Ensemble-variational assimilation with NEMOVAR Part 2: experiments with the ECMWF system

Toulouse, 20/06/2017

Marcin Chrust[1], Hao Zuo[1] and Anthony Weaver[2]

[1] ECMWF, UK
[2] CERFACS, FR

Marcin.chrust@ecmwf.int

ECMWF

Outline

- Ocean Data Assimilation at ECMWF
- Ensemble Generation
- Diagnostics
- Experiments
- Future directions

NEMOVAR

- NEMOVAR is developed by a consortium of ECMWF, Met Office, CERFACS and INRIA;
- NEMOVAR is an incremental variational data assimilation system for the NEMO ocean model
  - Solves a linearized version of the full non-linear cost function;
  - Incremental 3D-Var FGAT running operational, 4D-Var working on research model;

$$J[\delta \mathbf{w}] = \frac{1}{2}\delta \mathbf{w}^{\mathrm{T}} \mathbf{B}^{-1} \delta \mathbf{w} + \frac{1}{2}(\mathbf{G}\,\delta \mathbf{w} - \delta \mathbf{y}^{\mathrm{o}})^{\mathrm{T}} \mathbf{R}^{-1} (\mathbf{G}\,\delta \mathbf{w} - \delta \mathbf{y}^{\mathrm{o}})$$

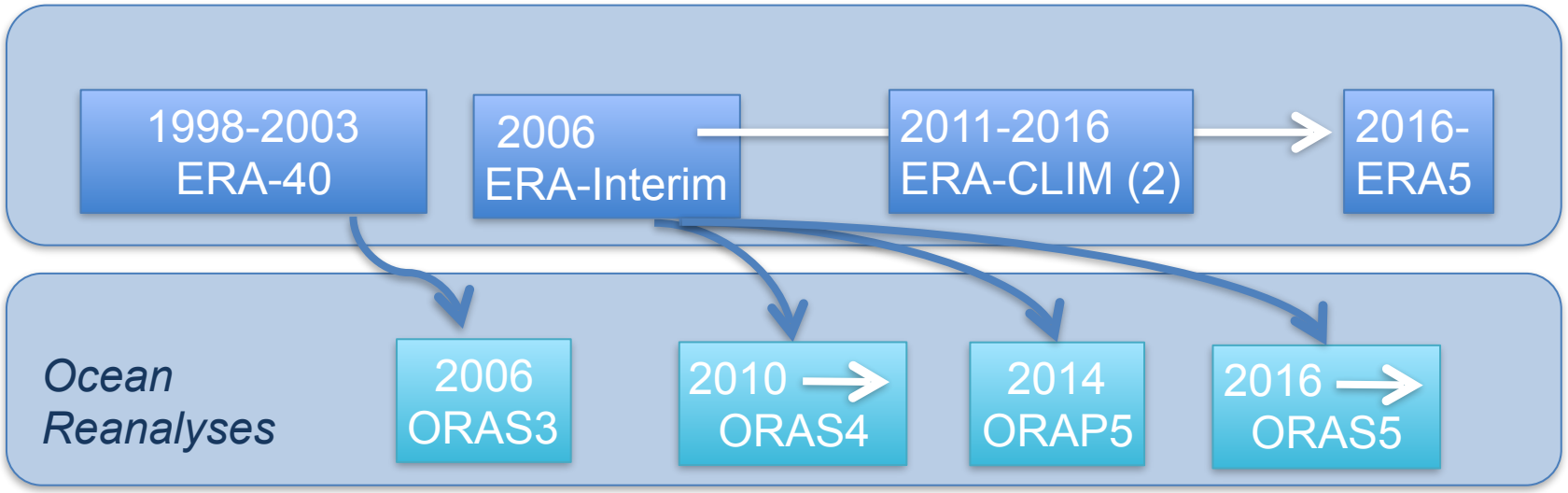
$$\mathbf{G} = \begin{pmatrix} \vdots \\ \mathbf{G}_i \\ \vdots \end{pmatrix} = \begin{pmatrix} \vdots \\ \mathbf{H}_i \mathbf{M}(t_i, t_0) \mathbf{K} \\ \vdots \end{pmatrix} \approx \begin{pmatrix} \vdots \\ \partial G_i / \partial \mathbf{w}|_{\mathbf{w}=\mathbf{w}^{\mathrm{b}}} \\ \vdots \end{pmatrix} \qquad \mathbf{G}^{\mathrm{T}} = (\cdots\ \mathbf{G}_i^{\mathrm{T}}\ \cdots) = (\cdots\ \mathbf{K}^{\mathrm{T}}\, \mathbf{M}(t_i, t_0)^{\mathrm{T}}\, \mathbf{H}_i^{\mathrm{T}}\ \cdots)$$

- Background errors are correlated between different variables through balance operator;

Summary of ECMWF Ocean (Re)analysis

The main purpose of the ocean (Re)analysis at ECMWF

- to provide initial conditions for coupled model forecasts;
- to provide initial conditions for the extended range forecasts (seasonal and monthly);
- ocean reanalyses are valuable resources for climate monitoring, climate variability studies, and verification of climate models;



ECMWF

Slides from M. Balmaseda

	ORAS4	ORAP5 (2015)	ORAS5 (2016)
Resolution	~1deg Vertical ~10m in upper 200m	~0.25 deg Vertical 75 levels	~0.25 deg Vertical ~ 1-10m in upper 200m
Model	NEMO V3.0. Prescribed Sea Ice	NEMO V3.4. LIM2 ice model +Wave effects	+ TKE mixing in partial ice cover + New z(lat) for wind enhanced mixing
Data assimilation	NEMOVAR V3.0 10 days assimilation window	NEMOVAR V3.4 Revised de-correlation scales 5 days assim window + Sea Ice assimilation	+Stability check for bias correction +modified QC +decreased deep ocean bck err +Change in vertical interpolation (spline)
Surface forcing	E4(<1989) EI (< 2010)- Ops Flux forcing	EI (1979-2013) Bulk formula +wave effects	E4 (<1979) EI (<2015) OPS Bulk formula + wave effects
Observations	T/S from EN3 XBT corr Wijffels et al 2008 Altimeter v3-v4-v5 SST E4-OIv2-OPS (0)	T/S from EN3 XBT corr Wijffels et al (2008) Altimeter: DUACS 2010 SST and SIC from OSTIA reanalyses	T/S from EN4 XBT corr. Gouretsky et al (2013) Altimeter: DUACS 2014 (new MDT) SST from HadISST SIC from OSTIA(1)
Time Span	1958-present	1979-2013	1979-present (back extend to 1955)
Ensembles	5 members. Pert: Ini: time sampling Stress: E4-NCEP Obs coverage	None	5 members. Pert: Ini: Different assim valid at 1975 New Forcing Pert (stress, sst, sic, emp, solar) Obs Perturbations

ORAS5 operational Implementation

The operational ORAS5 comprises NRT suite and BRT suite with a delay between two suites defined by $**NRTOFFSET = N**

- Behind Real Time (**BRT**) – run every 5 day with a delay of **N+1** days (BRT delay = Current day – last day in the BRT window) and provide initial condition for NRT system.
- Near Real Time (**NRT**) – run daily with variable assimilation window (**N+2** to **N+6**) in two chunks: first 5-day chunk initialized from delayed BRT system and second chunk with variable length, where the last day is effectively a forecast (with FC forcing fields and SST/SIC from previous day).

This operational ORAS5 task is carried out at **14Z** everyday and produce restart file from NRT system that will be used as ocean initial condition valid on **0Z** the following morning for the coupled forecasting system.

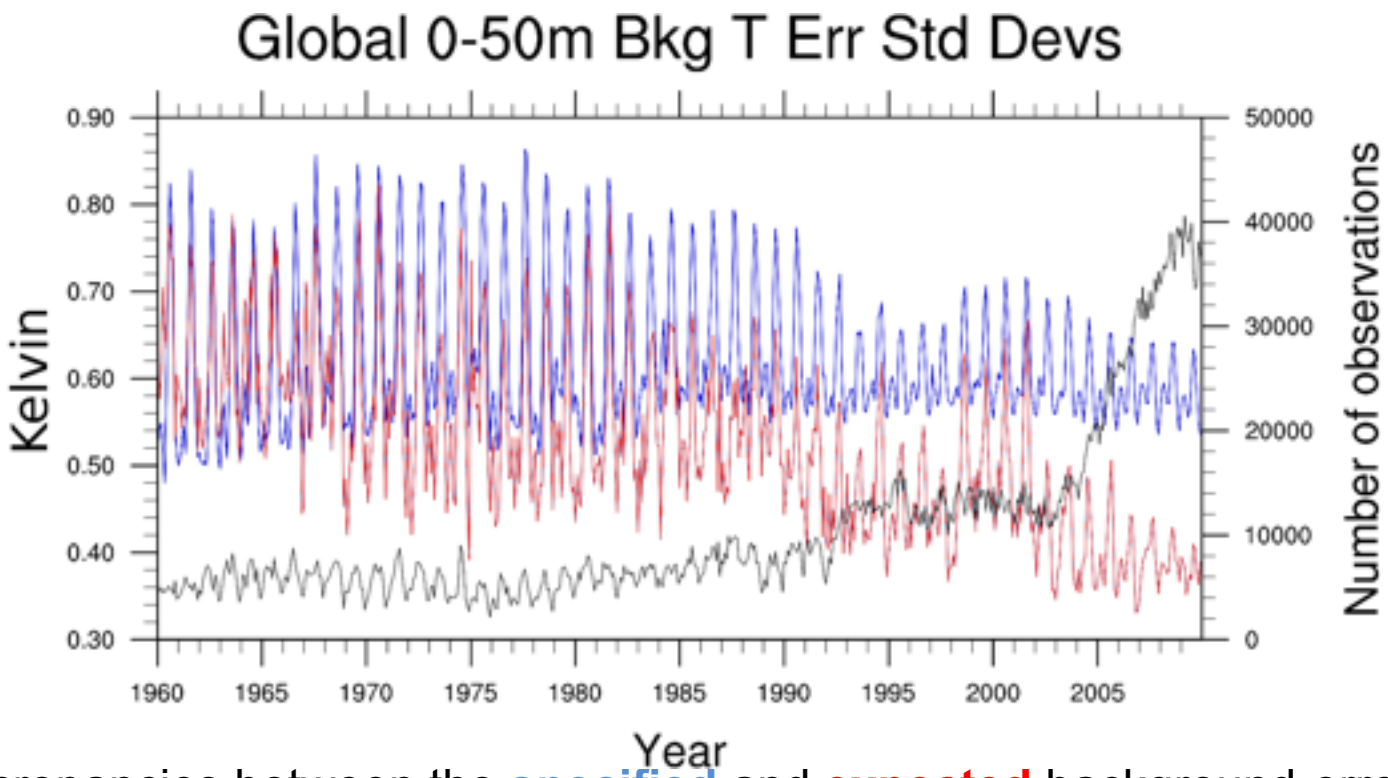
Output stream

	Stream-1	Stream-2	Stream-3
	3D daily	3D 5D-mean	2D daily
BRT	YES	YES	YES
NRT	YES	NO	YES

Outline

- Ocean Data Assimilation at ECMWF
- **Ensemble Generation**
- Diagnostics
- Experiments
- Future directions

Ensemble generation – why do we need ensembles?



- Discrepancies between the **specified** and **expected** background-error standard deviations are signal of sub-optimality in the covariance specifications.
- There is evidence of missing flow dependence in B, especially in response to the changing observing system (**black curve**)
- An ensemble-base B should be able to capture this flow dependence

Ensemble generation - observation representativeness error

Errors introduced by the interpolation operator and by the finite resolution of the model fields are often accounted for as **representativeness errors of observations**. It is introduced by degrading physical observations to model resolution. In ORAS5 we have implemented perturbation schemes for ocean in-situ and surface observations to account for their representativeness errors.

- Perturbation of geographical locations of in-situ profiles
- Perturbation of vertical depths of in-situ profiles
- Perturbation of gridded sea ice concentration data
- Perturbation of along-track SLA data

Perturbing geographic location of in-situ observations

A random noise (perturbation) can be added to the recorded geographic location of any given profile from the EN4 dataset. It is equivalent to shifting the observation profile horizontally in map within a pre-defined distance (**P**erturbation **D**istance). The random noise was generated using a random number generator and with probability density function of uniformed distribution. Ocean in-situ profiles in ORAS5 are horizontally perturbed to a random location within a circle, with PD=50 km.

