Implicitly localized MCMC sampler to cope with nonlocal/nonlinear data constraints in large-size inverse problems

Jean-Michel Brankart, Sammy Metref, P. Brasseur, E. Cosme

Institut des Géosciences de l'Environnement
CNRS, UGA, IRD

This presentation illustrates the method described in the paper:

https://www.frontiersin.org/articles/10.3389/fams.2019.00058

All codes necessary to reproduce the results are openly available from:

https://github.com/brankart/ensdam

Motivations for these developments

Solve inverse problems within the Bayesian framework,

Using an MCMC sampler
to have an explicit description of posterior uncertainties
going beyond the Gaussian assumption,

Coping with nonlinear/nonlocal data constraints,
for instance dynamical or observation constraints,

With good numerical efficiency,
to stay applicable to large size problems.

Approach:
Design an efficient proposal distribution,
which can be sampled at a very low cost,
by a multiple Schur product from a multiscale prior ensemble

Potential application to altimetry

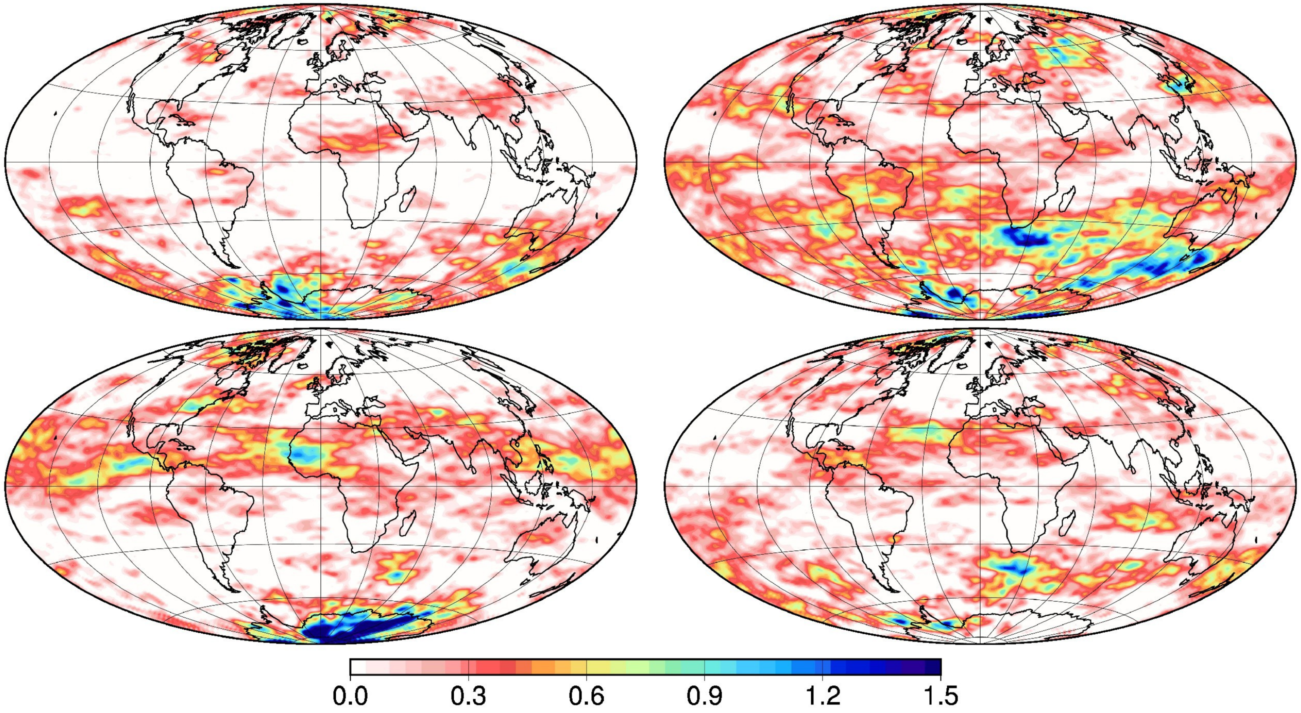
In the long term, the method should provide a model-free alternative to data assimilation.

The freedom in the observational constraint should make it easy to combine altimetry with complex data such as optical images to retrieve surface currents, or ice characteristics to estimate ice extent.

Application example

A positive 2D field on the sphere
with finite probability (~25%) to be equal to zero

Prior probability distribution known through an ensemble of size 100:

