Specific problems of Sequential Importance Resampling filter (SIRF) implementation in ecosystem modelling

Svetlana Losa, Gennady Kivman, Jens Schröter, Manfred Wenzel

Alfred Wegener Institute for Polar and Marine Research
Bremerhaven, Germany

AWI

Workshop on Data Assimilation applications in large-scale models, Delft, November 7-8 2006

Contents

Data assimilation in ecosystem modelling:
ECO MODEL uncertainties

SIRF description:

- initialization →
- model noise generation →
- resampling →
- parameter perturbation →
- spreading the ensemble

Examples

Outlook

Biogeochemical models' skills in reproducing the observed ecosystem dynamics strongly depends on the model biological parameter specification and, furthermore, on reliability mathematical descriptions of modeled biogeochemical processes.

Data assimilation in ecosystem modelling

parameter estimation

(model errors = uncertainties in parameters)

Strong constraint variational technique (VT)

Fasham and Evans, 1995
Matear, 1996
Prunet et al., 1996
Hart and Armstrong, 1996
Spitz et al., 1998, 2001
Fennel et al., 2001
Schartau et al., 2001

state estimation

(model errors = uncertainties in forcing, …)

global

Kagan et al., 1997
Natvik et al., 2001

sequential

Monte-Carlo methods

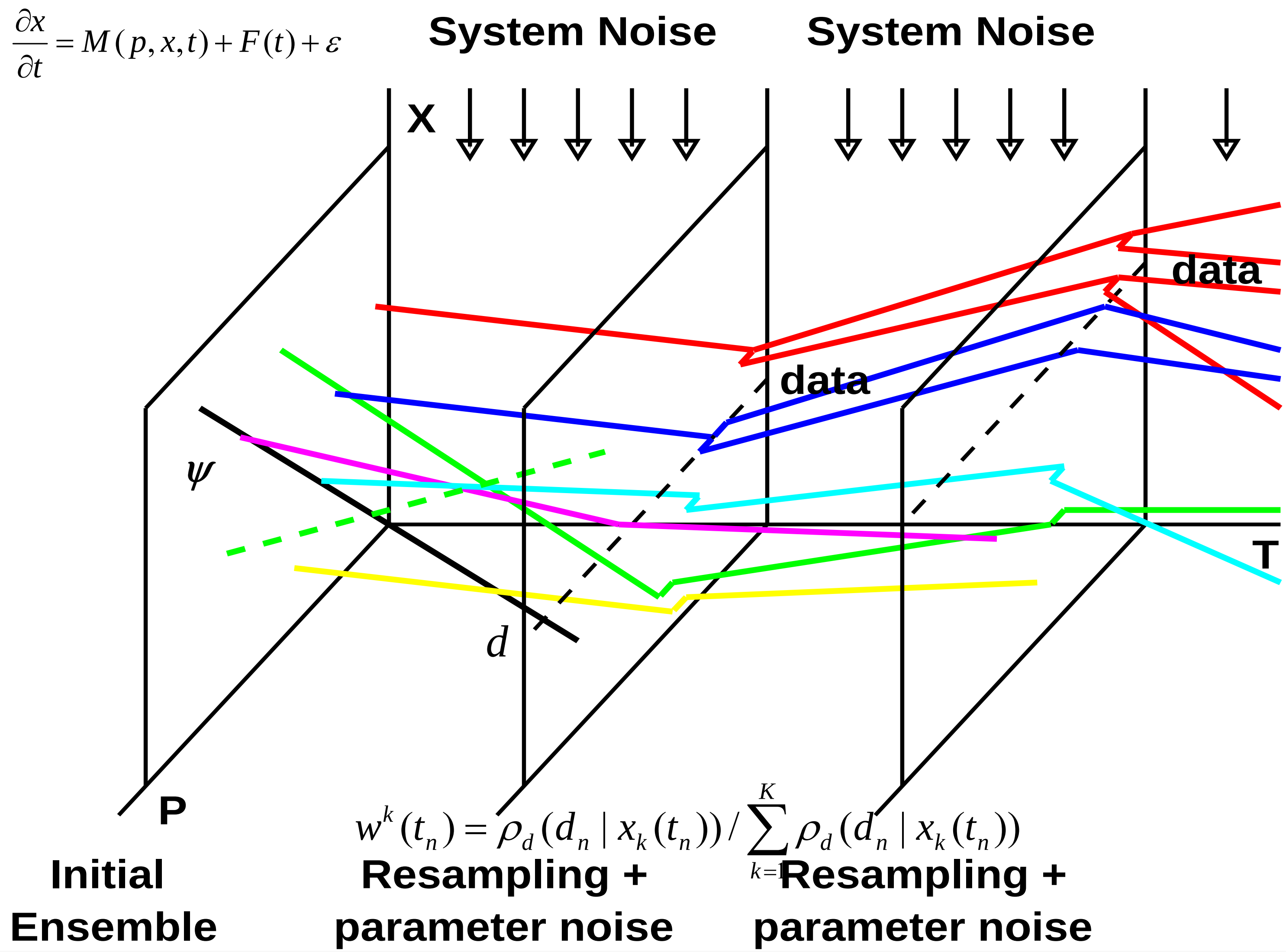
Eknes and Evensen, 2000
Carmillet et al., 2001
Natvik and Evensen, 2003
Nerger and Gregg, 2006

Weak constraint VT

Losa, Kivman and
Ryabchenko, 2004

Sequential Importance Resampling filter

Losa, Kivman, Schröter, Wenzel, 2003
Brasseur et al., 2006



$$\rho_t^f(x(t), p) = C\rho^f(x(t) \mid x(0), p)\rho(p)\rho_0(x(0))$$

$$\rho_t^a(x(t_1), p \mid d_1) = C\rho_d(d_1 \mid x(t_1))\rho_t^f(x(t_1), p)$$

The Sequential Importance Resampling filter has been first introduced by Rubin(1988), implemented for dynamical systems by Gordon et al. (1993).

The SIR filter is known to suffer from degeneration of the ensemble (van Leeuwen, 2003) if either the system noise does not provide sufficient spreading of states which are resampled several times or the ensemble badly approximates the true prior distribution (the distance between the best member and the true state is too big).

This problem is even more pronounced in the case of simultaneous state-parameter estimation where regenerating the number of samples in the parameter space is needed.

Ensemble Initialization

Spread of the initial ensemble reflects uncertainties in knowledge of *a prior* system and parameter pdf

$$\rho_0(x(0)) = K^{-1}\sum_{k=1}^{K}\delta(x(0)-x_k(0)),$$

$$\rho(p) = K^{-1}\sum_{k=1}^{K}\delta(p-p_k),$$

δ - Dirac functions

An ensemble of K members is generated from an exponential distribution

$$y = \bar{y}\exp(-\frac{y}{\bar{y}})$$

mean of the distribution is assumed to be a first guess.

Meaning of parameter perturbation

Physiological: biological parameters vary in space and time
Mathematical: avoiding the ensemble collapse

Meaning of model noise generating

With respect to SIRF algorithm: ensemble spreading
With respect to eco modelling : model errors identification,
more accurate parameter estimation

Model noise generation and jittering model parameters

Levels of the model noise *E* might be considered as additional parameters to be optimized $E \subset P$.

If, at an analysis step, parameter values are resampled *(r)* many times, a new parameter ensemble can be redrawn (West, 1993) from a smoothed approximation of the *posterior* probability density

$$\rho_t^a(p(t_n)) = K^{-1} \sum_{k=1}^{K} \delta(p(t_n) - p_a^r(t_n))$$

either from

a uniform distribution within the interval $p \pm \sigma_p$… one has to specify [p – nearest smaller value, p + nearest higher value];

or

a normal distribution with a variance… one has to specify;

…

Data and weighting

The Bermuda Atlantic Time-series Study:
measurements of nitrate, chlorophyll, dissolved organic nitrogen and carbon concentrations for the period December 1988 to January 1994.

