Assimilation diagnostics from a global 3D-Var system

A. Weaver[1], N. Daget[1], M. A. Balmaseda[2] and K. Mogensen[2]

[1] *CERFACS, Toulouse*
[2] *ECMWF, Reading*

Outline

Outline

1. The assimilation system
2. Diagnostics from a global 2^o reanalysis with an ensemble 3D-Var
3. Diagnostics from a global 1^o 3D-Var reanalysis
4. Conclusions

Basic features of the assimilation system

- The ocean model is the global 2^{o} configuration of OPA8.2 (Madec *et al.* 1998).
- The surface forcing fields are derived from ERA40 (Uppala *et al.* 2005).
- The assimilation method is a multivariate 3D-Var (Weaver *et al.* 2005).
- First-Guess at Appropriate Time (FGAT) and Incremental Analysis Updates (IAU) are employed.
- The data are quality-controlled temperature and salinity profiles from ENSEMBLES (EN3) data-base (Ingleby and Huddleston 2007).

$$J[\delta \mathbf{w}] \; = \; \frac{1}{2}\delta \mathbf{w}^{\mathrm{T}}\, \mathbf{B}_{(\mathbf{w})}^{-1}\, \delta \mathbf{w} + \frac{1}{2}(\mathbf{H}\delta \mathbf{w} - \mathbf{d})^{\mathrm{T}} \mathbf{R}^{-1} (\mathbf{H}\delta \mathbf{w} - \mathbf{d})$$

where

$$\mathbf{d} = \begin{pmatrix} \mathbf{d}_0 \\ \vdots \\ \mathbf{d}_i \\ \vdots \\ \mathbf{d}_N \end{pmatrix} = \begin{pmatrix} \mathbf{y}_0^o - \mathbf{H}_0 \mathbf{w}^b(t_0) \\ \vdots \\ \mathbf{y}_i^o - \mathbf{H}_i \mathbf{w}^b(t_i) \\ \vdots \\ \mathbf{y}_N^o - \mathbf{H}_N \mathbf{w}^b(t_N) \end{pmatrix} \quad \text{and} \quad \mathbf{H} = \begin{pmatrix} \mathbf{H}_0 \\ \vdots \\ \mathbf{H}_i \\ \vdots \\ \mathbf{H}_N \end{pmatrix}.$$

- $\delta \mathbf{w} = (\delta T, \delta S)^{\mathrm{T}}$ is the vector of temperature and salinity increments.
- $\mathbf{y}_i^o = (T_i^o, S_i^o)^{\mathrm{T}}$ is the vector of temperature and salinity observations.
- Increments for sea-surface height and velocity are obtained using balance constraints applied to the analysis increment $\delta \mathbf{w}^a$.

$$\mathbf{B}_{(\mathbf{w})} = \mathbf{K}_{(\mathbf{w})}\,\mathbf{D}^{1/2}_{(\widehat{\mathbf{w}})}\,\mathbf{F}_{(\widehat{\mathbf{w}})}\mathbf{F}^{\mathrm{T}}_{(\widehat{\mathbf{w}})}\,\mathbf{D}^{1/2}_{(\widehat{\mathbf{w}})}\mathbf{K}^{\mathrm{T}}_{(\mathbf{w})}$$

where

$$\mathbf{F}_{(\widehat{\mathbf{w}})} = \begin{pmatrix} \mathbf{F}_{TT} & 0 \\ 0 & \mathbf{F}_{S_U S_U} \end{pmatrix}, \quad \mathbf{D}^{1/2}_{(\widehat{\mathbf{w}})} = \begin{pmatrix} \mathbf{D}^{1/2}_{T} & 0 \\ 0 & \mathbf{D}^{1/2}_{S_U} \end{pmatrix}, \quad \mathbf{K}_{(\mathbf{w})} = \begin{pmatrix} \mathbf{I} & 0 \\ \mathbf{K}_{ST} & \mathbf{I} \end{pmatrix}$$

- $\widehat{\mathbf{w}} = (T, S_U)^{\mathrm{T}}$ where S_U corresponds to “unbalanced” salinity.
- $\mathbf{K}_{(\mathbf{w})}$ is a multivariate balance operator: $\widehat{\mathbf{w}} \mapsto \mathbf{w}$.
- $\mathbf{F}_{(\widehat{\mathbf{w}})}\mathbf{F}^{\mathrm{T}}_{(\widehat{\mathbf{w}})}$ is a quasi-Gaussian 3D univariate correlation operator, modelled using a diffusion operator.
- $\mathbf{D}_{(\widehat{\mathbf{w}})}$ is a variance matrix (for $\widehat{\mathbf{w}}$).