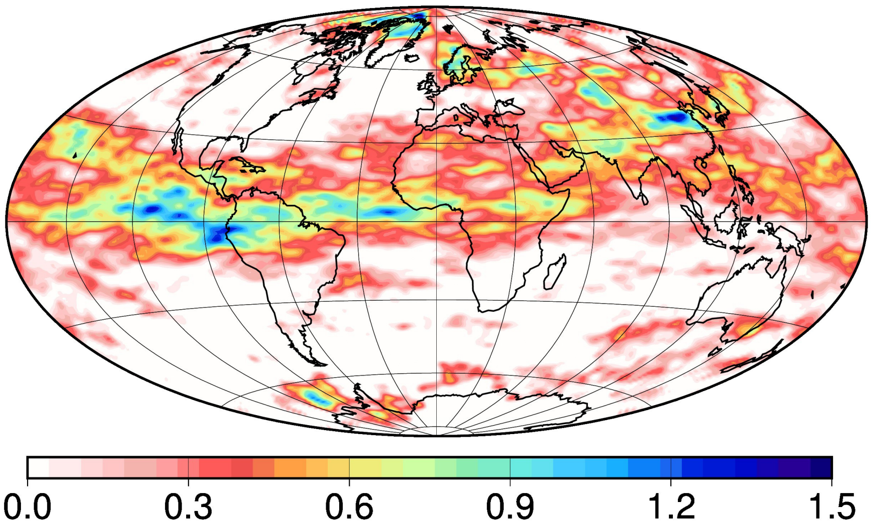


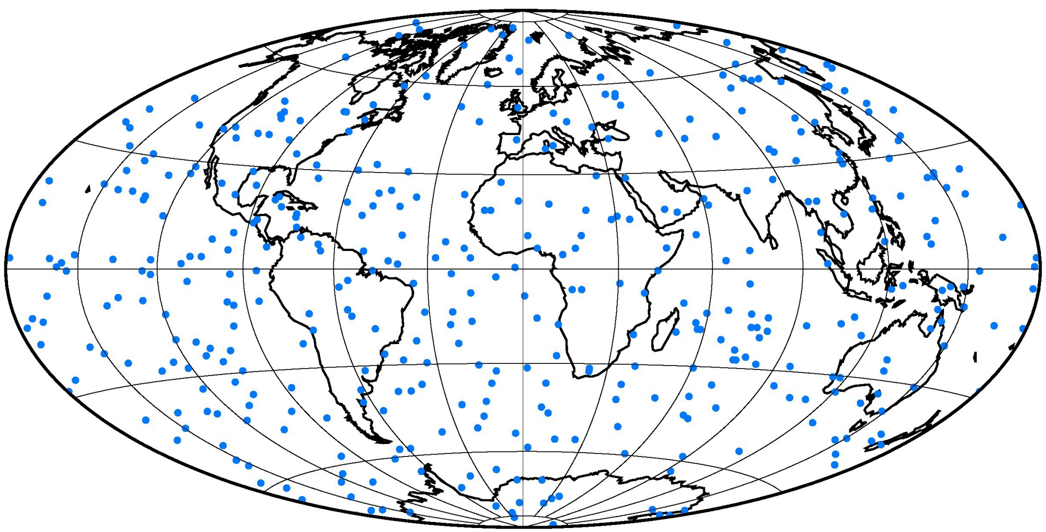
This can be for instance: precipitation, ice thickness, chlorophyll,...
This can be generalized to multivariate problems with more dimensions.

True state and observations

True state
Independent draw
from the same distribution
as the prior ensemble

Used to simulate the
synthetic observations and
to check the solution





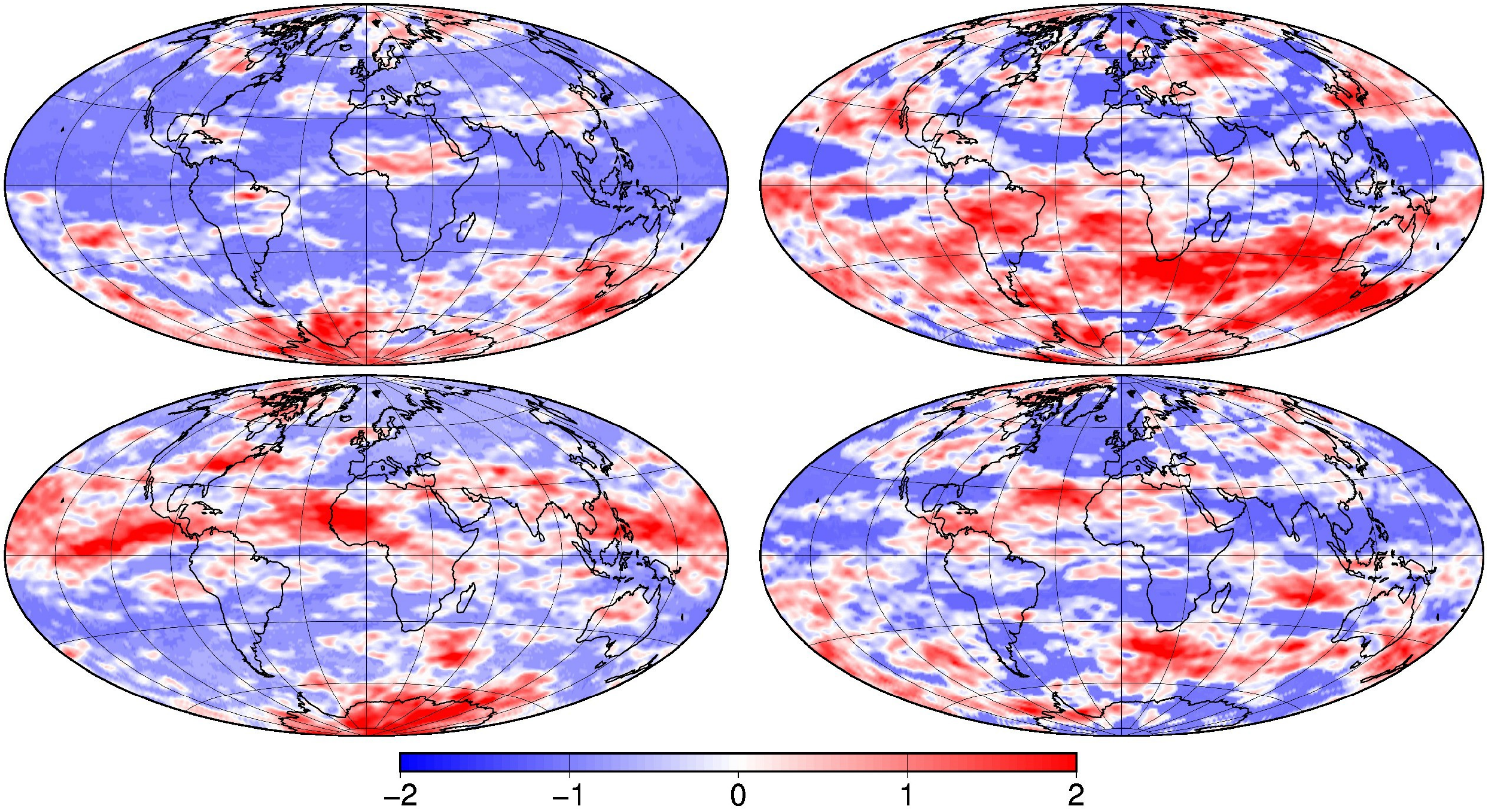
Observations
local (blue)dots
and non-local:

