentration: Michaelis-Menten (MM) kinetics.

However, Optimal Upake (OU) kinetics predicts that V_{max} (from short-term expts.) should increase hyperbolically with nutrient conc. (Smith et al. *MEPS* 2009).

In the near-surface ocean, *T* and Nutrient Conc. are strongly (negatively) correlated.

Field expts. observe the combined (net) effects.

Assumptions about Uptake Kinetics impact the interpretation of observations.

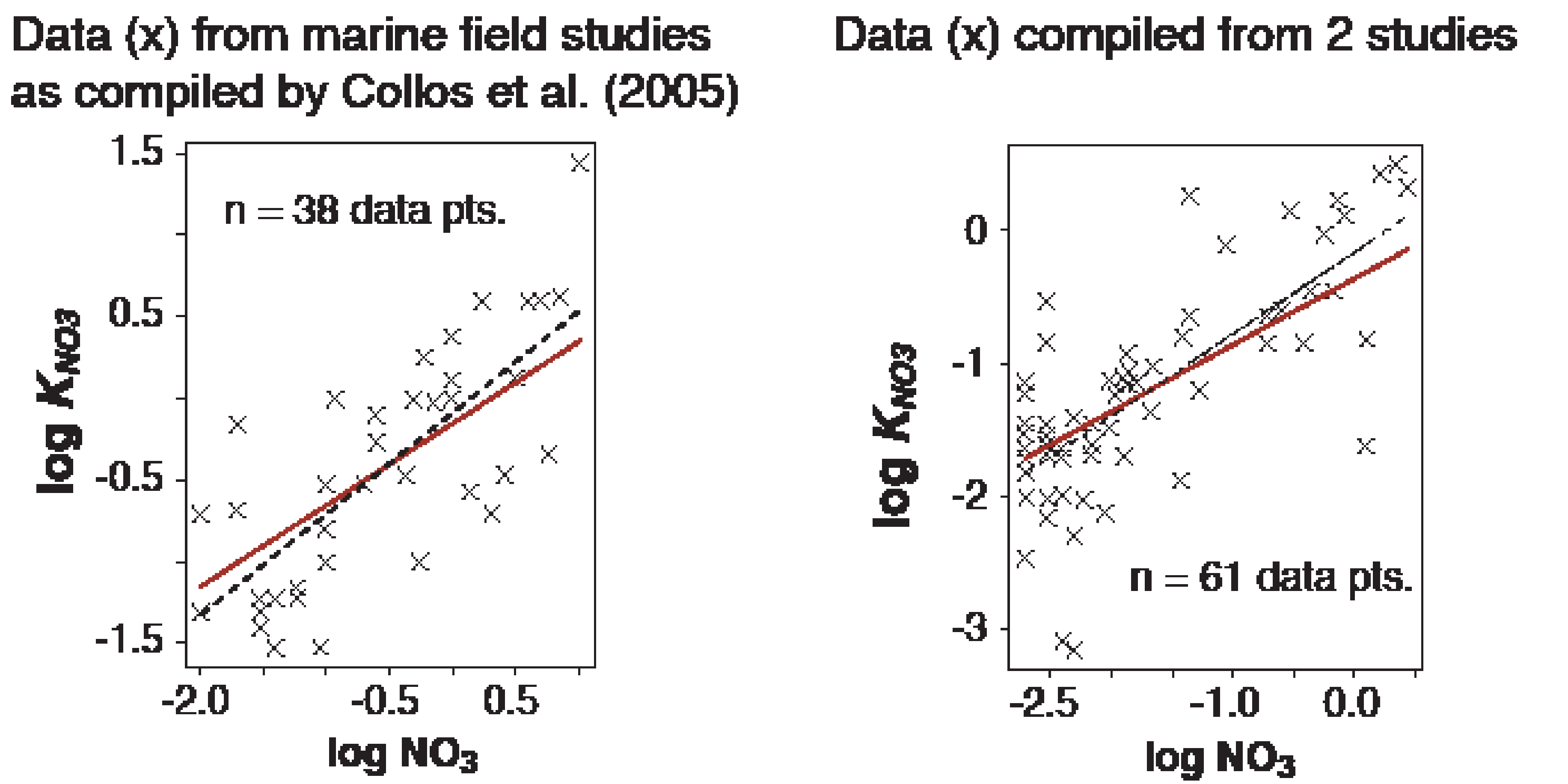


Smith (*Geophys. Res. Lett.* 2010)

K_{NO3} tends to increase with $[NO_3]$

The trend in field observations agrees with the prediction of Optimal Uptake kinetics, although there is wide scatter.

But does K_{NO3} not also depend on *T* ?



Least-squares fits to the data:
------ $\log K_s = -0.089 + 0.62 \log NO_3$
—— $\log K_s = -0.152 + 0.50 \log NO_3$

Least-squares fits to the data:
------ $\log K_s = -0.17 + 0.62 \log NO_3$
—— $\log K_s = -0.36 + 0.50 \log NO_3$

Red lines have the square root dependence predicted by Optimal Uptake kinetics (Smith et al. *MEPS*, 2009)

For one data set from the N. Pacific, Smith et al. (2009) found a weaker relationship with *T* than with [NO3].

Here I examine the T dependence of V_{max} and α, in the data of Harrison et al. (*L&O* 41, 1996)

$$\alpha = \frac{V_{max}}{K_s}$$

Dependence of V_{max} and α on T & Nutrient Concentration

for maximum uptake rate, V_{max}, as determined by short-term expts,

T only

$$V_{max} = V_0 \, e^{-E_{aV}/RT}$$

T & $[NO_3]$

$$V_{max} = \frac{\sqrt{[NO_3]_a A_0/V_0}}{1 + \sqrt{[NO_3]_a A_0/V_0}} V_0 \, e^{-E_{aV}/RT}$$

for α, as determined by short-term expts,

$$\alpha = A_0 \, e^{-E_{aA}/RT}$$

$$\alpha = \frac{1}{1 + \sqrt{[NO_3]_a A_0/V_0}} A_0 \, e^{-E_{aA}/RT}$$

4 parameters were fitted by Adaptive Monte Carlo to a data set for V_{max}, α, $[NO_3]_a$ & T, using both equations simultaneously.

- V_0 potential max. of V_{max}
- E_{aV} Energy of Activation for V_{max}
- A_0 potential max. of α
- E_{aA} Energy of Activation for α

3 parameter fits were also tested

assuming $E_{aV} = E_{aA} = E_a$

This ratio is indepdendent of T only if $E_{aA} = E_{aV}$.

This assumption agrees with the fits for K_{NO3} of Smith et al. (*MEPS*, 2009) and with fits to the data for V_{max} and α, using the data of Harrison et al. (1996).