- *P1: perturb the profile with random distance (0 to PD) and angle (0-360 degree)*
- *P2: perturb the profile with constant distance (PD) and random angle (0-360 degree)*
- *P3: perturb latitude of profile with random distance (0 to PD)*
- *P4: perturb longitude of profile with random distance (0 to PD)*
- *P5: perturb both latitude and longitude with the same random distance (0 to PD)*

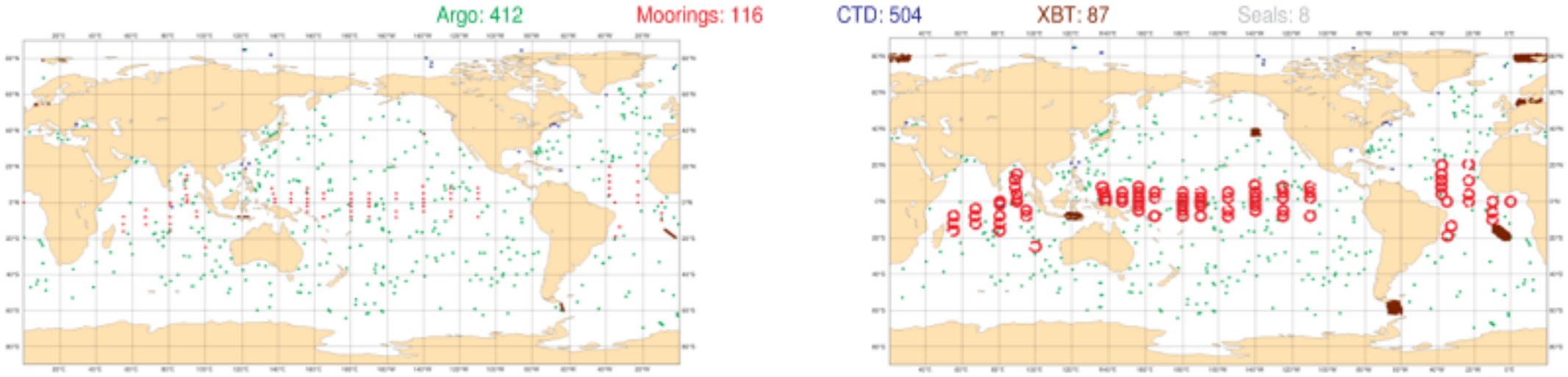
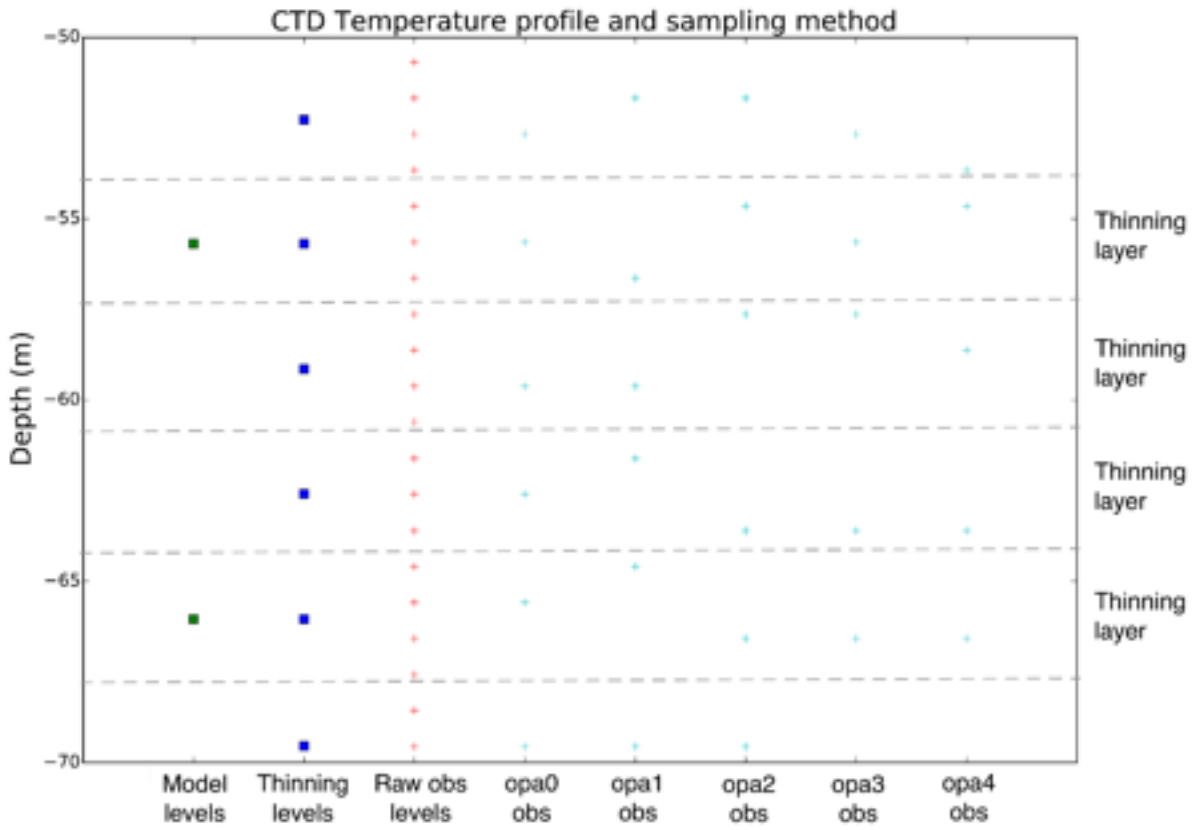


Fig. Geographic locations daily EN4 profiles in 20130101 and perturbed profile with 200 member ensemble; for illustration purpose only XBT and Moorings profiles are prtubed with a large PD value (200 km), using perturbing strategy P2 and P5, respectively. Source: Zuo et al., Tech. Memorandum No. 795

Perturbing vertical location of in-situ observations

Example of in-situ profile vertical perturbation
CTD temperature profile



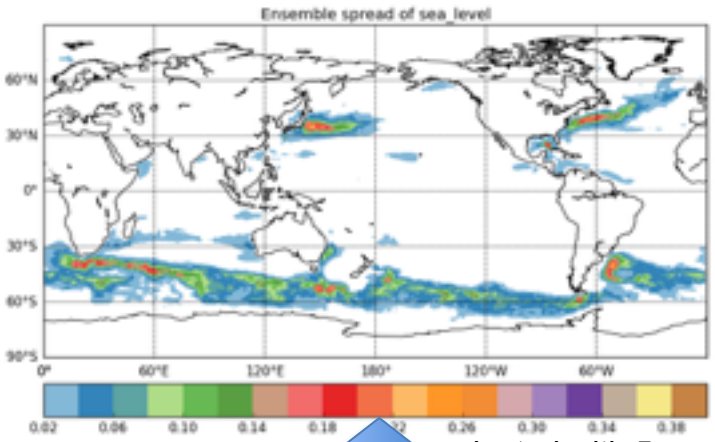
The representativeness error of in-situ observations at depth due to limited model vertical resolution can be represented by a stratified sampling method using pre-defined depth ranges as different sampling groups. In ORAS5, each model level are evenly separated into two thinning level. Contribution to the model uncertainty predominately comes from ocean domains below 1000m where model vertical resolution is reduced significantly

Fig. Example of stratified samling method used for vertical perturbation of in-situ observations, as demonstrated with a CTD profile from EN4 dataset and thinning factor N=3 with 5 ensemble members. Within each thinning layer only one observation, if exists, is selected using simple random sampling method. Source: Zuo et al., Tech. Memorandum No. 795

Perturbing along-track SLA observations

For along-track SLA observations, the super-observation scheme as implemented in ORAS4 and ORAP5 has been modified to include random sampling scheme.

- SLA observations within the same day and within each superob-grid (~100 km resolution) are randomly selected, with the same OBE as calculated using the superobing scheme (see Mogensen et al. 2012)

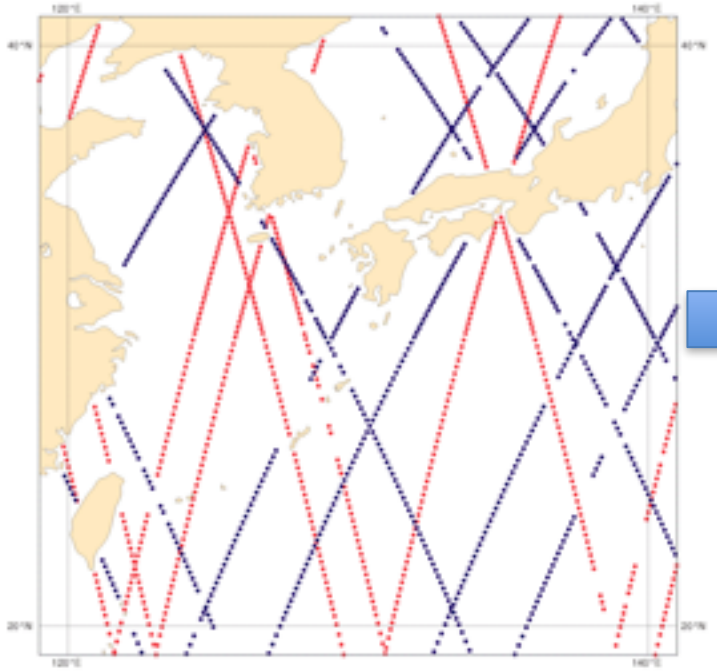


evaluated with 5 ensemble member in ¼ degree reanalysis for 5 yrs

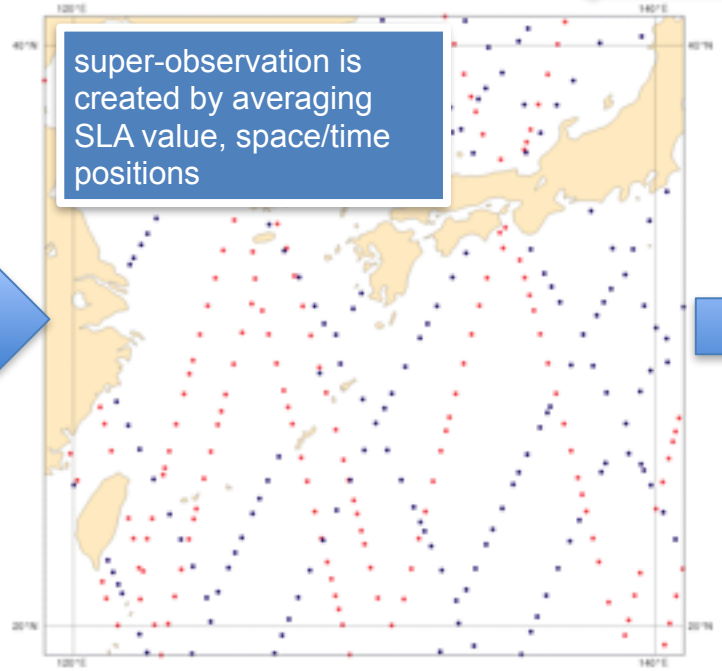
share the same obs number, superob grid and OBE

thinning

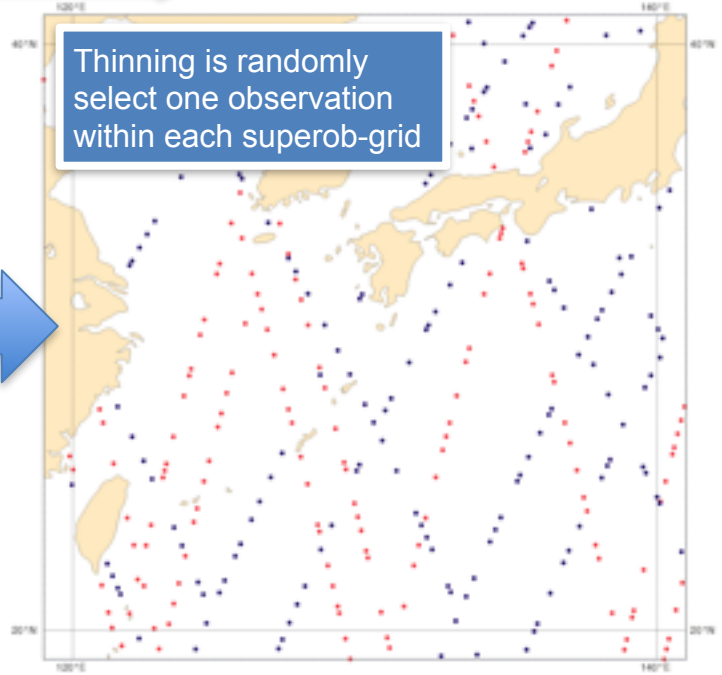
raw



superob



super-observation is created by averaging SLA value, space/time positions



Thinning is randomly select one observation within each superob-grid

ERS-2: 13005 Envisat: 0 Envisat N: 0
Jason-1 N: 0 Jason-2: 0 T/P: 13219

ECMWF

Slides from M. Balmaseda

Ensemble generations – other perturbations in ORAS5

Surface perturbations

Generic procedure for additive perturbations according to 4 criteria:

1. Variables: SST, SIC, P-E, solar, taux, tauy
2. Structural Uncertainty: diff between 2 analysis systems
   SST/SIC: OSTIA-ESACCI
   Stress: EI-OPS, Uncertainty in bulk formulation, E4-NCEP
3. Analysis Uncertainty:
   SST: from HadISST (10 ensemble member)
   P-E, solar, stress: from ERA-20C (5 ensemble member)
4. Temporal resolution: monthly and pentad band-pass filter. Mean removed
   Only monthly in ORAS5. Pentad can be superimposed to inflate ensemble when initializing the forecasts.

