Ensemble analysis and forecast using a newly developed MCMC sampler application to ocean colour satellite observations

Jean-Michel Brankart, Mikhail Popov, Emmanuel Cosme, Pierre Brasseur

IGE/MEOM
Grenoble

Context: SEAMLESS EU-H2020 project

Towards a **simplified** analysis and forecasting system based on a prior ensemble model simulation

Approach:

→ Perform a **prior ensemble simulation**, with a state-of-the-art coupled circulation/ecosystem model.

→ **Condition** this 4D ensemble **on ocean colour observations** to obtain the ensemble analysis and forecast.

Features:

→ **Decouple** the complex models **simulations and** the **inversion** problem → more flexibility in the system.

→ Focus the 4D inversion on a **specific region and time window.**

→ No need for full controllability of the complex model (with so many state variables, when so few are observed). No model restart from the analysis.

→ The complex model is not used anymore as a direct constraint in the inverse problem, but only indirectly through the prior ensemble

Inverse problem

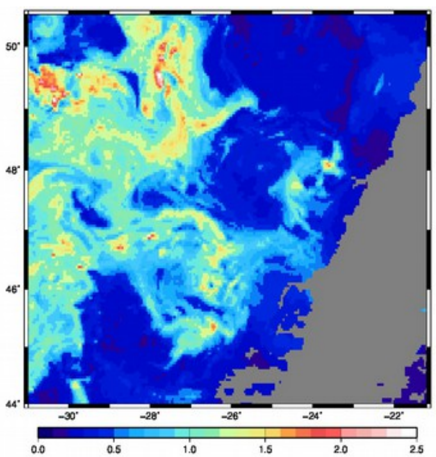
We **focus** on the small **4D subregion** (10° x 7°) at 1/4° resolution:

40 X 40 grid points (31°W-21°W, 44°N-51°N)
X 5 levels (depth: 0.5 m, 8 m, 23 m, 54 m, 108 m)
X 60 days (April 21 to June 19, 2019)
X 10 tracers (among 24 in PISCES) = **~ 5 x 10^6** variables
X 40 members = **~ 2 x 10^8** values

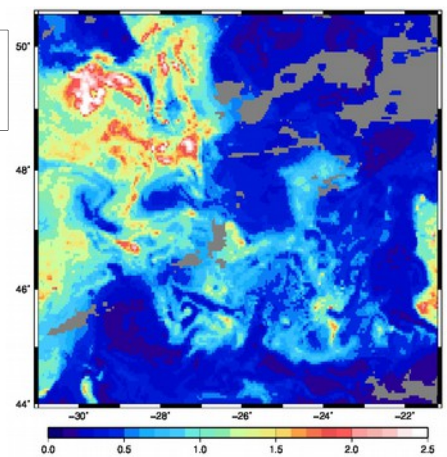
Observation system:

L3 chlorophyll product, between April 21 and June 19, 2019
Obs. error std: 30% = **~ 10^5** observations

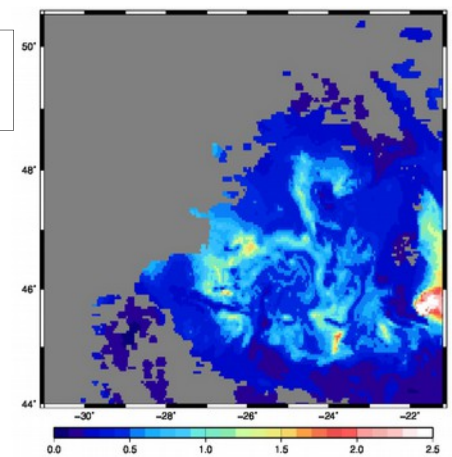
May 25



May 26



May 27



The prior ensemble simulation

Performed in the context of the SEAMLESS EU-H2020 project.

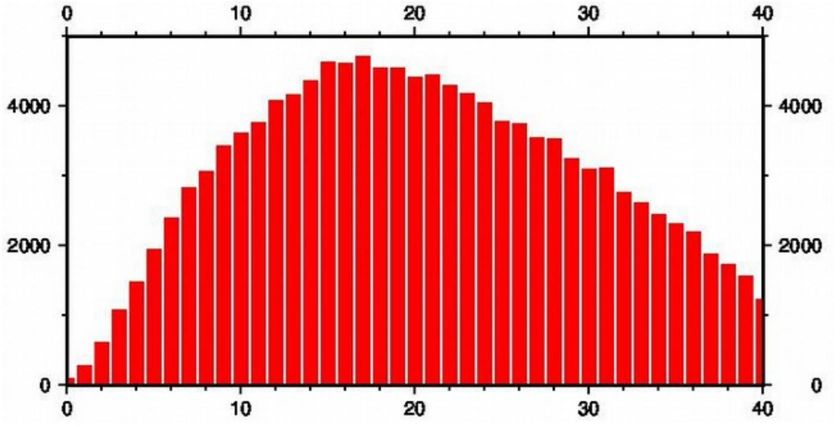
Using a global configuration of NEMO/PISCES at 1/4° resolution.

