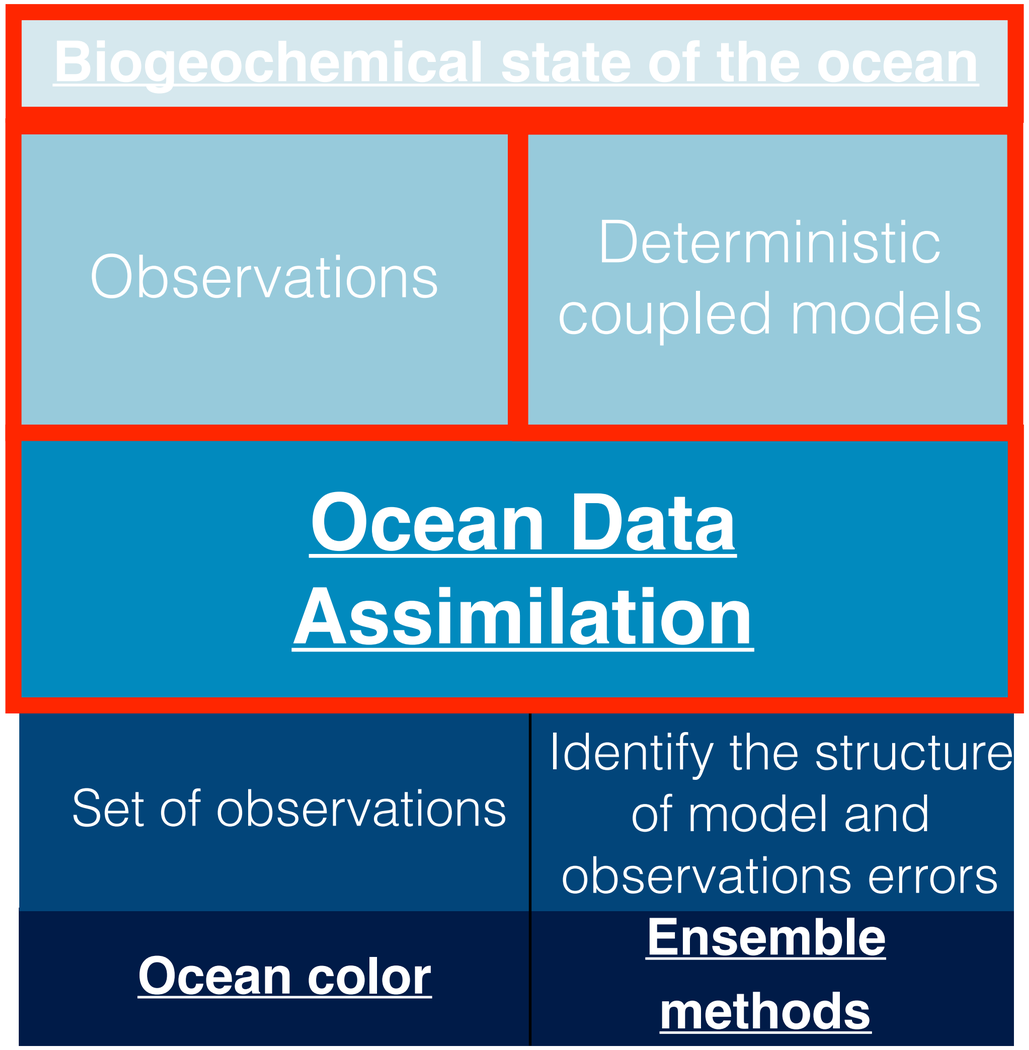
Towards the assimilation of satellite chlorophyll data into an ensemble simulation to infer the biogeochemical state of the North Atlantic