All the data were averaged over the ocean upper mixed layer (UML).
The UML thickness were estimated by means of an analysis of BATS temperature profiles for the same period. The UML depth is determined as the depth at which the temperature is 0.5^{0}C less than that at the surface.

The relative weights might be calculated under the assumption of Gaussian

$$\rho_d(d_n \mid x_k(t_n)) \approx \omega^k = C \exp[-0.5 (X^k - d)^2 / \sigma^2],$$

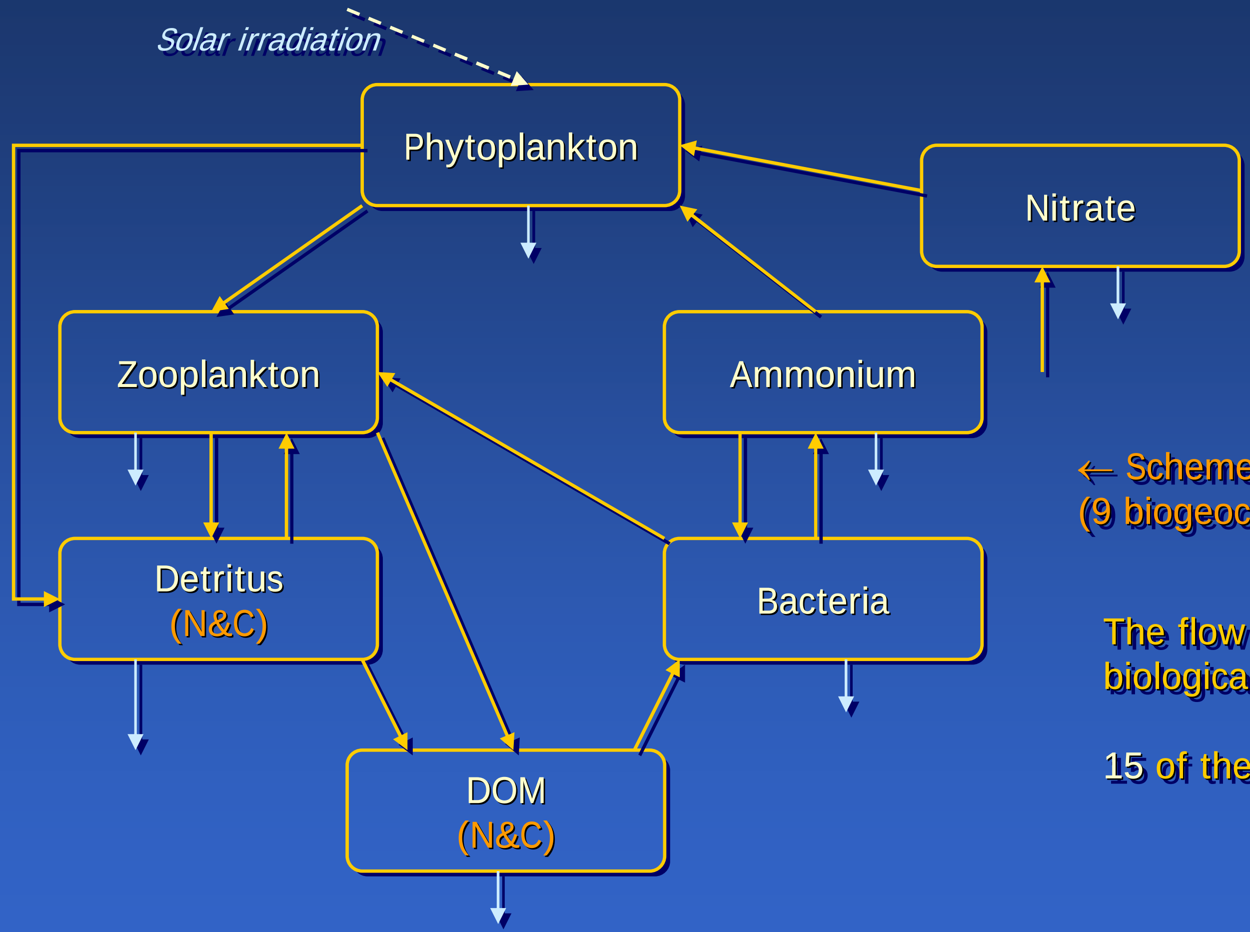
or Lorentz data errors

$$\omega^k = C/(1 + (X^k - d)^2 / \sigma^{-2}) \qquad \textit{(van Leeuwen, 2004)}$$

where σ^2 is the variance of the observation.

H. Drange's Ecosystem Model

(1996)



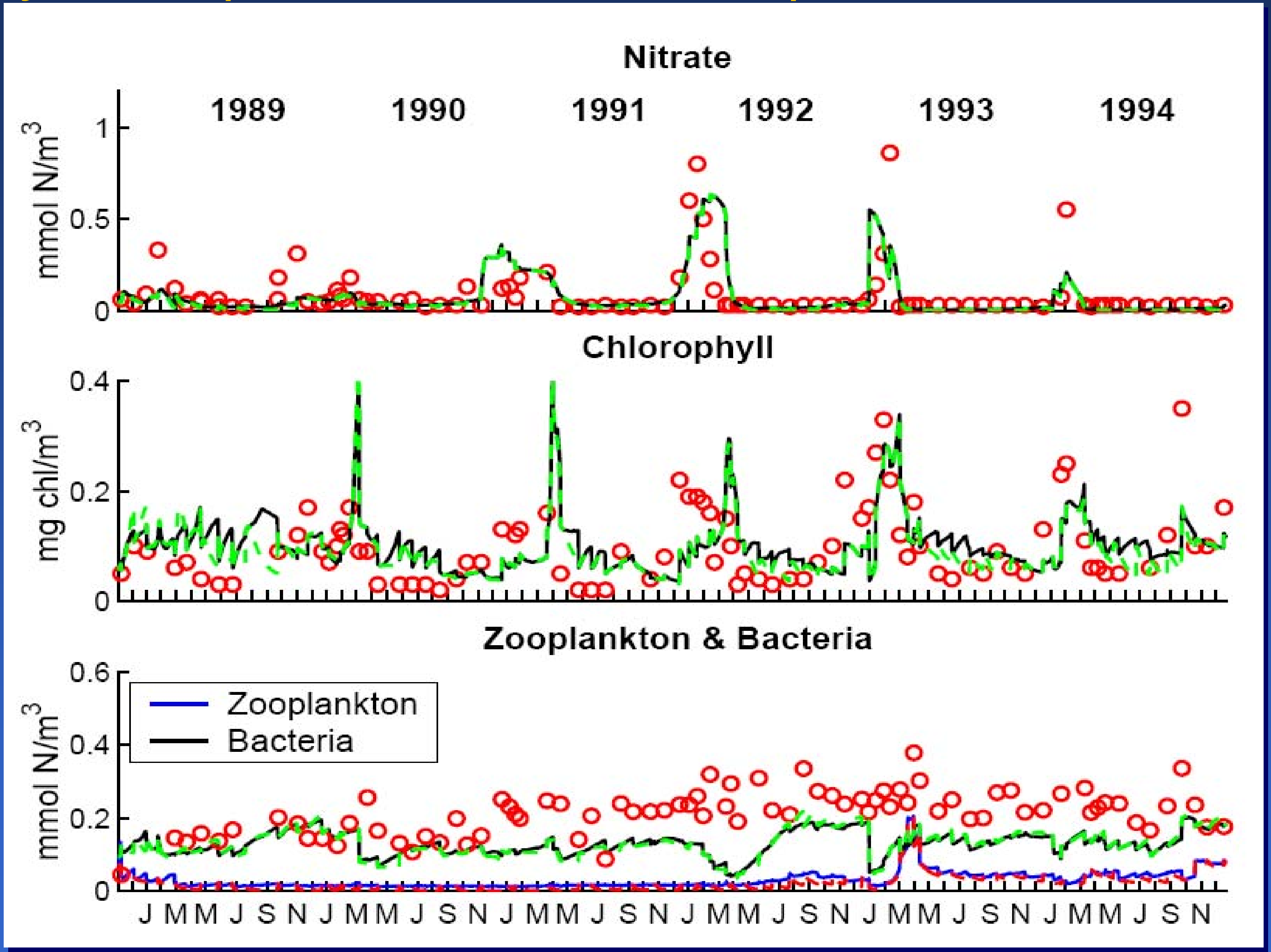
← Scheme of a reduced version (9 biogeochemical compartments)