→ 40 ensemble members.
→ Outputs every 5 days for the full model state.
→ Daily outputs for specific regions.

Probabilistic scores have been applied to evaluate this ensemble simulation using L3 ocean colour observations.

→ Example of rank histogram for the subregion used as an example below.

→ In the North Atlantic Drift: 31°W-21°W, 44°N-51°N, April 21 to June 19, 2019.



Method : an MCMC sampler based on the Metropolis/Hastings algorithm

Sample the posterior pdf
for the evolution of the system
$\mathbf{x}$ ($n \sim 5 \times 10^6$),
given observations $\mathbf{y}^o$ ($p \sim 10^5$)

$$p(\mathbf{x}|\mathbf{y}^o) \sim p^b(\mathbf{x}) \; p(\mathbf{y}^o|\mathbf{x})$$

Prior ensemble (m=40 members)

Observation constraint

Anamorphosis transformation

$\mathbf{x'} = A(\mathbf{x})$, $\mathbf{x} = A^{-1}(\mathbf{x'})$
to obtain marginally Gaussian $\mathbf{x'}$
(with mean=0 and variance=1):

$$p(\mathbf{x'}|\mathbf{y}^o) \sim p^b(\mathbf{x'}) \; p[\, \mathbf{y}^o \,|\, A^{-1}(\mathbf{x'}) \,]$$

We use local correlations only

Kept fully general

Iterative method in 2 steps:

1. **Propose** pseudo-random perturbation of **x'** (with cost linear in n)
→ by modulation of an ensemble member with large-scale signals
($\sim 10^{11}$ pseudo-random directions of perturbations)
→ equivalent to a localization of the prior ensemble covariance

2. **Accept/reject** according to cost function: $J^o = - \log p[\, \mathbf{y}^o \,|\, A^{-1}(\mathbf{x'}) \,]$

Brankart J;-M., 2019 : Implicitly Localized MCMC Sampler to Cope With Non-local/Non-linear Data Constraints in Large-Size Inverse Problems. Front. Appl. Math. Stat. 5:58.

Method : an MCMC sampler based on the Metropolis/Hastings algorithm

Sample the posterior pdf
for the evolution of the system
$\mathbf{x}$ ($n \sim 5 \times 10^6$),
given observations $\mathbf{y}^o$ ($p \sim 10^5$)

$$p(\mathbf{x}|\mathbf{y}^o) \sim p^b(\mathbf{x}) \; p(\mathbf{y}^o|\mathbf{x})$$

Prior ensemble (m=40 members)

Observation constraint

Anamorphosis transformation
$\mathbf{x}' = A(\mathbf{x}), \quad \mathbf{x} = A^{-1}(\mathbf{x}')$
to obtain marginally Gaussian $\mathbf{x}'$
(with mean=0 and variance=1):

$$p(\mathbf{x}'|\mathbf{y}^o) \sim p^b(\mathbf{x}') \; p[\, \mathbf{y}^o \,|\, A^{-1}(\mathbf{x}') \,]$$