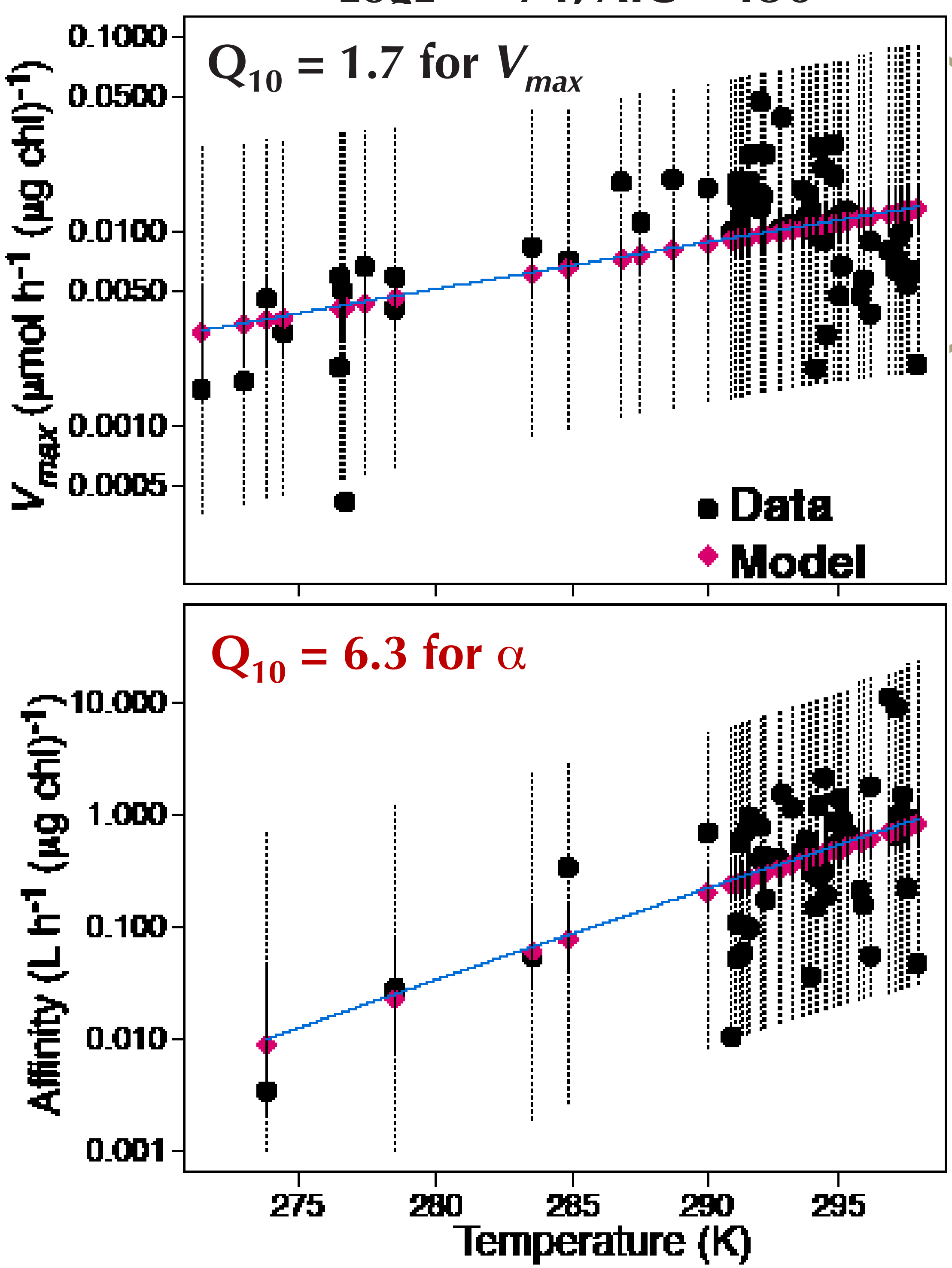
Smith (*J. Geophys. Res.* 2011)

Different Inferred Sensitivies to *T* (for field data from N. Pacific)

Affinity model

Assuming T dependence only

LogL = –74, AIC = 156

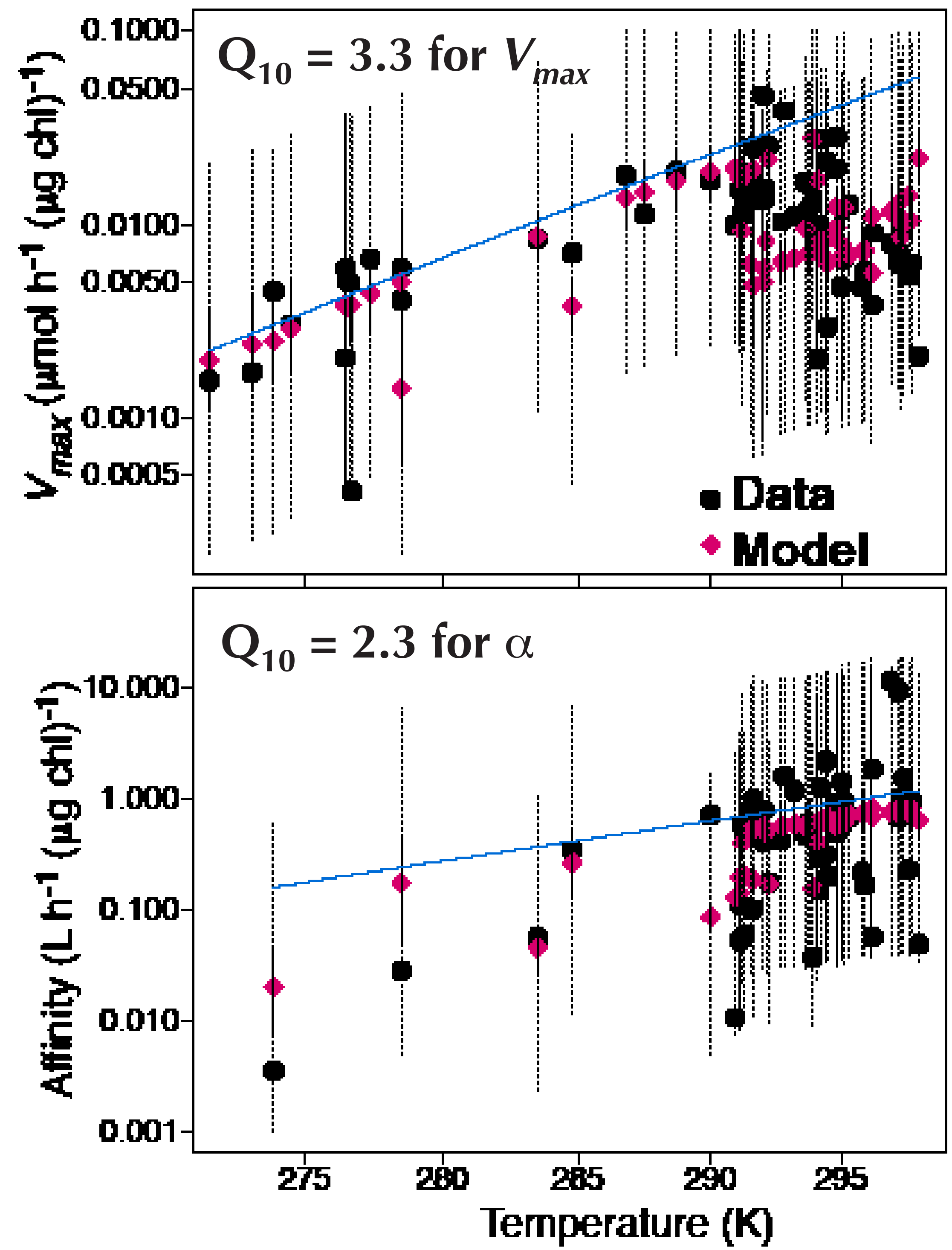


95% width of ensemble +/- 1.96σ_α => 95% of obs. should be in this range.