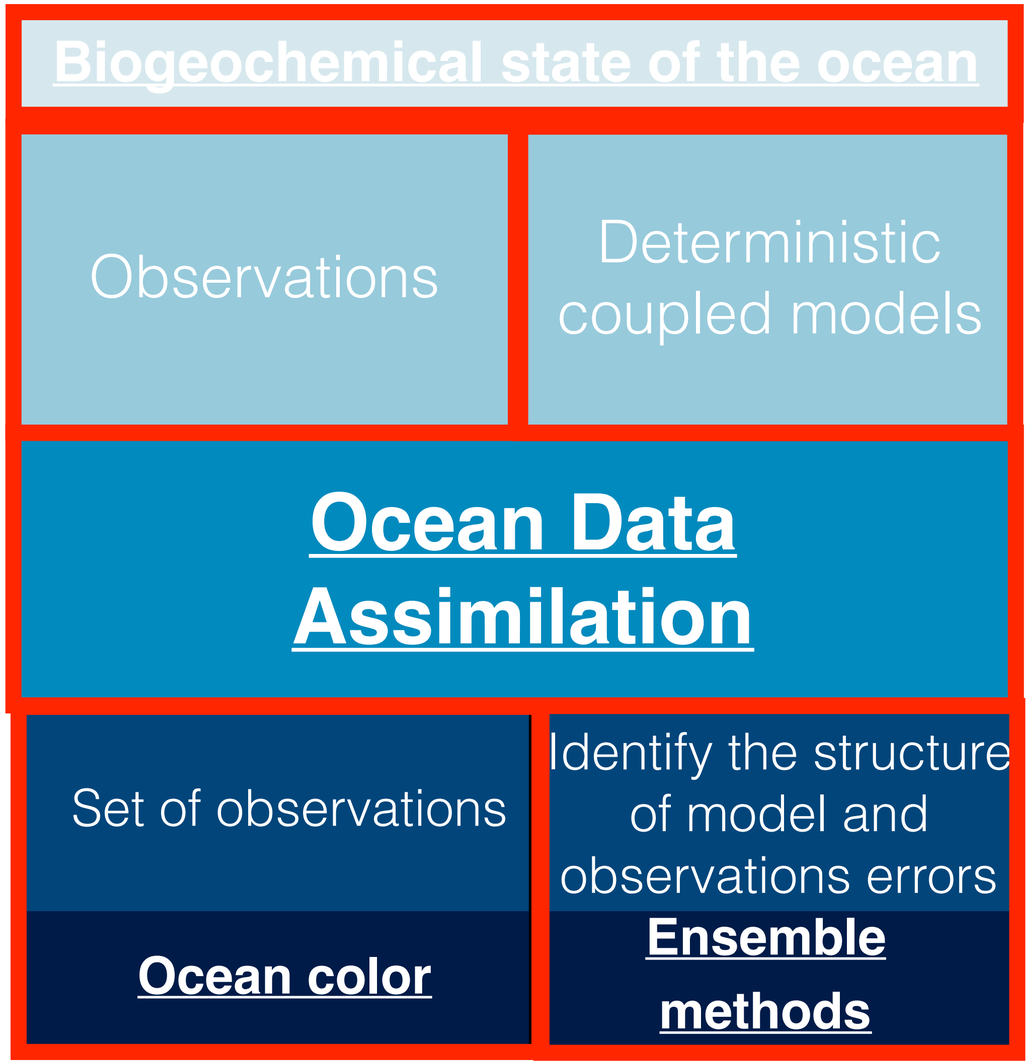
Insights from a 1-year experiment

Yeray Santana-Falcón, Pierre Brasseur, Jean-Michel Brankart, and Florent Garnier

1. Context and objectives



- **Ocean's key role in mediating global carbon stocks**
- **Detailed view but incomplete coverage**
- **Higher resolution but uncertain**
  - Vague description of physics
  - Reduced complexity
  - Empirical parameterisation
  - Unresolved scales
- **Ocean state depictions → integration of physical data**
- **To assimilate biogeochemical data is just starting to emerge**
  - No linearities
  - Lower availability of observations
- **CMEMS: to provide users with regular information in real time and delayed mode (reanalysis) about essential biogeochemical sate variables of the ocean, by combining models, satellite and *in situ* data**



Objective

Develop a statistically robust non-Gaussian coupled system to routinely estimate the biogeochemical state of the ocean through data assimilation

Strategy

Assimilating ocean color data into a probabilistic coupled configuration

- Only biogeochemical observations with routinely global coverage
- Describe the evolution of the probability density functions

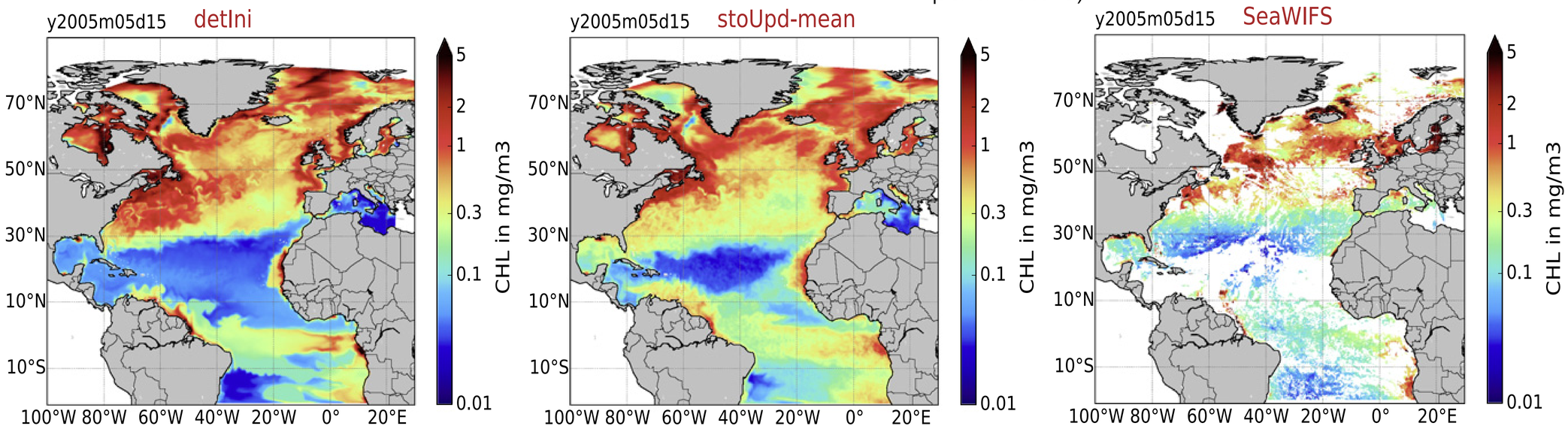
2. Methodology

Deterministic model

- Online coupled NEMO-PISCES
- North Atlantic basin
- 0.25° horizontal resolution (eddy-permitting)
- 46 geopotential levels

Probabilistic ensemble model

- Stochastic perturbations (Lognormal noises) to explicitly simulate model uncertainties
  - **Unresolved scales** in the presence of non-linear processes
  - Simplification of the true state by a **limited number of state vectors** (7 biogeochemical parameters)