The flow network possesses 29 biological parameters.

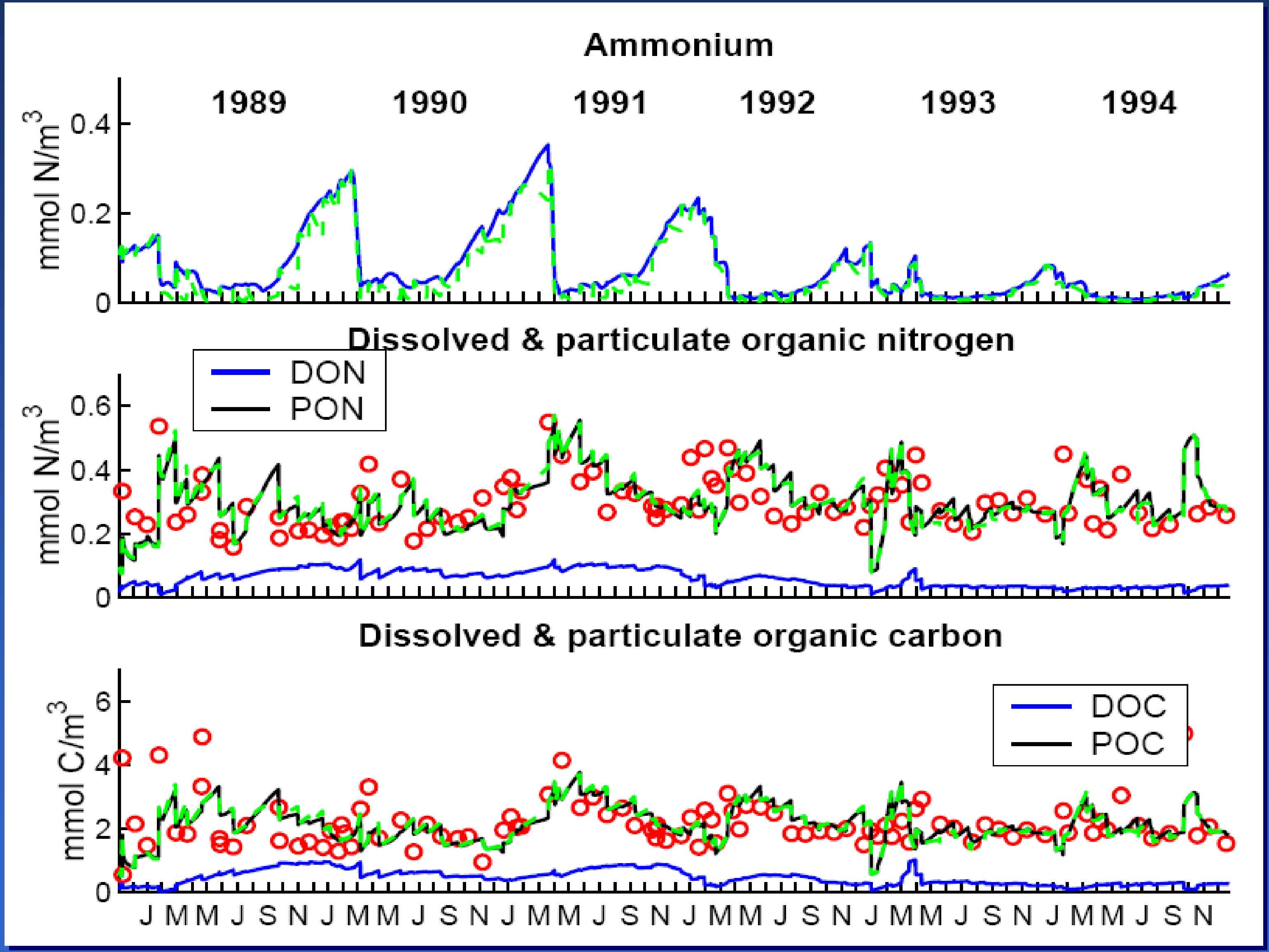
15 of them have been adjusted

The authors thank Helge Drange for the provided model code.

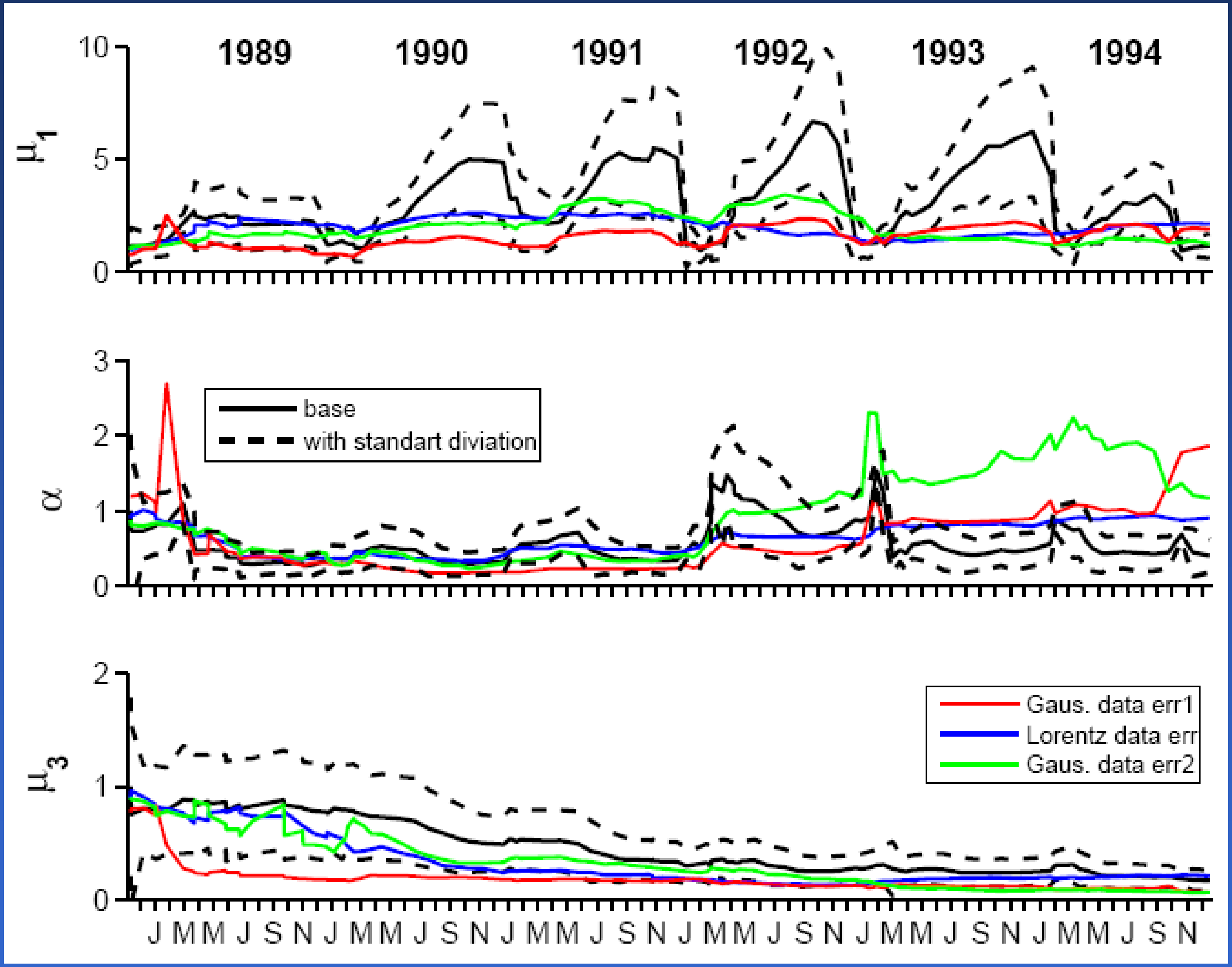
The evolution of the ecosystem components at the BATS obtained by the sequential weak constraint parameter estimation