Outline

General context

- An ensemble 3D-Var system was developed for the European project ENSEMBLES to provide multiple ocean analyses for sampling uncertainty in ocean initial conditions for probabilistic seasonal and decadal forecasts.
- An important feature of an ensemble data assimilation system is its capacity to provide flow-dependent information on analysis and background error.
  - This information can be exploited to improve the estimate of the background-error covariance matrix (**B**) on each assimilation cycle.
  - In the ENSEMBLES experiments, we made no attempt to use the ensemble to update **B**.
- The objective here is explore the possibility of using the ensemble 3D-Var system to improve **B** (Daget *et al.* 2009, QJRMS).

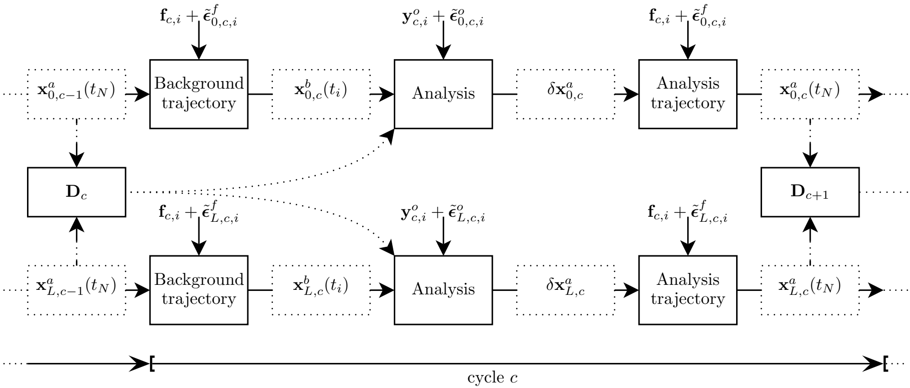
- Construct a low-rank approximation to **B** directly from the sample covariance of the ensemble of model forecast states.
  (Houtekammer and Mitchell 2001; Keppenne and Reinecker 2002; Ott *et al.* 2004; Buehner and Charron 2007; Oke *et al.* 2007).
  - Covariance localization is necessary to minimize spurious effects due to sampling error.

- or -

- Use the ensemble indirectly to define parameters of a (localized) covariance model in a full-rank (operator) representation of **B**.
  (Fisher 2003; Žagar *et al.* 2005; Belo Pereira and Berre 2006; Berre *et al.* 2006; Küçükkaraca and Fisher 2006).
  - A flexible covariance model (inhomogeneous, anisotropic) is required to make best use of the ensemble information.
  - Full-rank covariance matrix representations are important for fitting detailed structures in the observations.

Objective

- Here, we adopt the covariance model approach.
- In particular, we investigate the potential of an ensemble of ocean states to provide useful flow-dependent estimates of the background-error **variances** in the 3D-Var system.
- This approach will be compared with a simpler approach for incorporating flow dependence in the variances, based on a parameterization in terms of the background state.
- This study is a first step towards making more comprehensive use of an ensemble for specifying additional parameters of the covariance model.



- The background-error variance matrix ($\mathbf{D}_c$) used for the analysis on cycle c is estimated from the sample variance matrix computed from the ensemble of background states ($\mathbf{x}^b_{l,c}(t_0)$) at the start of cycle c.
- In our set-up, $\mathbf{x}^b_{l,c}(t_0) = \mathbf{x}^a_{l,c-1}(t_N)$.