Initial condition perturbations

Prepare for use EDA in Ocean data assimilation and Hybrid B for the next version of NEMOVAR

Outline

- Ocean Data Assimilation at ECMWF
- Diagnostics
- **Diagnostics**
- Experiments
- Future directions

Ensemble diagnostics - temperature

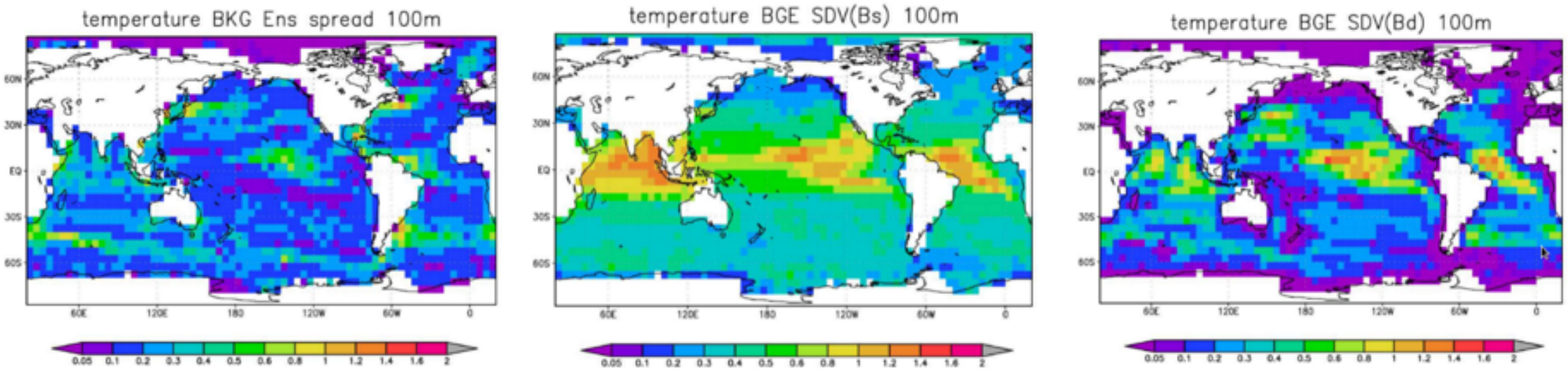


Fig. The ensemble spread (left), specified BGE standard deviations (middle) and diagnosed BGE standard deviations (right) for temperature. Ensemble spread has been computed as the standard deviations of temperature background fields from ORAS5 at 100m, temporally averaged over the 2010-2013 period, and spatially averaged into 5x5 degree grid boxes; source: Zuo et al., Technical Memorandum No. 795.

Ensemble diagnostics - temperature

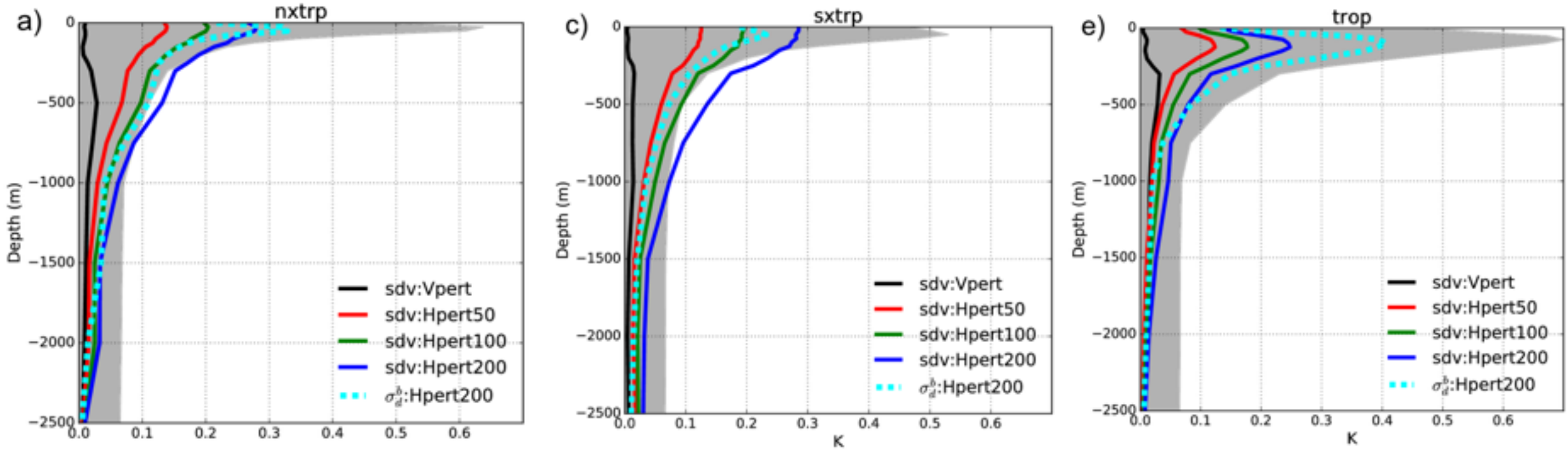


Fig. Vertical profiles of ensemble spread for temperature. Ensemble spreads have been computed as the standard deviations of temperature background fields, temporally averaged over the 2010-2013 period, and spatially averaged over different ocean domains for: northern extratropics (nxtrp: 30N to 70N), southern extratropics (sxtrp: 70S to 30S) and tropics (trop: 30S to 30N). σ_b^d is the diagnosed BGE standard deviation from ORAS5 (Cyan dashed lines) using Desroziers method. σ_s^d is the specified BGE standard deviation using empirical formulation (grey shaded areas) which serves as a reference: Zuo et al., Technical Memorandum No. 795.

Ensemble diagnostics - salinity

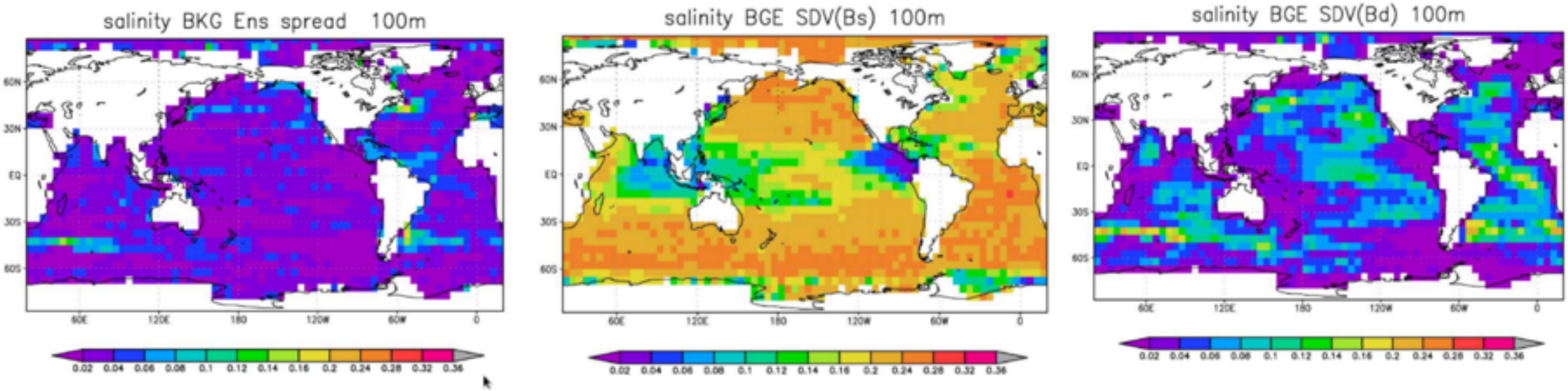


Fig. The ensemble spread (left), specified BGE standard deviations (middle) and diagnosed BGE standard deviations (right) for temperature. Ensemble spread has been computed as the standard deviations of salinity background fields from ORAS5 at 100m, temporally averaged over the 2010-2013 period, and spatially averaged into 5x5 degree grid boxes; source: Zuo et al., Technical Memorandum No. 795.

Ensemble diagnostics - salinity

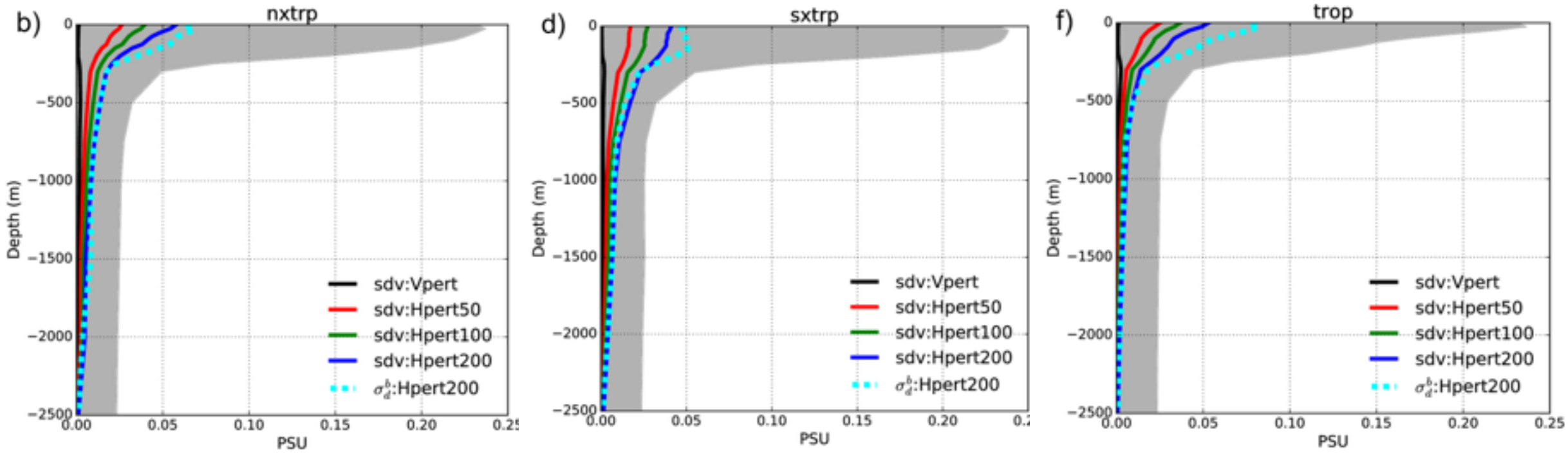


Fig. Vertical profiles of ensemble spread for salinity. Ensemble spreads have been computed as the standard deviations of temperature background fields, temporally averaged over the 2010-2013 period, and spatially averaged over different ocean domains for: northern extratropics (nxtrp: 30N to 70N), southern extratropics (sxtrp: 70S to 30S) and tropics (trop: 30S to 30N). σ_b^d is the diagnosed BGE standard deviation from ORAS5 (Cyan dashed lines) using Desroziers method. σ_s^d is the specified BGE standard deviation using empirical formulation (grey shaded areas) which serves as a reference: Zuo et al., Technical Memorandum No. 795.

Ensemble diagnostics – sensitivity to ensemble size and combined perturbations

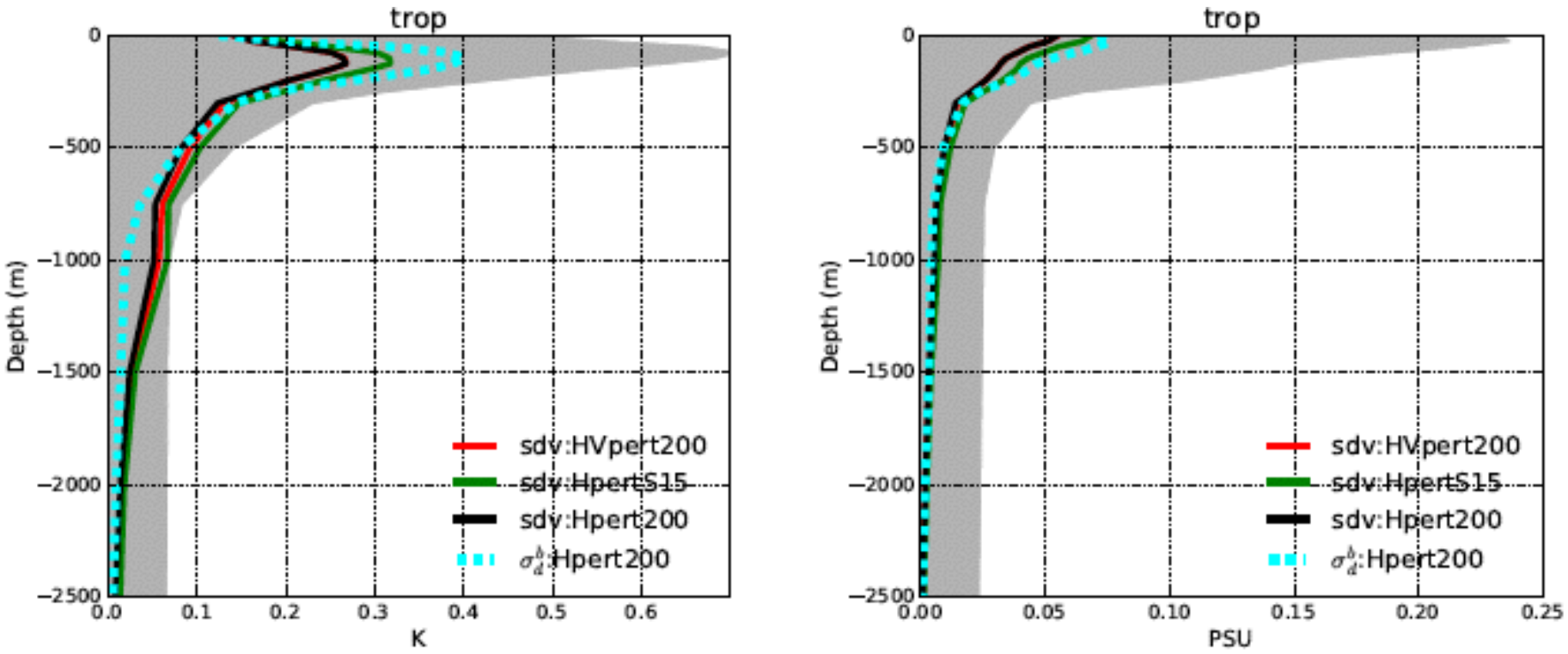


Fig. Vertical profiles of ensemble spread for temperature and salinity. Ensemble spreads have been computed as the standard deviations of temperature background fields, temporally averaged over the 2004-2006 period, and spatially averaged over different ocean domains for tropics (trop: 30S to 30N). σ_b^d is the diagnosed BGE standard deviation from ORAS5 (Cyan dashed lines) using Desroziers method. σ_s^d is the specified BGE standard deviation using empirical formulation (grey shaded areas) which serves as a reference.