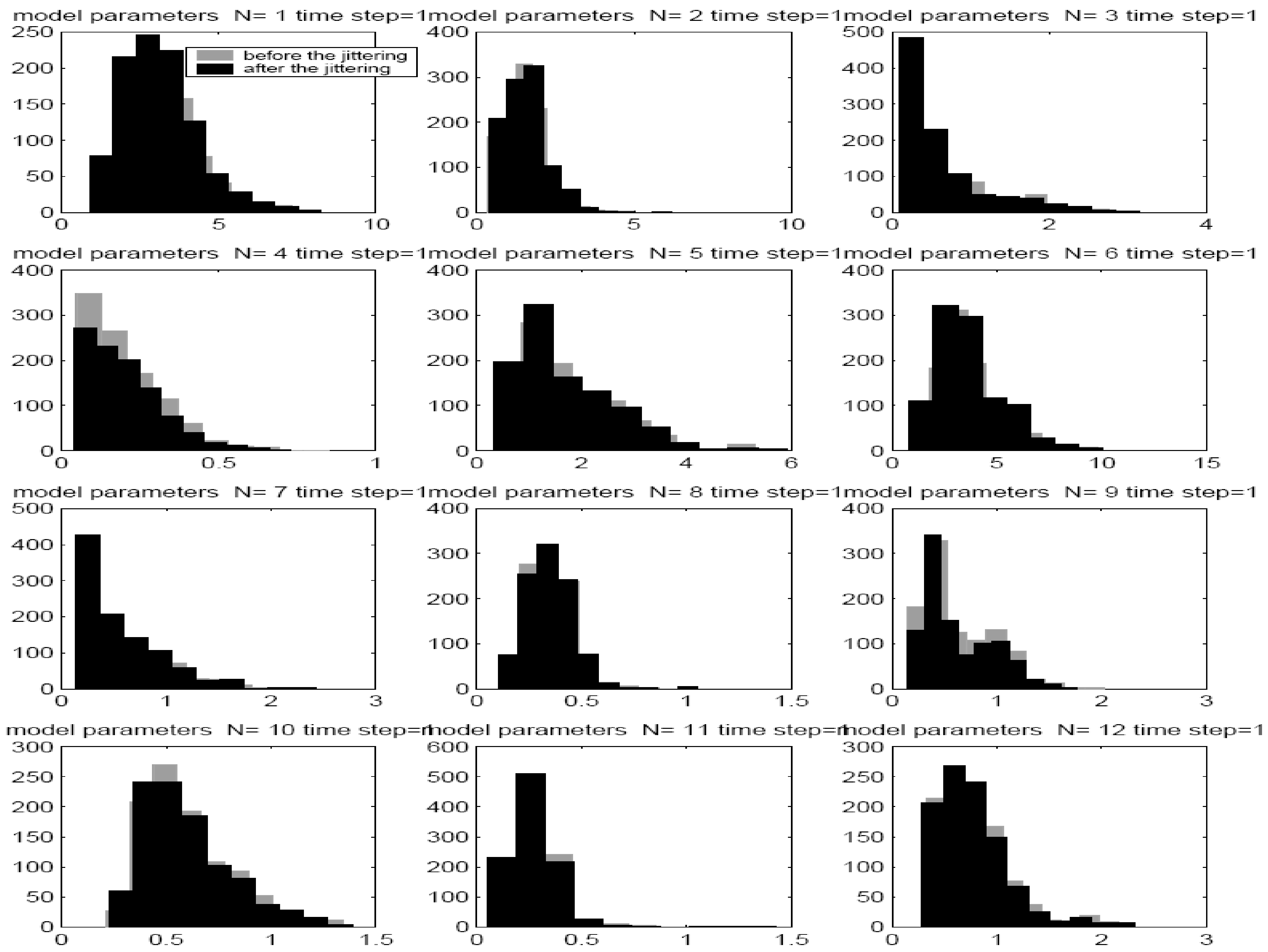


The evolution of the ecosystem components at the BATS obtained by the sequential weak constraint parameter estimation