Key features of the ensemble-generation strategy

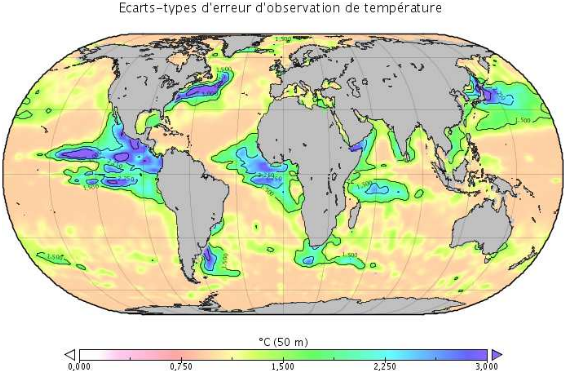
- A 9-member ensemble.
- The perturbed input parameters:
  - the surface forcing fields (heat flux, fresh-water flux, wind-stress);
  - the temperature and salinity observations;
  - the background state;
  - model error is neglected.
- Construction of the perturbations:
  - the forcing perturbations are derived from differences between different forcing analysis products (Balmaseda *et al.* 2008);
  - the observation perturbations are drawn from a Gaussian pdf with covariance matrix $\mathbf{R}$;
  - the background state is perturbed implicity via the cycling procedure;
- Reduction of sampling error:
  - A 90-day (9-cycle) sliding window is used, giving an effective ensemble size of 81 on each cycle for estimating σ^b.
  - Intraseasonal variability in σ^b is thus filtered out.

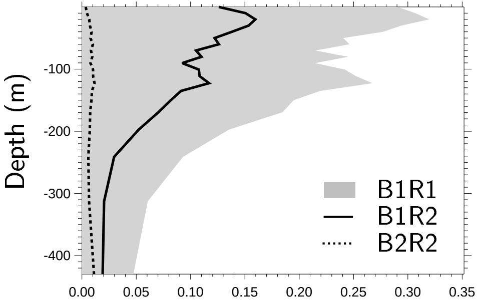
The numerical experiments

- The experimental design follows the common reanalysis procedures used in the ENSEMBLES and ENACT projects (Davey *et al.* 2006).
- The experiments are performed for the 9-year period from 1 January 1993 to 31 December 2001.
- A 10-day assimilation cycle is used.
- The experiments:
  - **CTL** : no data assimilation.
  - **B1R1** : parameterized σ^b, and σ^o defined using globally-averaged estimates from Ingleby and Huddleston (2007).
  - **B1R2** : parameterized σ^b, and σ^o estimated from Fu *et al.* method.
  - **B2R2** : ensemble σ^b, and σ^o estimated from Fu *et al.* method.
- Results will be displayed for temperature only and for the global ocean (results for salinity and in different regions are qualitatively similar).

Observation-error variances are estimated from a model-data comparison prior to assimilation.

Example of temperature σ^o at 50 m

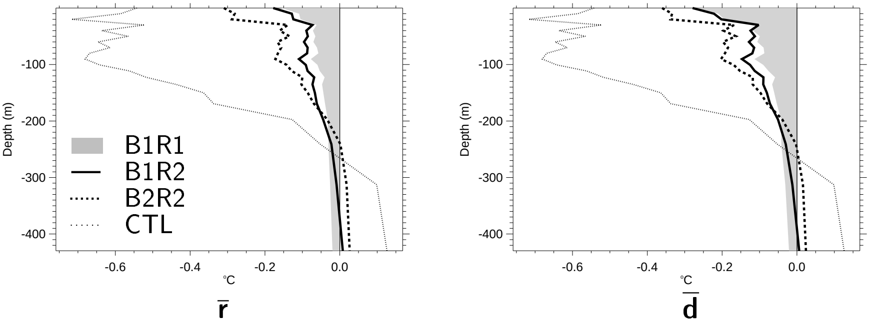




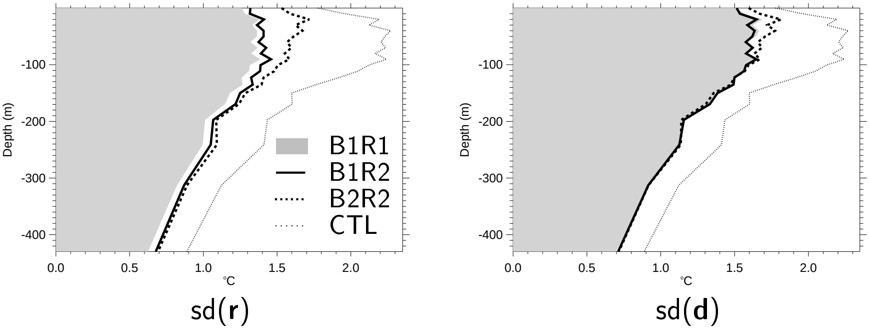
$$w_T = \frac{(\sigma_T^b)^2}{(\sigma_T^b)^2 + (\sigma_T^o)^2}$$

- Neglecting correlations, w_T is the average weight for an innovation.
- Both σ_T^b and σ_T^o have been computed at observation points, and averaged over the 1994-2000 period and the global domain.