Ensemble diagnostics - ensemble reliability

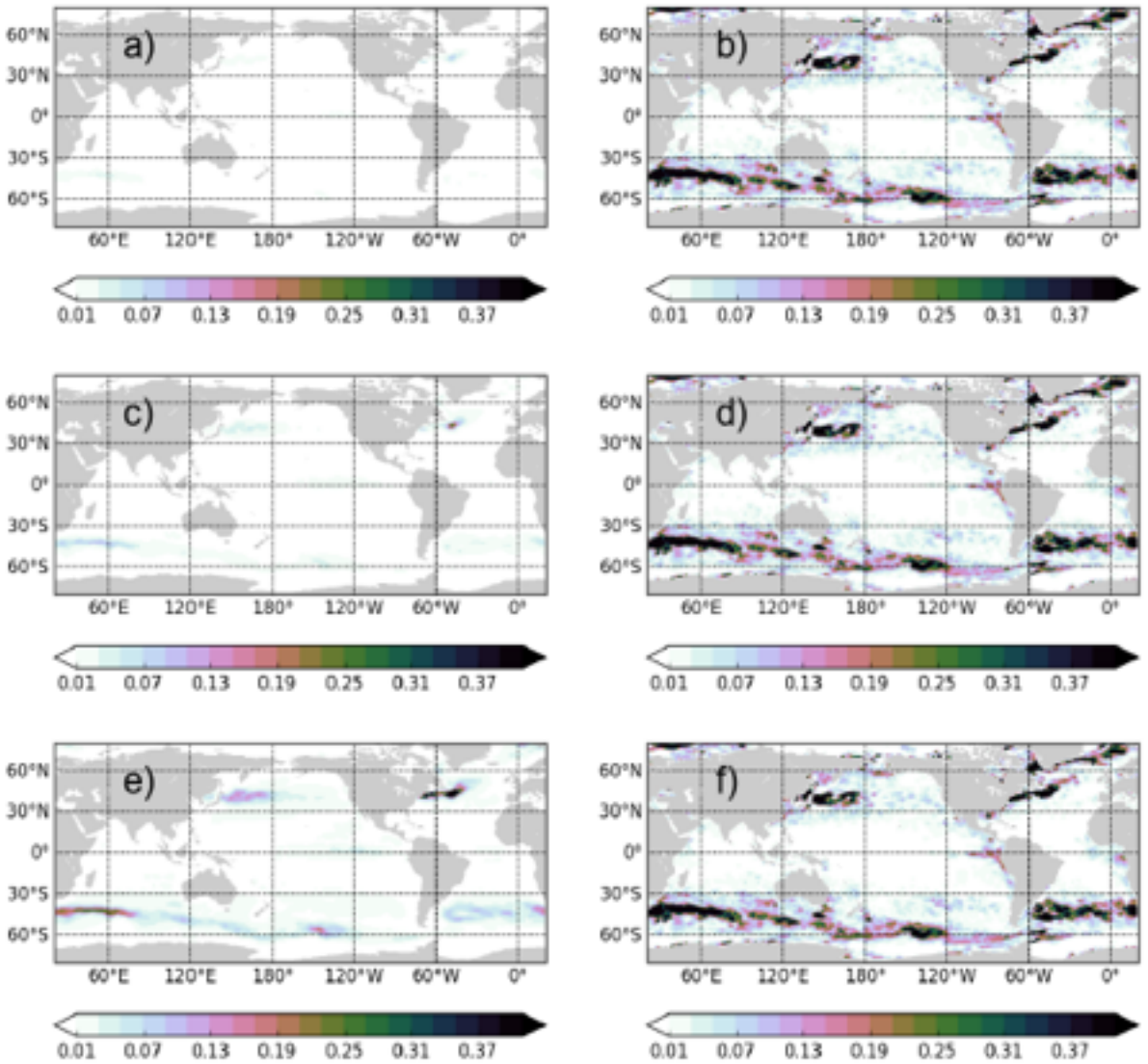


Fig. Geographical distribution of (a,c,e) SST ensemble variances (in K^2) and (b,d,f) squared ensemble mean departure against OSTIA SST for (a,b) Hpert50, (c,d) Hpert100, and (e,f) Hpert200. Both SST ensemble variances and squared errors have been computed using monthly mean SST from three perturbation experiments, and averaged over the 2004-2011 period. Here observation error variances from OSTIA SST have been subtracted from squared ensemble mean errors. Source: Zuo et al, Tech. Memorandum No. 795.

Ensemble diagnostics - ensemble reliability

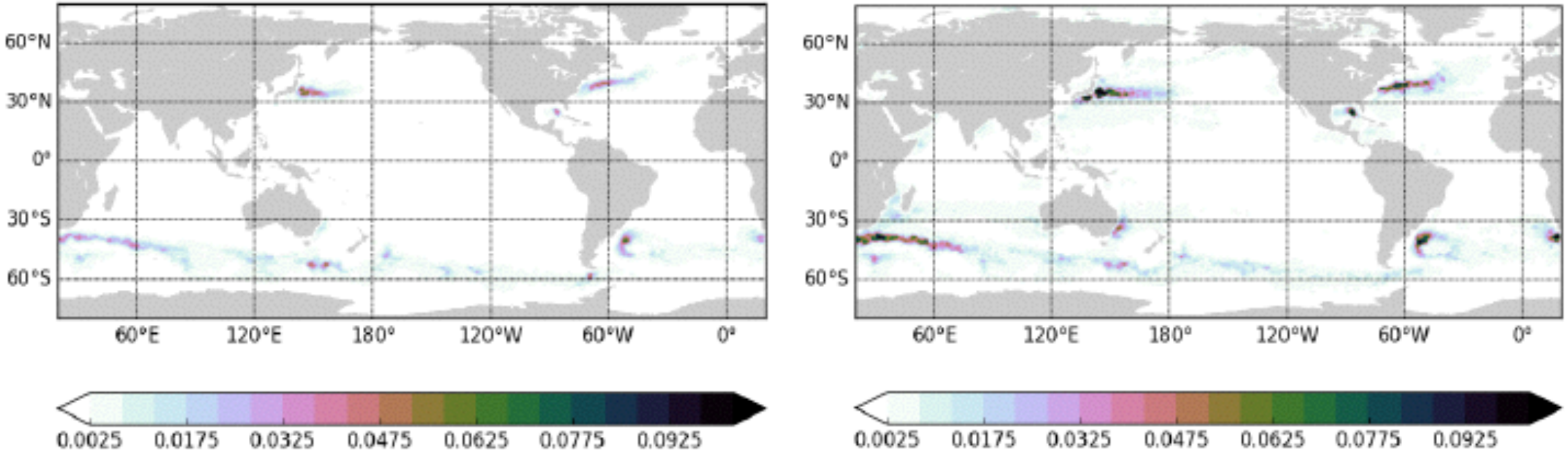


Fig. Geographical distribution of model (top) SLA ensemble variances (in m^2) and (bottom) squared ensemble mean departure against AVISO DUACS2014 MSLA. Both SLA ensemble variances and squared ensemble mean errors have been computed using monthly mean SLA analysis from SLApert, and averaged over the 2000-2004 period. Here squared observation error from AVISO MSLA have been subtracted from squared ensemble mean errors. Source: Zuo et al, Tech. Memorandum No. 795.

Outline

- Ocean Data Assimilation at ECMWF
- Ensemble Generation
- Hybrid Background Covariance
- **Experiments**
- Future directions

Hybrid Variances: preliminary experiments

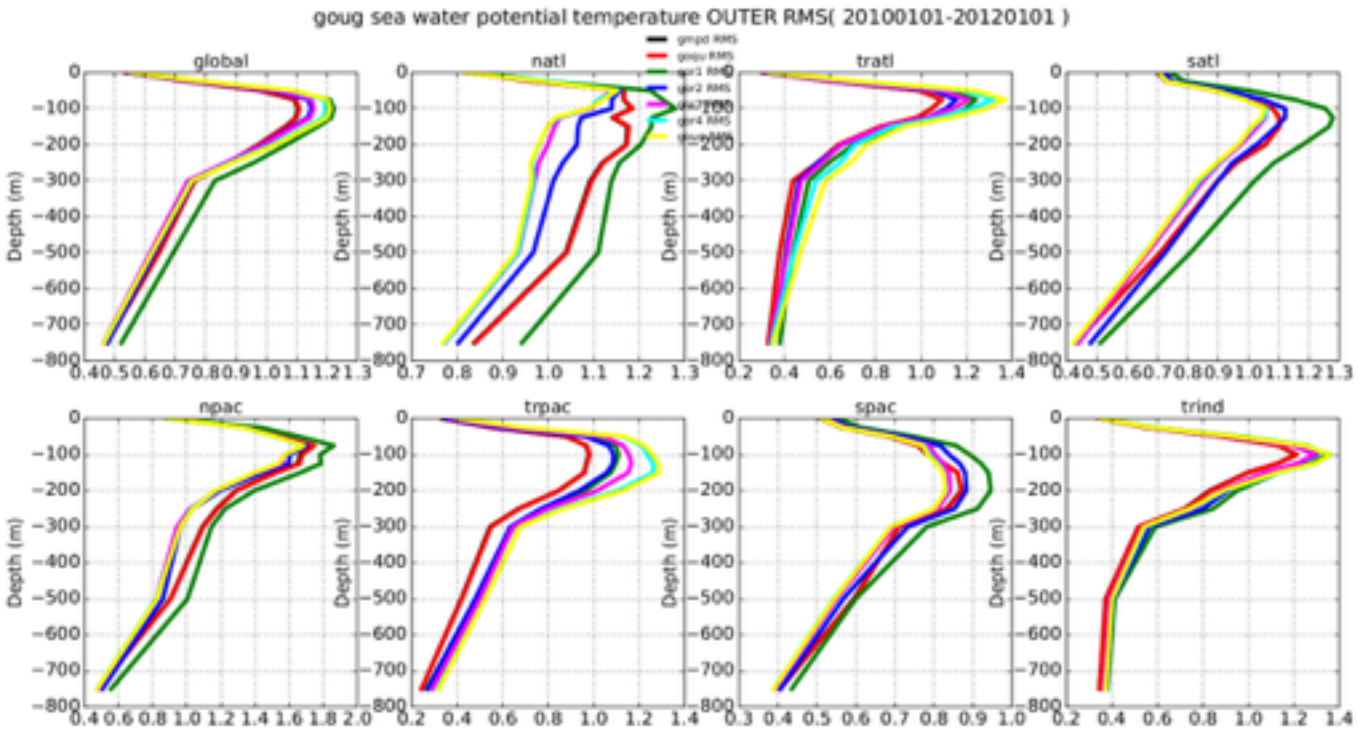


Fig. First experiments with ensemble based background error variances. ORCA1_Z42, 2 years of integration, O-B RMSE averaged profiles for T; gmpd – old NEMOVAR, goqu – new NEMOVAR, gor1 – ens. var., inf. fact. 1, gor2 – ens. var., inf. fact. 2, gor3 – ens. var., inf. fact. 3, gor4 – ens. var., inf. fact. 4, goug – ens. var., inf. fact. 5; gor1-goug purely ensemble based inflated filtered variances: significant improvement in extra-tropics and visible degradation in tropical regions

