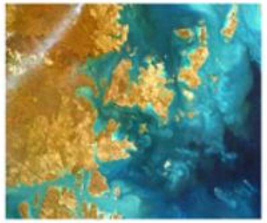
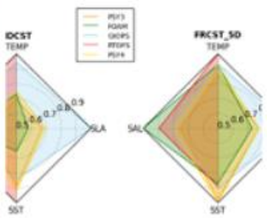
OceanPredict
Advancing the science of ocean prediction



Ensemble generation and evaluation for monitoring and forecasting the Green Ocean

Pierre Brasseur

*Institute for **G**eosciences and **E**nvironmental research (**IGE**), CNRS, France*

with contributions from a bunch of colleagues …

MEAP Task Team, 3rd video-meeting, June 10th 2021

IGE
Institut
des géosciences
et de l'environnement

cnrs

Outline

- ***Why do we need ensembles ?***
  Uncertainties in a chaotic ocean

- ***How to generate ensemble of 3D states / trajectories***
  Some examples using NEMO-PISCES

- ***Ensemble verification***
  Empirical metrics and skill scores

- **Challenges for the future**
  To open the discussion

- ***Why do we need ensembles ?***
  Uncertainties in a chaotic ocean

- ***How to generate ensemble of 3D states / trajectories***
  Some examples using NEMO-PISCES

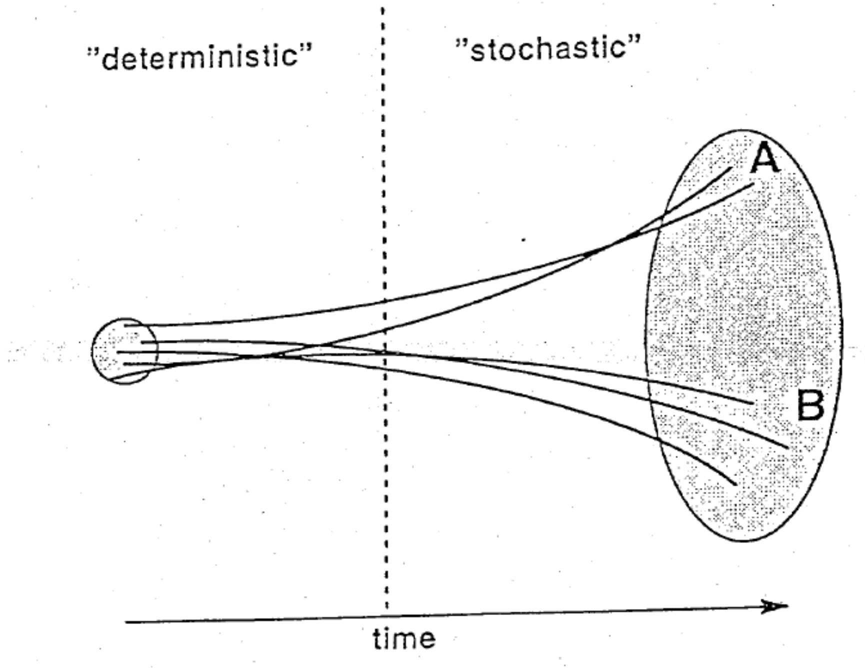
- ***Ensemble verification***
  Empirical metrics and skill scores

- **Challenges for the future**
  To open the discussion

Ensemble simulation: an old practice in operational NWP

A new era in NWP began on 7 December 1992 when NMC began performing daily ensemble predictions … In adopting the ensemble approach, we explicitly recognize that ***forecasts should be considered stochastic, not deterministic in nature****.*

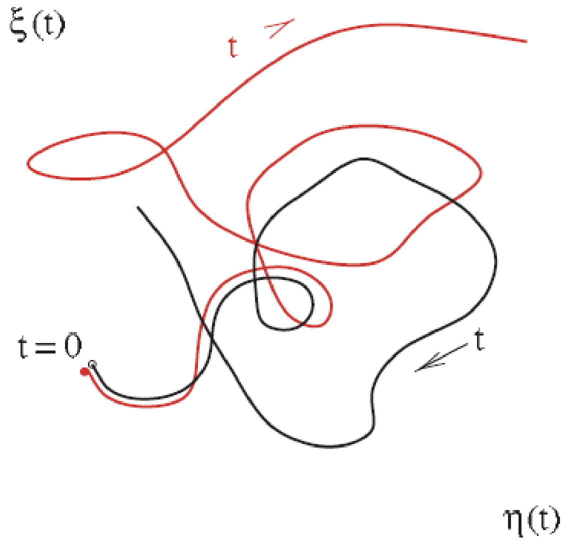
Tracton and Kalnay, 1993



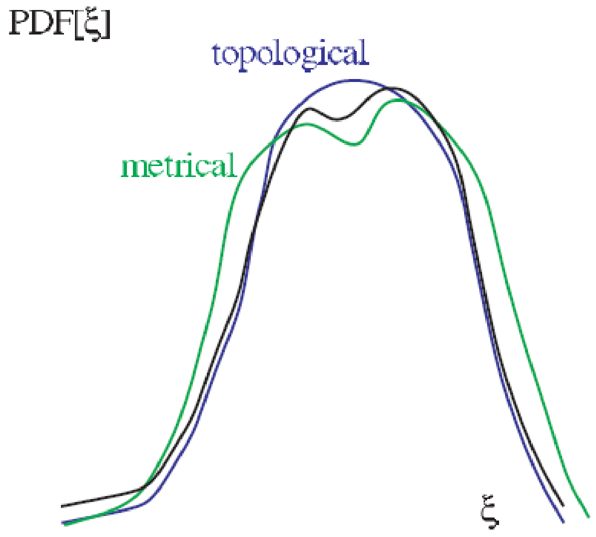
Ensemble simulation: an old practice in operational NWP

… also of interest to (operational) ***ocean modelling***

Atmospheric and oceanic computational simulation models often successfully depict chaotic space–time patterns … This success is accomplished through ***necessary but nonunique choices for discrete algorithms, parameterizations, and coupled contributing processes*** *that introduce structural instability into the model. Therefore, we should expect a degree of irreducible imprecision in quantitative correspondences with nature, even with plausibly formulated models and careful calibration.* ***Where precision is an issue (e.g., in a climate forecast), only simulation ensembles made across systematically designed model families allow an estimate of the level of relevant irreducible imprecision.***



Small changes in initial or boundary conditions imply limited predictability with exponential growth in phase space differences.



Small changes in model formulation alter the long-time probability distribution function (PDF) (i.e., the attractor).

Mc Williams, 2007

On the chaotic nature of ocean-atmosphere dynamics

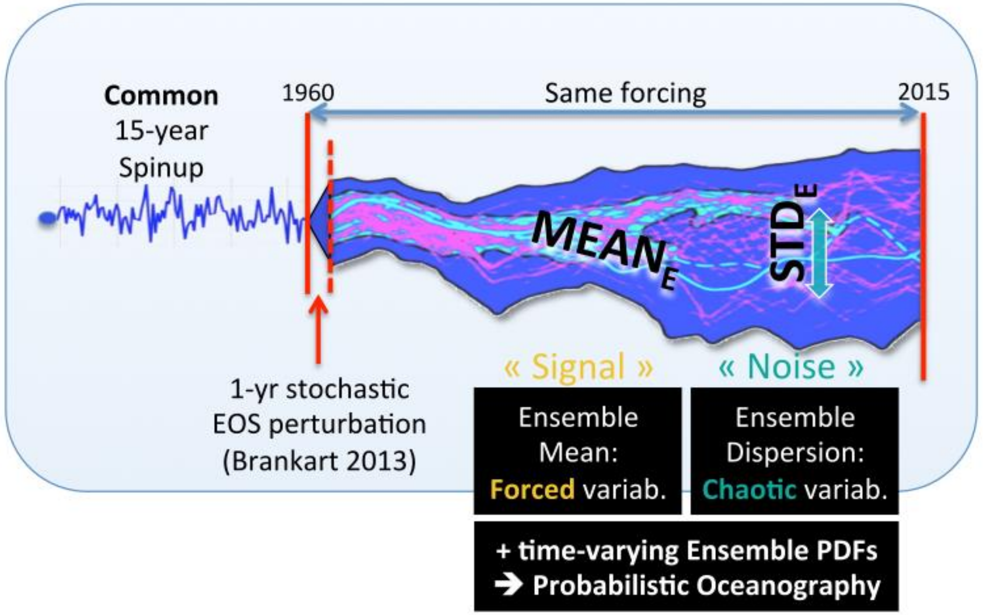
in a global eddy-permitting ocean circulation model (NEMO)