Analysis residual and innovation statistics for temperature

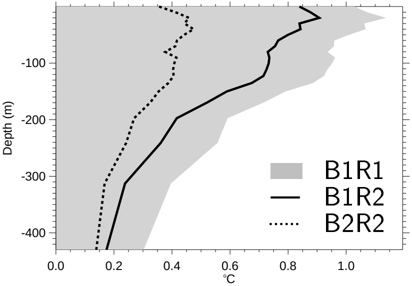


- $\mathbf{r} = \mathbf{d} - \mathbf{H}\delta\mathbf{w}^a$ (residual) and $\mathbf{d} = \mathbf{y}^o - \mathbf{H}\mathbf{w}^b$ (innovation).
- $\overline{\mathbf{z}}$ indicates spatial (global) and temporal (1994-2000) average.
- Mean bias in CTL is reduced substantially in all assimilation expts.

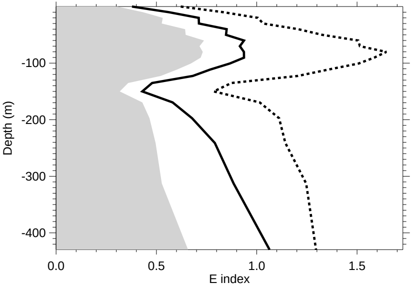


- $\mathbf{r} = \mathbf{d} - \mathbf{H}\delta\mathbf{w}^a$ (residual) and $\mathbf{d} = \mathbf{y}^o - \mathbf{H}\mathbf{w}^b$ (innovation).
- $\mathrm{sd}(\mathbf{z}) = \sqrt{\overline{(\mathbf{z} - \overline{\mathbf{z}})^2}}$
- All assimilation expts. improve the fit to the observed variability.
- The "error growth" in the 10-day forecast is smallest for B2R2.

An “efficiency” index



$\mathrm{rms}(\mathbf{H}\delta\mathbf{w}^a)$

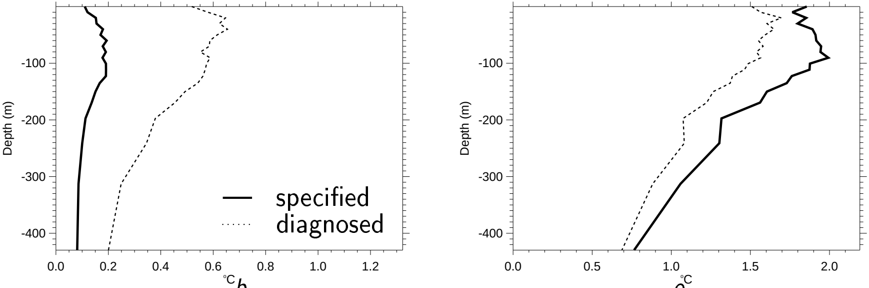


$$\mathrm{E} = \frac{\mathrm{rms}(\mathbf{d}_{\mathrm{CTL}}) - \mathrm{rms}(\mathbf{d})}{\mathrm{rms}(\mathbf{H}\delta\mathbf{w}^a)}$$

- $\mathrm{E} = \dfrac{\text{10-day forecast error from CTL – 10-day forecast error from assim.}}{\text{“work done” by assimilation method to reduce forecast error}}$
- $\mathrm{E} > 0$ ($\mathrm{E} < 0$) implies assimilation is beneficial (detrimental).
- E increases (decreases) if $\mathbf{d}$ or $\delta\mathbf{w}^a$ decreases (increases).