Hybrid Variances: preliminary experiments

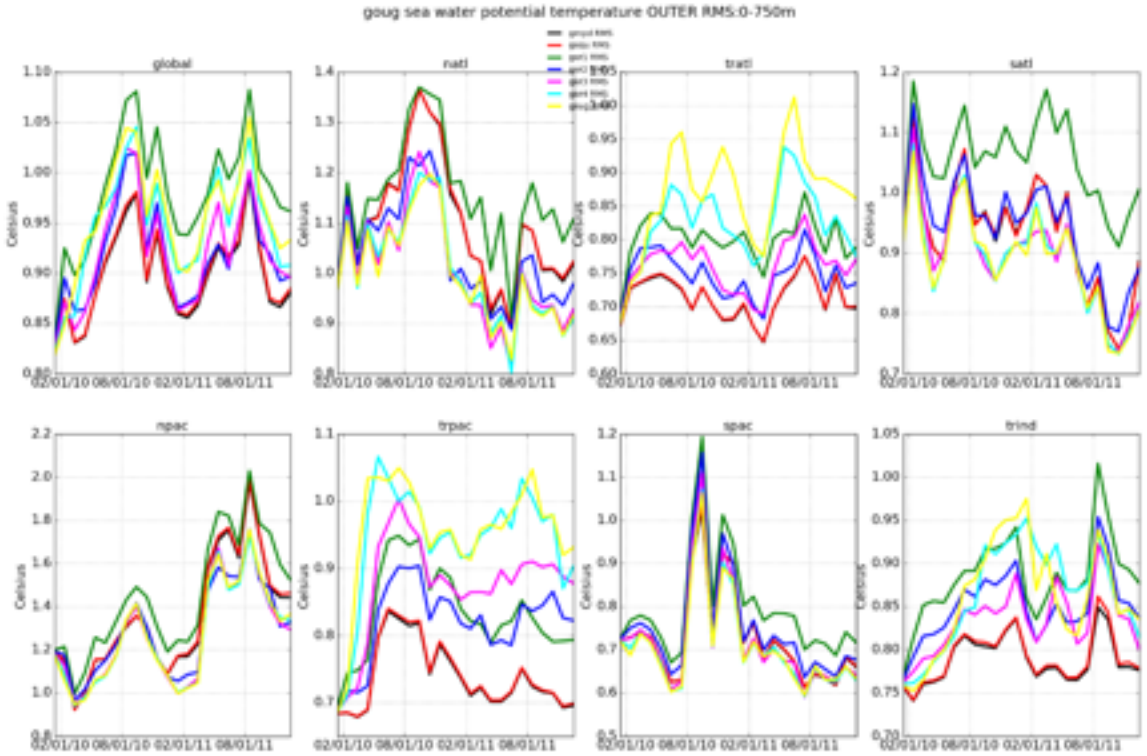


Fig. First experiments with ensemble based background error variances. ORCA1_Z42, 2 years of integration, O-B RMSE averaged profiles for T; gmpd – old NEMOVAR, goqu – new NEMOVAR, gor1 – ens. var., inf. fact. 1, gor2 – ens. var., inf. fact. 2, gor3 – ens. var., inf. fact. 3, gor4 – ens. var., inf. fact. 4, goug – ens. var., inf. fact. 5; gor1-goug purely ensemble based inflated filtered variances: significant improvement in extra-tropics and visible degradation in tropical regions

Hybrid Variances: preliminary experiments

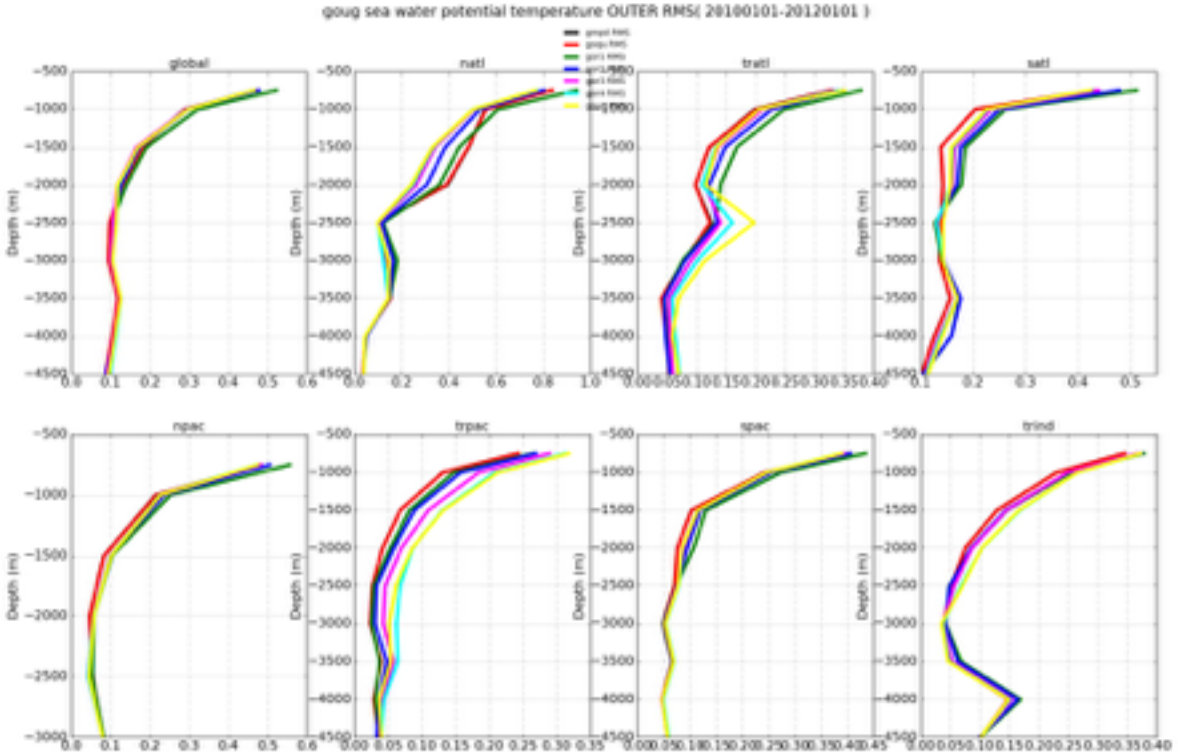
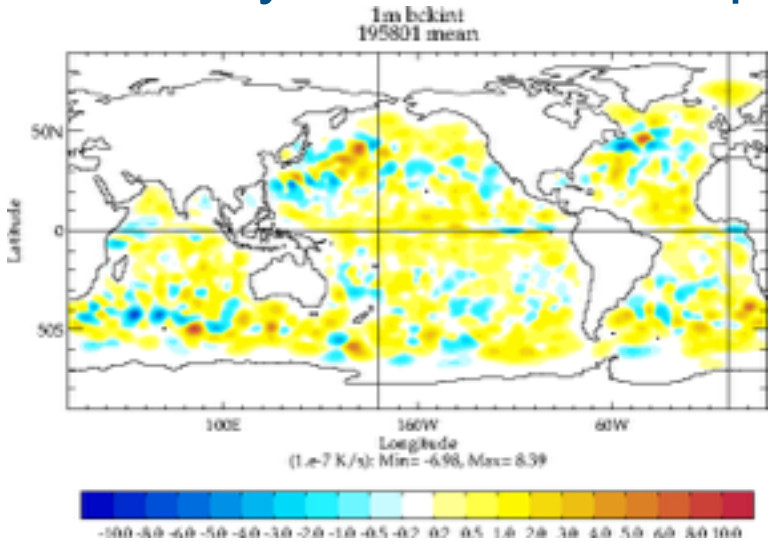
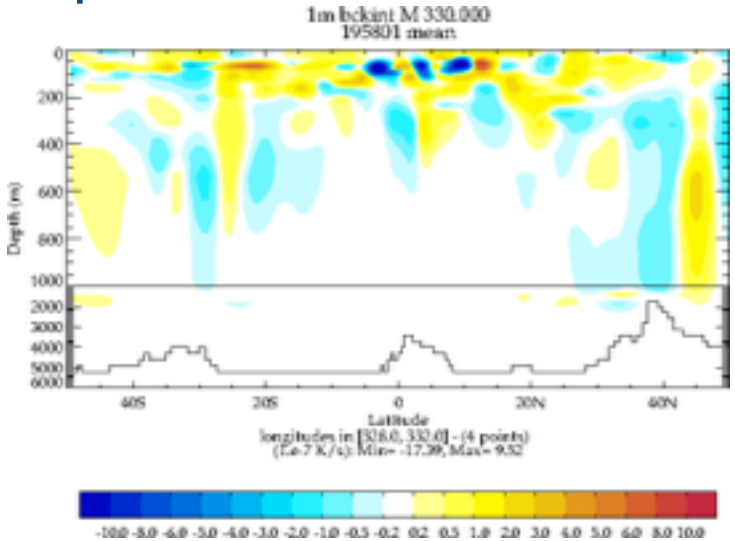


Fig. First experiments with ensemble based background error variances. ORCA1_Z42, 2 years of integration, O-B RMSE averaged profiles for T; gmpd – old NEMOVAR, goqu – new NEMOVAR, gor1 – ens. var., inf. fact. 1, gor2 – ens. var., inf. fact. 2, gor3 – ens. var., inf. fact. 3, gor4 – ens. var., inf. fact. 4, goug – ens. var., inf. fact. 5; gor1-goug purely ensemble based inflated filtered variances: significant improvement in extra-tropics and visible degradation in tropical regions

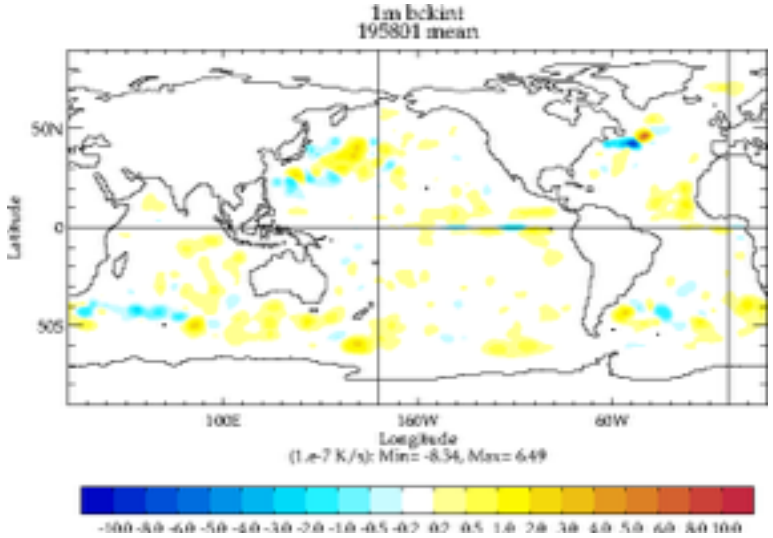
Hybrid Variances: preliminary experiments – T increments

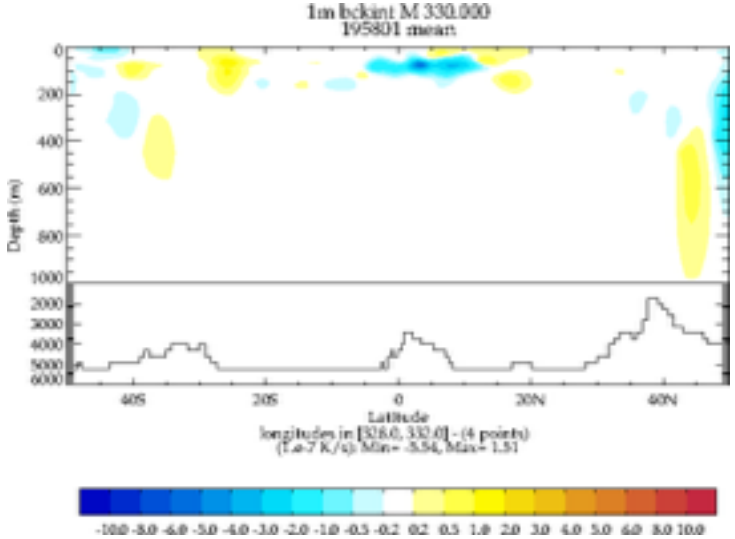
goqu





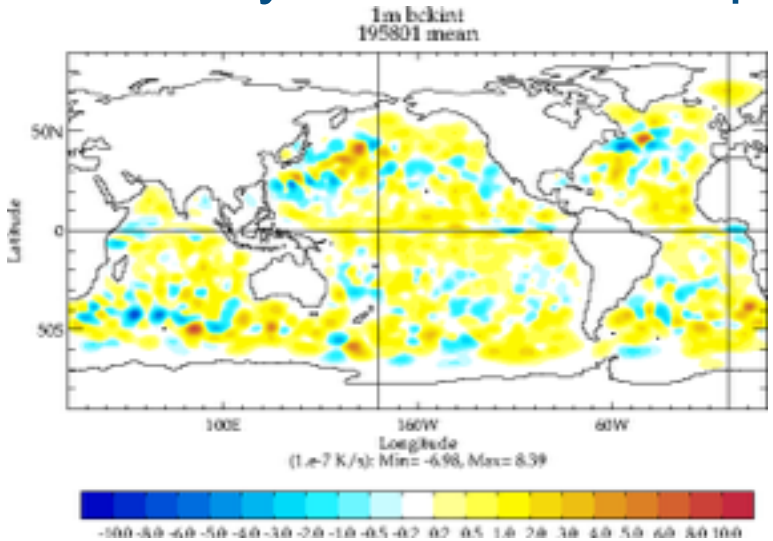
gor1

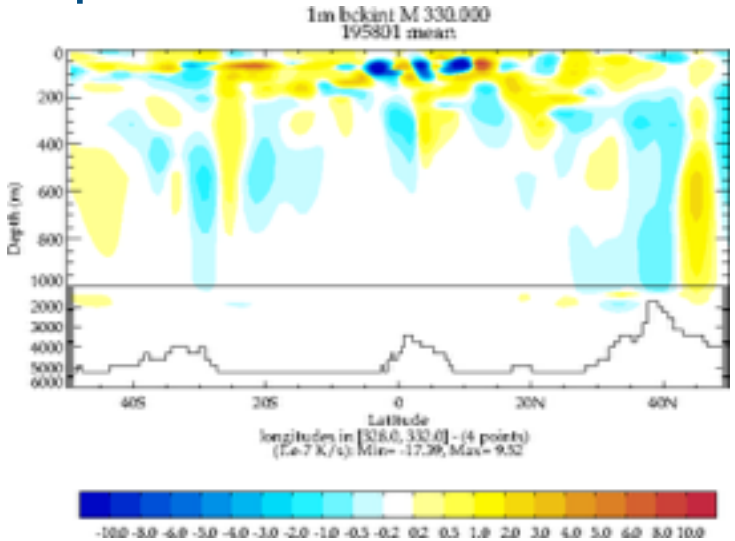




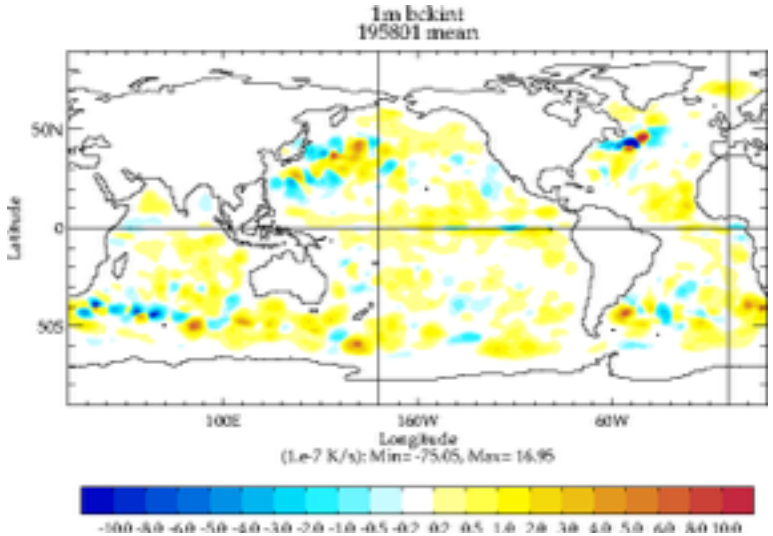
Hybrid Variances: preliminary experiments – T increments