We use local correlations only

Kept fully general

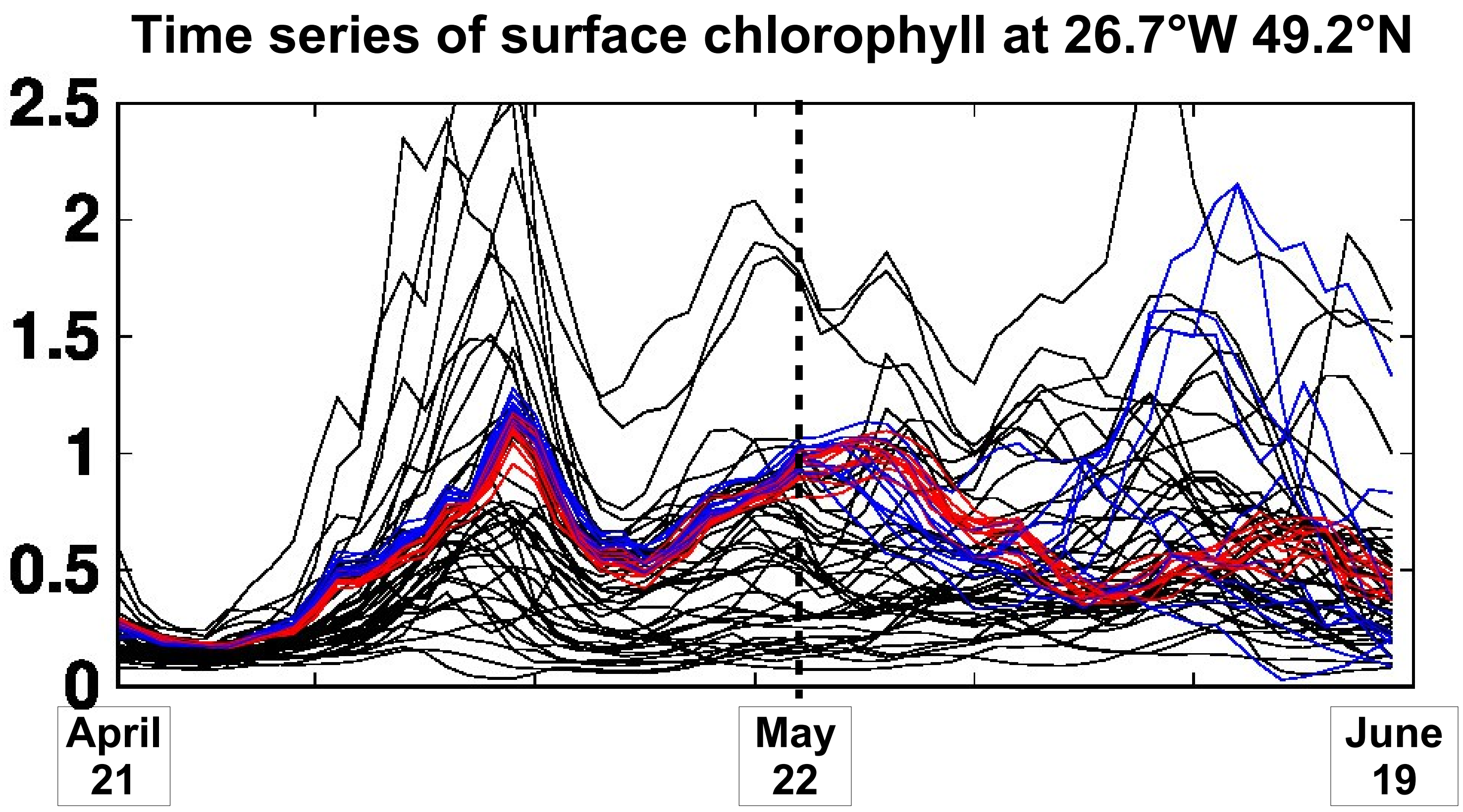
Iterative method in 2 steps:

1. **Propose** pseudo-random perturbation of **x'** (with cost linear in n)
→ by modulation of an ensemble member with large-scale signals
($\sim 10^{11}$ pseudo-random directions of perturbations)
→ equivalent to a localization of the prior ensemble covariance

2. **Accept/reject** according to cost function: $J^o = -\log p[\, \mathbf{y}^o \,|\, A^{-1}(\mathbf{x}') \,]$

Brankart J;-M., 2019 : Implicitly Localized MCMC Sampler to Cope With Non-local/Non-linear Data Constraints in Large-Size Inverse Problems. Front. Appl. Math. Stat. 5:58.

Results from the MCMC sampler: ensemble analysis and forecast



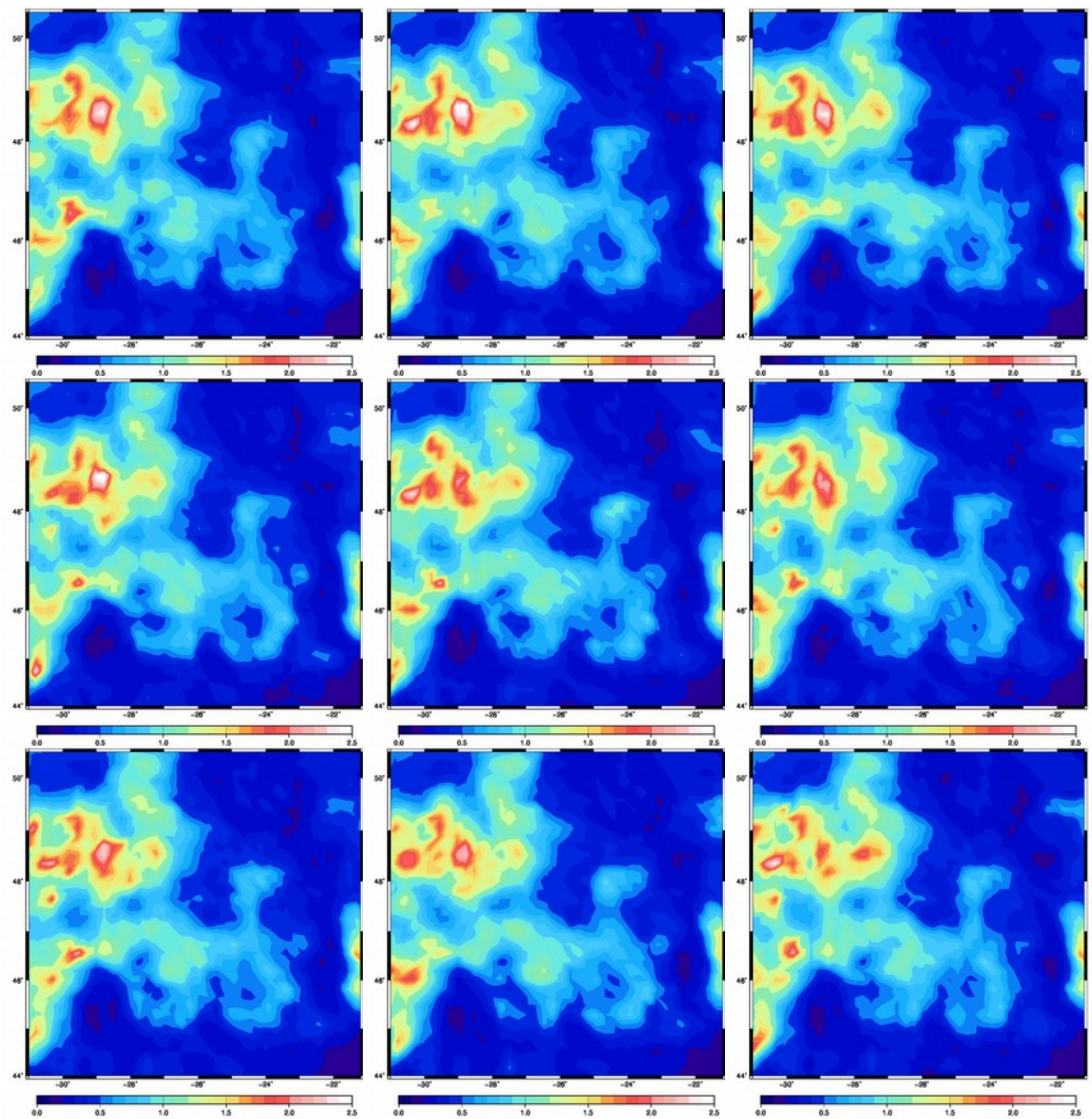
In black: prior ensemble simulations from NEMO/PISCES