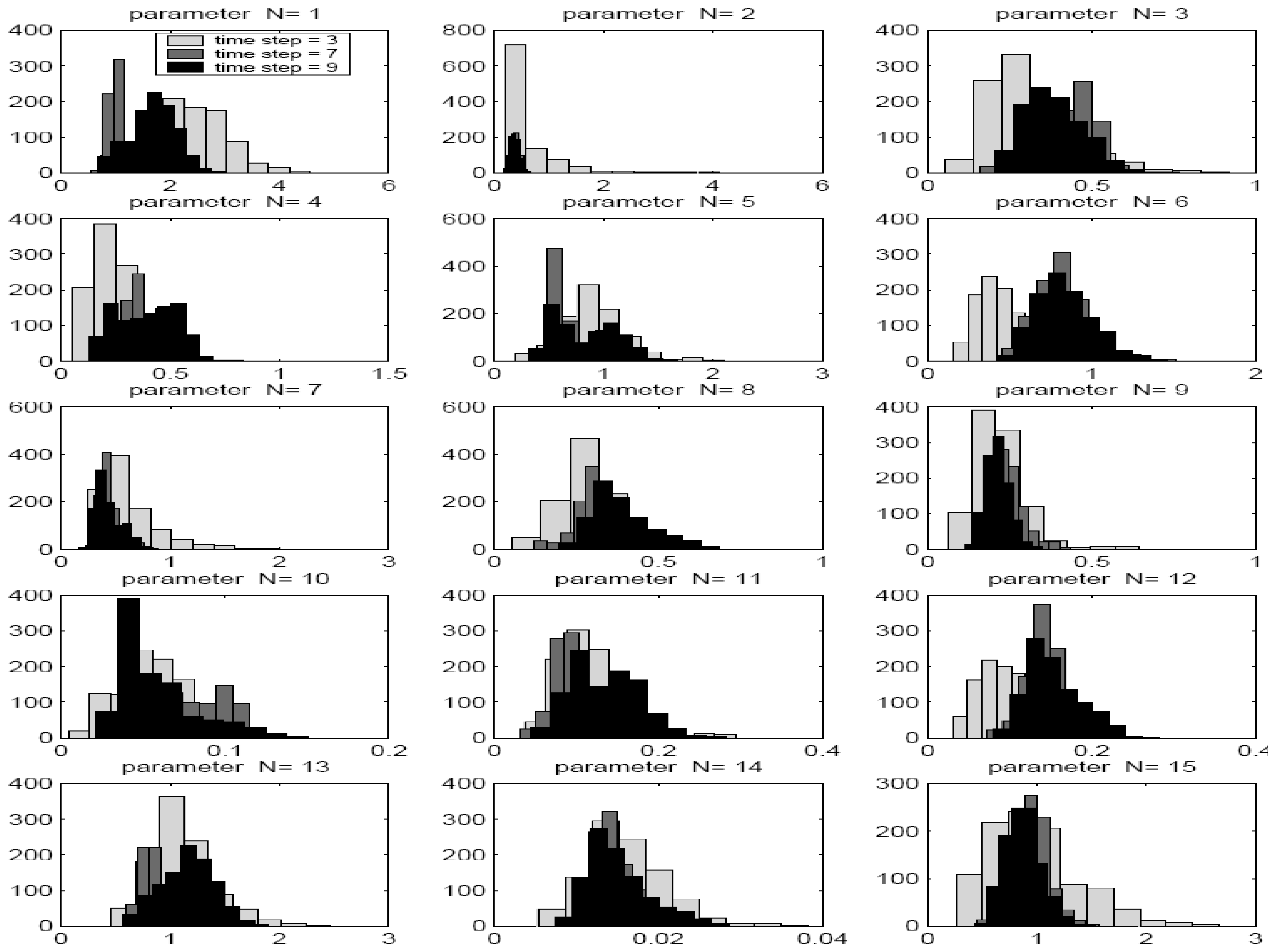


The evolution of the biological parameters at the BATS obtained by the sequential weak constraint parameter estimation



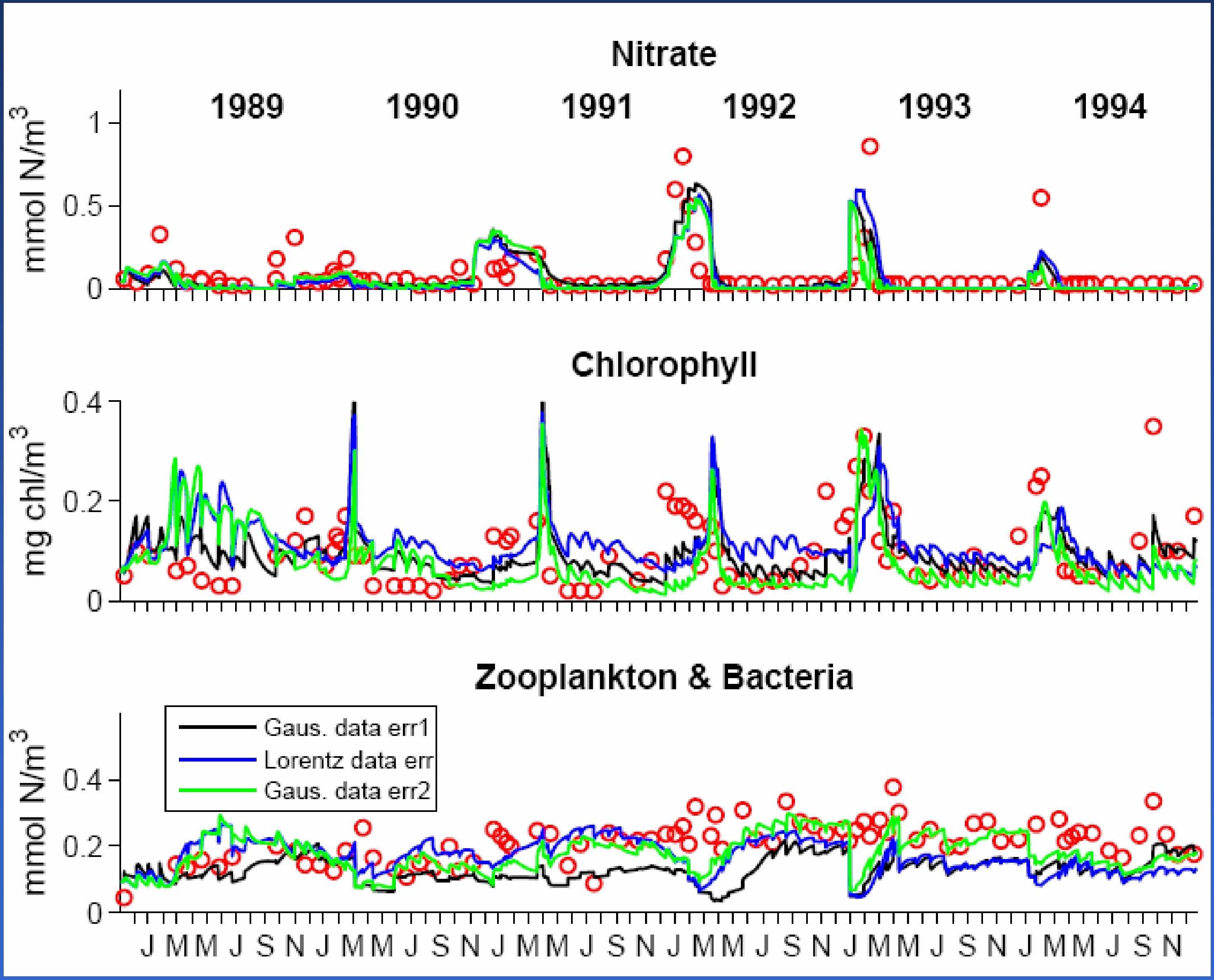


model parameters N= 1 time step=1
before the jittering
after the jittering
model parameters N= 2 time step=1
model parameters N= 3 time step=1
model parameters N= 4 time step=1
model parameters N= 5 time step=1
model parameters N= 6 time step=1
model parameters N= 7 time step=1
model parameters N= 8 time step=1
model parameters N= 9 time step=1
model parameters N= 10 time step=1
model parameters N= 11 time step=1
model parameters N= 12 time step=1