Solid vertical lines show width of model predictions only (not including error).

OU model

Both T & Conc. Dependence



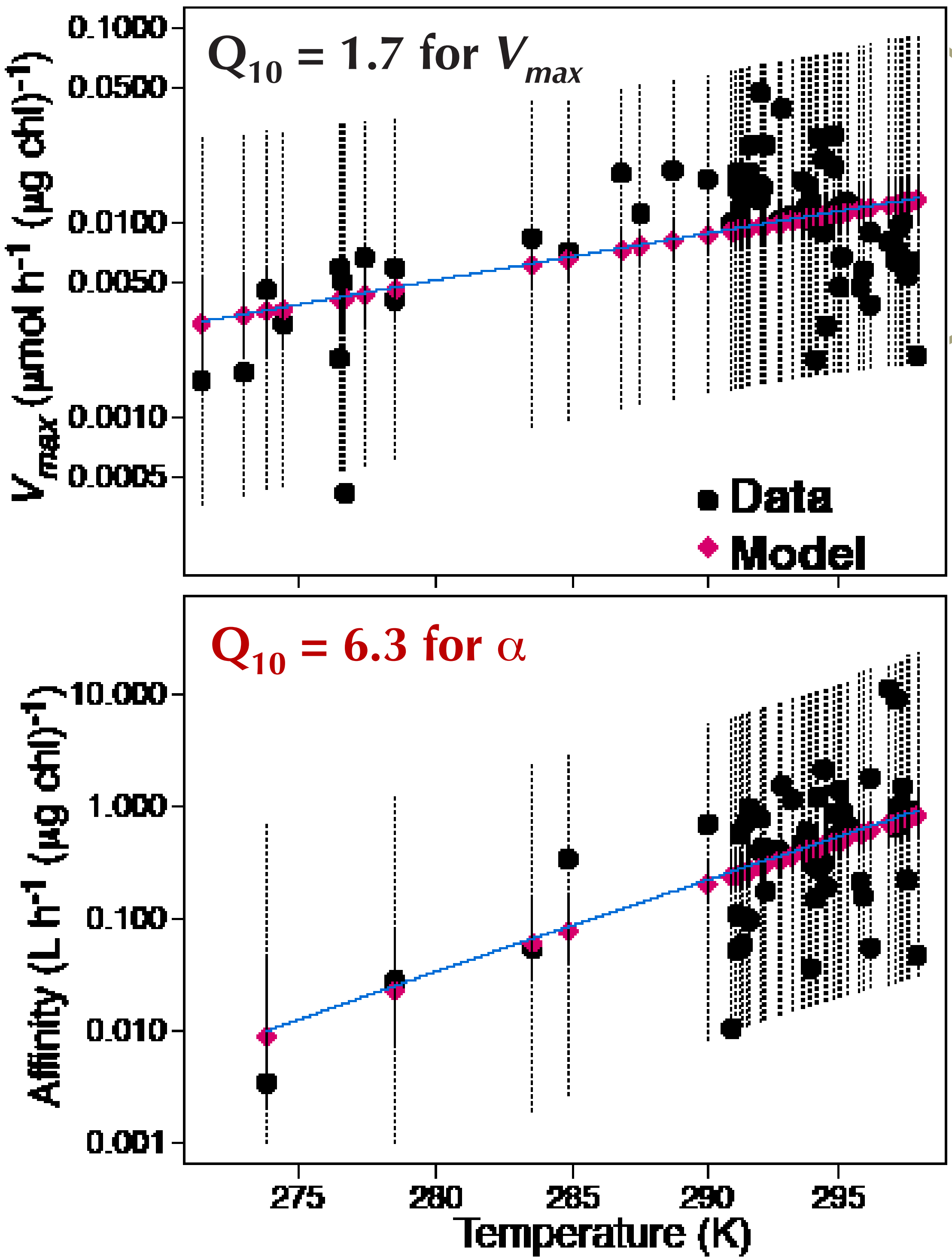
Adaptive Monte Carlo fits of equations for V_{max} and α for Nitrate

Different Inferred Sensitivies to *T* (for field data from N. Pacific)