Brankart, J. M. et al. (2015)
Garnier, F. et al. (2016)

Deterministic model

- Online coupled NEMO-PISCES
- North Atlantic basin
- 0.25° horizontal resolution (eddy-permitting)
- 46 geopotential levels

Probabilistic ensemble model

- Stochastic perturbations (Lognormal noises) to explicitly simulate model uncertainties
  - **Unresolved scales** in the presence of non-linear processes
  - Simplification of the true state by a **limited number of state vectors** (7 biogeochemical parameters)

24-members ensemble simulation

- **Perform data assimilation**
- **Includes uncertainties**
- **Objectively comparable with observations**

- Daily 4-km surface chl-*a* from MERIS
- Avoids interpolation before being integrated
- Partial coverage

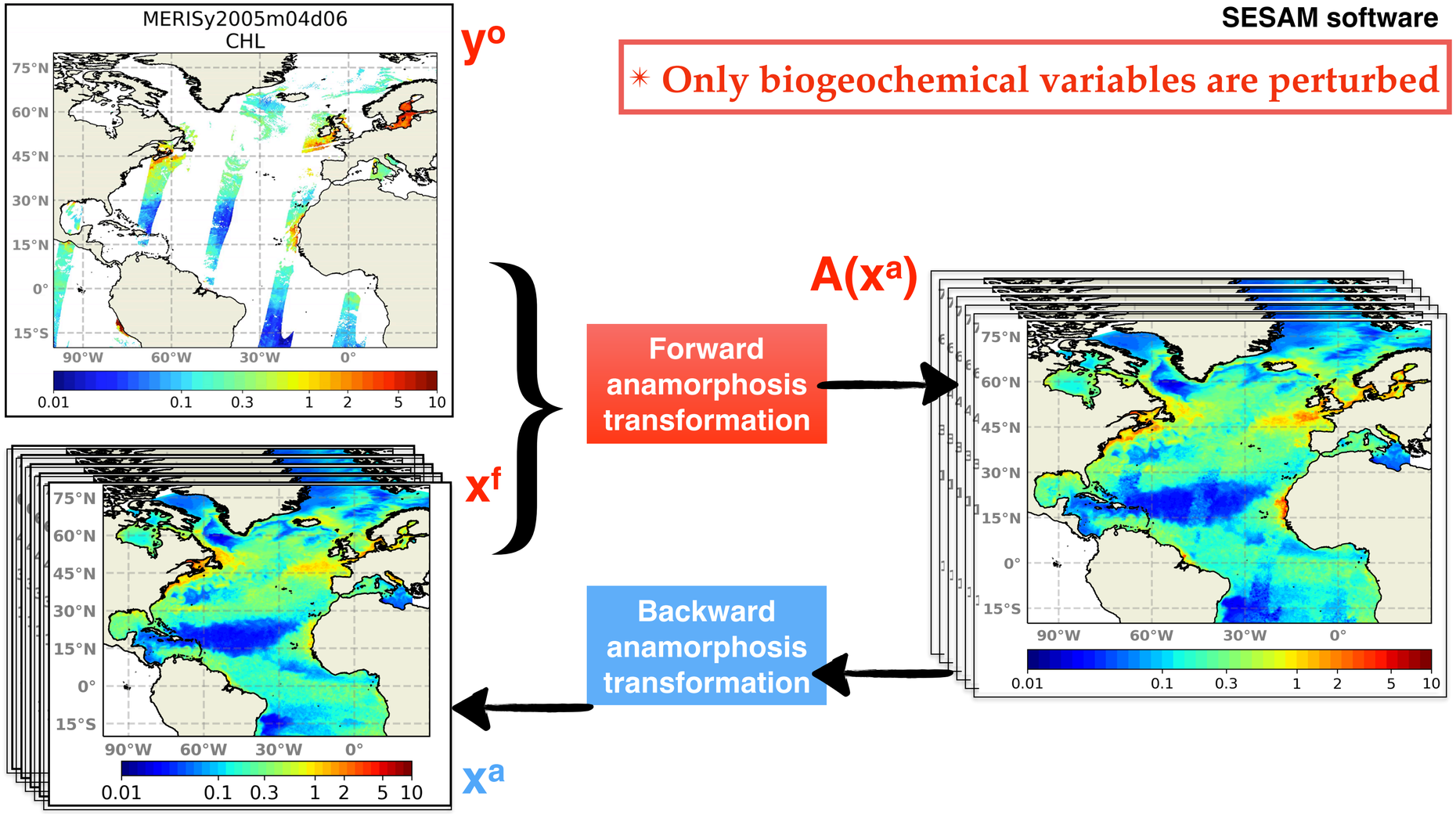
Square root observational update

(Brasseur and Verron, 2006)