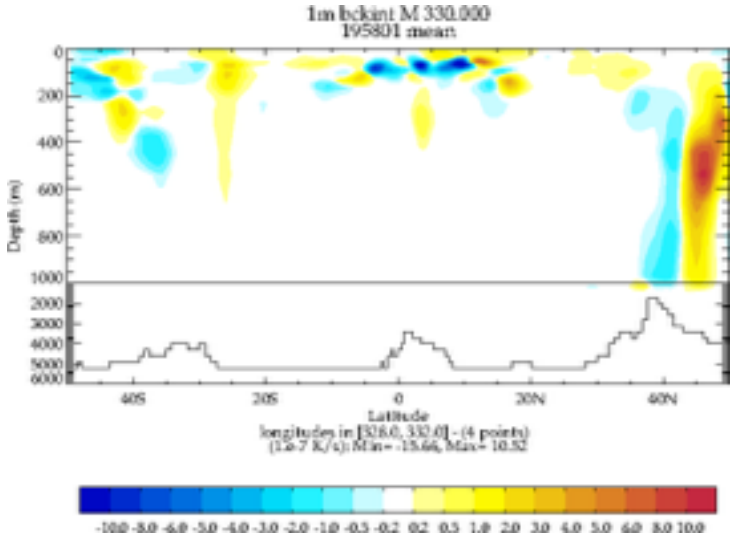
goqu





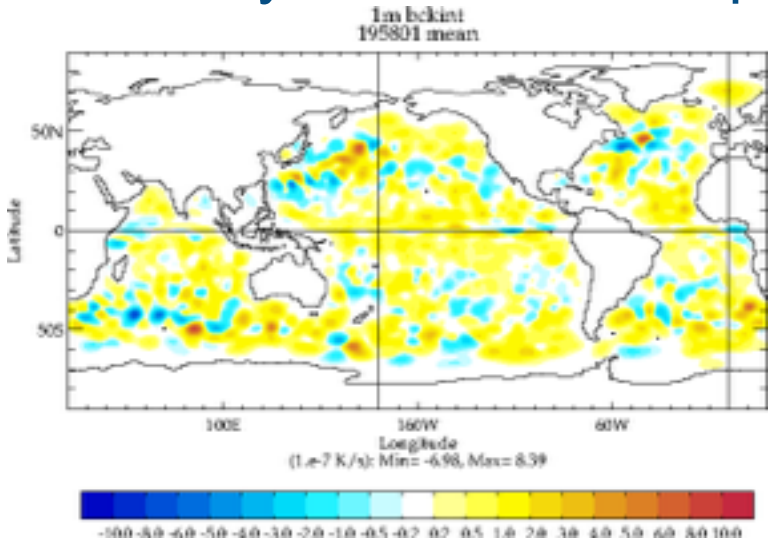
gor2

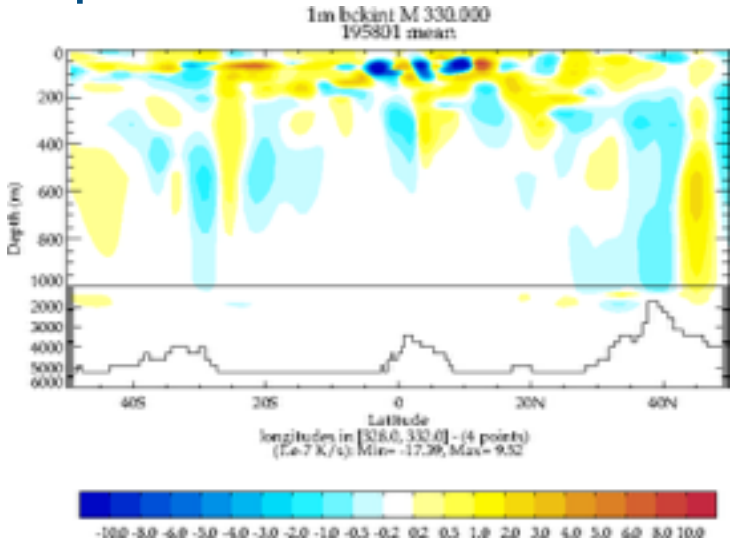




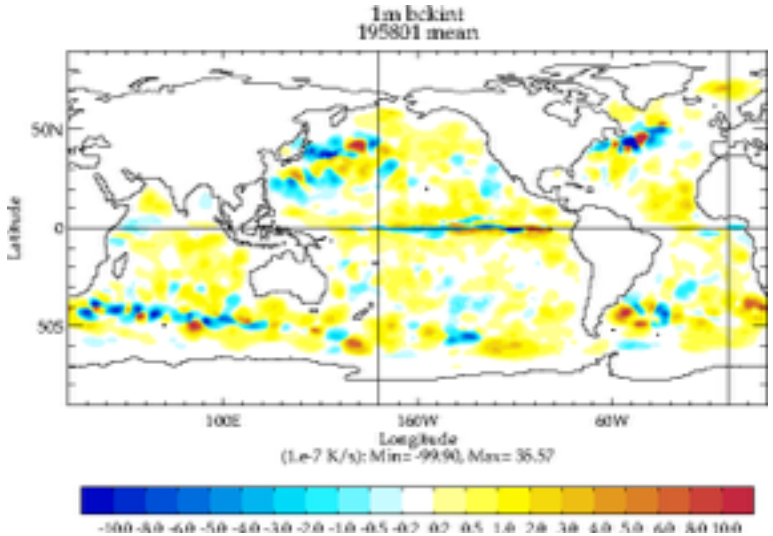
Hybrid Variances: preliminary experiments – T increments

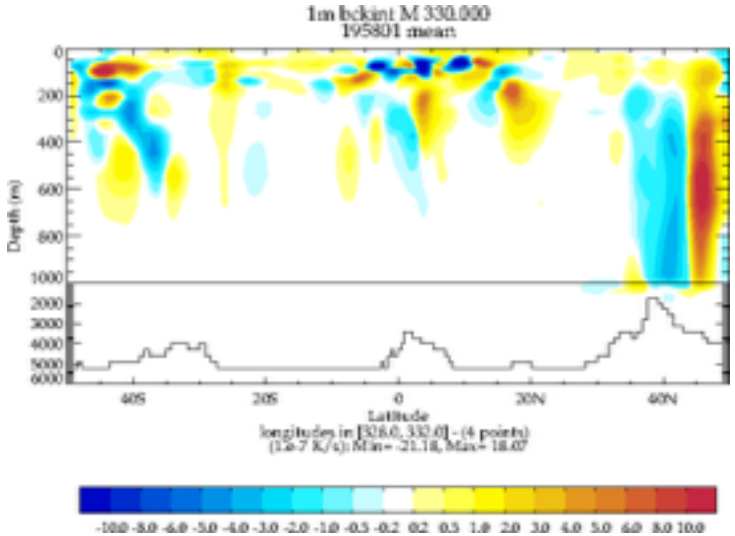
goqu





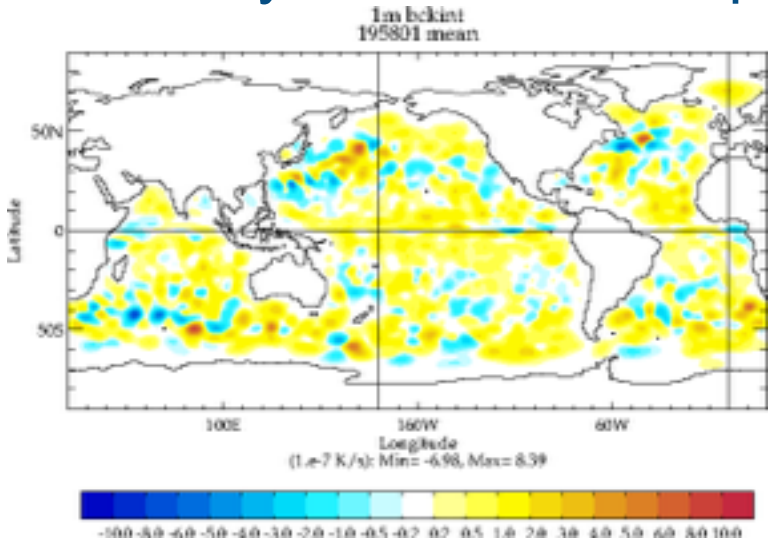
gor3

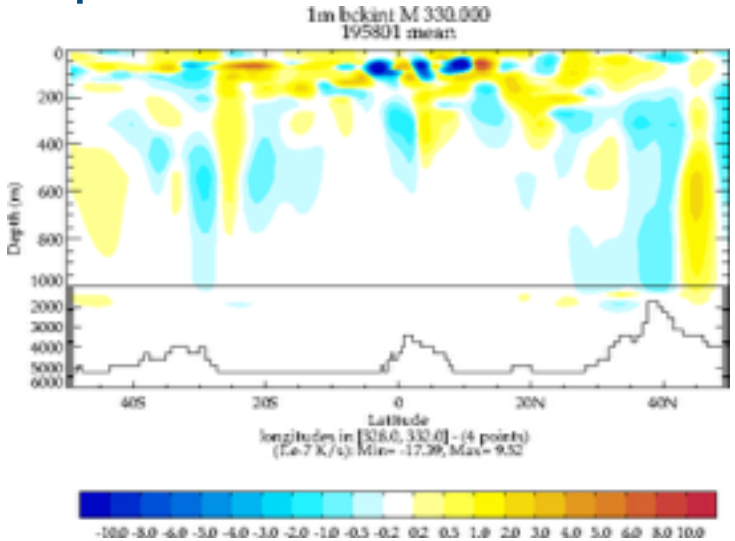




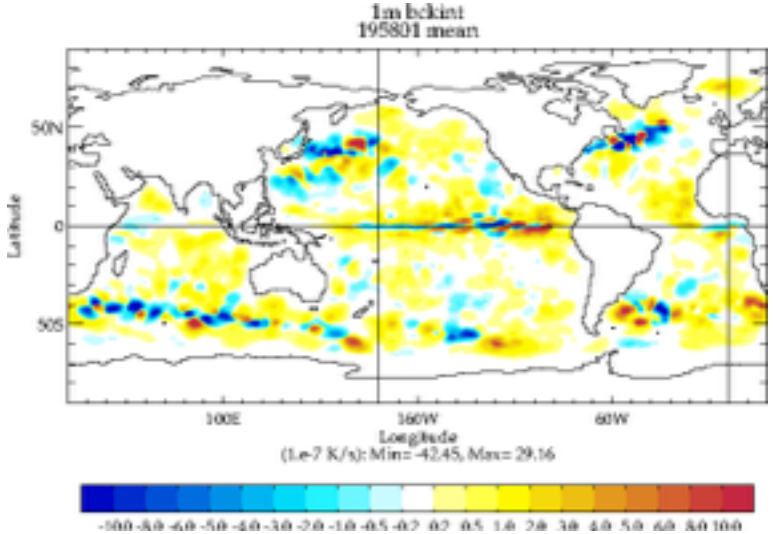
Hybrid Variances: preliminary experiments – T increments