In red: ensemble analysis using all L3 observations

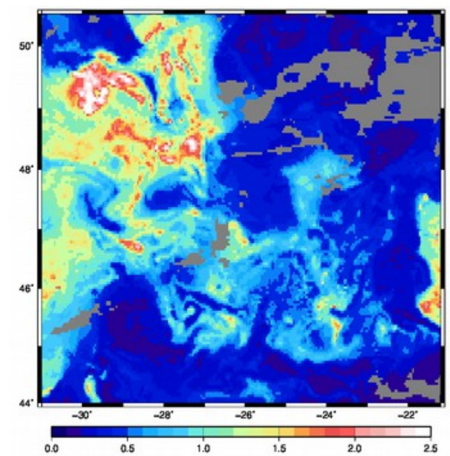
In blue: ensemble analysis and forecast using L3 observations until May 22 → some forecast skill for about 10 days

[Localization scales: ~0.8° on the horizontal and ~10 days in time]

Results from the MCMC sampler: **ensemble analysis for May 26, 2019** (using past and future observations)



L3 chlorophyll product

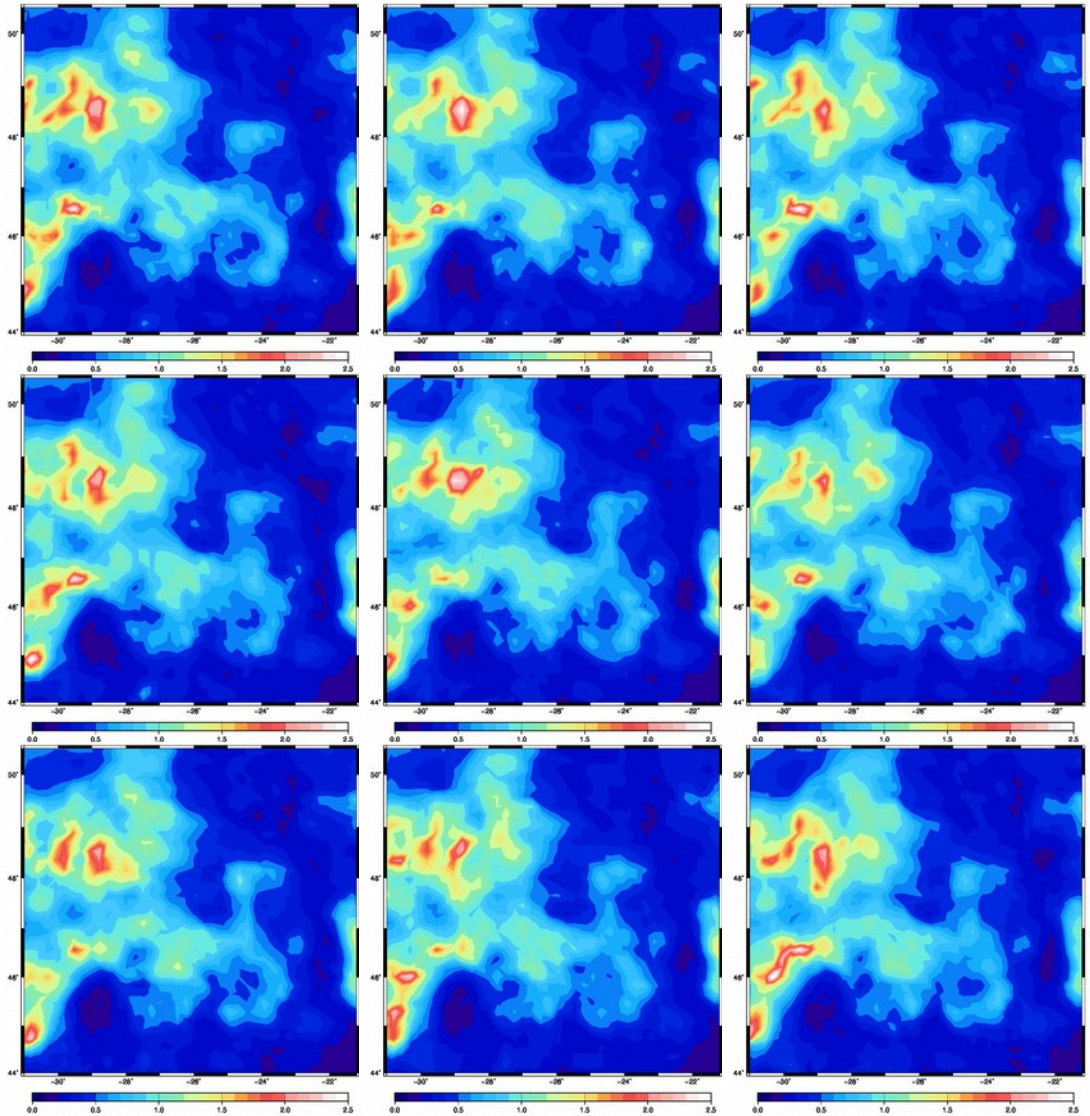


Good fit to observations
(within obs. error bar)

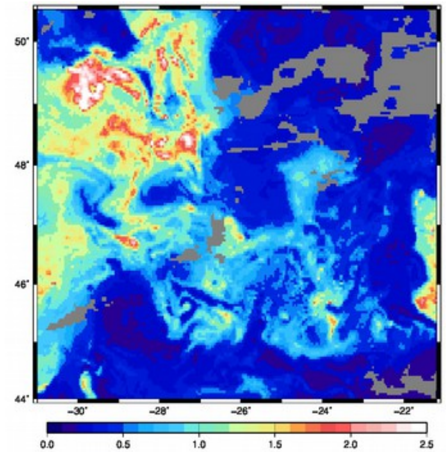
CRPS resolution:
0.121 mg/m3

Optimality score: 1.02

Results from the MCMC sampler:
ensemble analysis for May 26, 2019
(leaving out observation of May 26, 2019)



L3 chlorophyll product



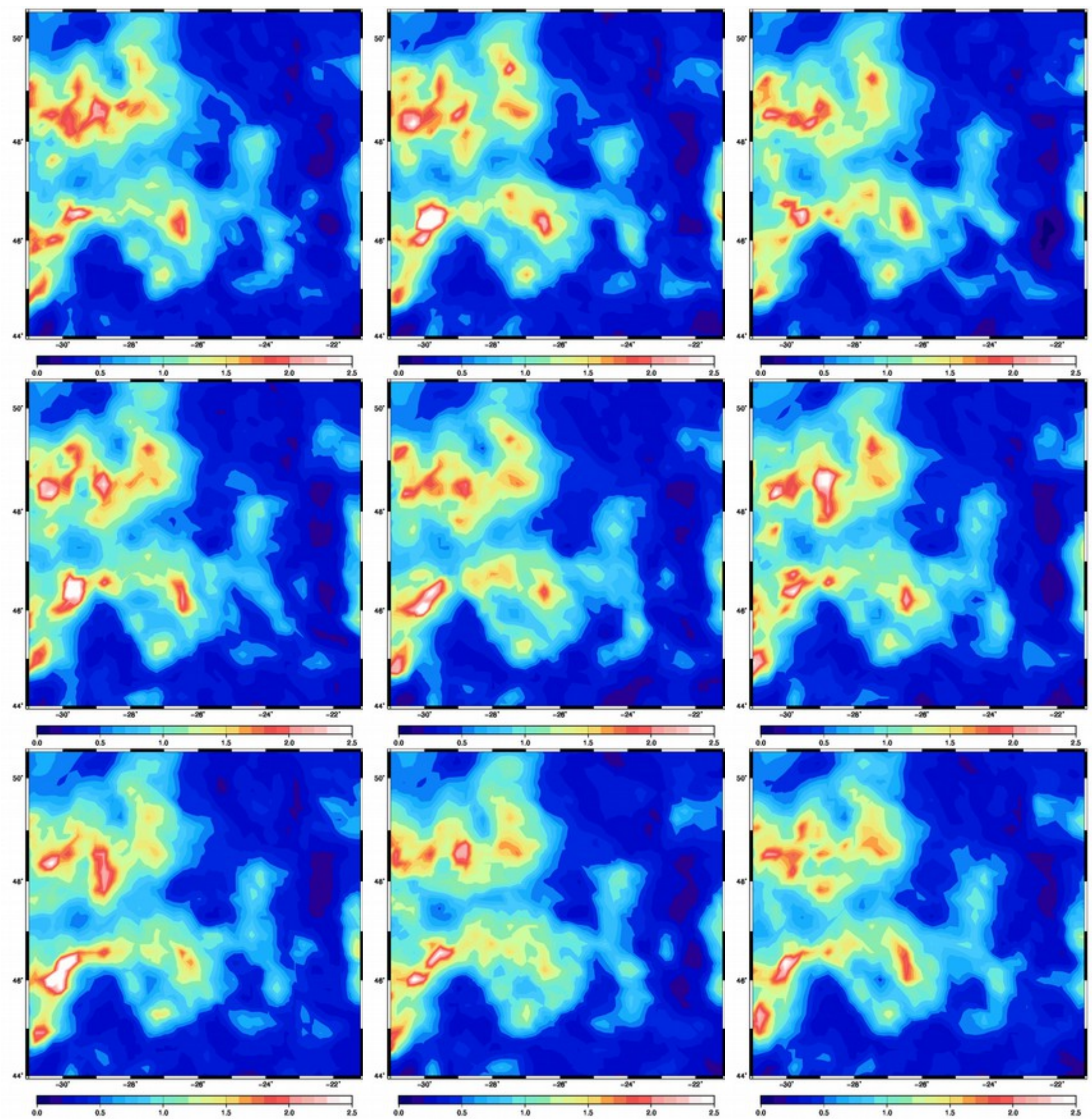
Obs. is independent.