SEEK algorithm

SESAM software

✳ Only biogeochemical variables are perturbed



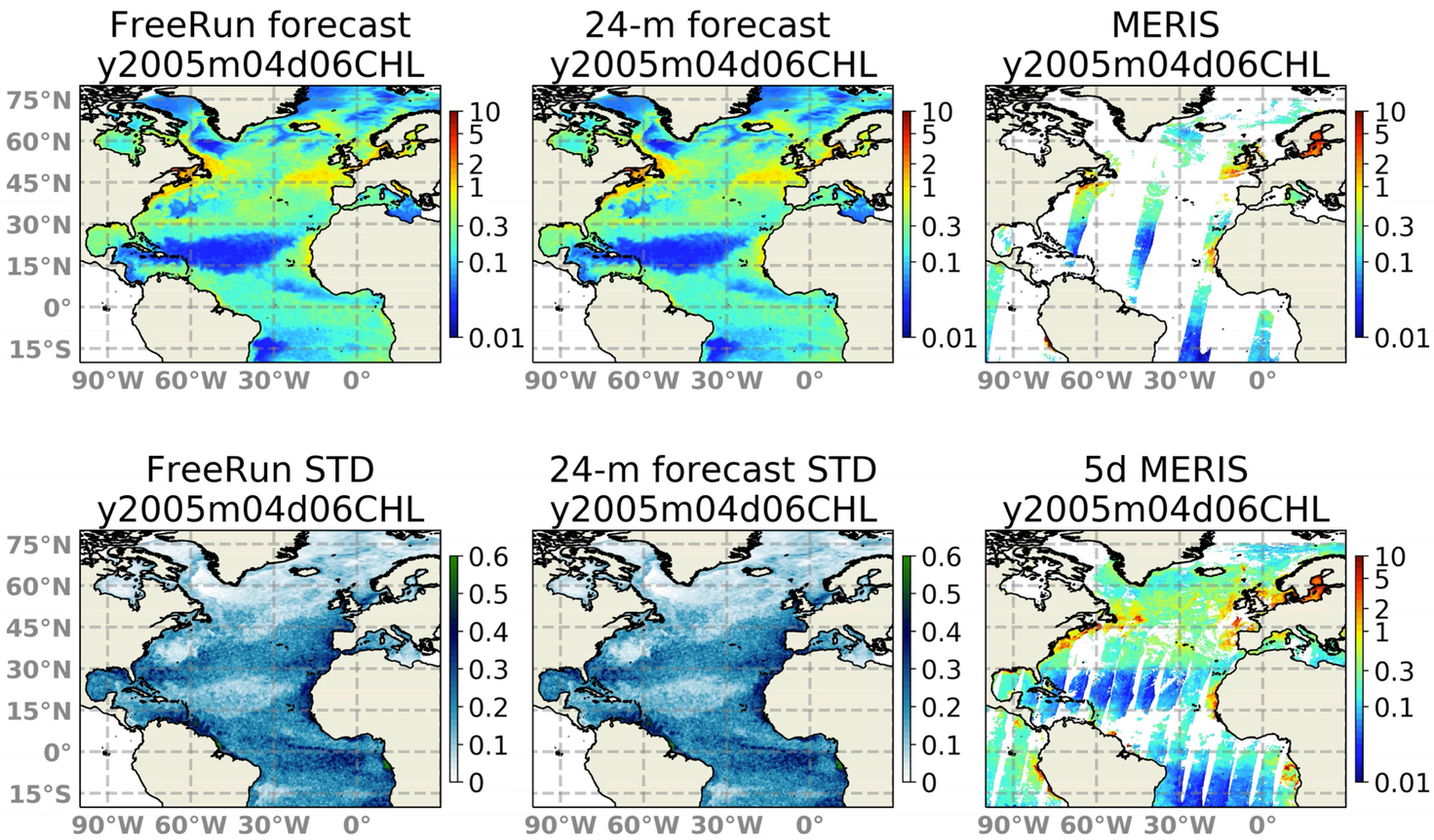
3. Global diagnosis of a 1-year assimilation experiment

Daily 24-members ensemble median

Merged daily-integrated satellite product

✳ **Non-assimilated analogue Free Run**

- Large-scale spatial pattern is reproduced
- Good performance at highly productive regions
- Forecast restores the uncertainty of the system to match the satellite's
- Improvements over the free run simulation
  - Underestimation of oligotrophic region
  - Overestimation of Gulf Stream and high latitudes
- Strong transition between oligotrophic to temperate waters
  - Misfits increase during autumn