Affinity model

Assuming T dependence only

LogL = –74, AIC = 156



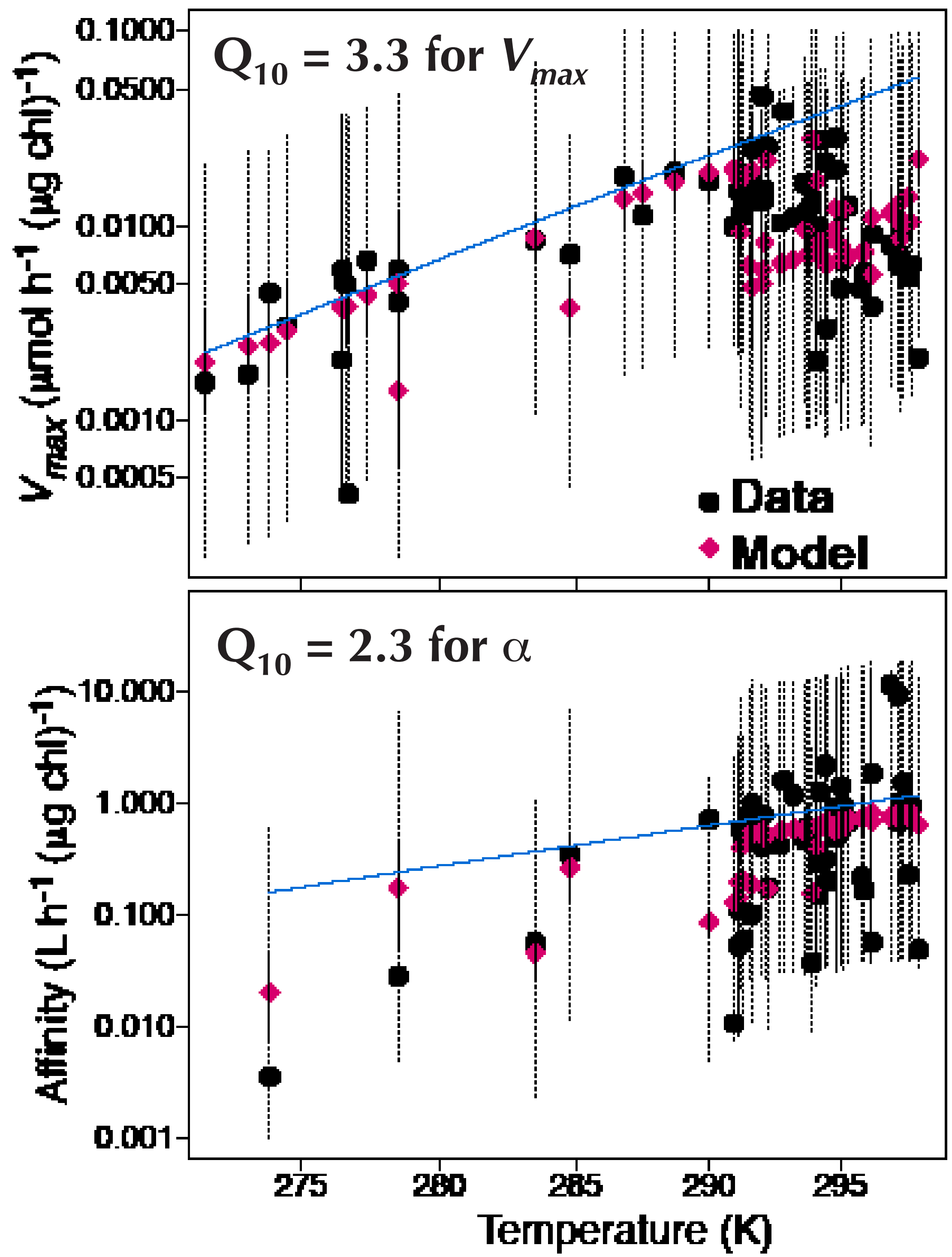
95% width of ensemble +/- 1.96σ_α

=> 95% of obs. should be in this range.

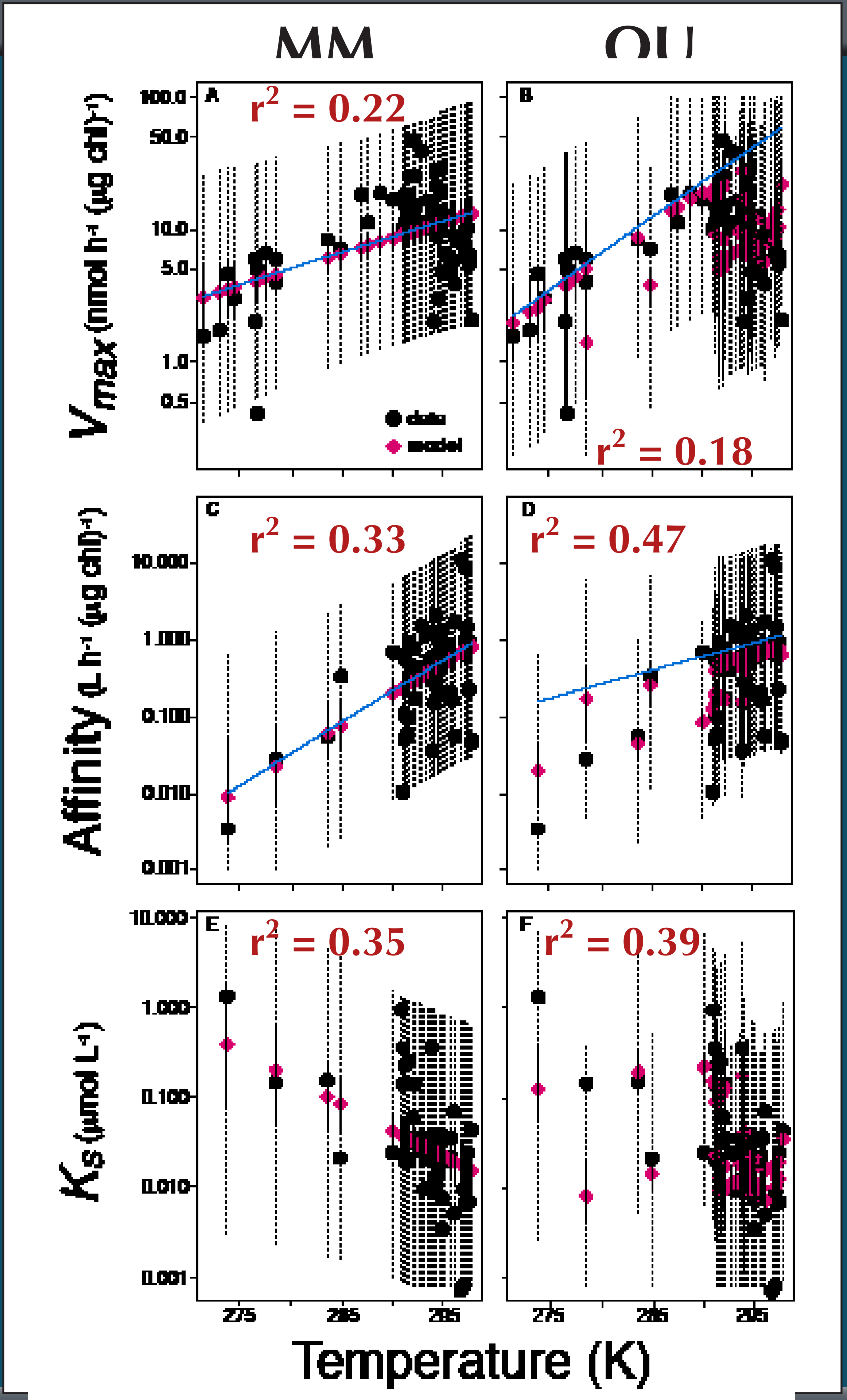
Solid vertical lines show width of model predictions only (not including error).

OU model

Both T & Conc. Dependence



In terms of *MM*, the strong increase in α with T causes K_s to decrease strongly with increasing T.