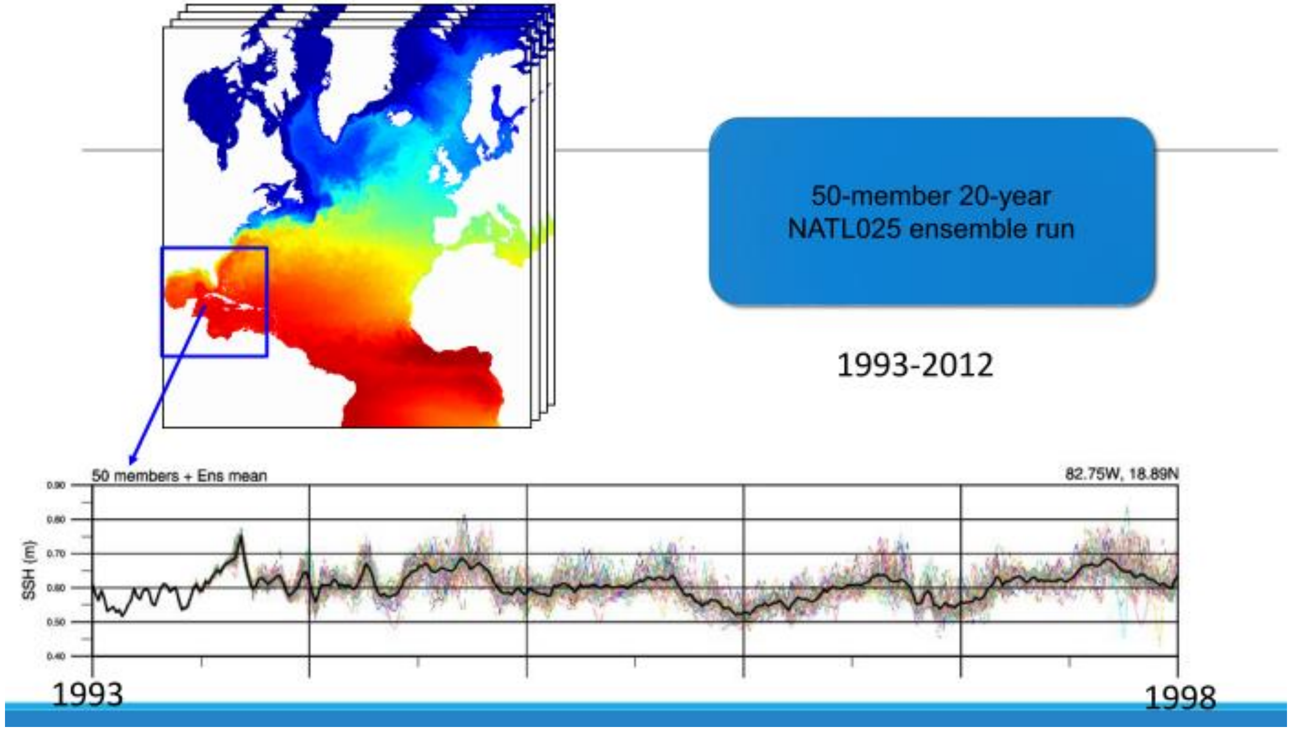
OCCIPUT ensemble simulations

50-member ¼° ensemble hindcasts :

- Global ocean (56 years)

- North Atlantic (20 years)





Penduff et al (2014)
Bessières et al (2016)

Ensembles: needed to solve a variety of problems

in NWP, ocean physics, marine biogeochemistry and ecology etc.

when precision is an issue

- **Sensitivity** to initial conditions and chaoticity (Popova et al., 1997)
- **Predictability** studies (Seferian et al., 2014)
- Probabilistic modelling and **risk assessment for decision-making** (Meier et al., 2019)
- Quantification of chaotic **intrinsic variability** modes (Gehlen et al., 2020)

model-only

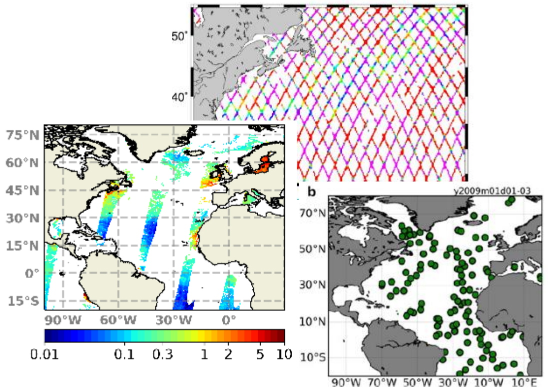
- **Ocean BGC state estimation** through Data Assimilation (Carmillet et al., 2001)
- **Parameter estimation** incl. regionalization (Doron et al., 2013)
- **Hypothesis testing** to identify error sources (Schartau et al., 2017)
- OSSEs and **observing system design** (Germineaud et al., 2019)
- …

model-data combinations

Generation methods under consideration depend on the purpose

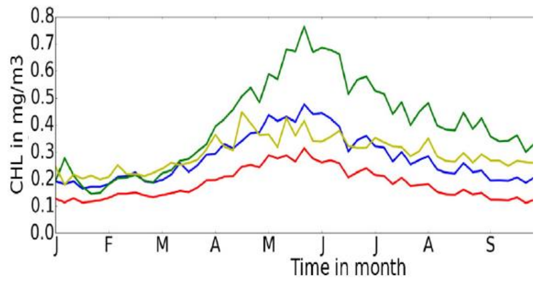
Ensemble strategies for DA and indicator estimations

The SEAMLESS concept



Observations (OC, SLA, Argo)
inaccurate, incomplete

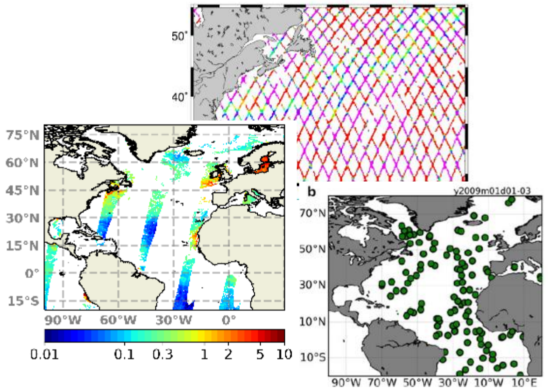
models

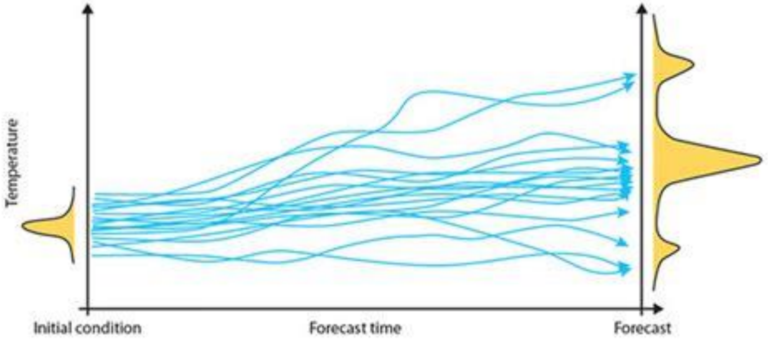


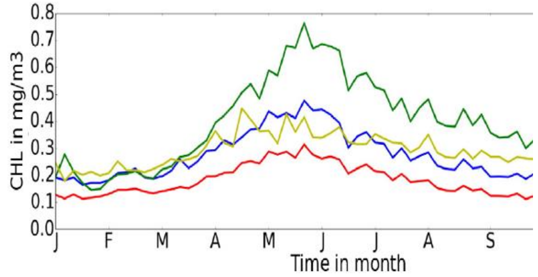
Indicators (e.g. phenology, POC …)
<u>and</u> associated uncertainty

Ensemble strategies for DA and indicator estimations

The SEAMLESS concept







Observations (OC, SLA, Argo)
inaccurate, incomplete

Probabilistic
models

Indicators (e.g. phenology, POC …)
<u>and</u> associated uncertainty

Ensembles of BGC variables usually depict <u>non-gaussian PDFs</u>

- ***Why do we need ensembles ?***
  Uncertainties in a chaotic ocean

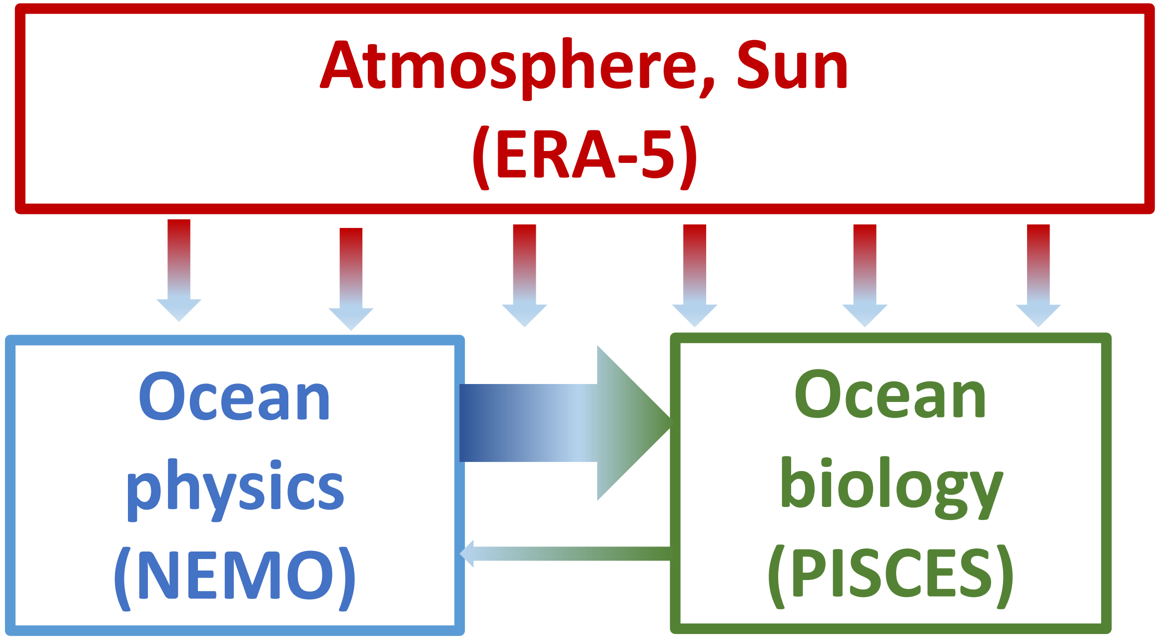
- ***How to generate ensemble of 3D states / trajectories***
  Some examples using NEMO-PISCES

- ***Ensemble verification***
  Empirical metrics and skill scores

- **Challenges for the future**
  To open the discussion

Coupled physical-BGC models: system components

Approximations resulting from the system design



Uncertainty sources:

- *external forcings, air-sea fluxes*
- *physics – biological « coupling »*
- *unresolved scales, unresolved bio-diversity*
- *physical, optical processes and parametizations*
- *biological processes and parameterizations*
- *numerical schemes, discretizations etc.*

Ensembles are assumed to sample the PDFs resulting from these many sources of uncertainty