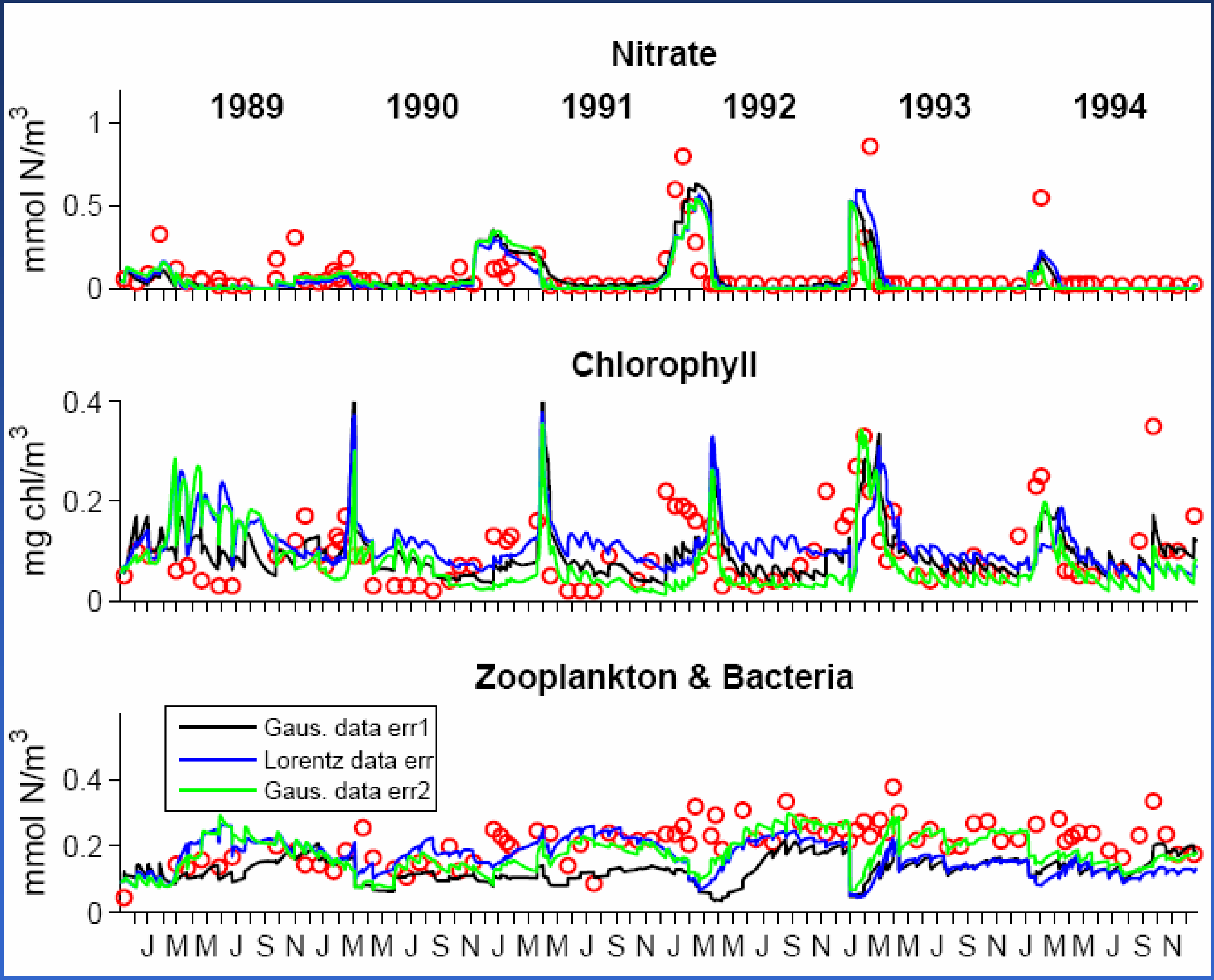


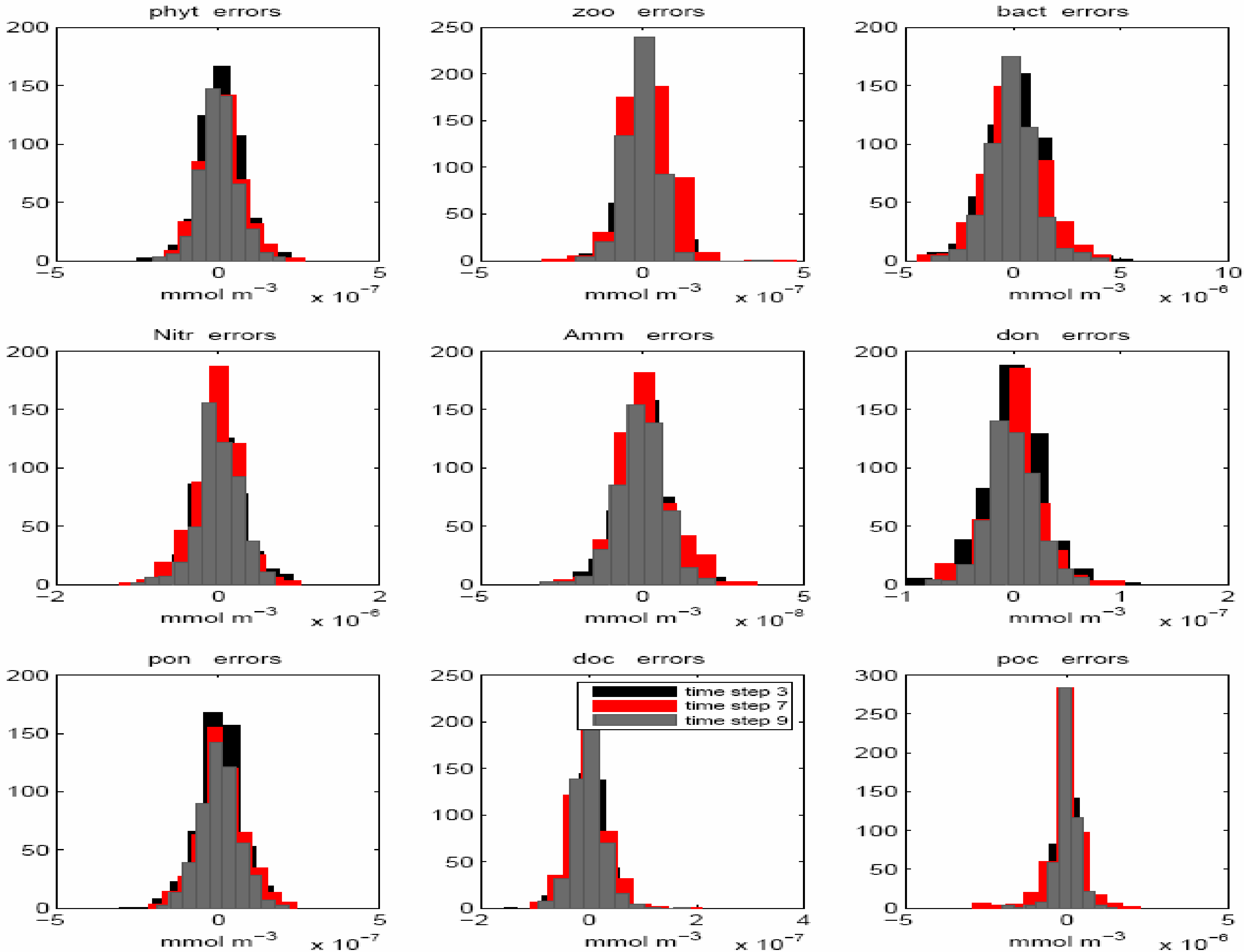
parameter N= 1
parameter N= 2
parameter N= 3
parameter N= 4
parameter N= 5
parameter N= 6
parameter N= 7
parameter N= 8
parameter N= 9
parameter N= 10
parameter N= 11
parameter N= 12
parameter N= 13
parameter N= 14
parameter N= 15
time step = 3
time step = 7
time step = 9

The evolution of the ecosystem components at the BATS obtained by the sequential weak constraint parameter estimation