position of the maximum

**fraction of the sphere
where the field
is equal to zero**

Anamorphosis

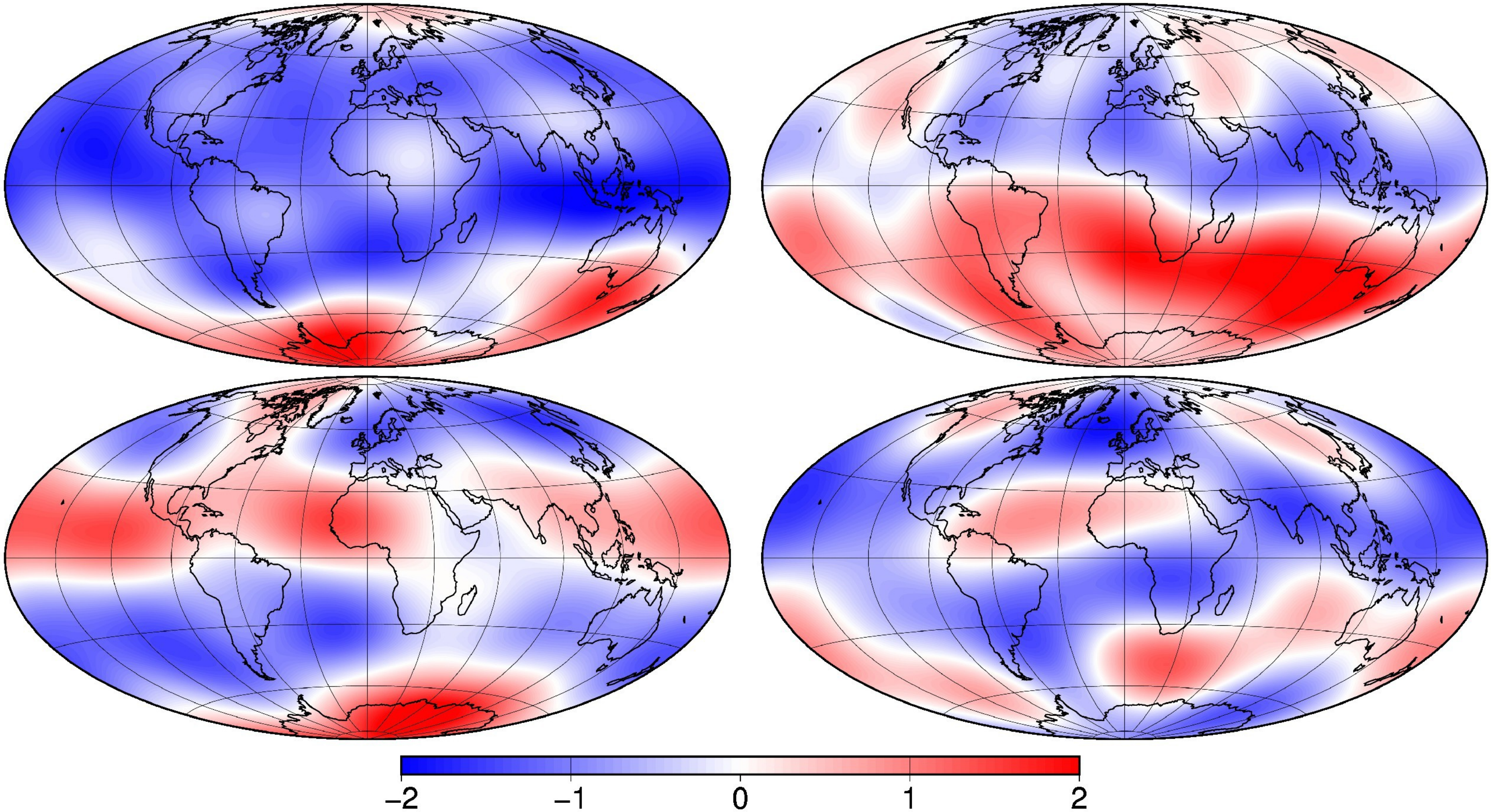
Nonlinear transformations to have Gaussian marginal distributions



A stochastic transformation is used where the field is equal to zero to cope with the concentration of probability