Ensemble generation with NEMO and/or PISCES

Exemples

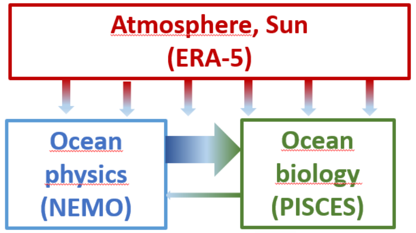
- Unresolved sub-grid scale fluctuations in PHY or BIO (Brankart et al., 2015)
- Uncertain BIO model parameters and unresolved sub-grid scales fluctuations in BIO (Garnier et al., 2016; Santana-Falcon et al., 2020)
- Impact of intrinsic dynamical variability on CO2 air-sea fluxes (Gehlen et al., 2020)
- Perturbation of numerics in coupled PHY/BIO simulations (to account for location uncertainties) (Leroux et al., 2021)
- Perturbation of atmospheric forcings (Vervatis et al., 2021)

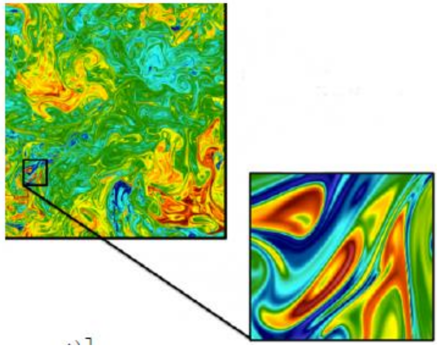
… and combinations

NEMO ensemble generation based on Brankart et al. (2015)

Uncertainties in the computation of density – unresolved sub-grid-scale fluctuations

$$\rho = \frac{1}{2}\Big\{\rho\big[T+\Delta T, S+\Delta S, p_o(z)\big] + \rho\big[T-\Delta T, S-\Delta S, p_o(z)\big]\Big\},$$



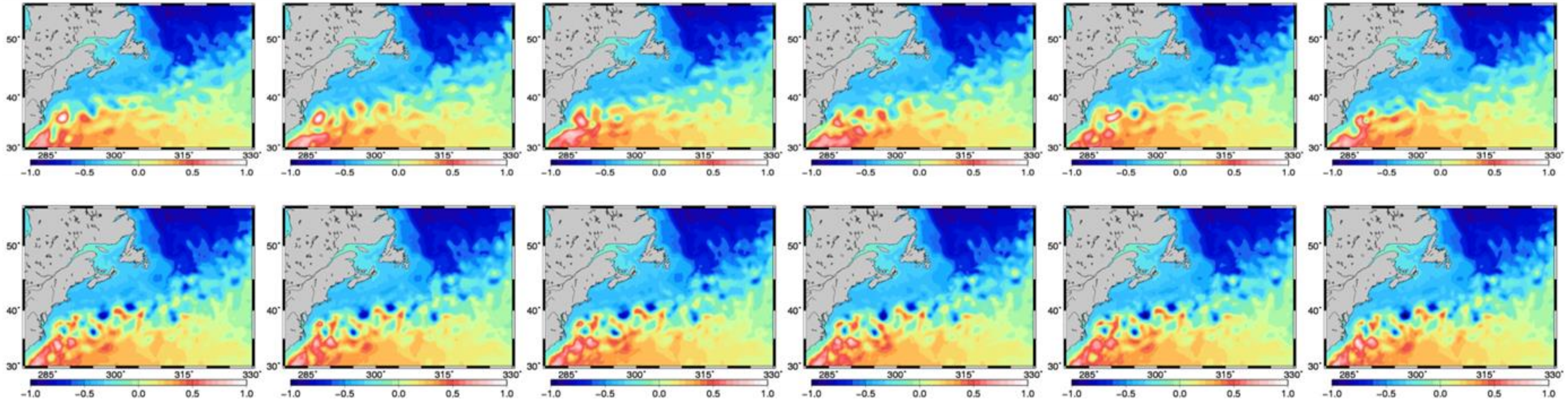


$$\left.\begin{aligned} \Delta T &= \xi \cdot \nabla T \\ \Delta S &= \xi \cdot \nabla S, \end{aligned}\right\}$$ random fluctuations (based on AR1 processes)

https://zenodo.org/record/61611

No effect if equation of state (EOS) is linear

Ensemble spread of the eddy field (SSH) over the Gulf Stream



From a 96-member ensemble after 6 months

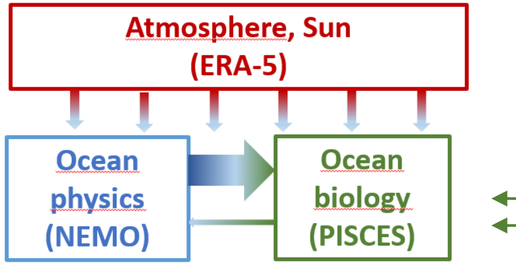
NEMO-PISCES ensemble generation based on Garnier et al. (2016)

Uncertainties in BGC model parameters p + unresolved, sub-grid scale fluctuations

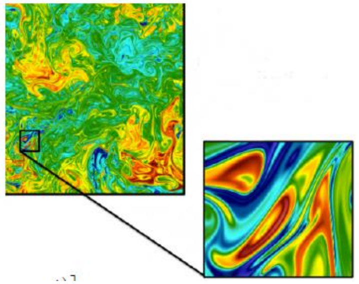
$$\left.\frac{\partial C}{\partial t}\right|_{bio} = \mathrm{SMS}\,(C, u, p \cdot \exp[\xi(t)], t)$$

$$p' = p \cdot \exp[\xi(t)] \sim \log\mathrm{N}(\mu = 0, \sigma = 0.3) \approx \mathrm{N}(\mu = 1, \sigma = 0.3)$$

Autoregressive processes : $\xi(t_{n+1}) = a\xi(t_n) + bw + c$



$$\left.\frac{\partial C}{\partial t}\right|_{bio} = \frac{1}{2}\Big[\mathrm{SMS}\,(C + C\xi(t), u, p, t) + \mathrm{SMS}\,(C - C\xi(t), u, p, t)\Big],$$

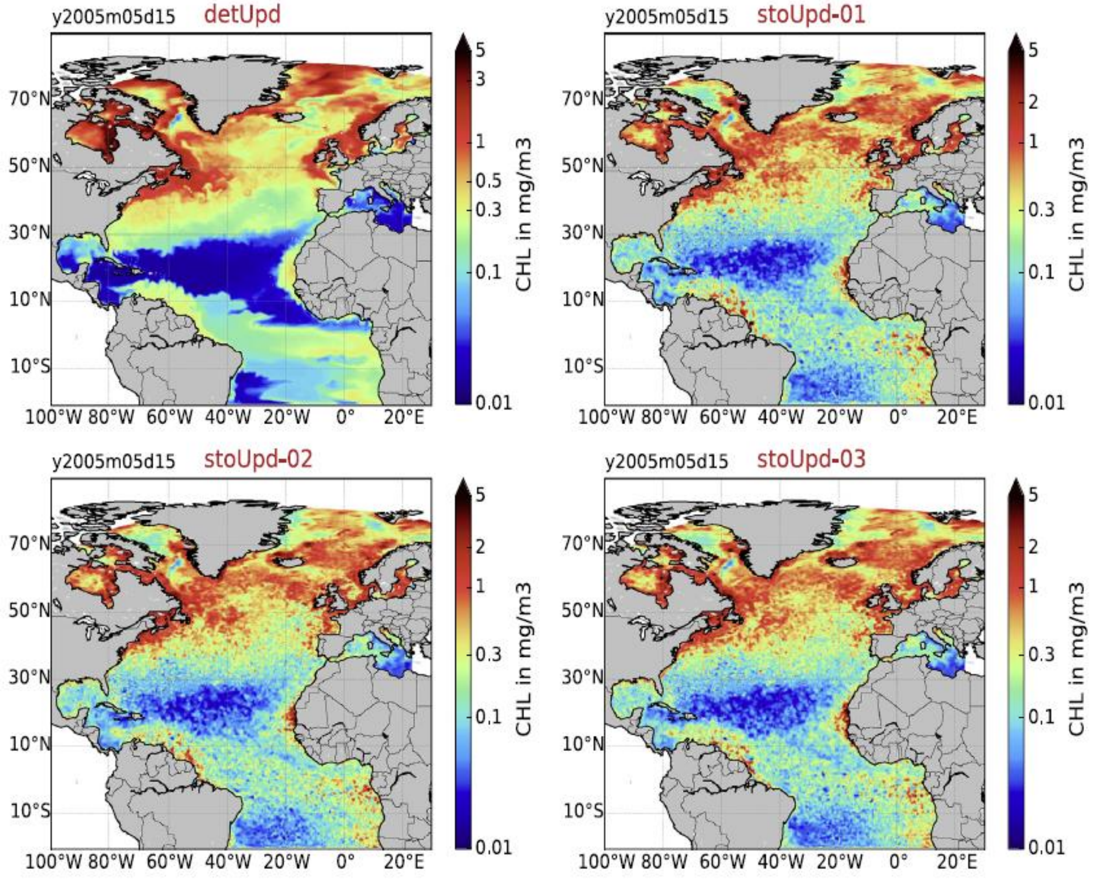


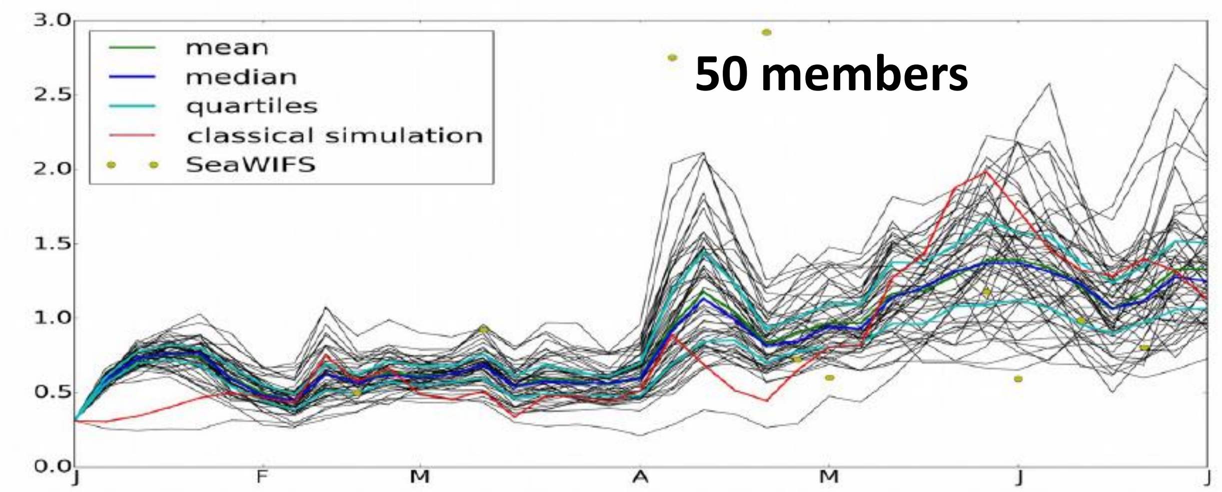
$C\xi(t) \equiv \delta C(t)$ – fluctuations, not resolved by the mesh.