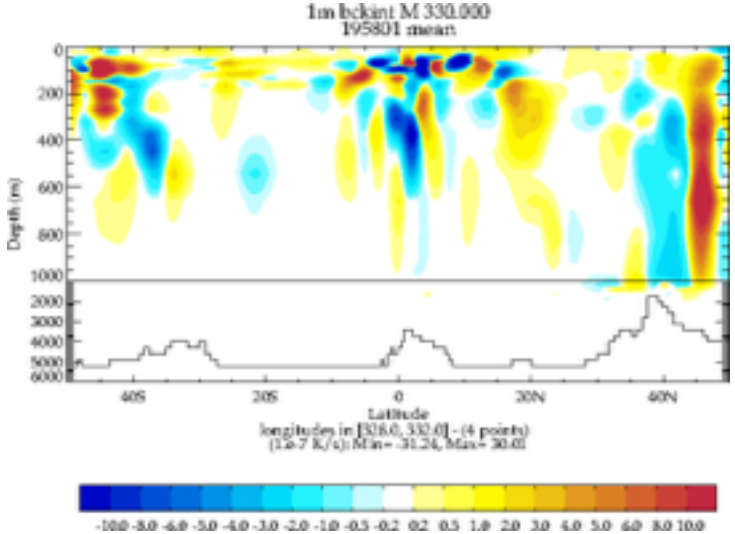
goqu





gor4





Hybrid Variances: preliminary experiments

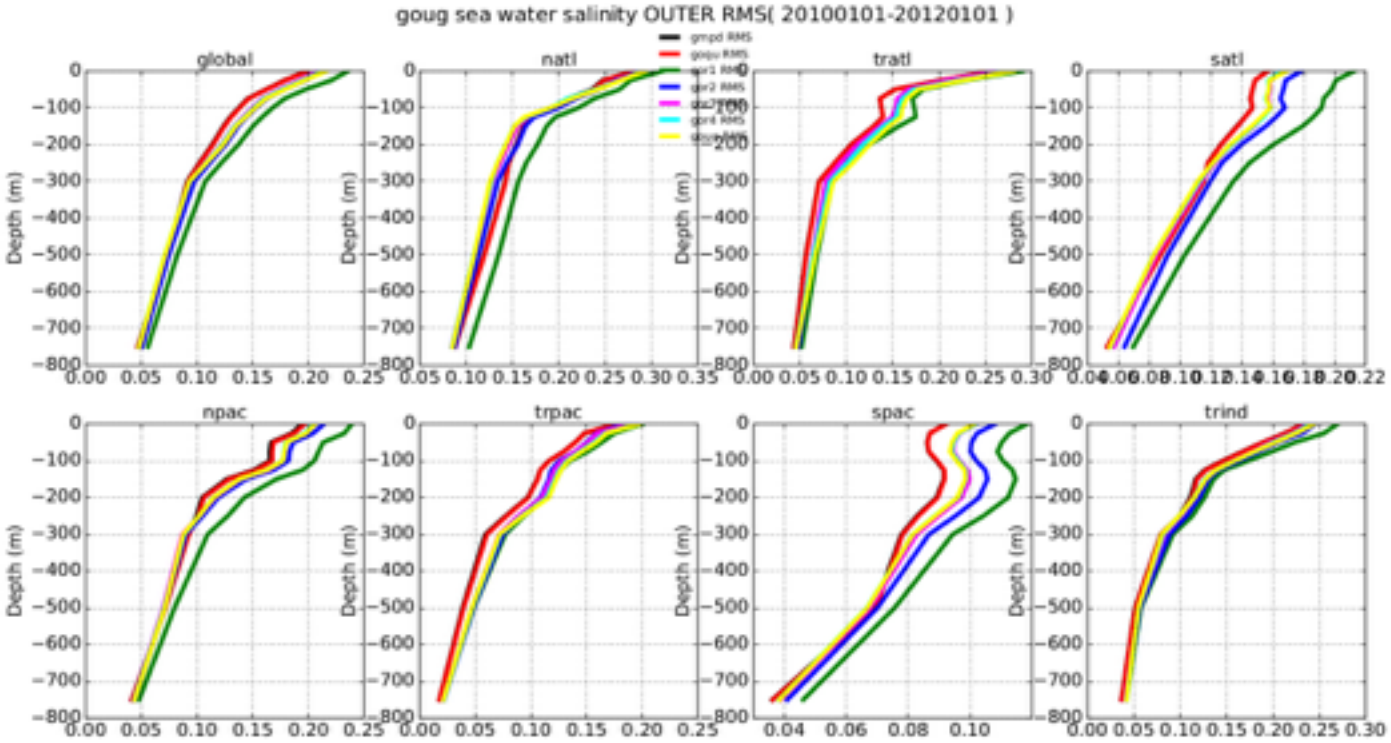


Fig. First experiments with ensemble based background error variances. ORCA1_Z42, 2 years of integration, O-B RMSE averaged profiles for S; gmpd – old NEMOVAR, goqu – new NEMOVAR, gor1 – ens. var., inf. fact. 1, gor2 – ens. var., inf. fact. 2, gor3 – ens. var., inf. fact. 3, gor4 – ens. var., inf. fact. 4, goug – ens. var., inf. fact. 5; gor1-goug purely ensemble based inflated filtered variances: significant improvement in extra-tropics and visible degradation in tropical regions

Hybrid Variances: preliminary experiments

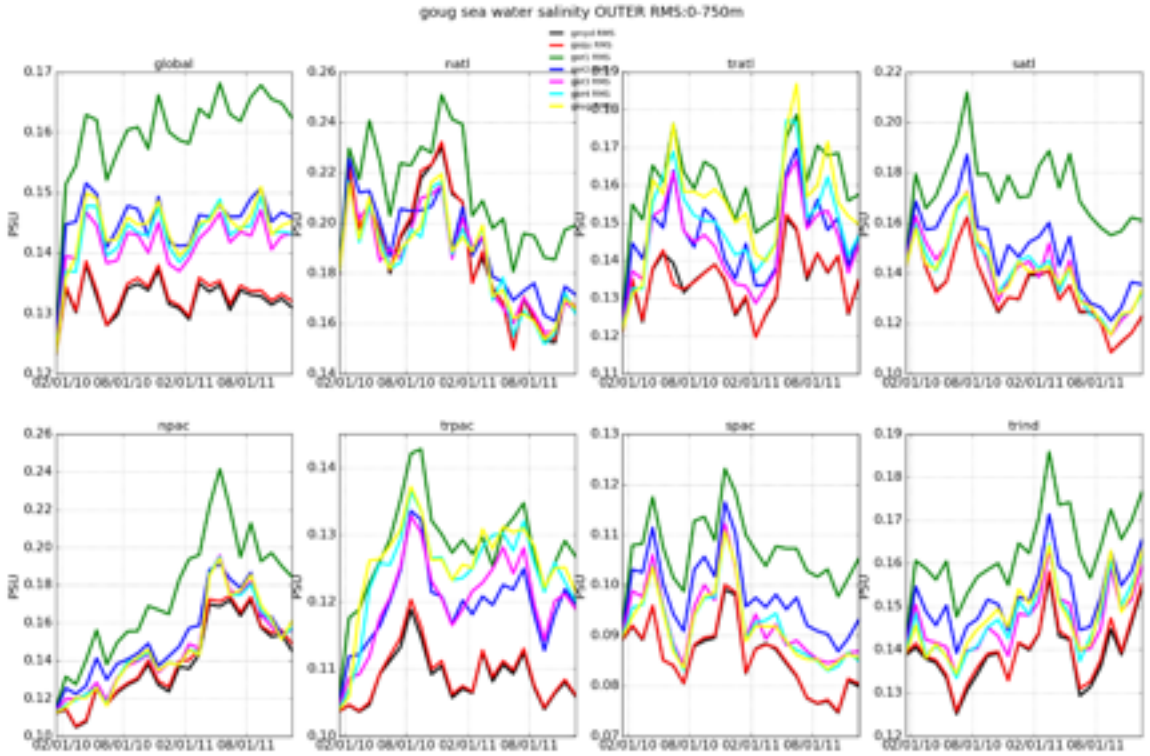


Fig. First experiments with ensemble based background error variances. ORCA1_Z42, 2 years of integration, O-B RMSE averaged profiles for S; gmpd – old NEMOVAR, goqu – new NEMOVAR, gor1 – ens. var., inf. fact. 1, gor2 – ens. var., inf. fact. 2, gor3 – ens. var., inf. fact. 3, gor4 – ens. var., inf. fact. 4, goug – ens. var., inf. fact. 5; gor1-goug purely ensemble based inflated filtered variances: significant improvement in extra-tropics and visible degradation in tropical regions

Hybrid Variances: preliminary experiments

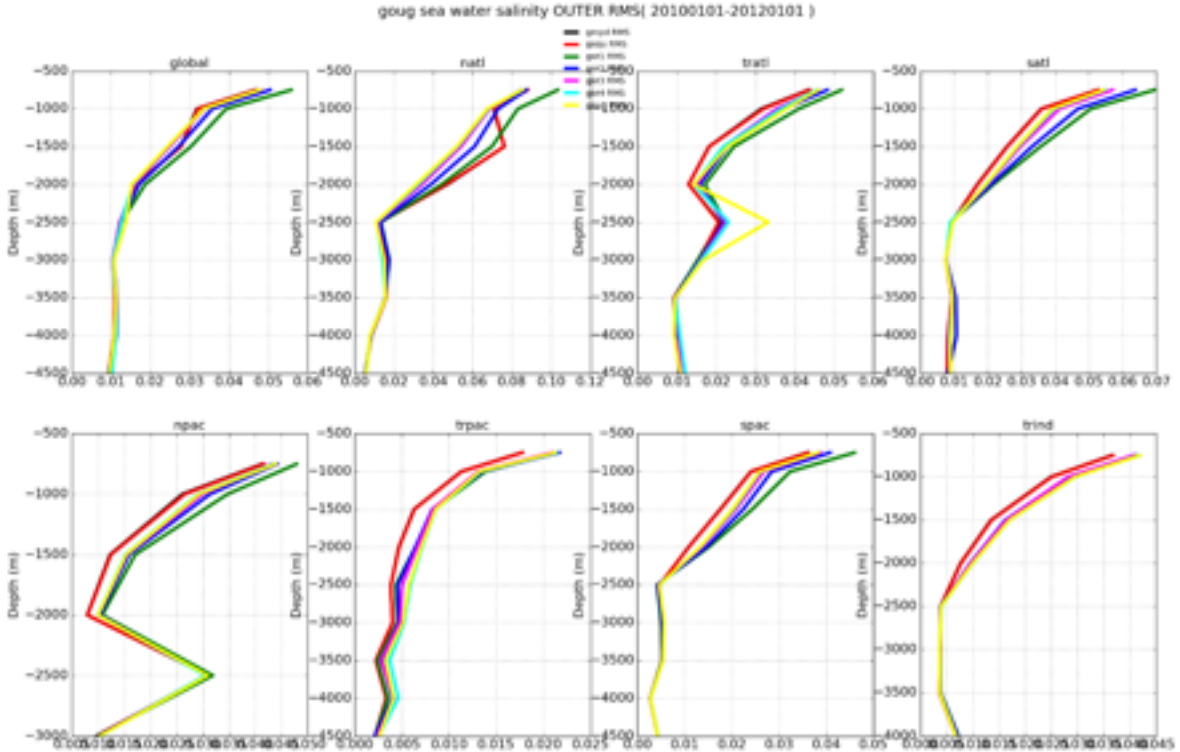
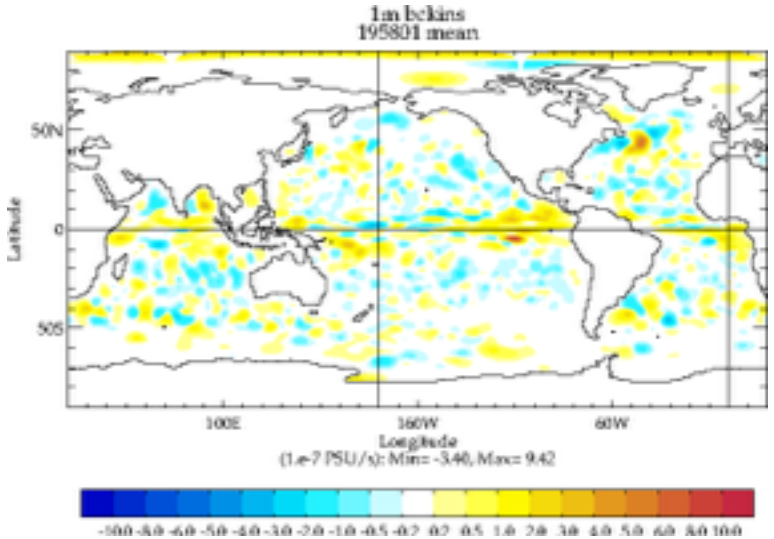
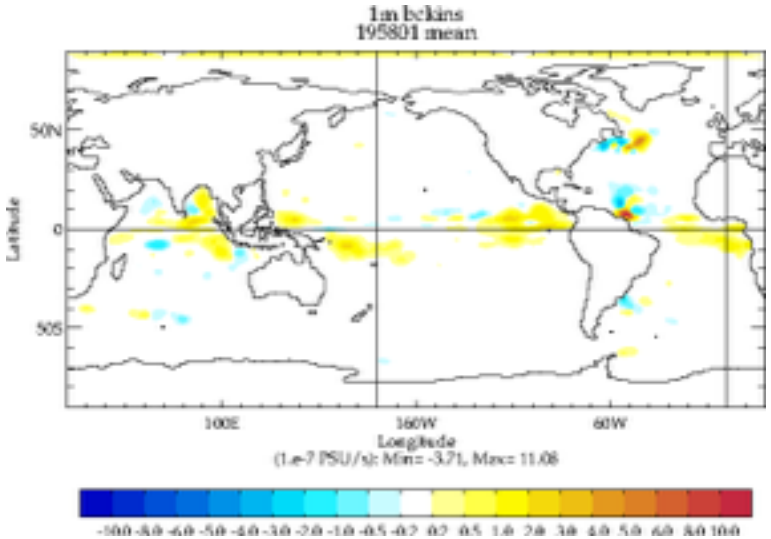