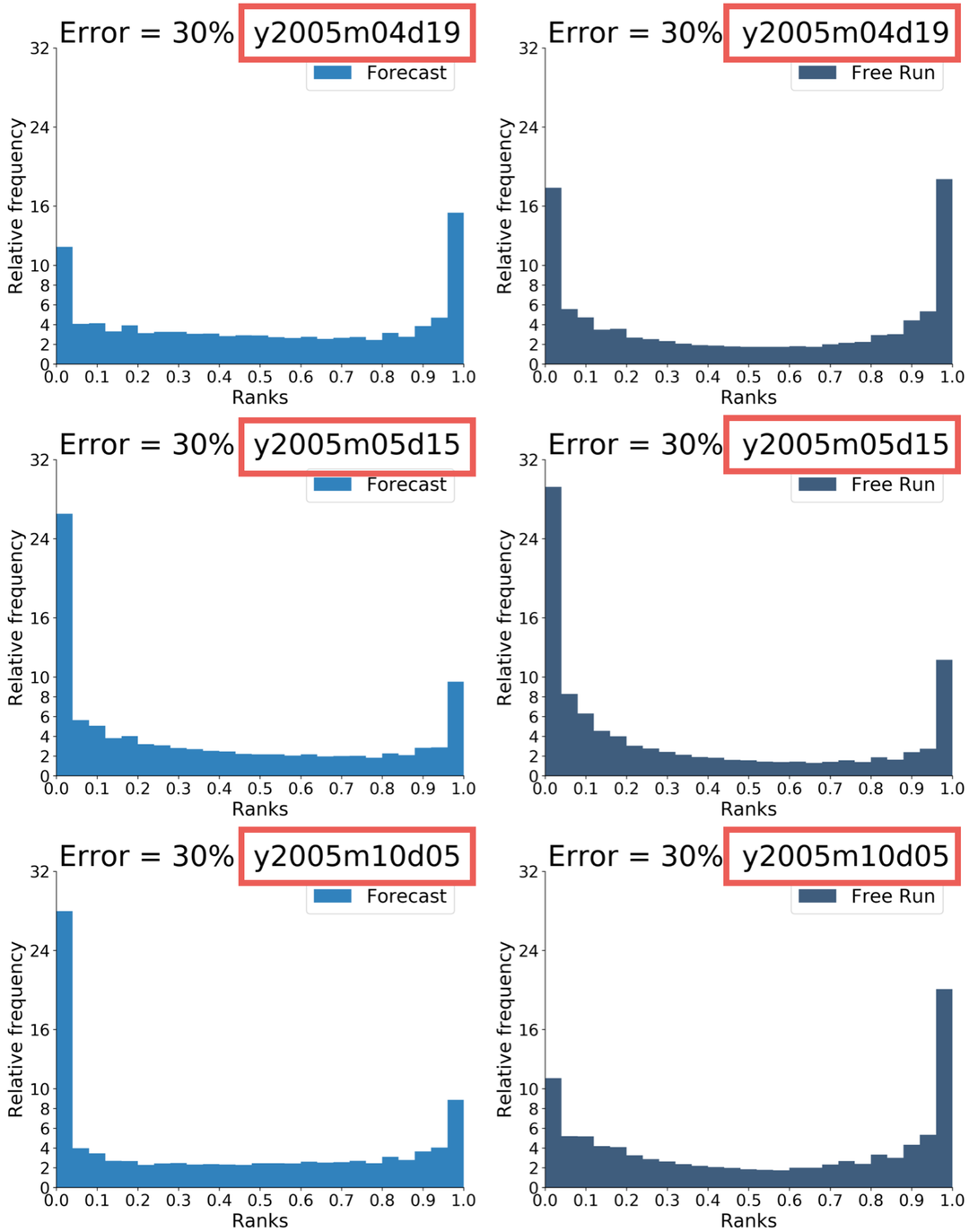


Daily rank histograms

SeaWiFS data with 30% of error

✳ **Non-assimilated analogue Free Run**

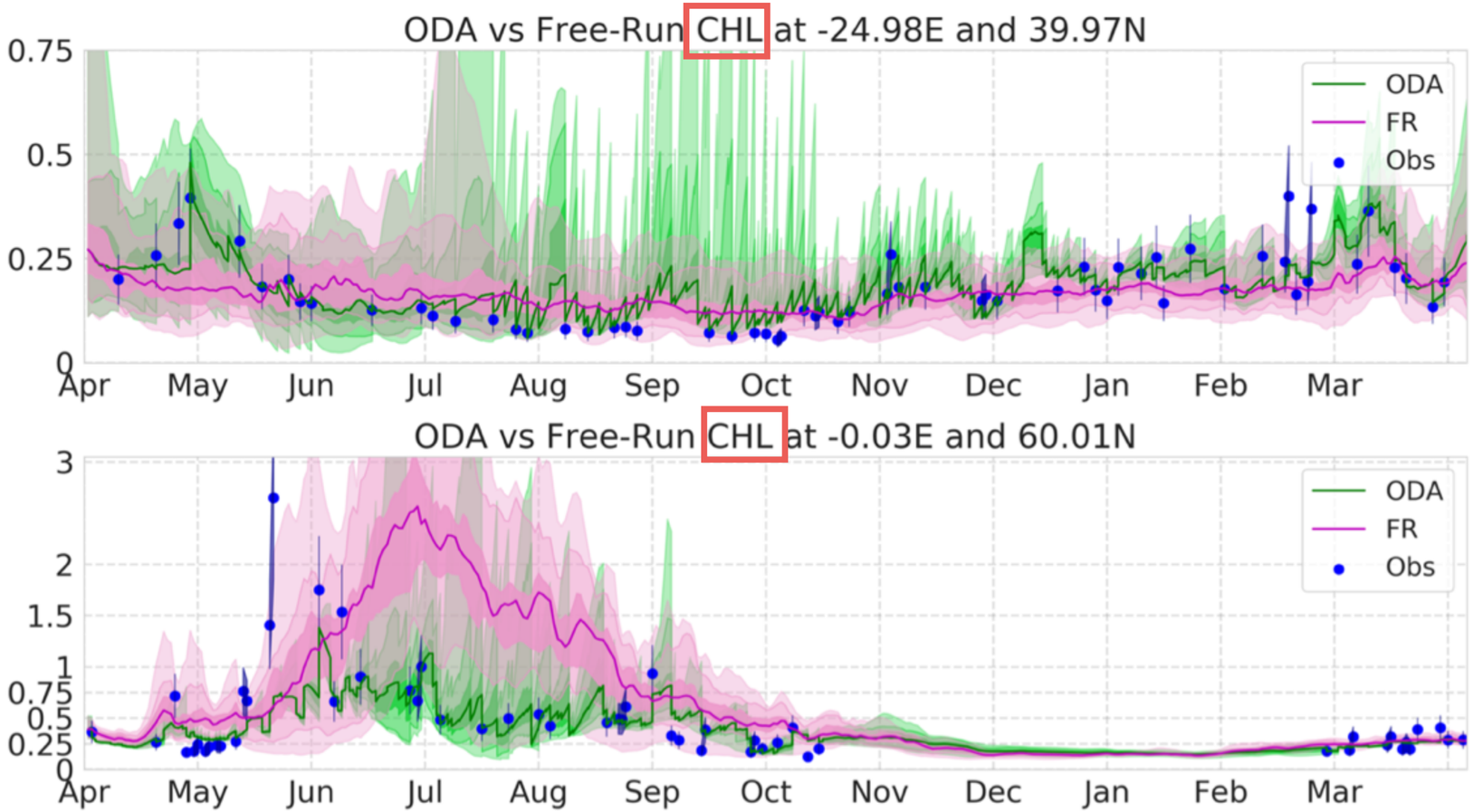
- Accumulation of ranks at the extremes; insufficient number of members
- Overpopulated left side → positive bias; chlorophyll can never become negative
- Ensemble envelope covers a major part of the given observations (60 to 80%)
- Uniform distribution of ranks; relatively good reliability