Stochastic processes $\xi(t)$ are applied to all 24 passive tracers.

NEMO-PISCES ensemble generation based on Garnier et al. (2016)

deterministic vs. stochastic





50 members

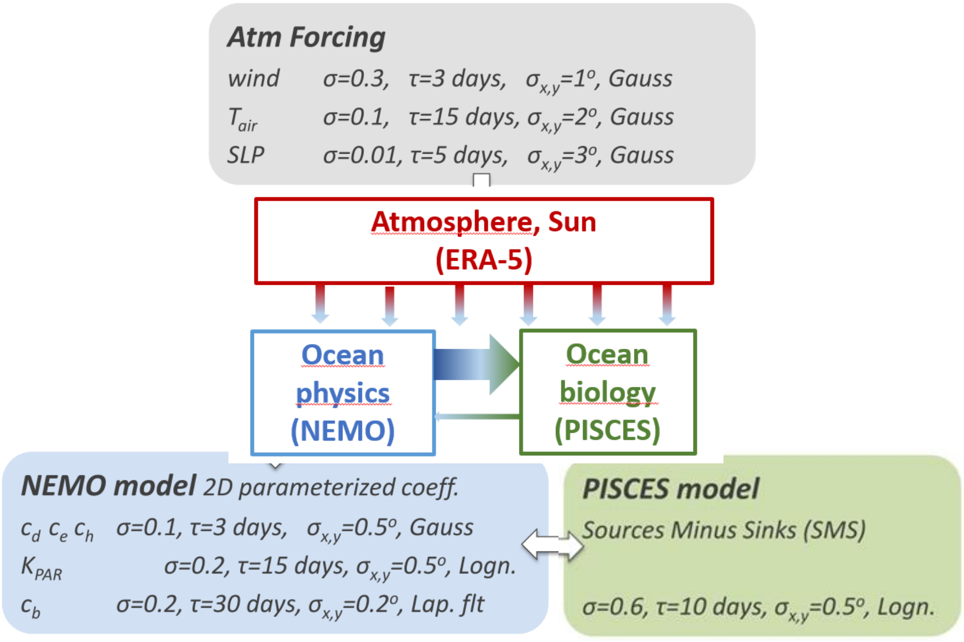
Time evolution of the phytoplankton PDF

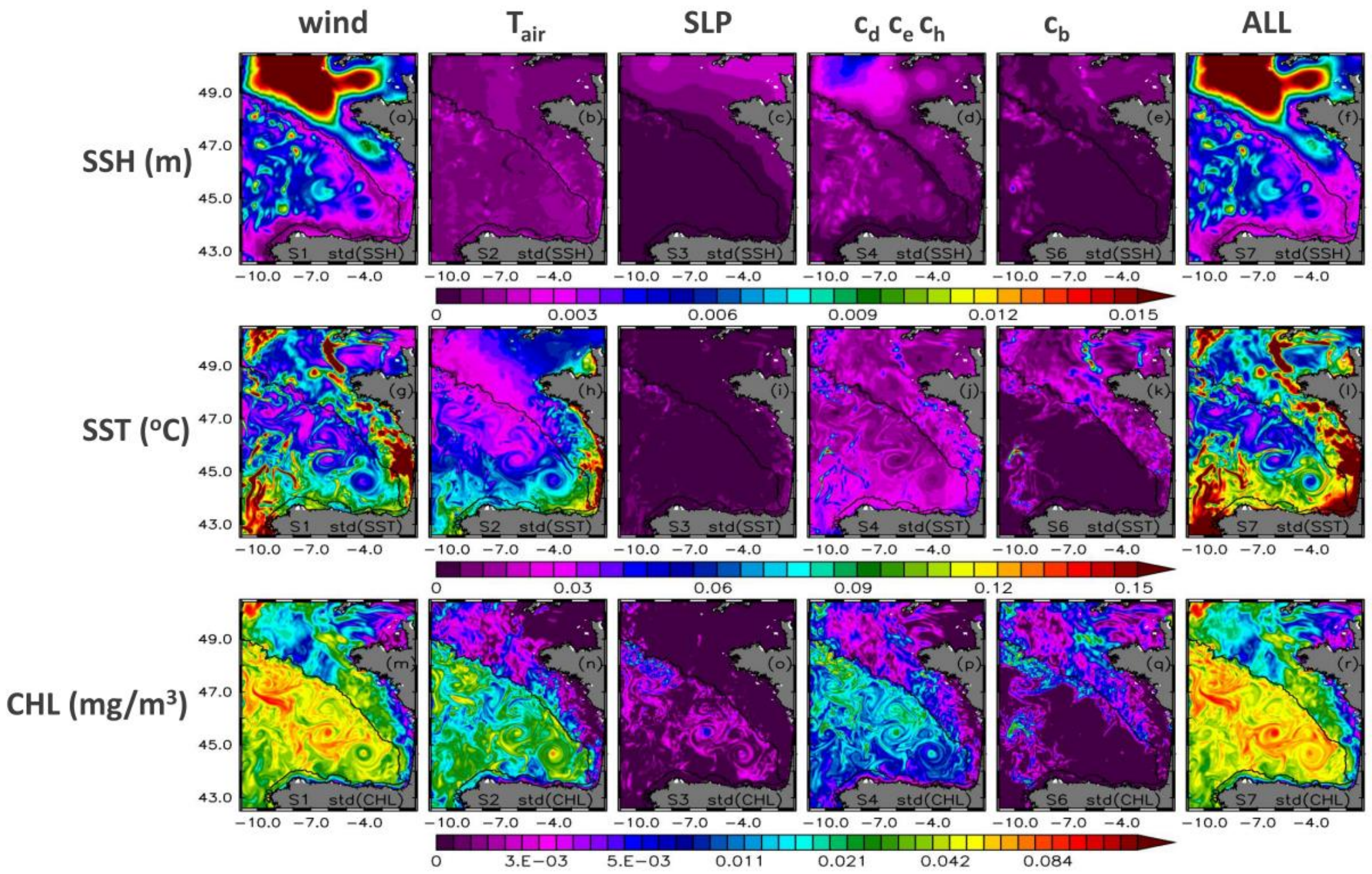
from January to June 2005

(no perturbation of the dynamics)

NEMO-PISCES ensemble generation - Vervatis et al. (2021)

Uncertainties in physical model parameters, unresolved, sub-grid scale BGC fluctuations and atmospheric forcings (Bay of Biscay)





https://zenodo.org/record/2556530

(De Mey et al., 2016 ; Vervatis et al., 2021)

Wind uncertainties dominate

- ***Why do we need ensembles ?***

  Uncertainties in a chaotic ocean

- ***How to generate ensemble of 3D states / trajectories***

  Some examples using NEMO-PISCES

- ***Ensemble verification***

  Empirical metrics and skill scores

- **Challenges for the future**

  To open the discussion

Verification metrics and skill scores

First applications to BGC ensembles in FP7 SANGOMA projet

- **Main idea**: consider the probability distribution as described by the ensemble, not only the mean and standard deviation (→ deals with non-Gaussian behaviours)

- **Key properties**:
  - ***Reliability*** = system's ability in producing PDFs in agreement with obs distribution
  - ***Resolution*** = system's ability in discriminating distinct observed situations (how informative is the system)

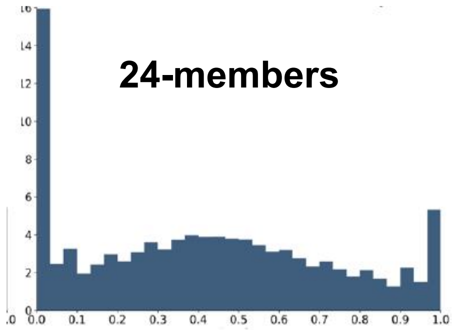
- **Commonly used metrics** (mostly from NWP)
  - Rank Histograms (**RH**, or *Talagrand Diagrams*, Candille et al., 2015)
  - Reduced Centered Random Variable (**RCRV**)
  - Continuous Rank Probability Scores (**CRPS**) : measures the global skill of a probabilistic system by evaluating both reliability and resolution
  - Brier Score (**BS**): a restriction of the CRPS to the probability space
  - Entropy (**EN**): measures the information content of the system, closely related to resolution