Scale separation

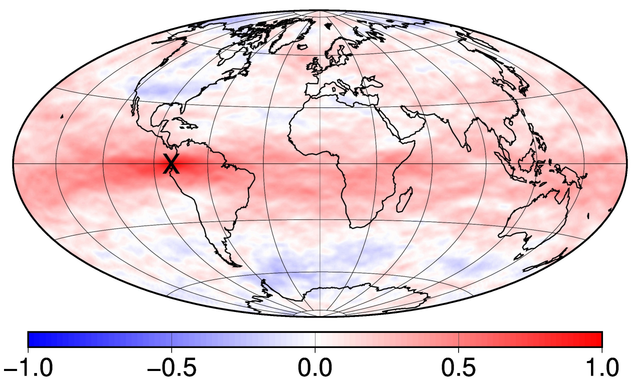
A multiscale ensemble is produced by extracting the large-scale component of each ensemble member



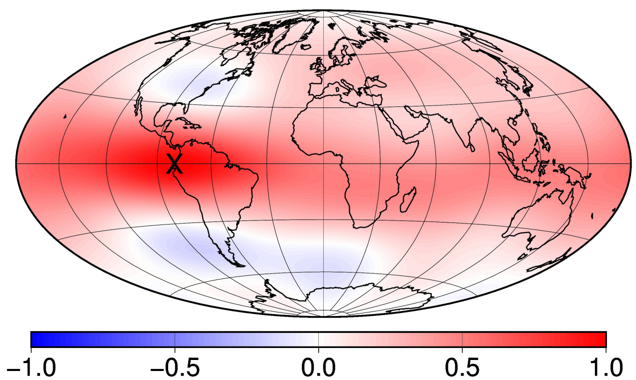
Localization

The ensemble covariance is localized by considering Schur products of one of the ensemble member with the large-scale component of p other members (here, p=4)