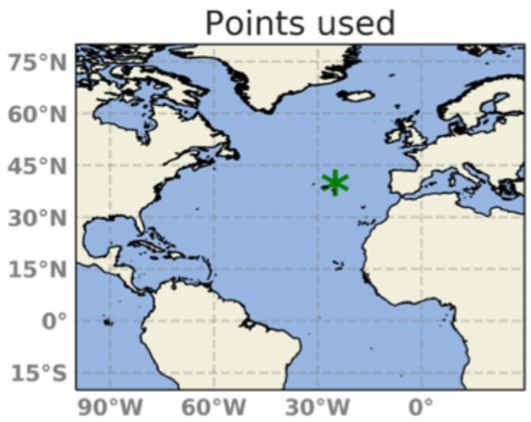


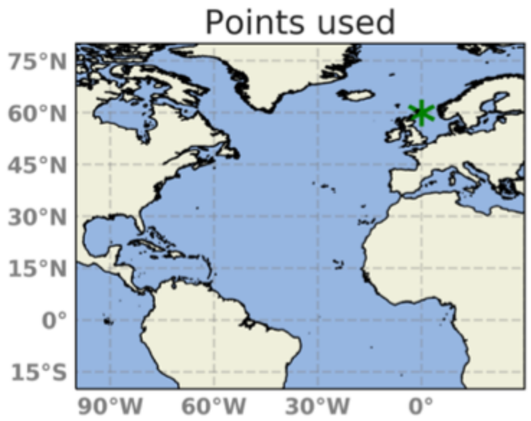
4. Regional diagnosis

Yearly time-series of quantiles of the assimilation and free run ensembles

Available observations are included



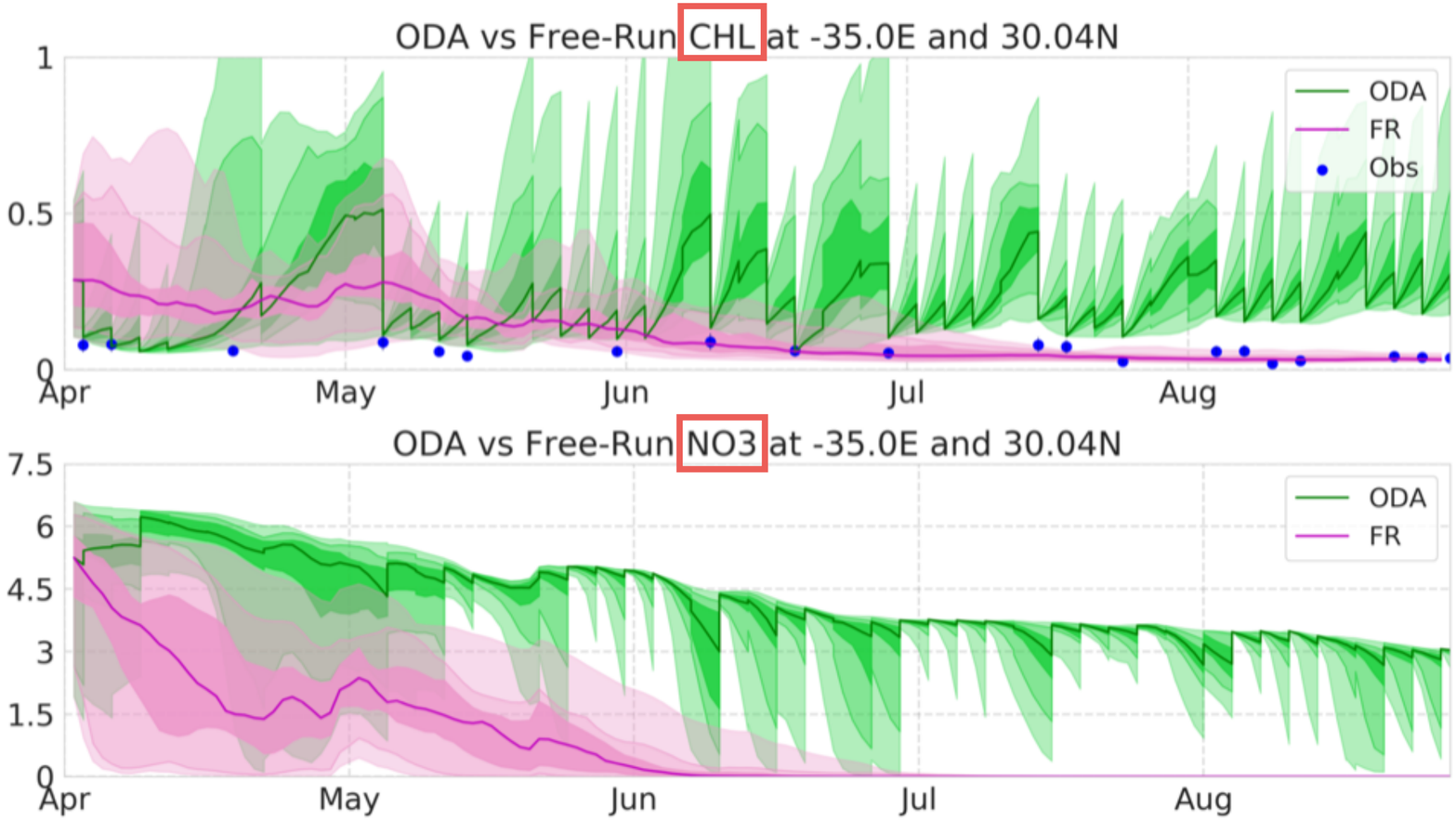


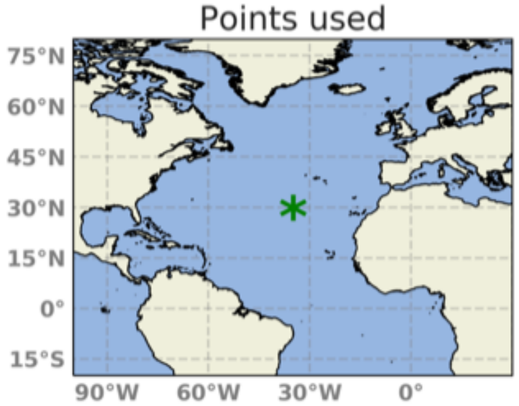


- Relative agreement between the assimilation system and observations