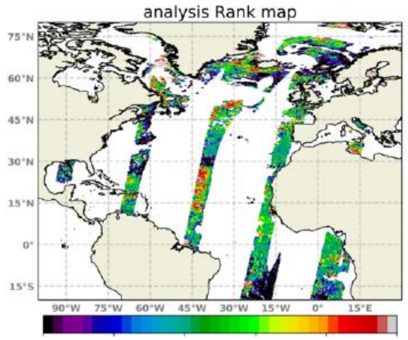
▶ **Code available :** github.com/brankart/ensdam (EnsDAM library), in src/EnsScores

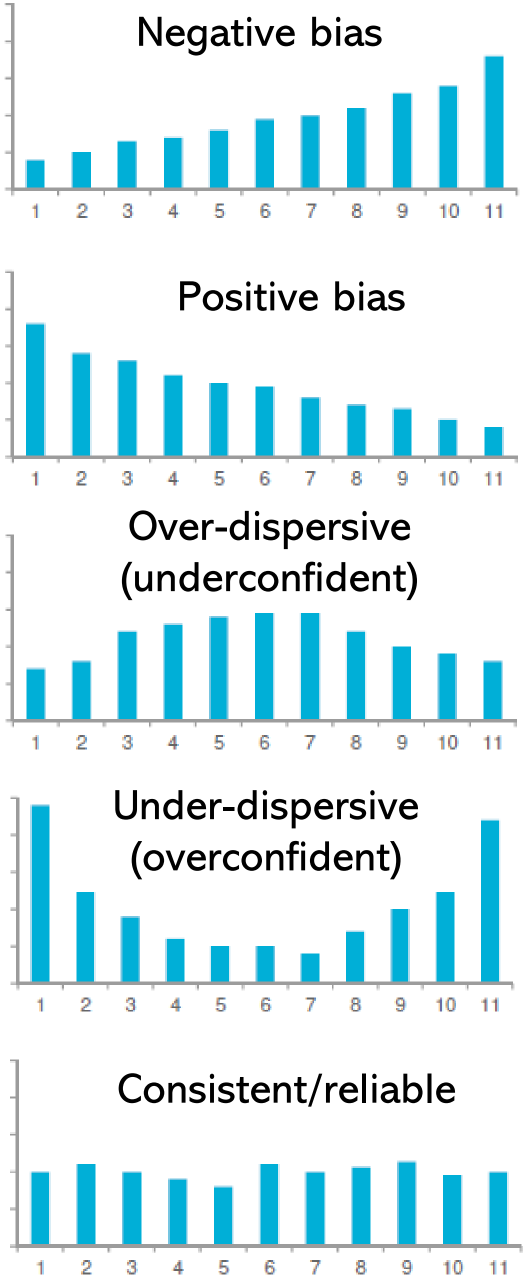
Rank histograms

Consistency w.r.t. verification (i.e. independent) data

- **Principle**: for each observation, sort the N ensemble members from lowest to highest, and identify the rank of observation in the sorted ensemble
- **Interpretation**: a well-calibrated ensemble leads to a uniform rank histogram - deviations from uniformity indicate miscalibration



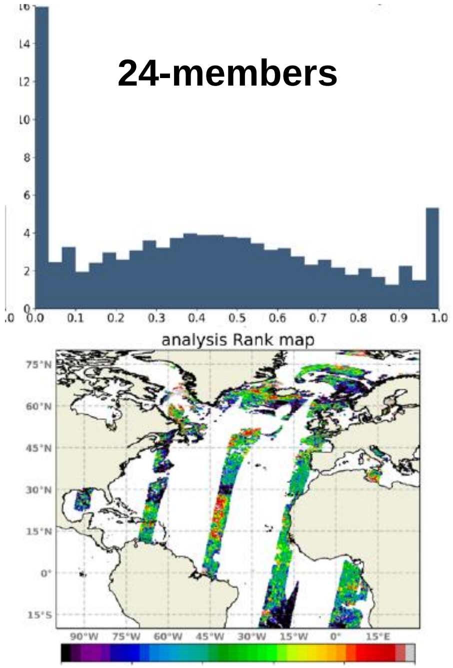


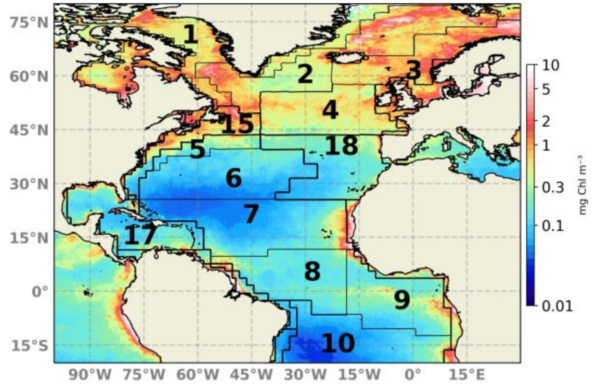


Rank histograms

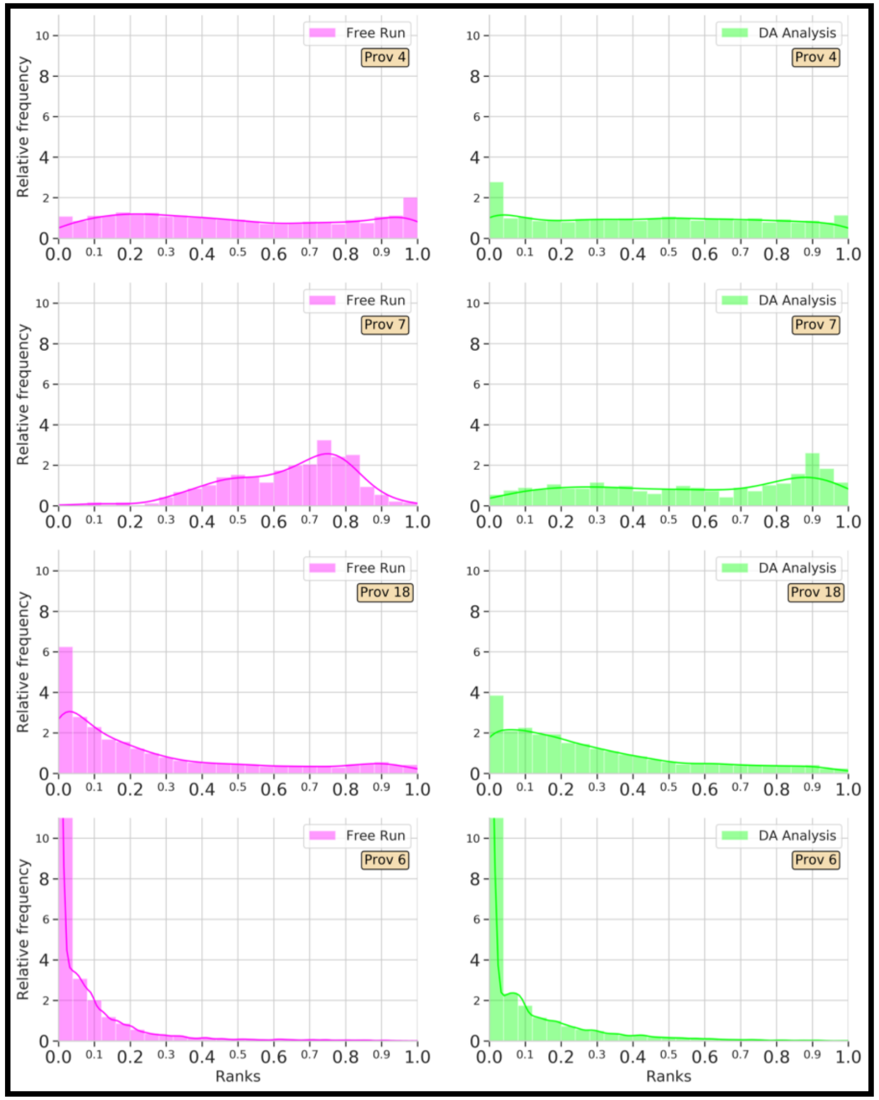
Consistency w.r.t. verification (i.e. independent) data

- **Principle**: for each observation, sort the N ensemble members from lowest to highest, and identify the rank of observation in the sorted ensemble
- **Interpretation**: a well-calibrated ensemble leads to a uniform rank histogram - deviations from uniformity indicate miscalibration





Longhurst Provinces in the North Atlantic



Assimilation impact assessment

(Santana-Falcon et al., 2020)

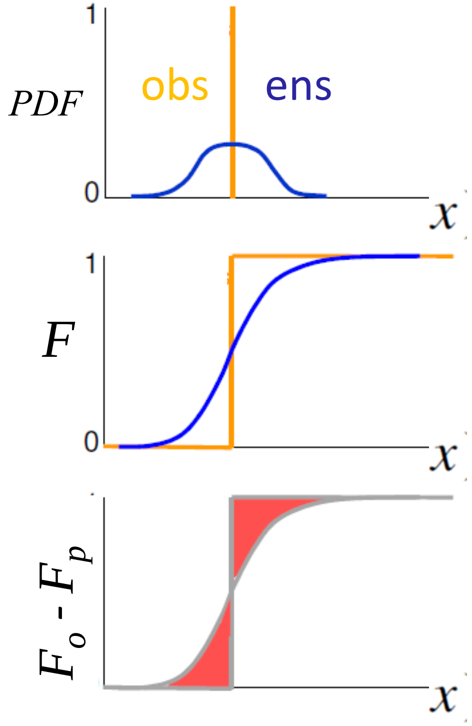
CRPS, reliability, resolution

Global skill of probabilistic system

$$\mathrm{CRPS} = E\left[\int_{\mathbb{R}} \left(F_{\mathrm{p}}(x) - F_o(x)\right)^2 \mathrm{d}x\right]$$