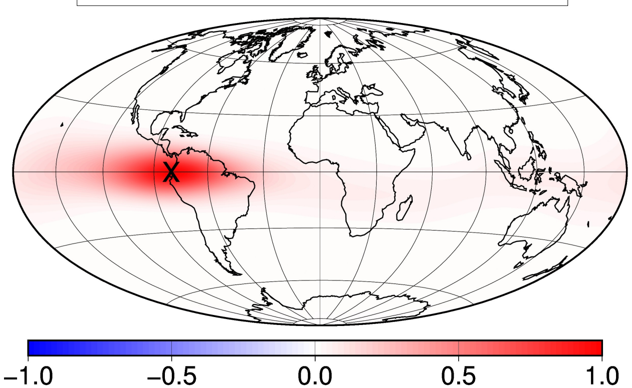
Ensemble covariance **C**



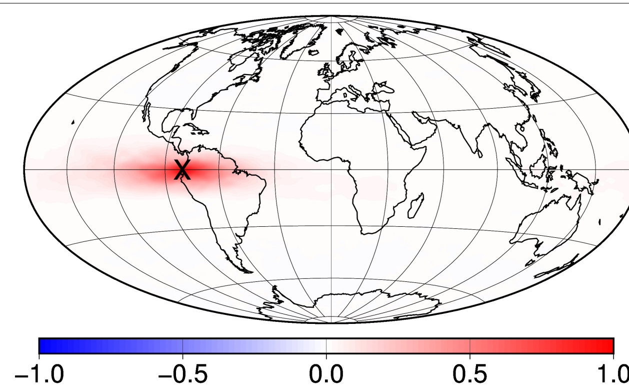
Larg-scale covariance $\mathbf{C}^1$



$\mathbf{C}^1 \circ \mathbf{C}^1 \circ \mathbf{C}^1 \circ \mathbf{C}^1$

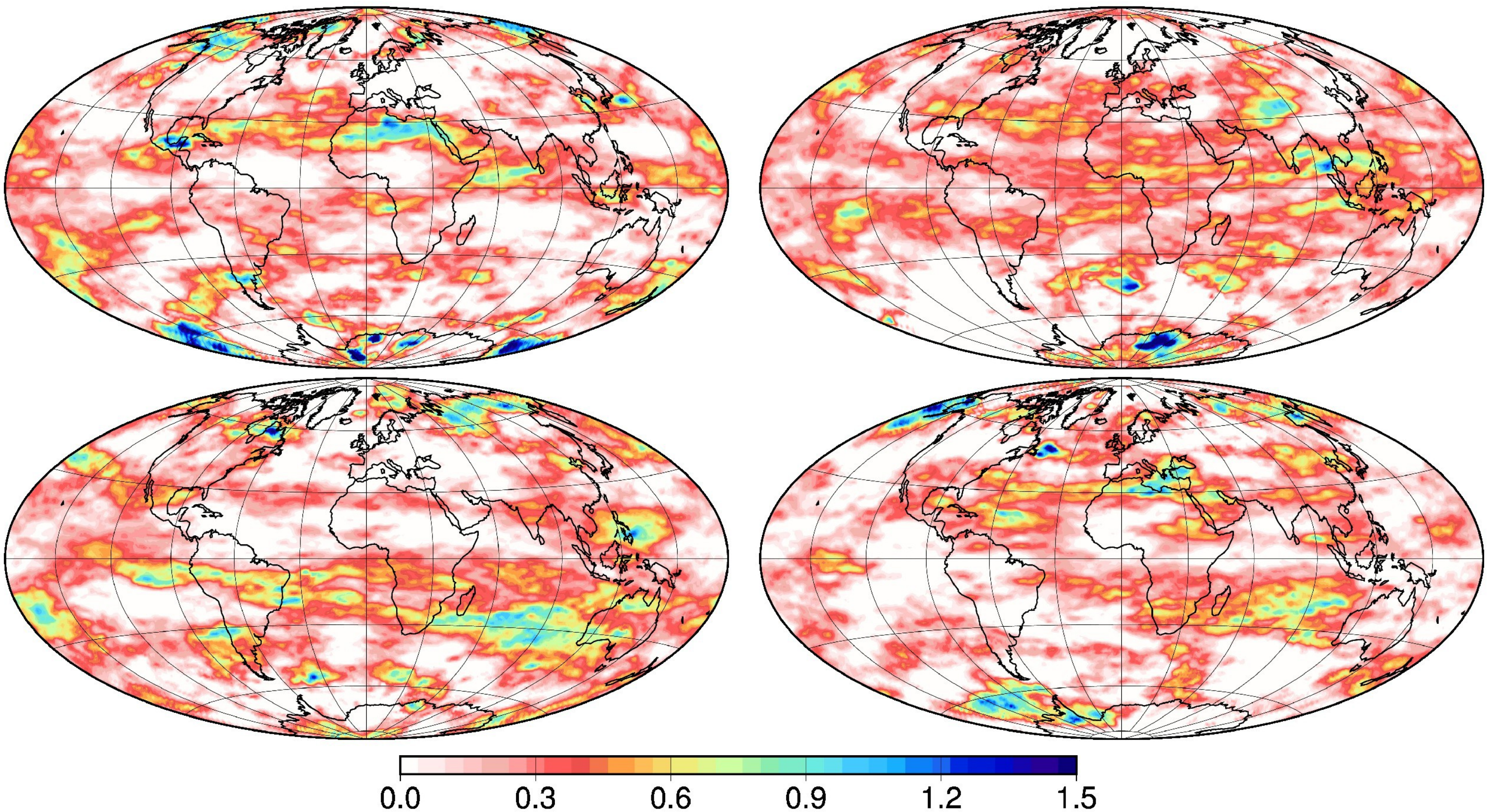


covariance after localization



Ensemble augmentation

New ensemble members with the same local covariance structure as the prior ensemble can then be generated by randomly combining the Schur products using Markov chains

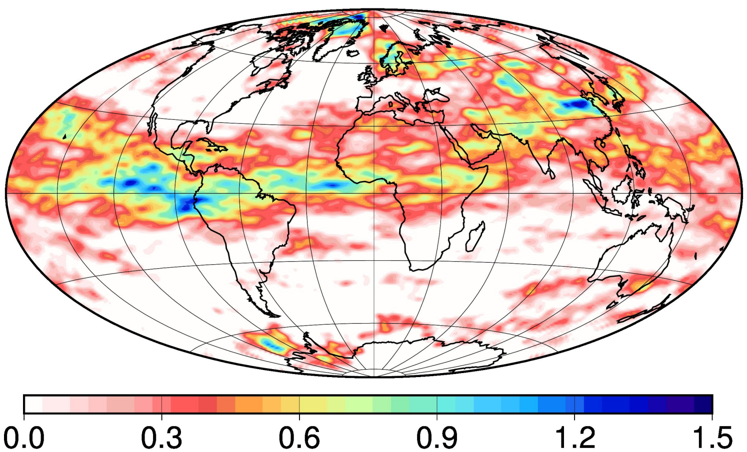


$$\mathbf{x}_i^{(0)} = 0\,; \quad \mathbf{x}_i^{(K+1)} = \alpha_K\, \mathbf{x}_i^{(K)} + \beta_K\, w_i^{(K)}\, \tilde{\mathbf{x}}_{\boldsymbol{\pi}_i(K)}$$

Conditioning the ensemble to observations

Conditions to observations can then be applied by including an acceptance probability (in the same Markov chains) decreasing with the distance to observations (cost function)

True state



Markov chain

$$\mathbf{x}_i^{(0)} = 0\,; \quad \mathbf{x}_i^{(K+1)} = \alpha_K\, \mathbf{x}_i^{(K)} + \beta_K\, w_i^{(K)}\, \tilde{\mathbf{x}}_{\boldsymbol{\pi}_i(K)}$$