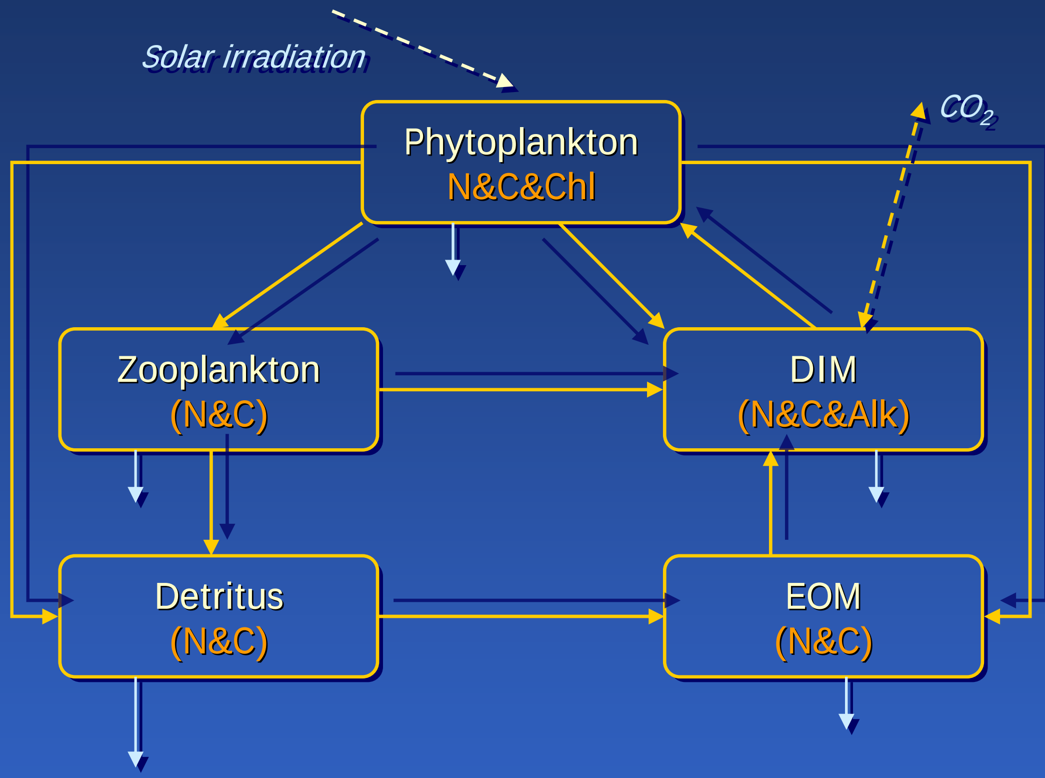
The evolution of the ecosystem components at the BATS obtained by the sequential weak constraint parameter estimation





phyt errors
zoo errors
bact errors
Nitr errors
Amm errors
don errors
pon errors
doc errors
poc errors
time step 3
time step 7
time step 9
mmol m^{-3}
x 10^{-7}
x 10^{-6}
x 10^{-8}

1D version of M. Schartau's Ecosystem Model



The authors thank Markus Schartau
for providing the model code.

Assimilated data:

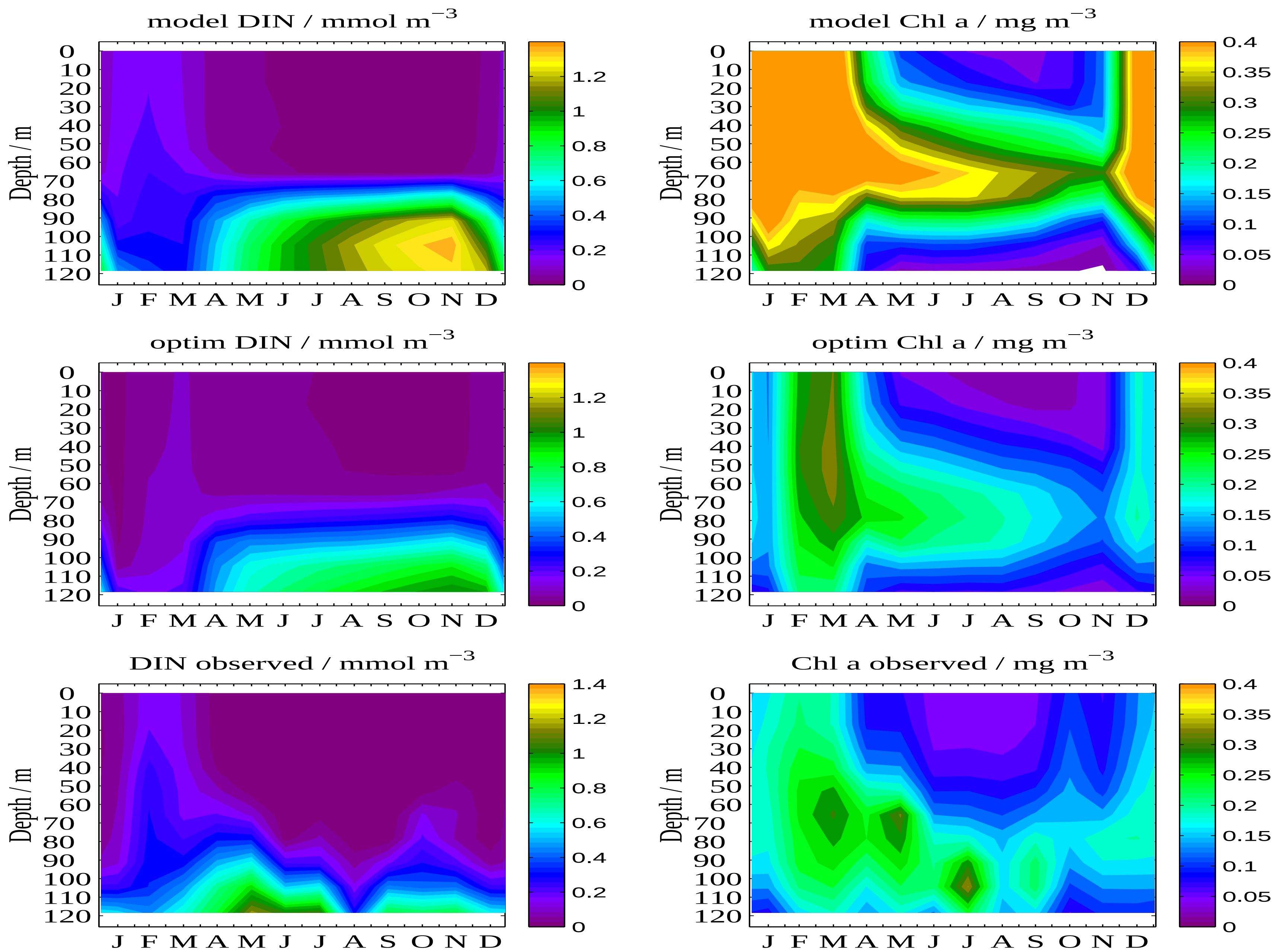
Monthly mean BATS chlorophyll and niutrient vertical profiles.

Method : SIR smoother

The flow network between 12 biogechemical components possesses ~30 biological parameters.

13 of them have been adjusted

Monthly means of chlorophyll “a” and dissolved inorganic nitrogen at BATS site (REcoM)



Few notes

- The system noise generation (with noise level optimization) has allowed us to obtained more accurate parameter estimates,

  $\Rightarrow$ to improve the forecast.
- However the model errors averaged over the considered integration sub-period have appeared to be very small (with 0 mean).