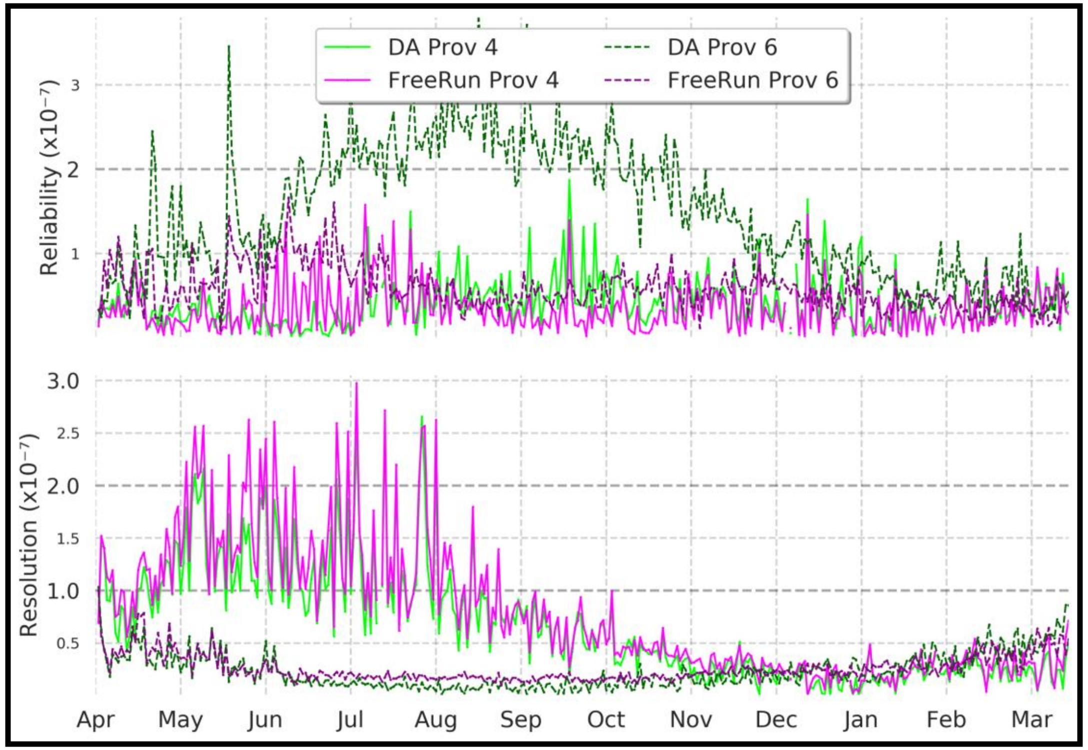
$$= \mathrm{Reli} + \mathrm{Resol}$$

based on Hersbach (2000)



A skillful probabilistic system must satisfy two criteria:

- to be reliable (*Reli close to 0)*
- F_p to be as sharp as possible compared to F_o



Assimilation impact assessment

(Santana-Falcon et al., 2020)

Entropy-based skill score

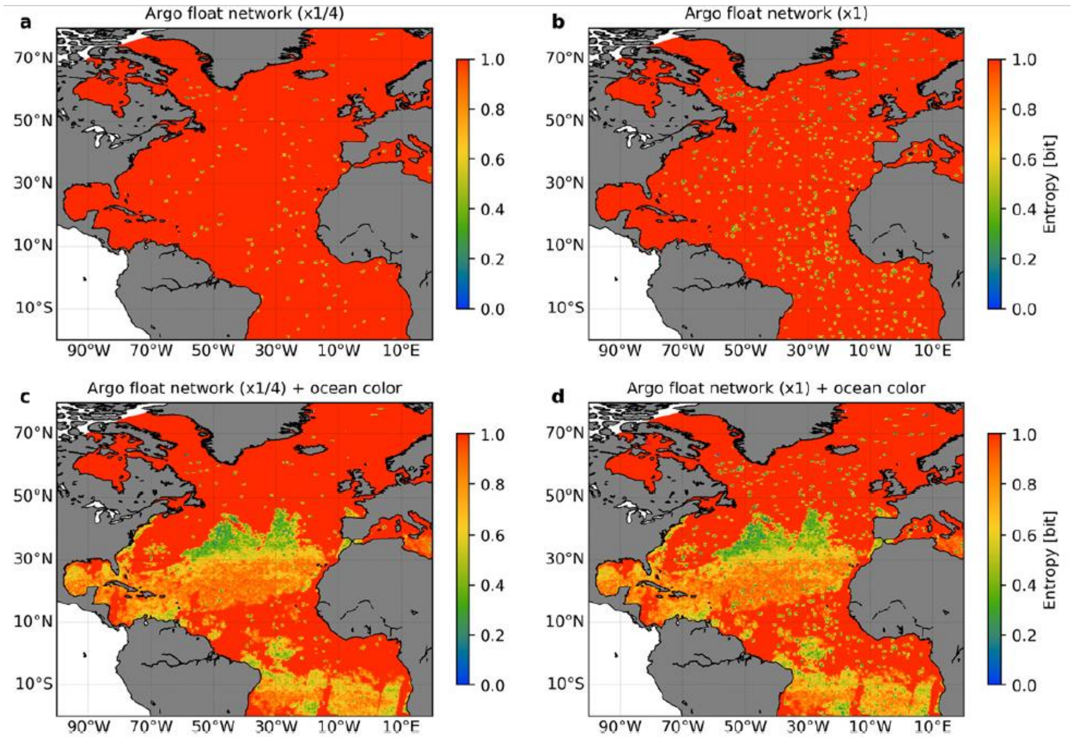
Comparison between systems (true, estimated)

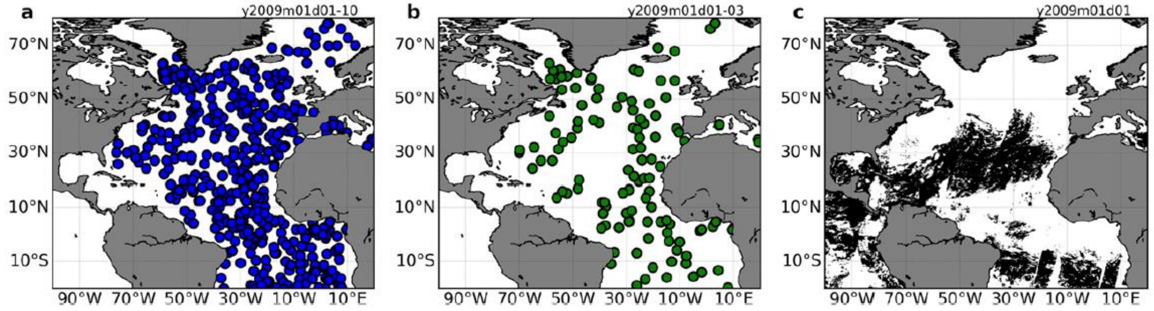
- **Entropy** H: only measures the information content of the system (closely related to the resolution as evaluated through the Brier score)
- ***Ignorence*** score: developed for ensemble-based OSSEs that consider a PDF of true states $\mathbf{p}$ (rather than a single true), using cross-validation
- **Normalized IGN** defined in [0, 1] with $\mathbf{q}$ being the PDF of the updated ensemble:

$$\mathrm{IGN}_n = \frac{H(\mathbf{p})}{H(\mathbf{p},\mathbf{q})}$$

nature-run entropy

cross-entropy



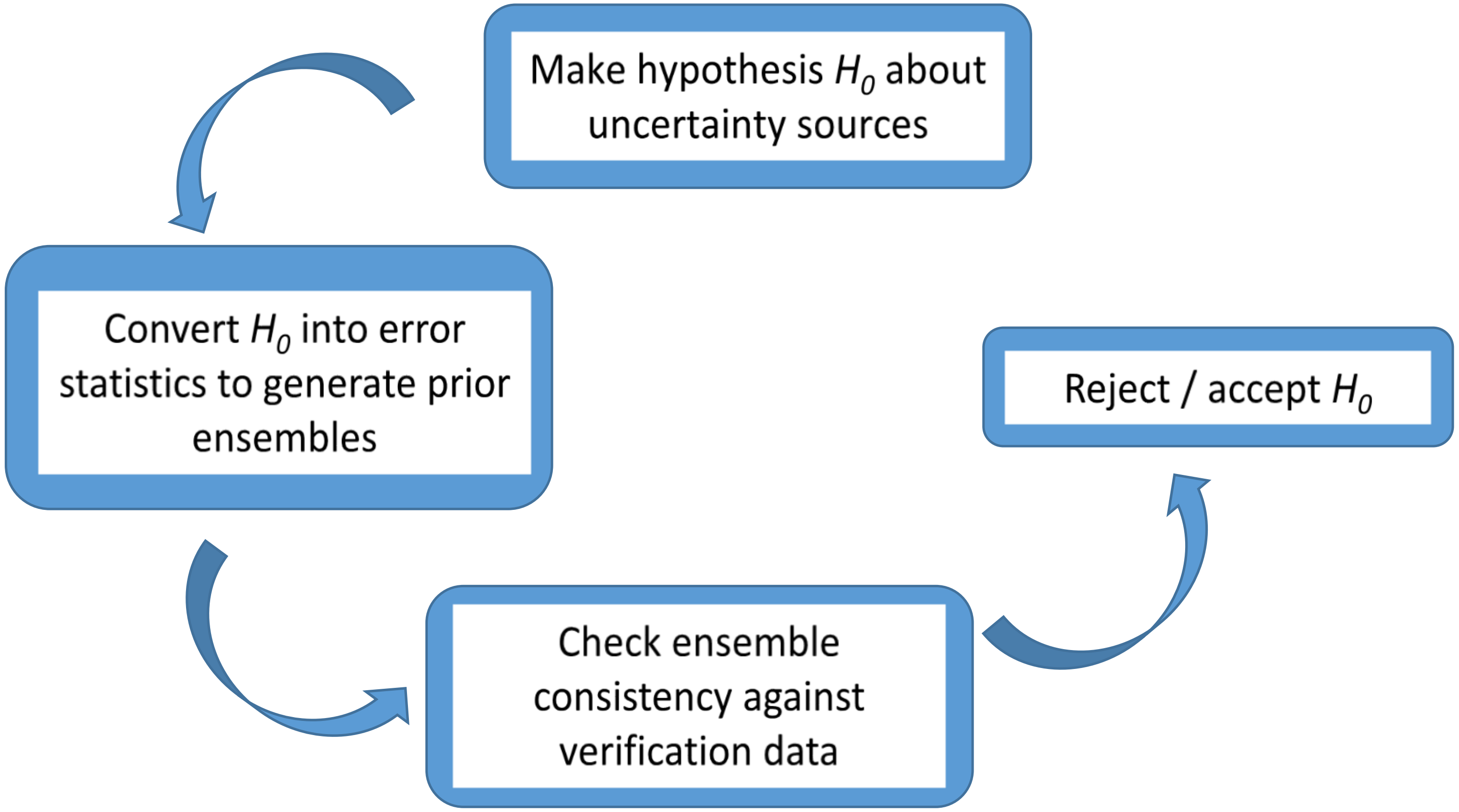


BGC ARGO + satellite OC scenarii

(Germineaud et al., 2020)