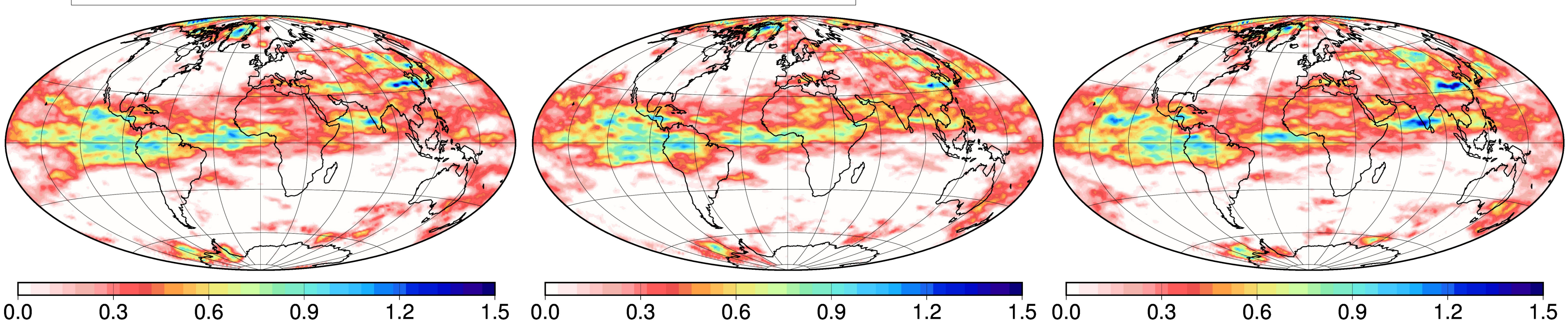
Acceptance probability

$$\theta^a = \min\left[\exp(\delta J^o), 1\right]$$

with

$$\delta J^o = J^o(\mathbf{x}^{(K+1)}) - J^o(\mathbf{x}^{(K)})$$

Posterior ensemble

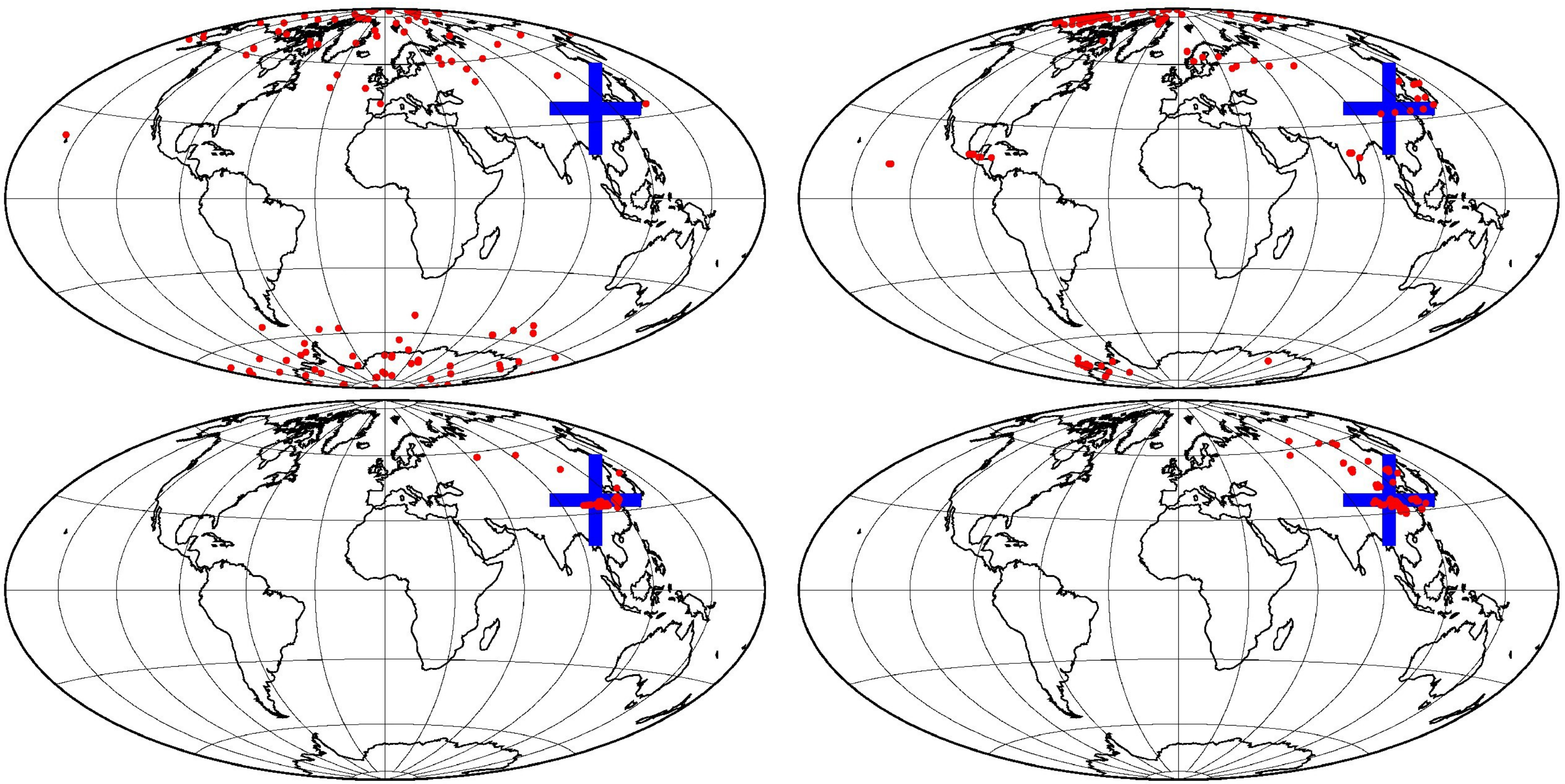


Nonlocal and nonlinear observations

Nonlocal and nonlinear constraints can be included, as illustrated here for the position of the maximum (blue cross for the true state)

Prior ensemble

Conditioned
to local observations only



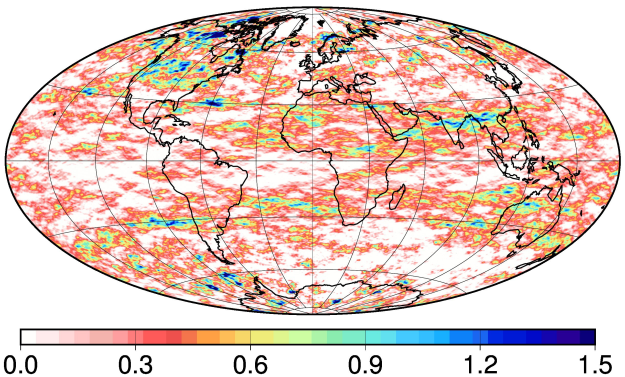
Conditioned to local observations
and to the position of the maximum