- The assimilation system reproduces the regional seasonal variability
- MERIS info is integrating into the system every 6-7 days

5-months time-series of quantiles of the assimilation and free run ensembles

Chlorophyll and nitrate are represented



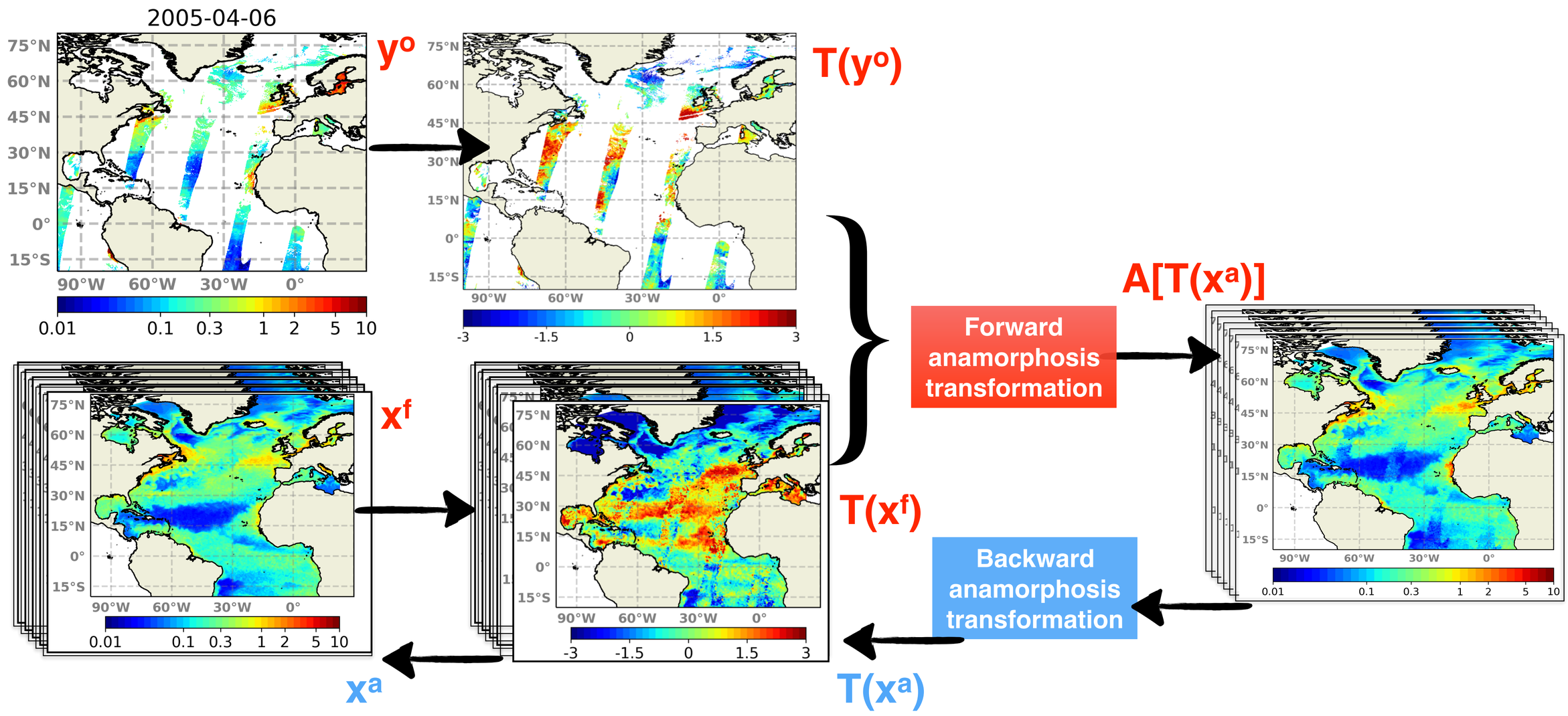


- Imbalances between the observed and non-observed variables (mainly nutrients)