$$K_s = \frac{V_{max}}{\alpha}$$

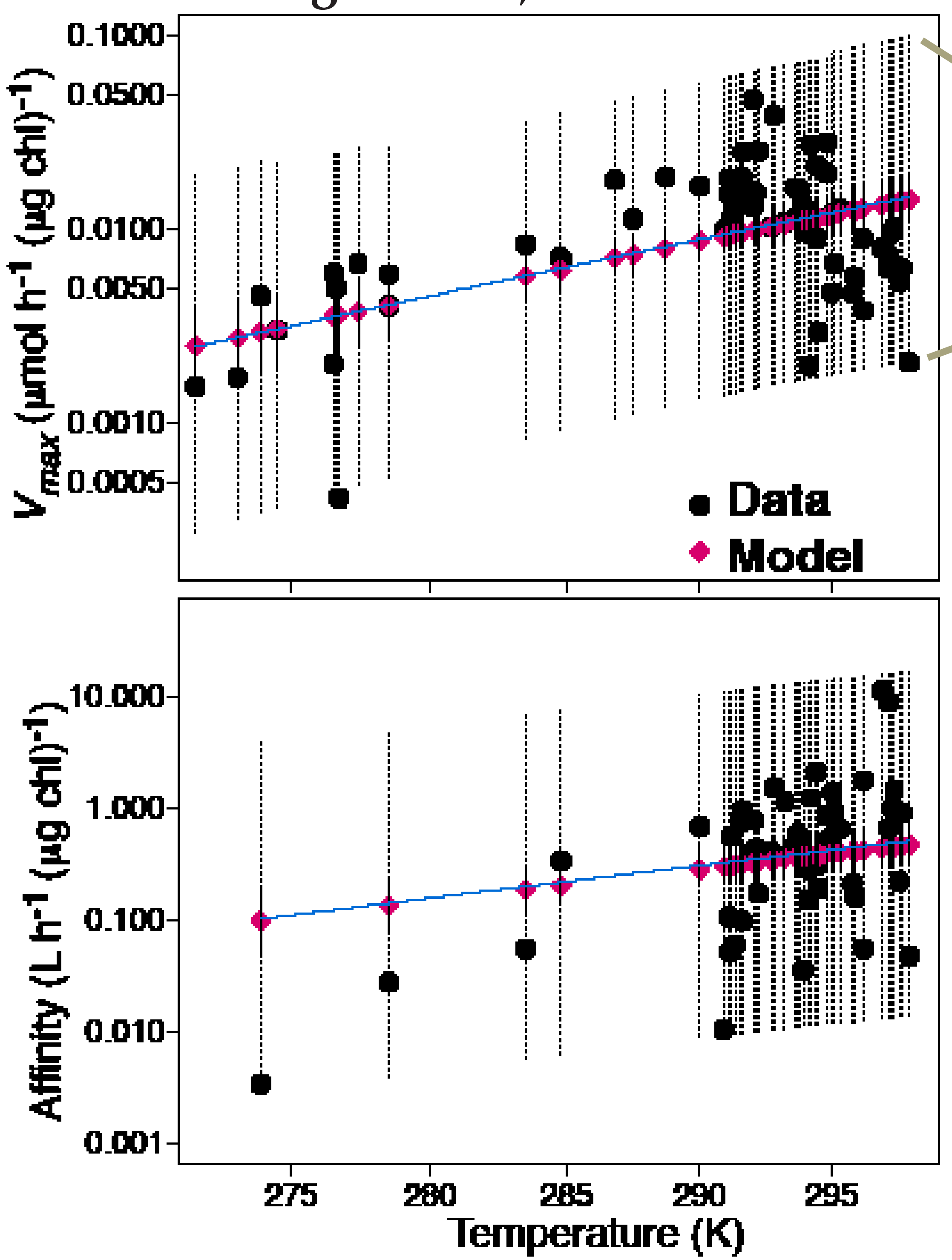
Adaptive Monte Carlo fits of equations for V_{max} and α for Nitrate

Assuming the same T sensitivy (E_a) for both V_{max} and α

Affinity model

Assuming T dependence only

LogL = –73, AIC = 151



3 param. fits with $E_{aV} = E_{aA}$

95% width of ensemble +/- 1.96σ$_\alpha$

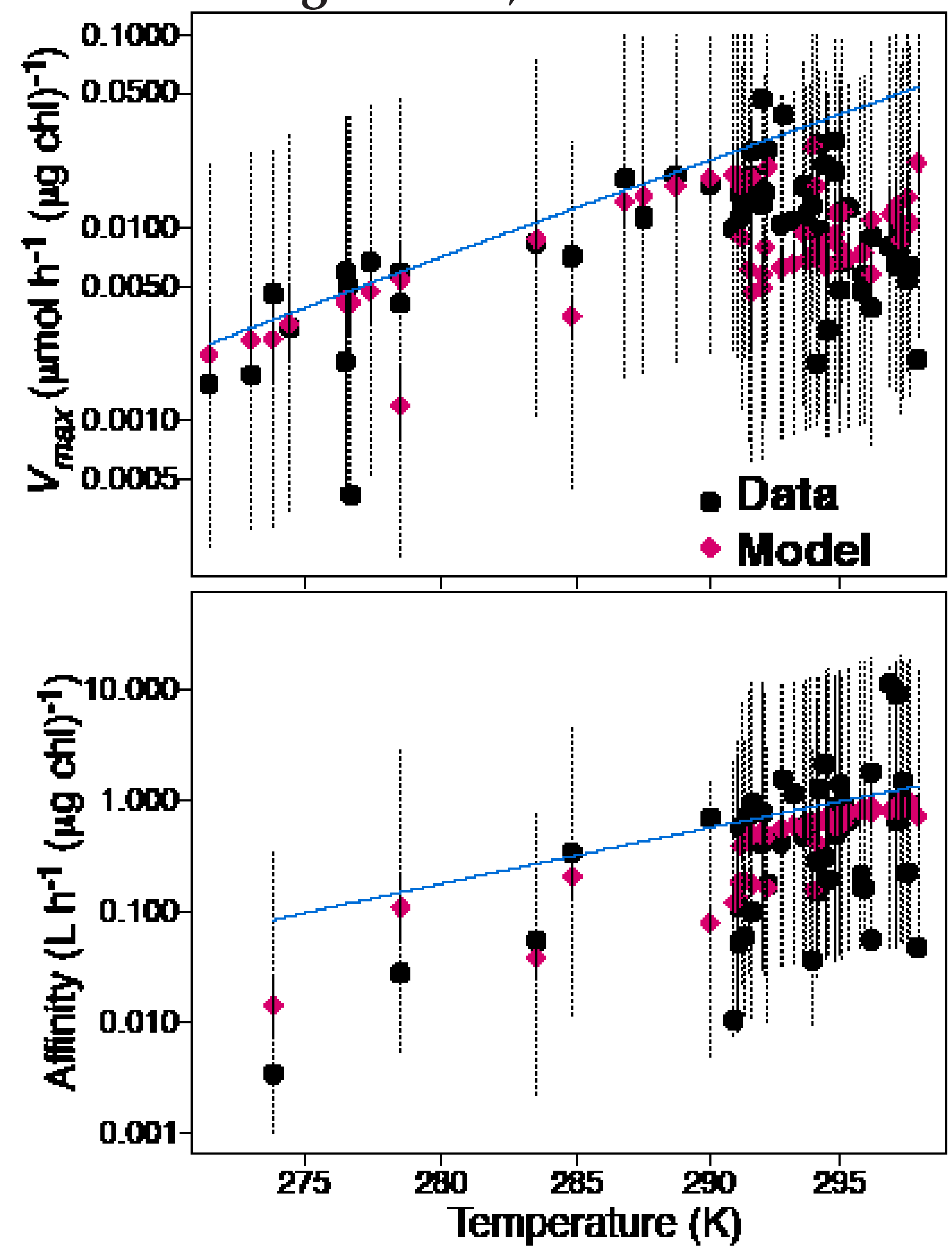
=> 95% of obs. should be in this range.

Solid vertical lines show width of model predictions only (not including error).