Conditioned to the position
of the maximum only

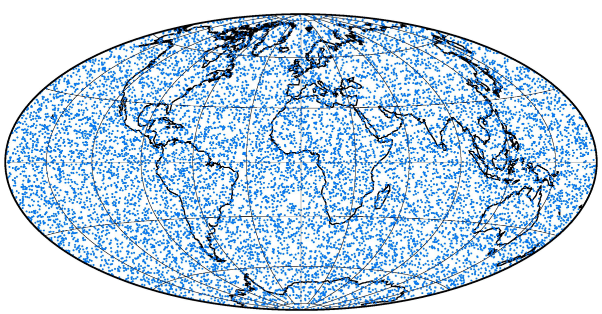
Scalability

The algorithm is directly parallelizable and the cost is linear w.r.t. the size of the problem

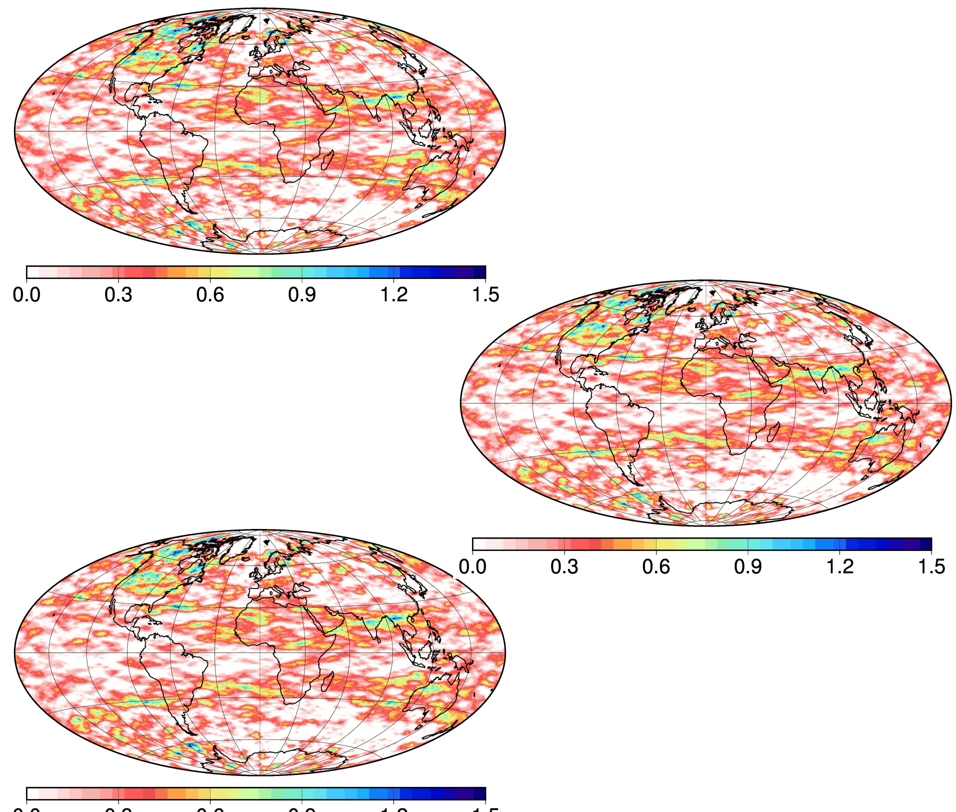
Cost = Number of iterations x Size of the problem x Size of the ensemble