- Observations not included in the assimilation envelope
- Assimilation simulation systematically overestimates chlorophyll
- The assimilation system is unable to reproduce the regional seasonal variability

5. Corrections on the fluctuating component

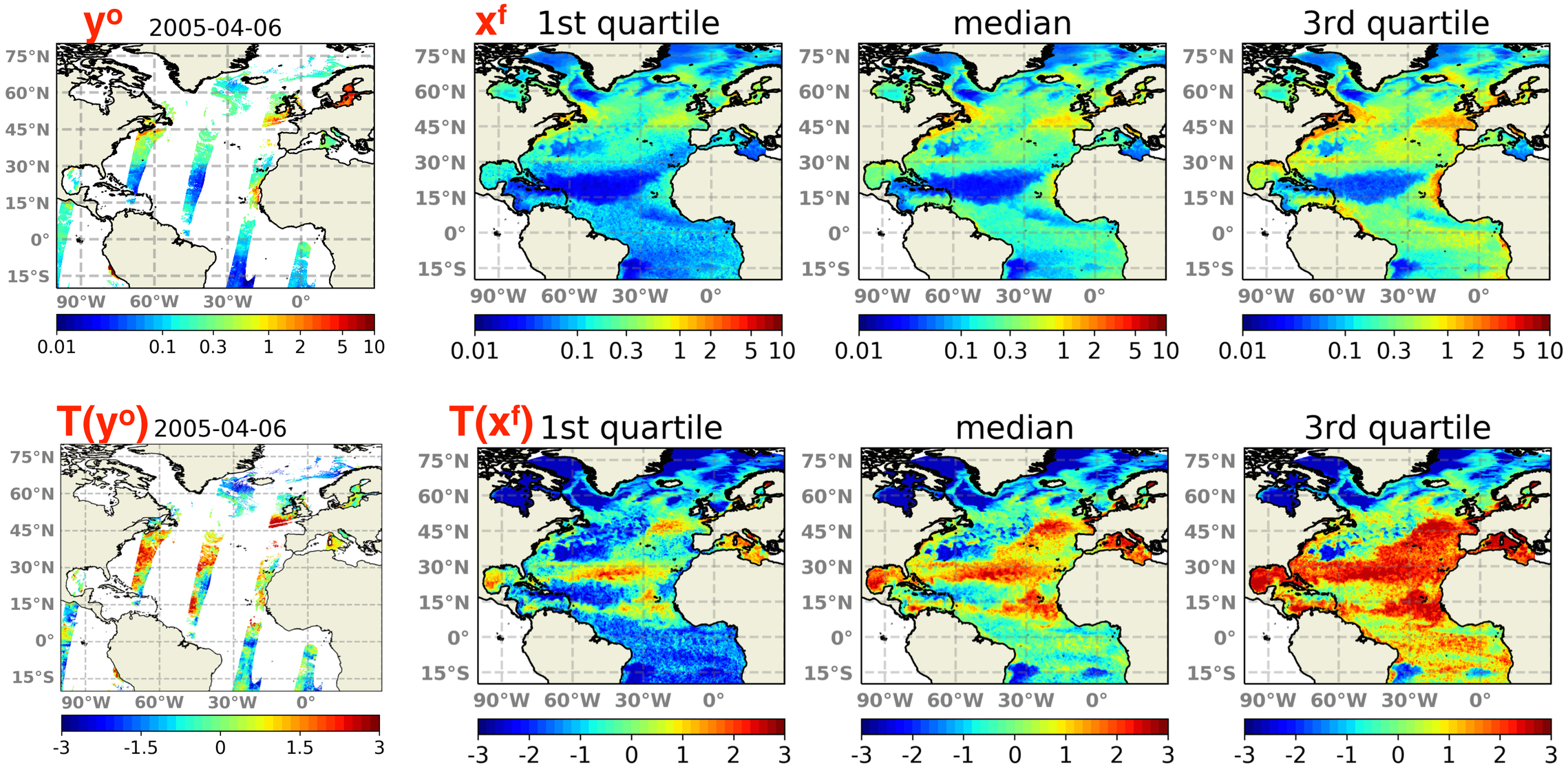
Time-independent transformations are applied prior the analysis update



Observations and forecast before and after climatological transformations

Corrections avoid imbalances due to model formulation attractors

Distribution of the ensemble should include observations



6. Summary and perspectives

Non-Gaussian ensemble coupled assimilation system

Assimilation problem in probabilistic terms

Parameterisations are valid for a major part of the domain

Improves several misfits associated with a non-assimilated simulation

Main problem arises in the transition zone between the oligotrophic and temperate waters

Imbalances between chlorophyll and nutrients' increments

Assimilation applied only to the fluctuating component

We expect to alleviate inconsistencies due to the model strong attractors

Short term

Insights from applying corrections only to the fluctuating component

Publication

Santana-Falcón, Y., Brankart, J. M., Brasseur, P., and Garnier, F. Ensemble data assimilation of satellite chlorophyll to operationally infer the biogeochemical state of the North Atlantic, Ocean Science *(in preparation)*

Long term

Sophistication of the system

Assimilation of physical variables, *in situ* data from Bio-Argo profiles

Re-structure of the methodology

Use of *AI* for developing simpler coupled assimilation systems