Fig. First experiments with ensemble based background error variances. ORCA1_Z42, 2 years of integration, O-B RMSE averaged profiles for S; gmpd – old NEMOVAR, goqu – new NEMOVAR, gor1 – ens. var., inf. fact. 1, gor2 – ens. var., inf. fact. 2, gor3 – ens. var., inf. fact. 3, gor4 – ens. var., inf. fact. 4, goug – ens. var., inf. fact. 5; gor1-goug purely ensemble based inflated filtered variances: significant improvement in extra-tropics and visible degradation in tropical regions

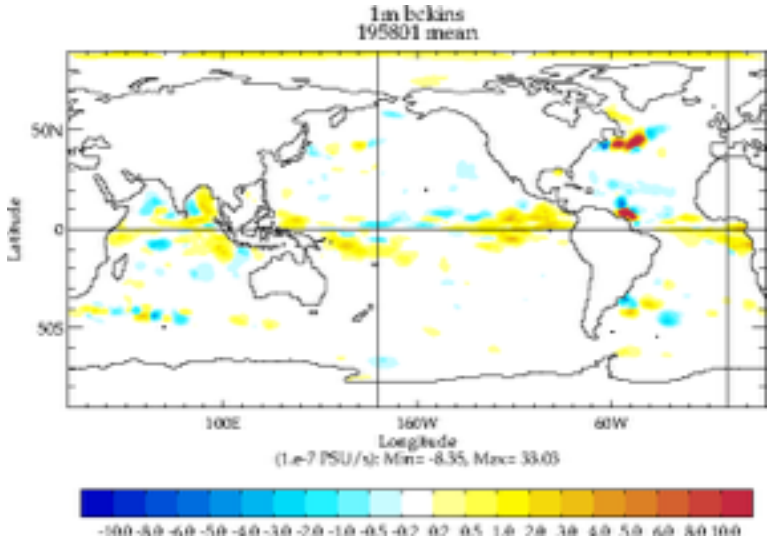
Hybrid Variances: preliminary experiments – salinity increments



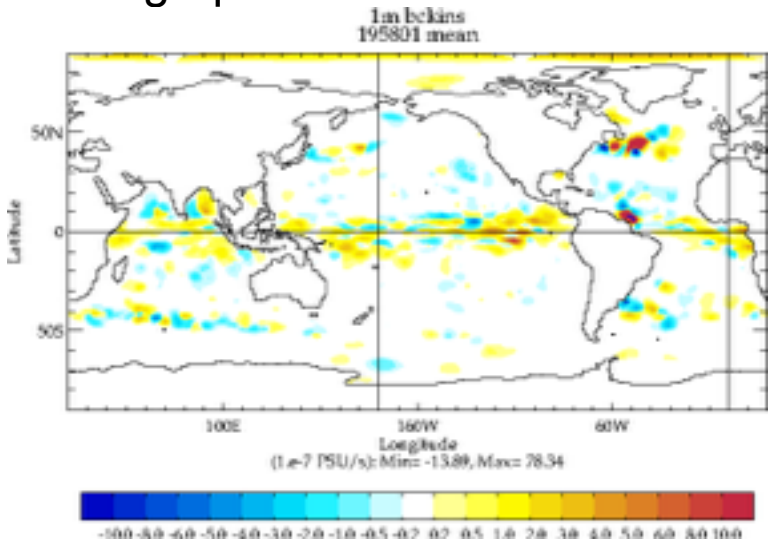
goqu



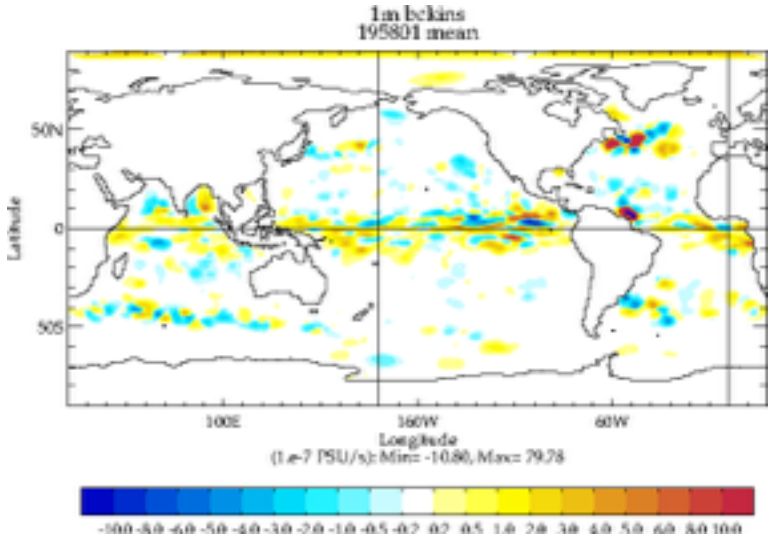
gor1



gor2

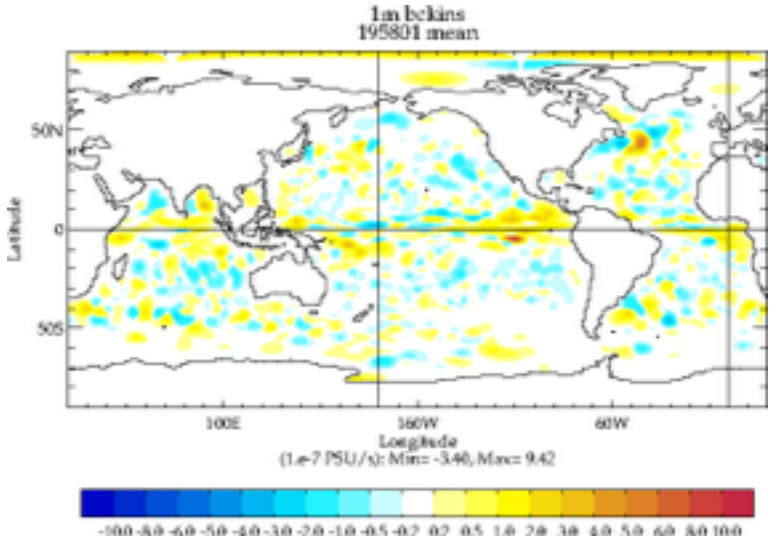


gor3

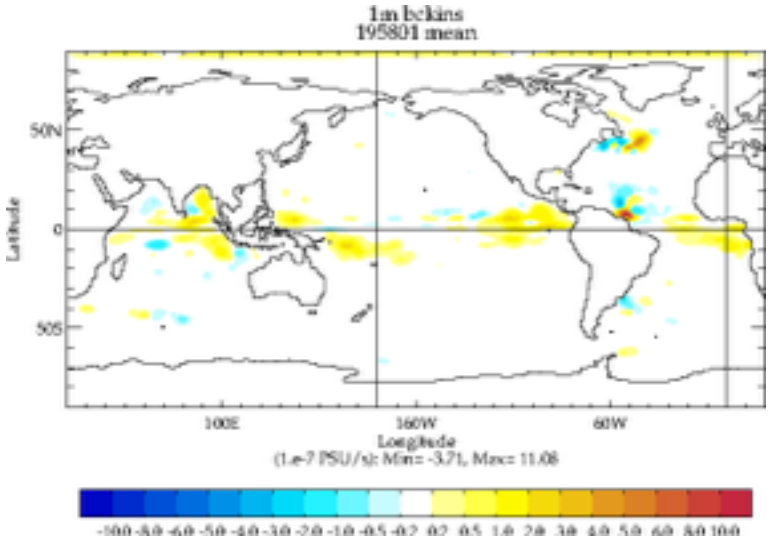


gor4

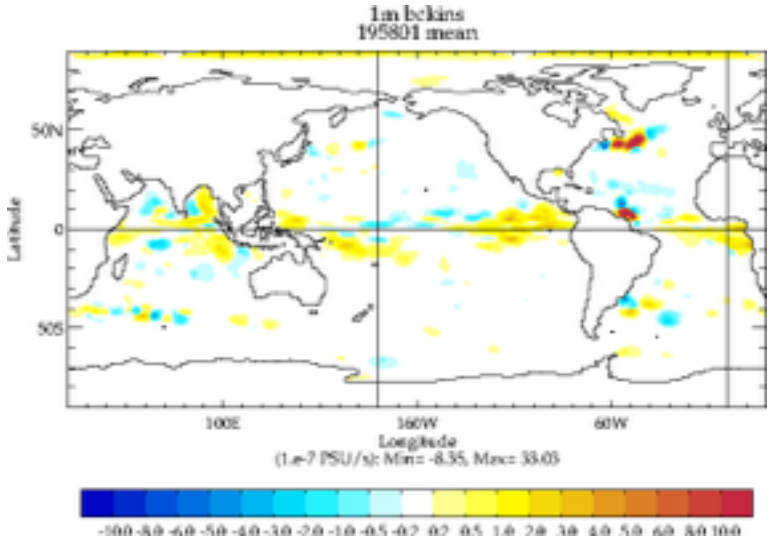
Hybrid Variances: preliminary experiments



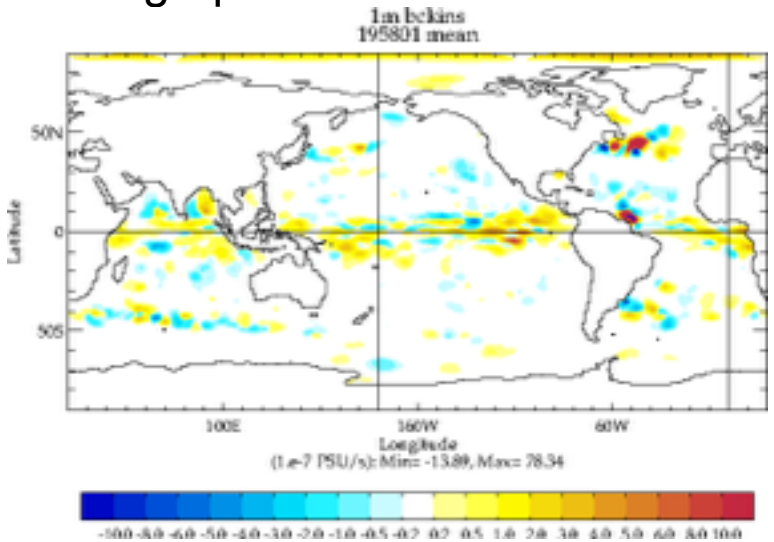
goqu



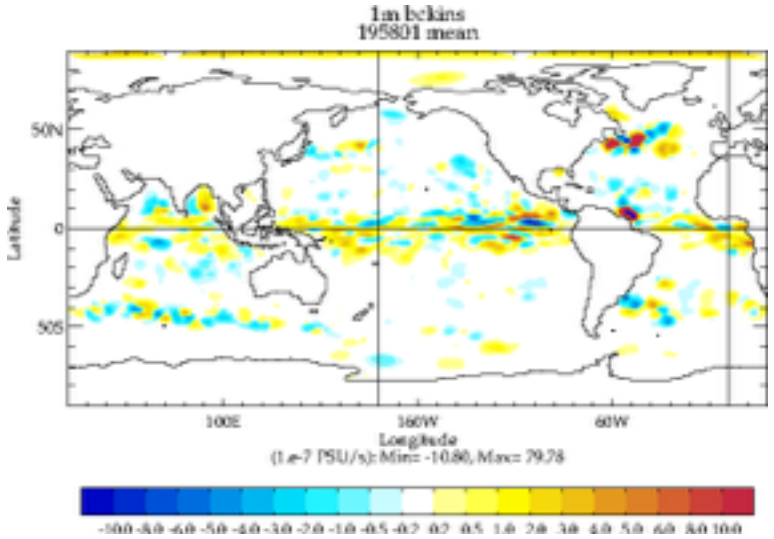
gor1



gor2



gor3



gor4

Outline

- Ocean Data Assimilation at ECMWF
- Ensemble Generation
- Hybrid Background Covariance
- Experiments
- **Future directions**

Summary

- The new generic ensemble generation scheme including both perturbations of assimilated observations and surface forcing fields has been developed;
- The ensemble spread captures the uncertainty in the background errors in eddy active regions;
- The temperature ensemble spread compared to the diagnosed temperature background error standard deviation is too low especially in the tropical regions;
- The geographical distribution of the salinity spread is consistent with the diagnosed salinity background error standard deviations but the magnitude is too low;
- The SLA ensemble spread is under-dispersive
- Deep ocean temperature and salinity spread collapses.

Future directions

- Attempt to combine the specified/parametrized background error standard deviations with the ensemble based accounting for shortcomings of the ensemble;
- Regional dependency of the weighting/inflation factors as well as depth dependency seems to be required;
- Carry out experiments at 1/4 degree and evaluate the impact of the ensemble size;
- Evaluate errors of the day;