OU model

Both T & Conc. Dependence

LogL = –69, AIC = 143



	AIC	Δ	Akaike weight, w
Affinity model			
sep. T sens.	156	12.4	0.002
same T sens.	151	7.6	0.02
OU model			
sep. T sens.	157	13.4	0.001
same T sens.	144	0	0.975

For Michaelis-Menten, $Q_{10} = 1.9$

very close to the value applied in most models (Eppley. 1972)

For Optimal Uptake, $Q_{10} = 3.1$

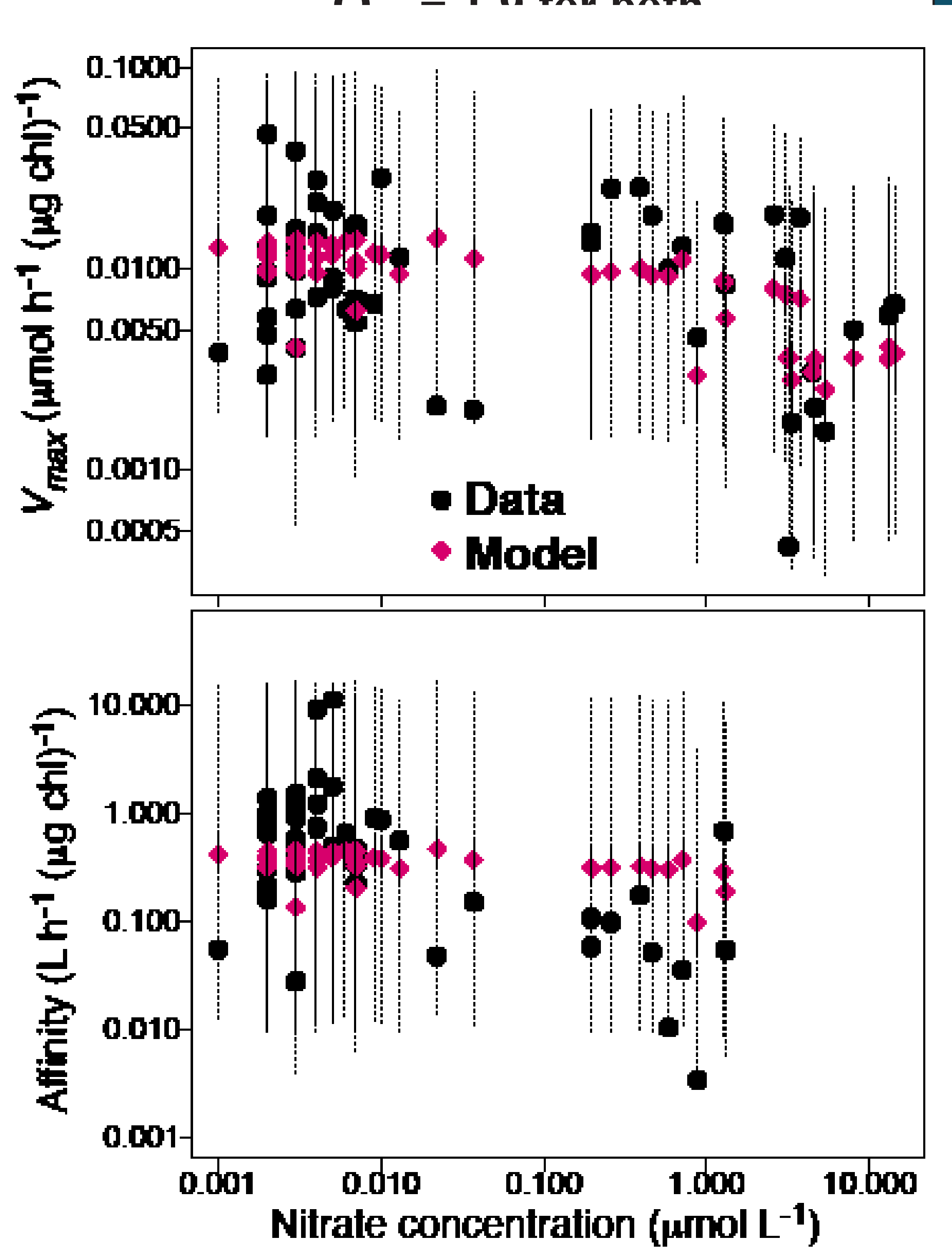
more sensitive to temperature, and agrees better with the data

close to the previous estimate of 3.4 for V_{max} alone (Smith. *GRL* 2010).

Adaptive Monte Carlo fits of equations for V_{max} and α for Nitrate

Here plotted versus Concentration

Assuming T dependence only
LogL = −73, AIC = 151



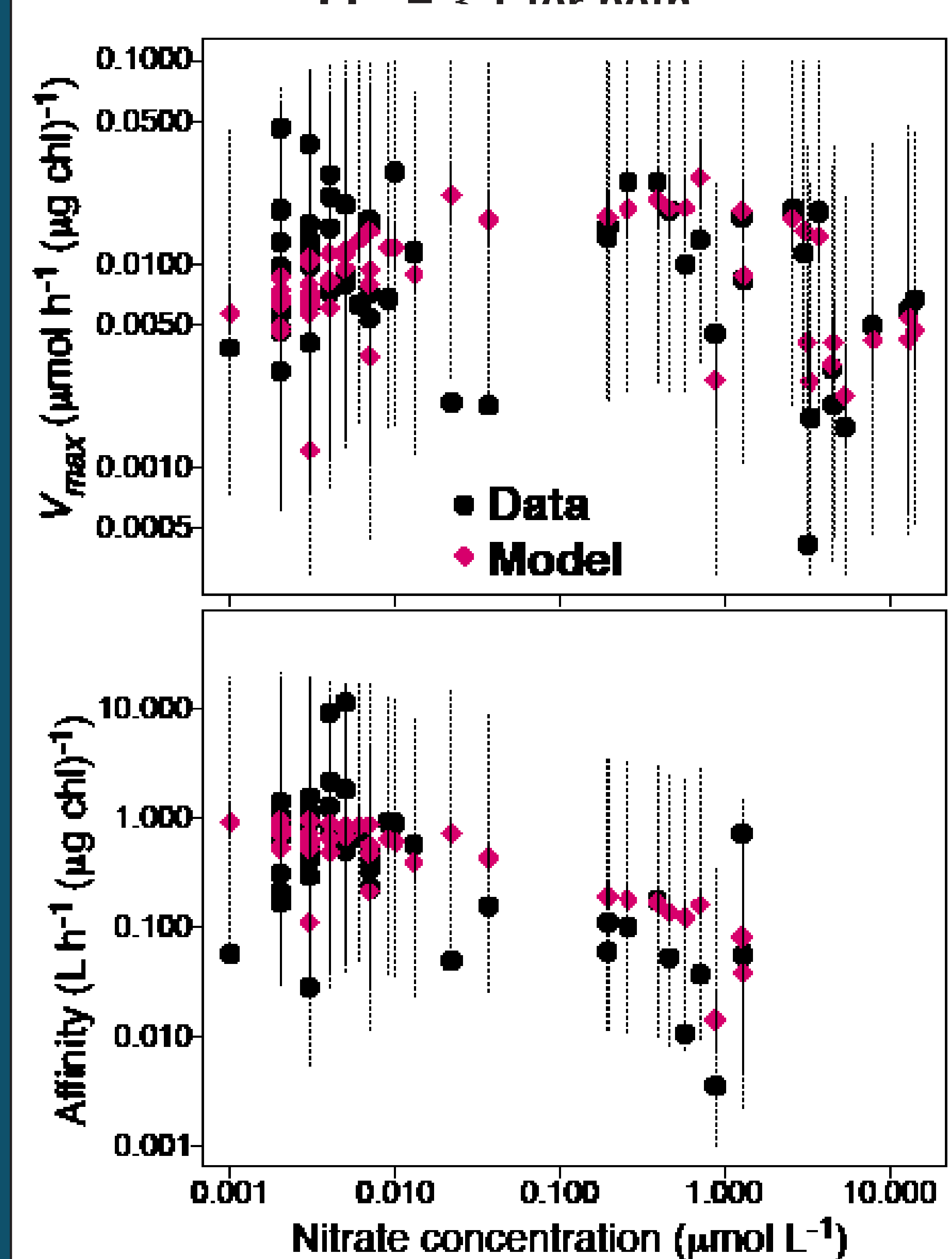
3 param. fits with $E_{aV} = E_{aA}$

The pattern is more complex for V_{max}.