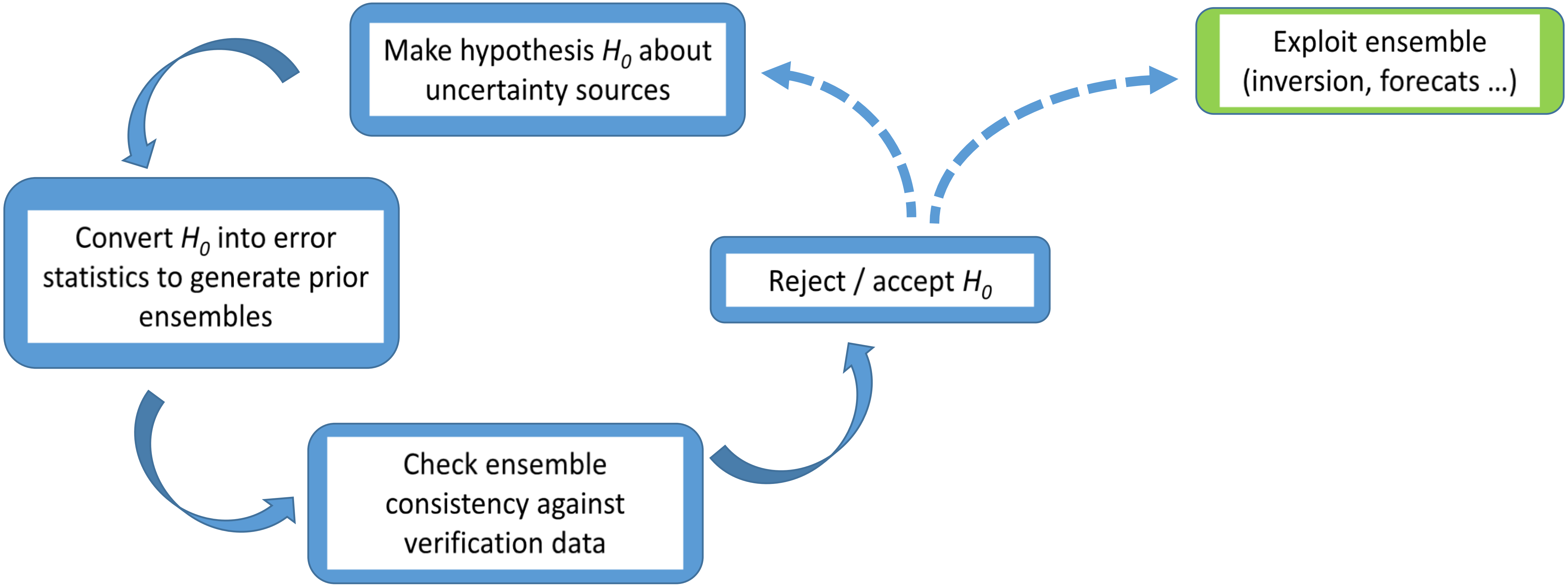
Best-practices for ensemble generation and use in Data Assimilation / Inverse problems ?

To open the discussion !



Best-practices for ensemble generation and use in Data Assimilation / Inverse problems ?

To open the discussion !



- ***Why do we need ensembles ?***
  Uncertainties in a chaotic ocean

- ***How to generate ensemble of 3D states / trajectories***
  Some examples using NEMO-PISCES

- ***Ensemble verification***
  Empirical metrics and skill scores

- **Challenges for the future**
  To open the discussion

Challenges for the future

Furthering the discussion

- How to avoid capturing all uncertainty sources to get consistent ensembles ?
- How to ballance ensemble size, model complexity, resolution ….
- Design of integrated systems driven by what has to be delivered
- Parameterization of statistical model involved in stochastic parameterizations ?

References

- Bessières, L., Leroux, S., Brankart, J.-M., Molines, J.-M., Moine, M.-P., Bouttier, P.-A., Penduff, T., Terray, L., Barnier, B., and Sérazin, G., 2017: Development of a probabilistic ocean modelling system based on NEMO 3.5: application at eddying resolution, Geosci. Model Dev., 10, 1091-1106.
- Brankart J.-M., Cosme E., Testut C.-E., Brasseur P. and Verron J., 2011 : Efficient local error parameterizations for square root or ensemble Kalman filters : application to a basin-scale ocean turbulent flow, Mon. Weather Rev., 139(2), 474-493.
- Brankart J.-M., Candille G., Garnier F., Calone Ch., Melet A., Bouttier P.-A., Brasseur P. and Verron J., 2015: A generic approach to explicit simulation of uncertainty in the NEMO ocean model, Geophysical Model Development, 8, 1285–1297.
- 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.
- Carmillet V., Brankart J.M., Brasseur P., Drange H., Evensen G. and Verron J., 2001 : A Singular Evolutive Extended Kalman filter to assimilate ocean colour data in a coupled physical-biochemical model of the North Atlantic ocean, Ocean Modelling, 3, 167-192.
- Candille, G., Brankart, J. M., and Brasseur, P., 2015: Assessment of an ensemble system that assimilates Jason-1/Envisat altimeter data in a probabilistic model of the North Atlantic ocean circulation, Ocean Sci, 11, 425–438, doi:10.5194/os-11-425-2015.
- Doron M., Brasseur P., Brankart J.-M., Losa S. and Melet A., 2013: Stochastic estimation of biogeochemical parameters from Globcolour ocean colour satellite data in a North Atlantic 3D coupled physical-biogeochemical model, J. Marine Syst., 117–118, 81–95
- Garnier F., Brankart J.-M., Brasseur P. and Cosme E., 2016: Stochastic parameterizations of biogeochemical uncertainties in a 1/4° NEMO/PISCES model for probabilistic comparisons with ocean color data, J. Mar. Systems, 155, 59-72.
- Gehlen, M. Berthet, S., Séférian, R., Ethé, C., and T. Penduff, 2020: Quantification of Chaotic Intrinsic Variability of sea-air CO2 Fluxes at Interannual Timescales. Geophysical Research Letters. https://doi.org/10.1029/2020GL088304.
- Germineaud C., Brankart J.M. and Brasseur P., 2019: An ensemble-based probabilistic score approach to compare observation scenarios: an application to biogeochemical-Argo deployments, J. Atmos. Oceanic Technol., https://doi.org/10.1175/JTECH-D-19-0002.1

References

- Leroux S., Brankart J.M., Albert A., Molines J.M., Brodeau L., Penduff T., Le Sommer J., and Brasseur P., 2021: Ensemble quantification of short-term predictability of the ocean fine-scale dynamics, in preparation.
- McWilliams, J.C., 2007. Irreducible imprecision in atmospheric and oceanic simulations, *PNAS*, 104 (21) 8709-8713.
- Meier H, Edman M, Eilola K, Placke M, Neumann T, Andersson HC, Brunnabend S-E, Dieterich C, Frauen C, Friedland R, Gröger M, Gustafsson BG, Gustafsson E, Isaev A, Kniebusch M, Kuznetsov I, Müller-Karulis B, Naumann M, Omstedt A, Ryabchenko V, Saraiva S and Savchuk OP (2019) Assessment of Uncertainties in Scenario Simulations of Biogeochemical Cycles in the Baltic Sea. Front. Mar. Sci. 6:46.
- Penduff, T., G. Sérazin, S. Leroux, S. Close, J.-M. Molines, B. Barnier, L. Bessières, L. Terray, and G. Maze., 2018. Chaotic variability of ocean heat content: Climate-relevant features and observational implications. Oceanography 31(2).
- Popova E., M.J.R. Fasham, A.V.Osipov and V.A. Ryabchenko, 1997. Chaotic behaviour of an ocean ecosystem model under seasonal external forcing. J. Plankton Res., 19(10), 1495-1515.
- Santana-Falcon Y., Brasseur P., Brankart J.M. and Garnier F., 2020: Assimilation of chlorophyll data into a stochastic ensemble simulation for the North Atlantic ocean, Ocean Sci., 16, 1297–1315, 2020, https://doi.org/10.5194/os-16-1297-2020.
- Séférian R, Bopp L, Gehlen M, Swingedouw D, Mignot J, Guilyardi E, Servonnat J. Multiyear predictability of tropical marine productivity. Proc Natl Acad Sci U S A. 2014 Aug 12;111(32):11646-51
- Tracton S. and Kalnay E., 1993. Operational ensemble prediction at the NMC: Practical aspects. *Weather and Foecasting*, Vo.l 8(3), 379-398.
- Vervatis, V.D., De Mey-Frémaux, P., Ayoub, N., Karagiorgos, J., Ghantous, M., Kailas, M., Testut, C.-E., and Sofianos, S., 2021. Assessment of a regional physical-biogeochemical stochastic ocean model. Part 1: ensemble generation, *Ocean Modell*., 160, 101781, https://doi.org/10.1016/j.ocemod.2021.101781
- Vervatis, V.D., De Mey-Frémaux, P., Ayoub, N., Karagiorgos, J., Ciavatta, S., Brewin, R., Sofianos, S., 2021. Assessment of a regional physical-biogeochemical stochastic ocean model. Part 2: empirical consistency. *Ocean Modell*. 160 (4), 101770. http://dx.doi.org/10.1016/j.ocemod.2021.101770.