CRPS reliability:
0.0047 mg/m3

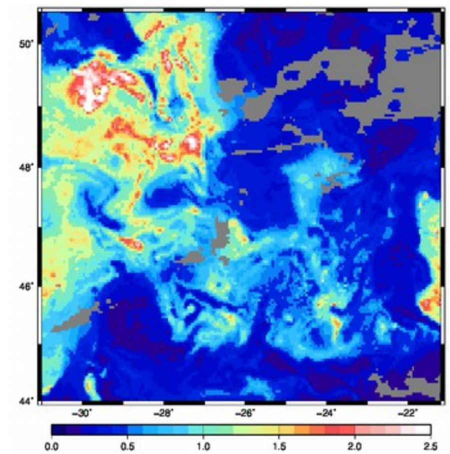
CRPS resolution:
0.132 mg/m3

Optimality score: 1.01

Results from the MCMC sampler:
1-day ensemble forecast for May 26, 2019
(i.e. using past observations only)



L3 chlorophyll product



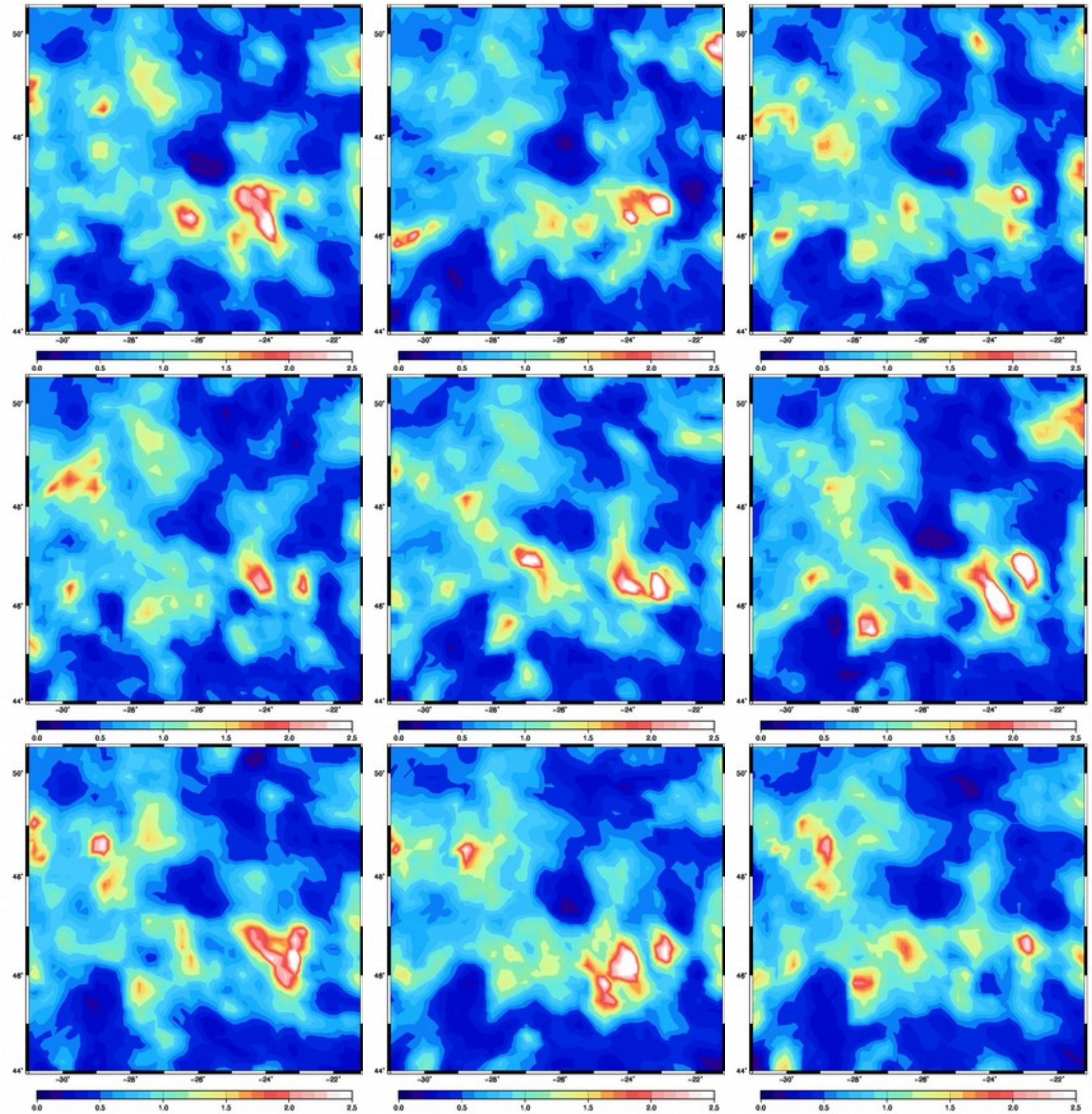
More uncertain...