Specified versus diagnosed σ^b and σ^o for temperature in B2R2

(method of Desroziers *et al.* 2005)



σ_T^b σ_T^o

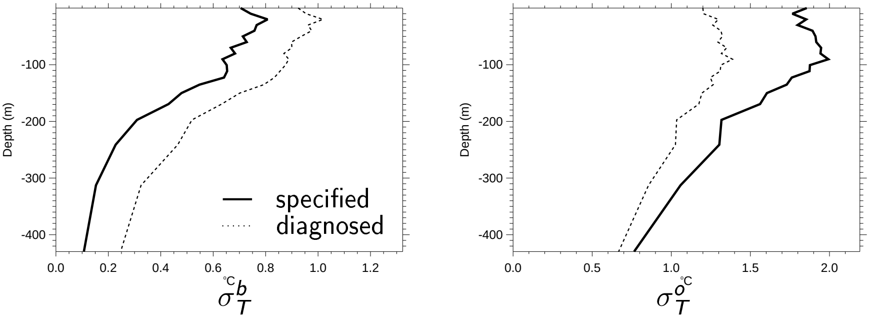
- If **B** and **R** are good estimates of the true background- and observation-error covariance matrices then

$$E[\mathbf{d}\,(\mathbf{H}\delta\mathbf{w}^a)^{\mathrm{T}}] \approx \mathbf{H}\mathbf{B}_{(\mathbf{w})}\mathbf{H}^{\mathrm{T}}$$

$$E[\mathbf{d}\,(\mathbf{d} - \mathbf{H}\delta\mathbf{w}^a)^{\mathrm{T}}] \approx \mathbf{R}$$

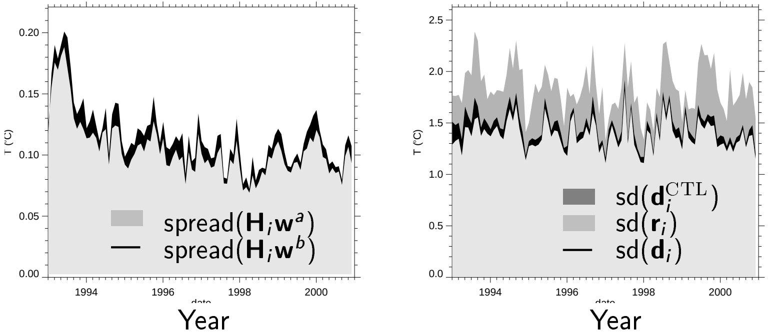
- Here, σ_T^b is underestimated, and σ_T^o is overestimated.

(method of Desroziers *et al.* 2005, QJRMS)



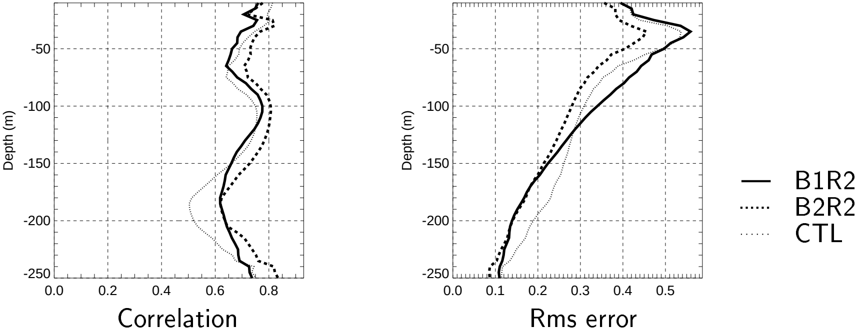
- σ^b_T is also underestimated (to a lesser extent than in B2R2).
- σ^o_T is also overestimated (to a greater extent than in B2R2).

Experiment B2R2



- $\mathrm{spread}\{\mathbf{H}_i\mathbf{w}^{a,b}\} = \sqrt{\frac{1}{L-1}\sum_{l=0}^{L-1}\left(\mathbf{H}_i\mathbf{w}_l^{a,b}(t_i) - \frac{1}{L}\sum_{l=0}^{L-1}\mathbf{H}_i\mathbf{w}_l^{a,b}(t_i)\right)^2}$
- Spread of the analysis $<$ spread of the background.
- No evidence of ensemble collapse.
- Spread($\mathbf{H}_i\mathbf{w}^{a,b}$) is approximately a factor 10 smaller than sd($\mathbf{r}_i$), sd($\mathbf{d}_i$).