Additional slides

Autoregressive processes (1)

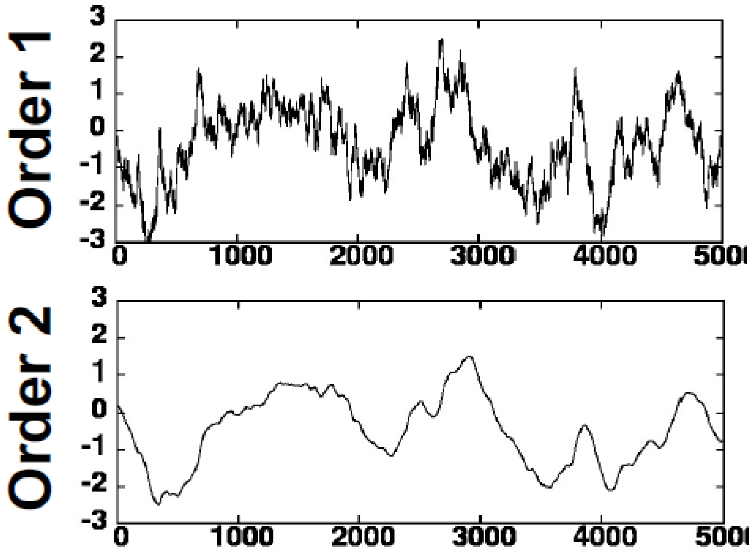
At every model grid point (in 2D or 3D), generate a set of **independent Gaussian autoregressive processes:**

$$\xi(t_k) = a\,\xi(t_{k-1}) + b\,w + c$$

where ***w*** is a Gaussian white noise (→ order 1 process) or an autoregressive process of order n-1 (→ order n process)

Parameters ***a, b, c*** to specify:

mean, standard deviation and correlation timescale



Final meeting (Brankart 2015)

Autoregressive processes (2)

Introduce a spatial correlation structure

by applying a spatial filter to the map of autoregressive processes:

$\tilde{\boldsymbol{\xi}} = \mathcal{F}[\boldsymbol{\xi}]$ (filtering operator)

$\mathcal{L}[\tilde{\boldsymbol{\xi}}] = \boldsymbol{\xi}$ (elliptic equation)

which can easily be made flow dependent if needed

Modify the marginal probability distributions

by applying anamorphosis transformation to every individual Gaussian variable:

$\tilde{\boldsymbol{\xi}} = \mathcal{T}[\boldsymbol{\xi}]$ (nonlinear function)

for instance to transform the Gaussian variables into lognormal or gamma variables if positive noise is needed

→ This provides a generic technical way of implementing a wide range of stochastic parameterizations

Final meeting
(Brankart 2015)

Technolgical approach: a new module in NEMO

These processes are generated using a **new module in NEMO**,
and **can be used in any component** of the model:
circulation model, ecosystem model, sea ice model

Algorithm 1 sto_par

for all (map $i = 1, \ldots, m$ of autoregressive processes) **do**
  Save map from previous time step: $\xi_- \leftarrow \xi_i$
  if (process order is equal to 1) **then**
    Draw new map of random numbers w from $\mathcal{N}(0,1)$: $\xi_i \leftarrow w$
    Apply spatial filtering operator $\mathcal{F}_i$ to ξ_i: $\xi_i \leftarrow \mathcal{F}_i[\xi_i]$
    Apply precomputed factor f_i to keep SD equal to 1: $\xi_i \leftarrow f_i \times \xi_i$
  else
    Use previous process (one order lower) instead of white noise: $\xi_i \leftarrow \xi_{i-1}$
  end if
  Multiply by parameter b_i and add parameter c_i: $\xi_i \leftarrow b_i \times \xi_i + c_i$
  Update map of autoregressive processes: $\xi_i \leftarrow a_i \times \xi_- + \xi_i$
end for

Algorithm 2 sto_par_init

Initialize number of maps of autoregressive processes to 0: $m \leftarrow 0$
for all (stochastic parameterization $k = 1, \ldots, p$) **do**
  Set m_k, the number of maps of autoregressive processes required for this parameterization
  Increase m by m_k times the process order o_k: $m \leftarrow m + m_k \times o_k$
end for
for all (map $i = 1, \ldots, m$ of autoregressive processes) **do**
  Set order of autoregressive processes
  Set mean (μ_i), standard deviation (σ_i) and correlation timescale (τ_i) of autoregressive processes
  Compute parameters a_i, b_i, c_i as a function of μ_i, σ_i, τ_i
  Define filtering operator $\mathcal{F}_i$
  Compute factor f_i as a function of $\mathcal{F}_i$
end for
Initialize seeds for random number generator
for all (map $i = 1, \ldots, m$ of autoregressive processes) **do**
  Draw new map of random numbers w from $\mathcal{N}(0,1)$: $\xi_i \leftarrow w$
  Apply spatial filtering operator $\mathcal{F}_i$ to ξ_i: $\xi_i \leftarrow \mathcal{F}_i[\xi_i]$
  Apply precomputed factor f_i to keep standard deviation equal to 1: $\xi_i \leftarrow f_i \times \xi_i$
  Initialize autoregressive processes to $\mu + \sigma \times w$: $\xi_i \leftarrow \mu + \sigma \xi_i$
end for
if (restart file) **then**
  Read maps of autoregressive processes and seeds for the random number generator form restart file (thus overriding the initial seed)
end if

→ Generic and flexible technological approach
→ Model independent implementation
→ Possible to simulate many kinds of uncertainty

SANGOMA

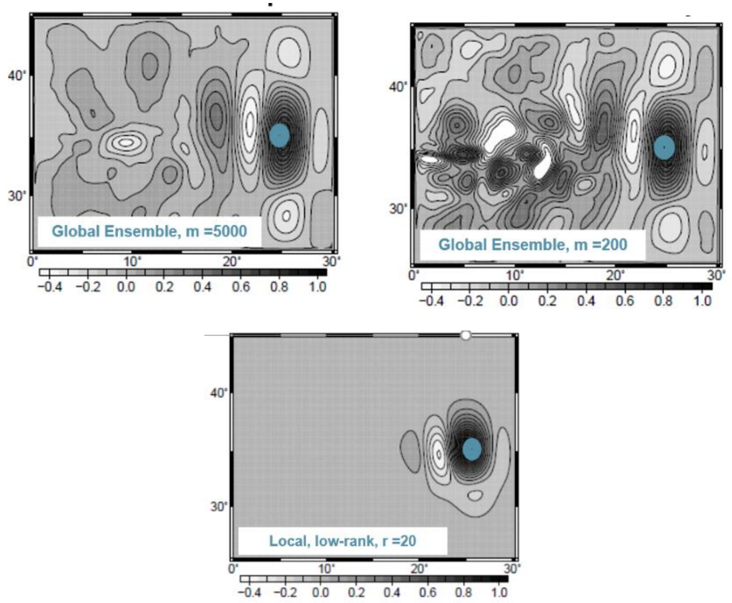
Final meeting
(Brankart 2015)

Limited ensemble size for very high dimensional systems

Strategies

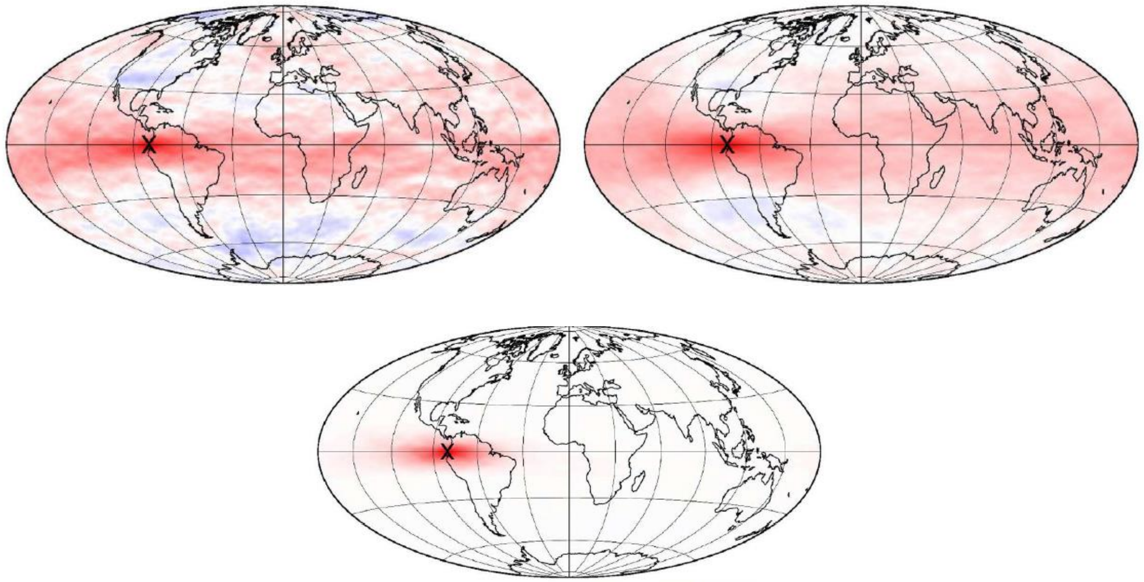
- additional constraints to reduce the number of dof (e.g. multivariate relationships)
- specific procedures to augment artificially the ensemble size (e.g. localization)

Correlations in double-gyre model



Local-support localization
(Brankart et al., 2011)

Correlations on the sphere



Localization by Schur products with large-scale patterns
(Brankart 2019)