CRPS reliability: 0.0089 mg/m3

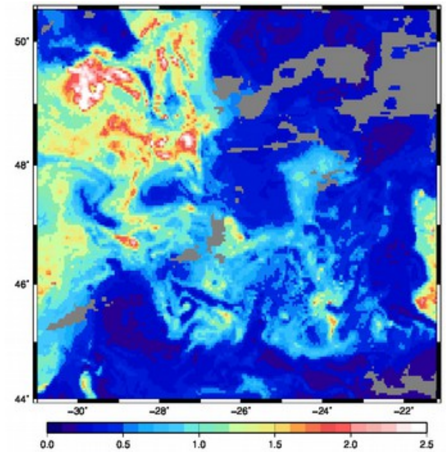
CRPS resolution: 0.145 mg/m3

Optimality score: 0.99

Results from the MCMC sampler:
4-day ensemble forecast for May 26, 2019
(i.e. using observations until May 22, 2019)



L3 chlorophyll product



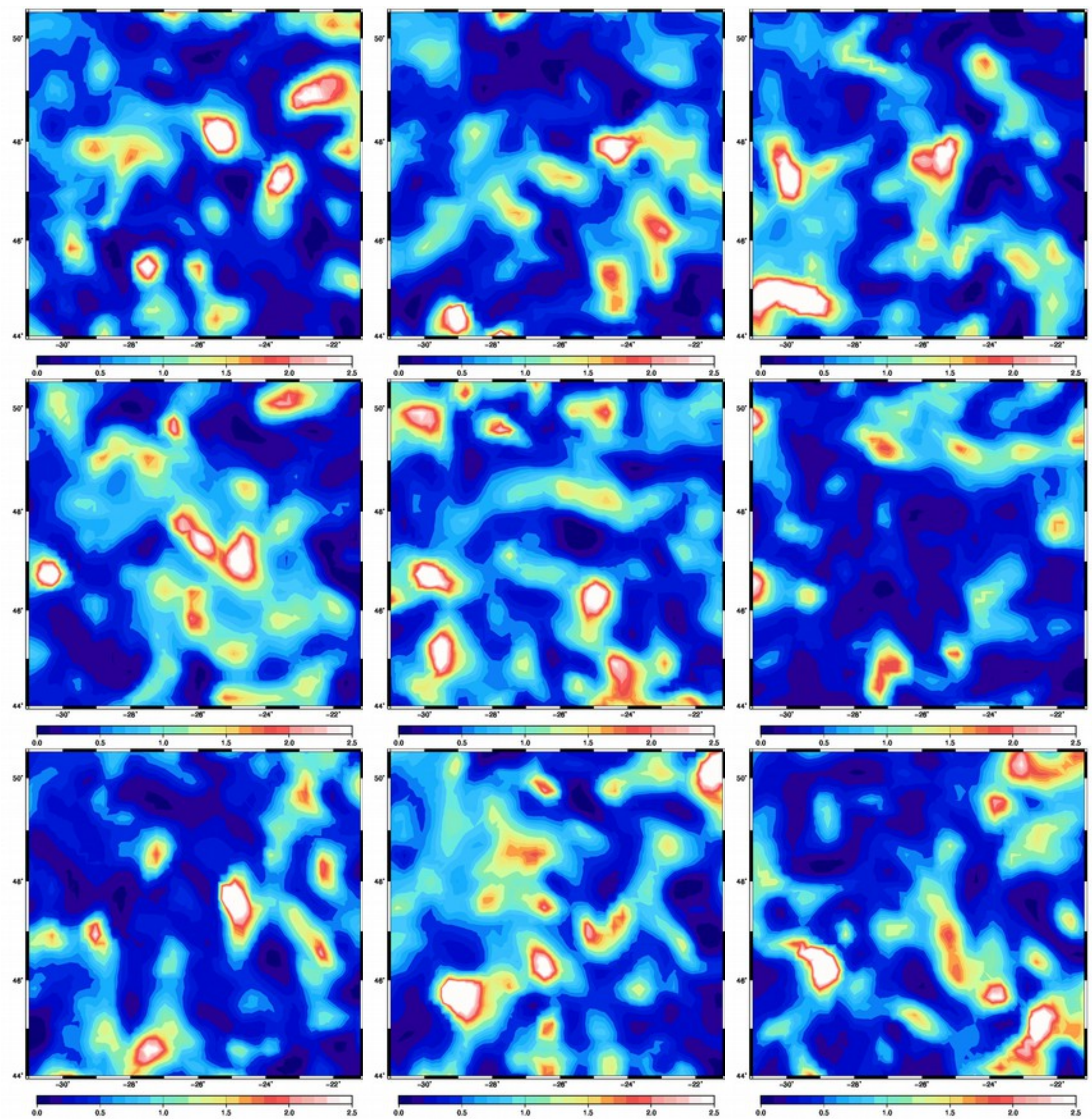
Even more uncertain...