α clearly tends to decrease with [NO3].

Modeled dependence on conc. is weaker than estimated from K_{NO3} alone, but it is still evident, particularly for α.

Both T & Conc. Dependence
LogL = −69, AIC = 143



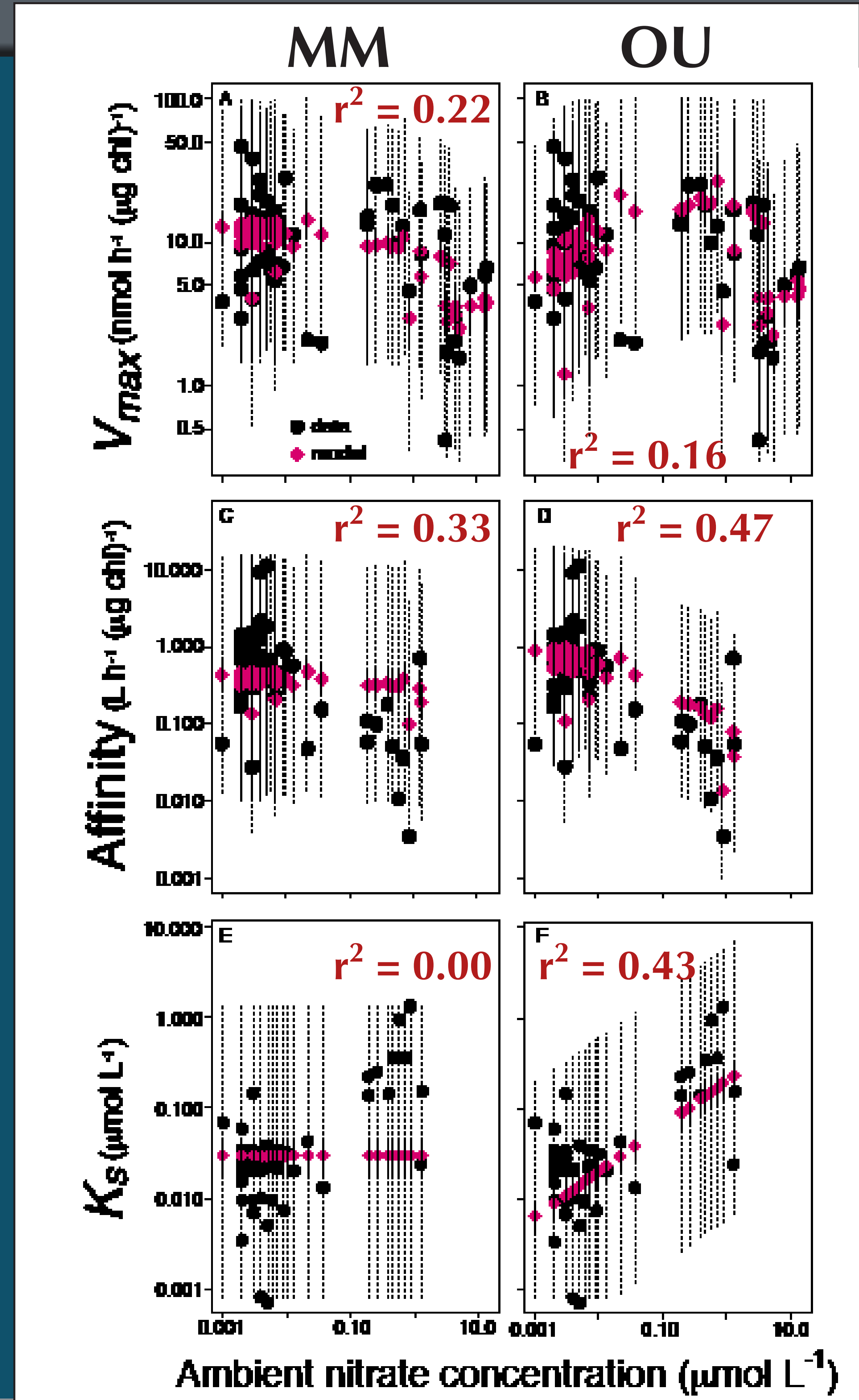
What's going on here?

In terms of MM, the concentration, this explains the increase in K_s with ambient nutrient concentration,

as observed in multiple data sets, for both saltwater and freshwater.

i.e., if K_s depended on temperature, different patterns would be observed in different oceanic regions vs. freshwater.

(Smith. *JGR* 2011)



$$K_s = \frac{V_{max}}{\alpha}$$

No evidence that V_{max} and α have different sensitivities to Temperature

Greater likelihoods for the assumption that they have the same sensitivity, with either uptake kinetics,

i.e., there is no evidence that K_{NO3} depends on T.

Recall that

$$K_s = \frac{V_{max}}{\alpha}$$

This is consistent with findings of a robust relationship between K_{NO3} and $[NO_3]$, for natural assemblages in freshwater and seawater, spanning different combinations of temperature and nitrate conc.

(Collos et al. *J. Phycol.* 41, 2005; Smith et al. *MEPS* 384, 2009).