  $\Rightarrow$
- When applying a SIR smoother, one can expect a solution to be dependent on the smoothing period which biological parameters are assumed to be constant for.
- Lorentz data error statistics assumption leads to less variable (in time) parameter estimates (but produced larger forecast errors)

Outlook

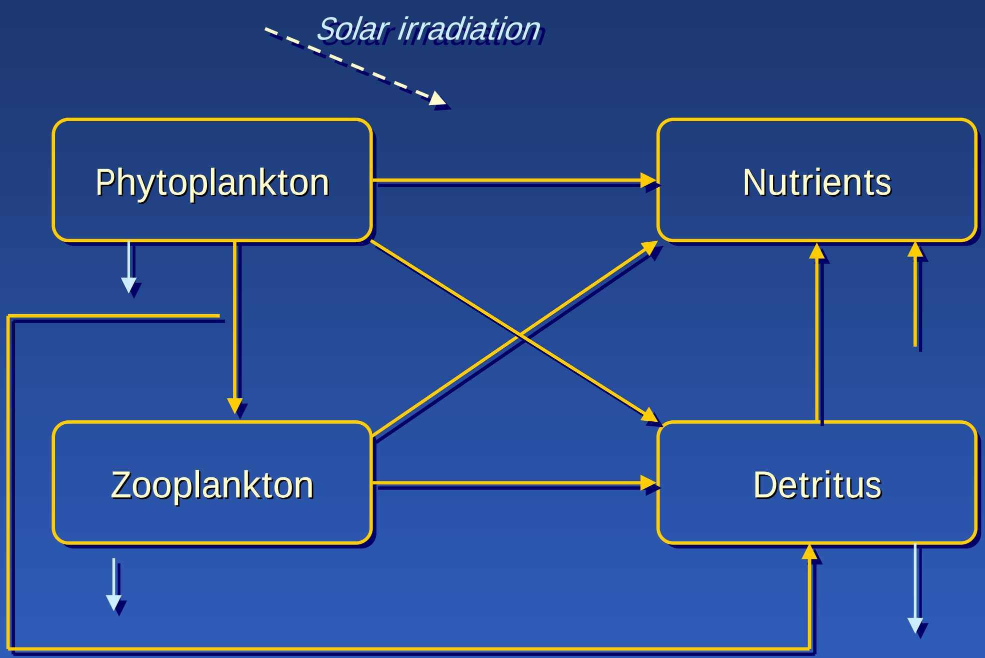
- Procedure of parameters' *posterior* probability density smoothing is still under development.

- SIRF has not been implemented yet for assimilating data into basin or large scale ecosystem models.

It will have to be a local

Popova's Ecosystem Model

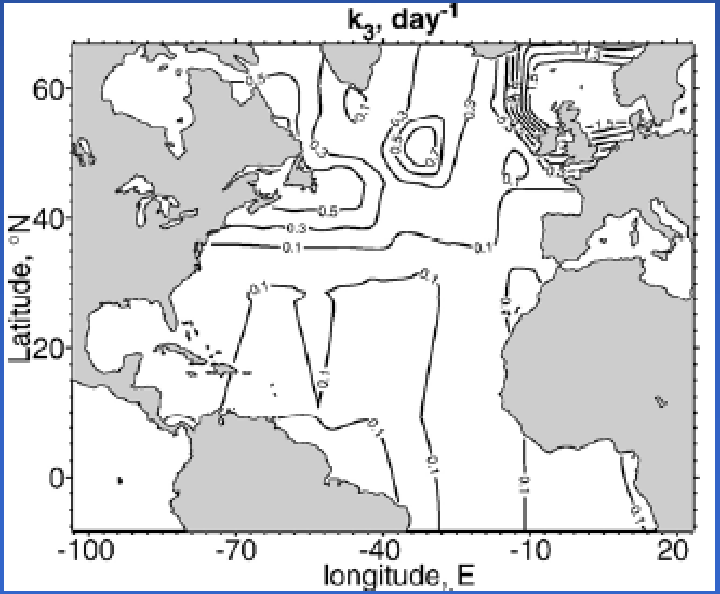
(1995)



Assimilated data:

Monthly mean satellite CZCS surface chlorophyll averaged over 1979 – 1985.

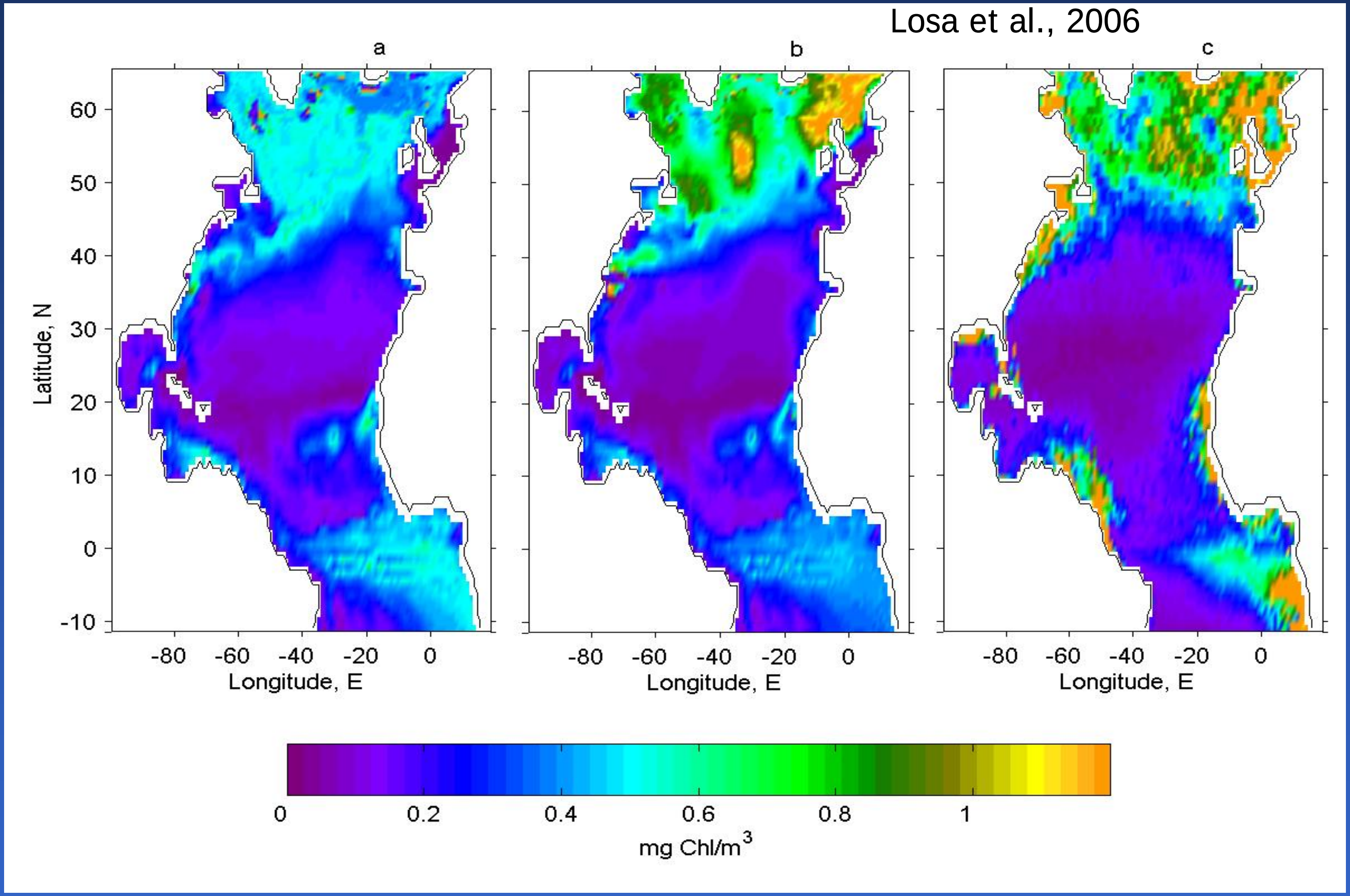
Method : a weak constraint variational technique (Losa et al, 2004)



The flow network between 4 biogechemical components possesses 19 biological parameters.

6 of them have been adjusted for each cell of 5^0x5^0 grid covering the North Atlantic

August horizontal distribution of the surface chlorophyll "a" concentration (mgChl m^{-3}) in the North Atlantic



a) the model solution obtained with constant biological parameters; b) the model solution obtained with spatially variable biological parameters (Losa et al., 2004) and c) SeaWiFS (http://seawifs.gsfc.nasa.gov/SEAWIFS.html) data averaged over 1997-2003.