CRPS reliability:
0.020 mg/m3

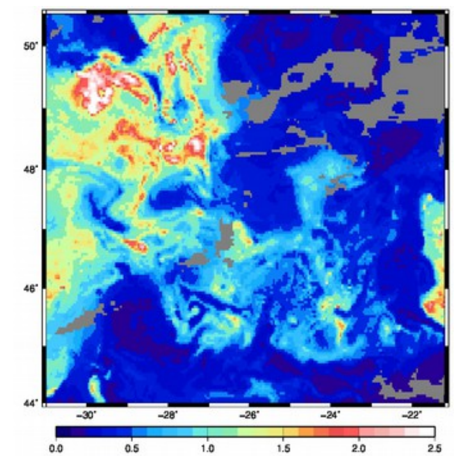
CRPS resolution:
0.191 mg/m3

Optimality score: 0.98

Results from the MCMC sampler:
prior ensemble for May 26, 2019
(i.e. from the model only, without observations)



L3 chlorophyll product



Very uncertain,
but quite reliable

CRPS reliability:
0.0077 mg/m3

CRPS resolution:
0.261 mg/m3

Conclusions

A practical method to perform **4D ensemble analyses and forecasts**

→ based on prior ensemble statistics
(marginal pdfs and local rank correlations)

→ coping with fully general observation constraint p($\mathbf{y}^o|\mathbf{x}$)
(nonlinear, non-Gaussian, nonlocal)

The focus is on **sampling possibilities** consistent with the observations.

<u>To explore in SEAMLESS:</u>

→ Results for **non-observed variables** (surface and subsurface)
→ Results in different regions/seasons
→ How to produce a yearly solution

Perspectives (more or less remote)

Reconstruction of surface circulation:

→ from altimetric observations

→ possibly introducing a weak dynamical constraint in the cost function:

$$J^c = \frac{1}{\sigma_\psi^2} \int_\Omega \left[\frac{D(\Delta\psi + f)}{Dt} \right]^2 d\Omega$$

Conservation of potential vorticity

Joint reconstruction of circulation and tracers:

→ from joint observation system, and

→ possibly introducing a joint dynamical constraint:

$$J^c = \frac{1}{\sigma_\psi^2} \int_\Omega \left[\frac{D(\Delta\psi + f)}{Dt} \right]^2 d\Omega + \frac{1}{\sigma_c^2} \int_\Omega \left[\frac{DC}{Dt} \right]^2 d\Omega$$

Including a tracer with an unknown parameter in the dynamics

$$J^c = \frac{1}{\sigma_\psi^2} \int_\Omega \left[\frac{D(\Delta\psi + f)}{Dt} \right]^2 d\Omega + \frac{1}{\sigma_c^2} \int_\Omega \left[\frac{DC}{Dt} - \mu C \right]^2 d\Omega + \frac{1}{\sigma_\mu^2} \int_\Omega \left[\frac{D\mu}{Dt} \right]^2 d\Omega$$

→ most simple joint physical/biogeochemical data assimilation
with joint state and parameter estimation