True state

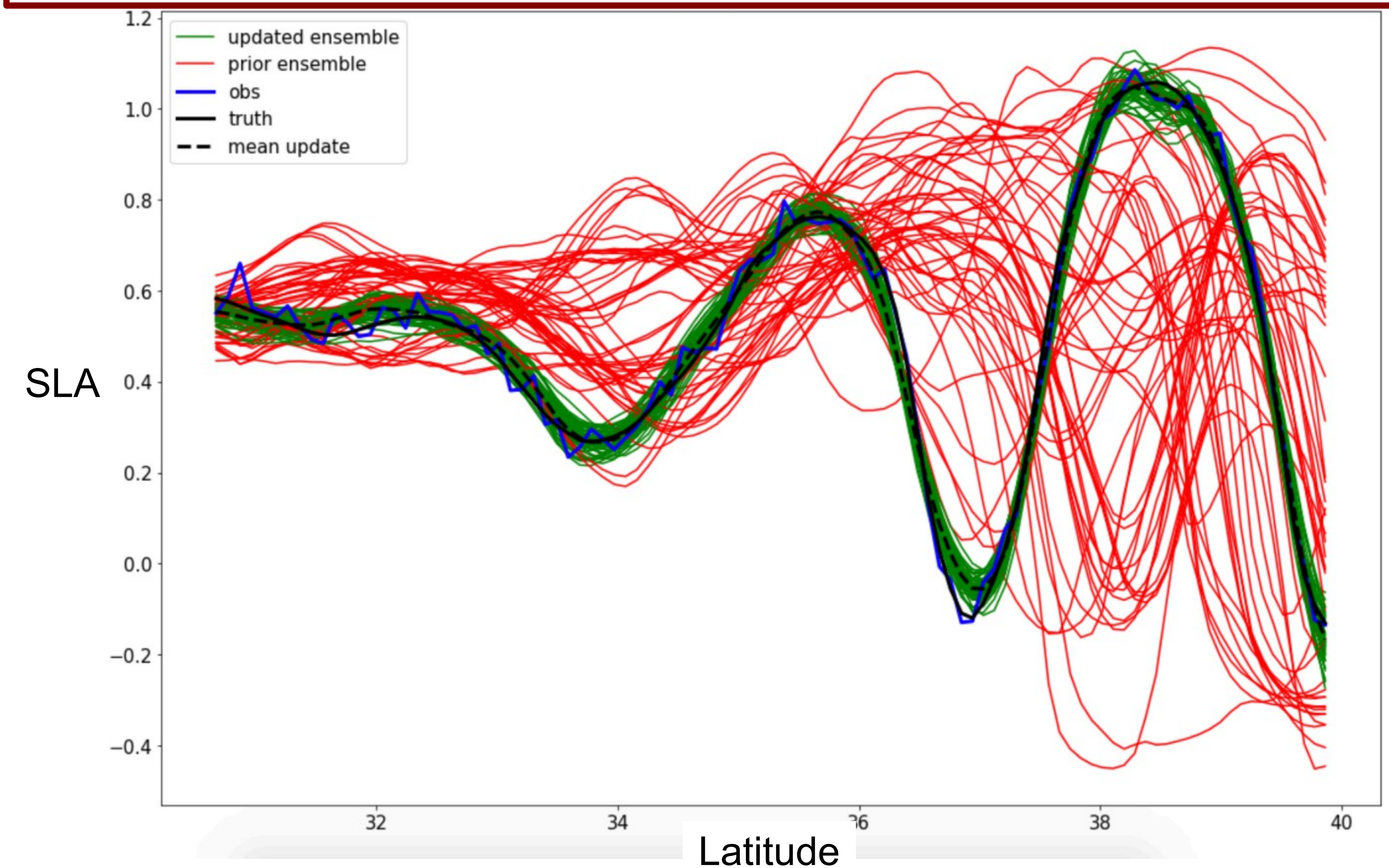


Position of the observations



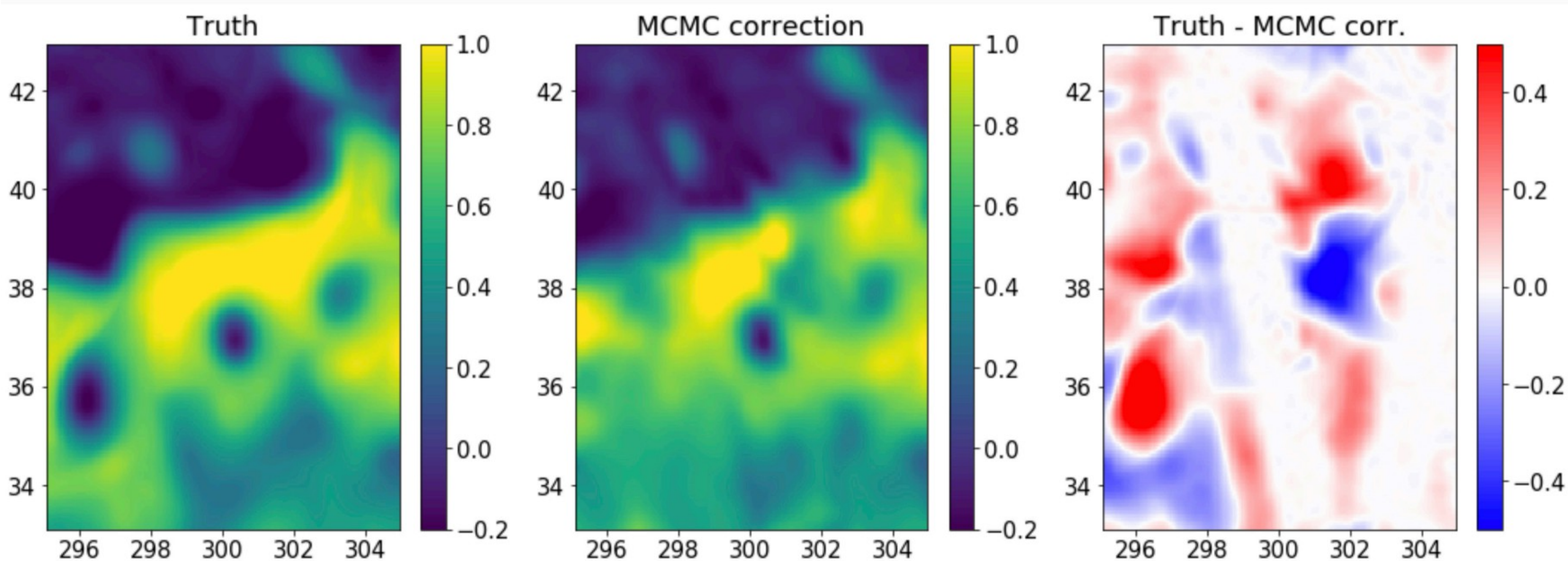
Posterior ensemble

First attempt with nadir altimetry



Black: “True” along-track SLA (simulated with NEMO model). Blue: Simulated observation (truth+noise). Red: Prior ensemble. Green posterior ensemble.

First attempt with wide-swath altimetry



Pseudo-observations from SWOT are simulated from a NEMO (NATL60) model SLA field (left). The MCMC sampler is run constrained with the observations (centre). Right: difference.

Conclusions and perspectives

A generic approach that is applicable to several disciplines

The method is able to generate random fields subjected to structural constraints, dynamical constraints and/or observational constraints

This can be an alternative to Gaussian ensemble data assimilation approaches at a cost that remains about the same in many situations

With the possibility to cope with nonlinear and nonlocal observation operators

Practical implementation for altimetry is work in progress.