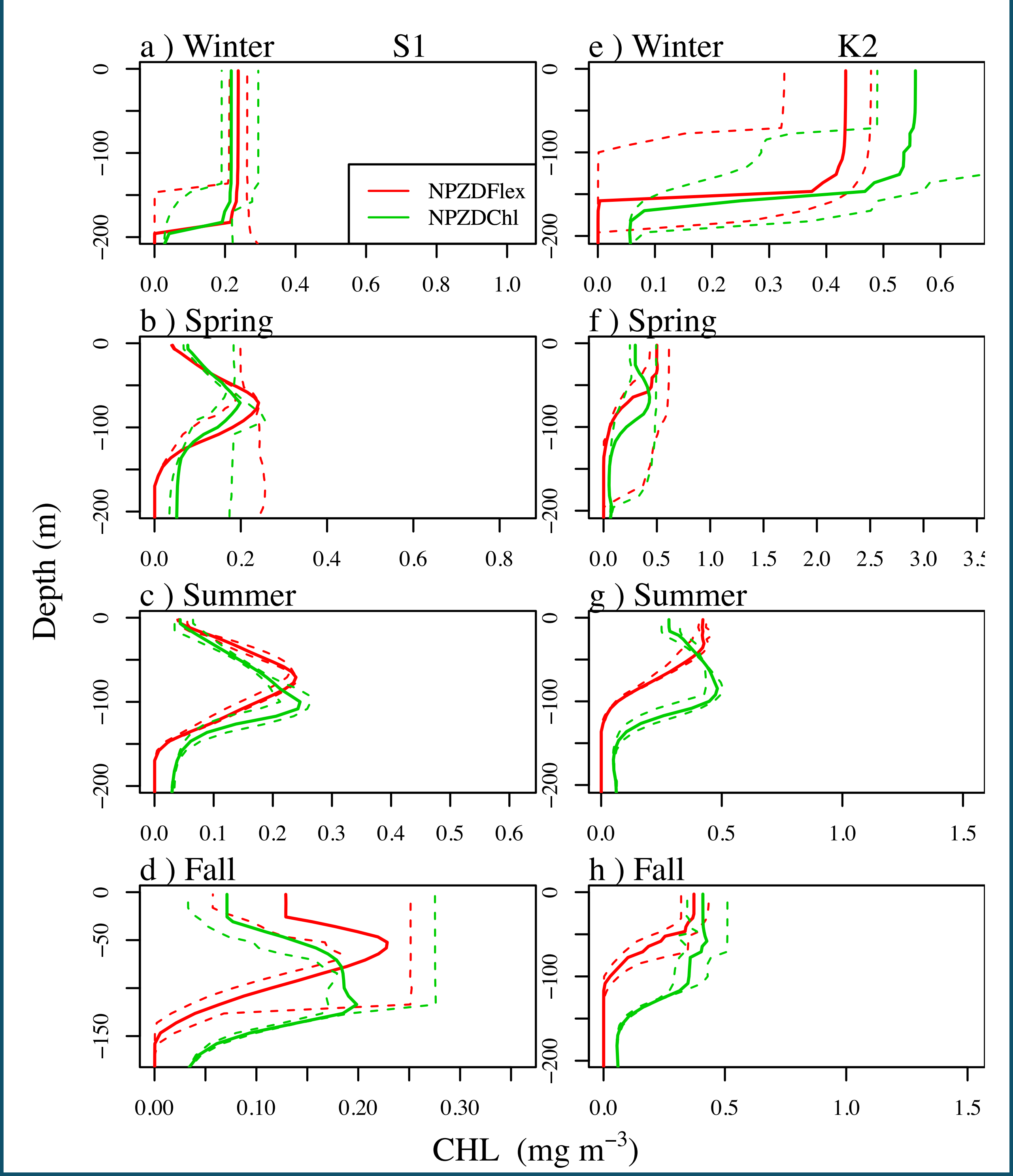
However, note that this contrasts with the general (but not universal) tendency for K_s to increase with T in controlled single-species expts.

(Eppley et al. *Limnol. Oceanogr.* 14, 1969; Dauta. *Ann. Limnol.* 18, 1982)

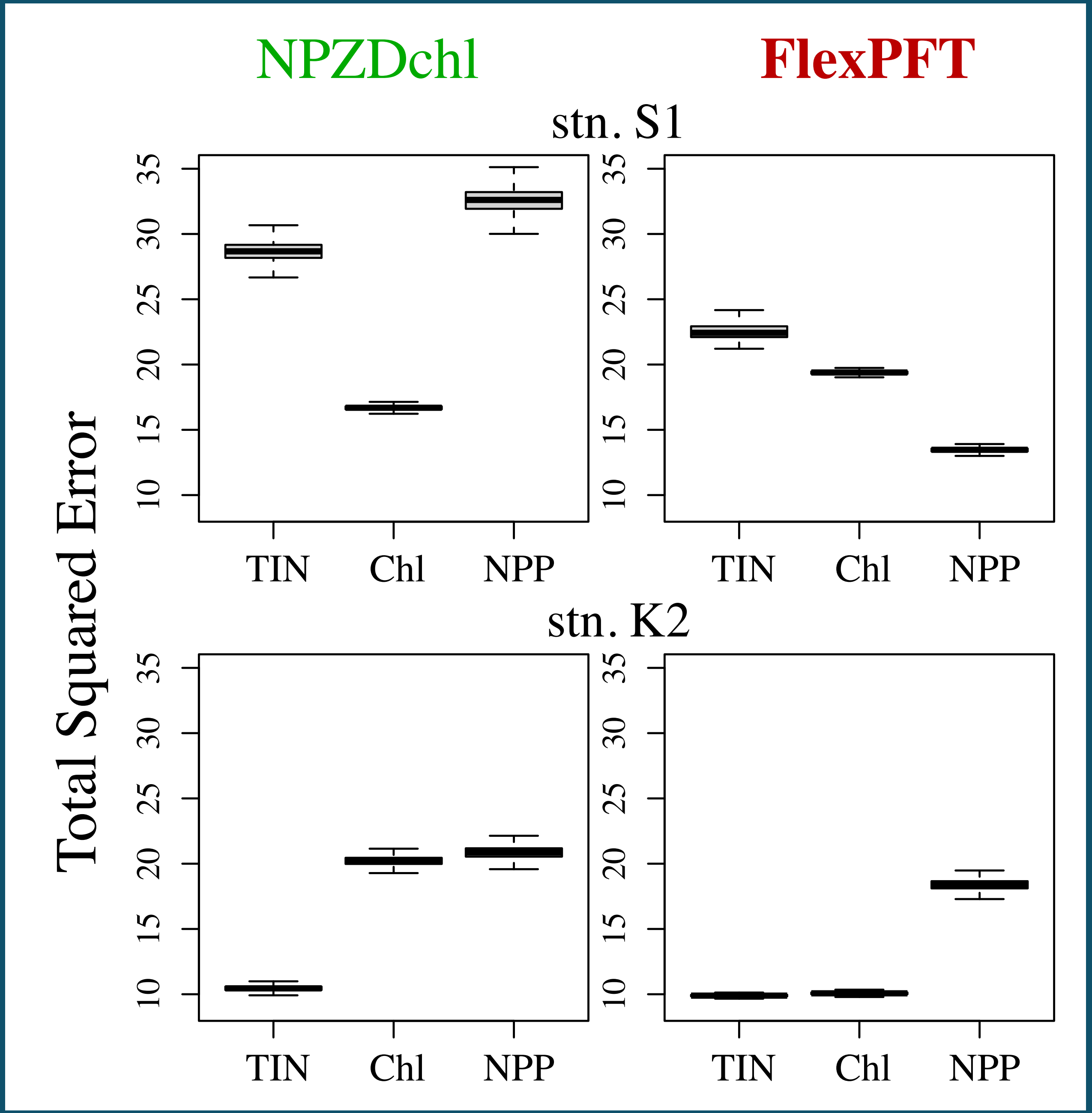
Data Assimilation using a large data set from obs. of TIN, chl, Primary Prod (NPP)

see also Chen & Smith, *Geosci. Model Devel.* 2018)

For example, vertical profiles of chl



FlexPFT performs better, except for chl @ S1



AM & other Metropolis algorithms are now practically useful!

Bayesian Statistics + Fast Computers allow:

- **More Meaningful Model-Data Comparisons & Model Selection**
- **Extracting more Information from Data**

Coding the complicated algorithms is tedious, but it's not necessary!

Various Software is freely available

- **Marko Laine's MCMC toolbox for MatLab**
  https://mjlaine.github.io/mcmcstat/
- **OpenBUGS runs on Windows, Linux, MacOS**
  http://openbugs.net/w/FrontPage
- **Bingzhang Chen's FORTRAN code** (Chen & Smith *GMD*, 2018)
  https://github.com/BingzhangChen/citrate

Thanks for your Attention!