Example from the eastern Pacific (110°W)



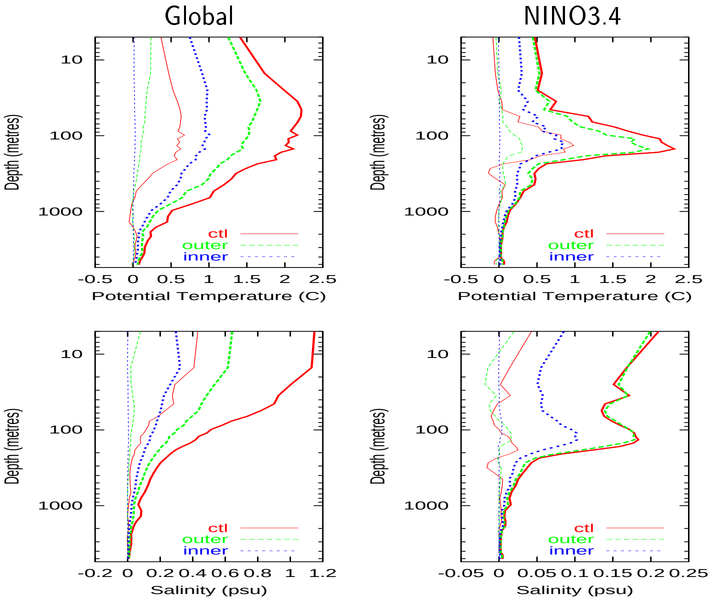
- B2R2 outperforms B1R2 (and B1R1) at all moorings.
- B2R2 outperforms CTL in the central and eastern Pacific, but slightly worse in the western Pacific.

Outline

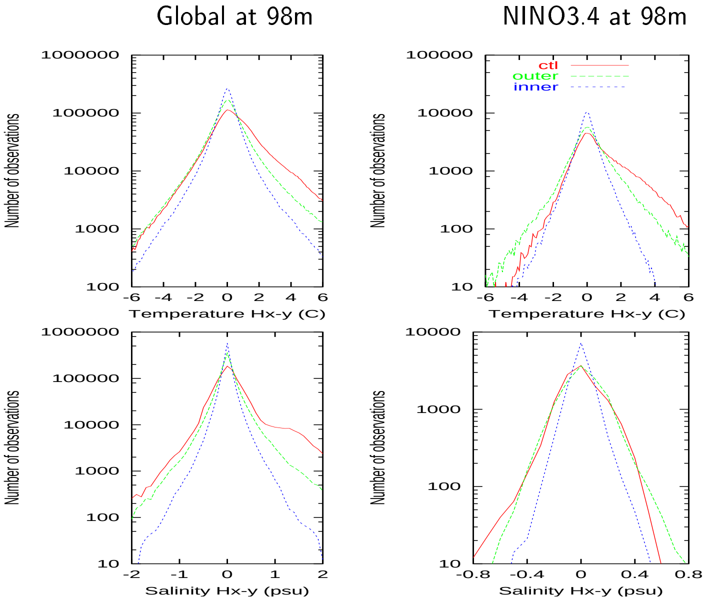
The numerical experiments

- Employs a 3D-Var FGAT version of the NEMOVAR system, developed for NEMO v3.
- Employs a global 1^o configuration, with increased meridional resolution near the equator.
- ERA-Interim forcing fluxes.
- QC T and S profiles from EN3 v1d.
- The experiments are performed for the 18-year period from 1 January 1989 to 31 December 2006.
- A 10-day assimilation cycle.

RMS (thick) and mean (thin) errors of T and S (model-obs) misfits

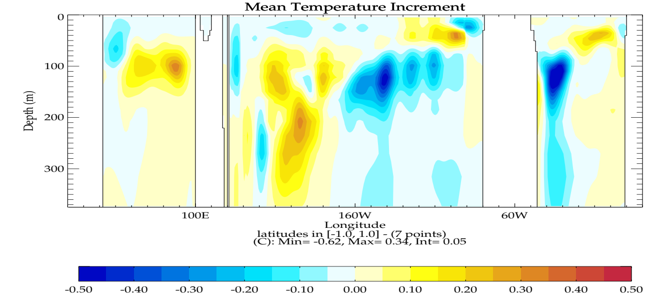


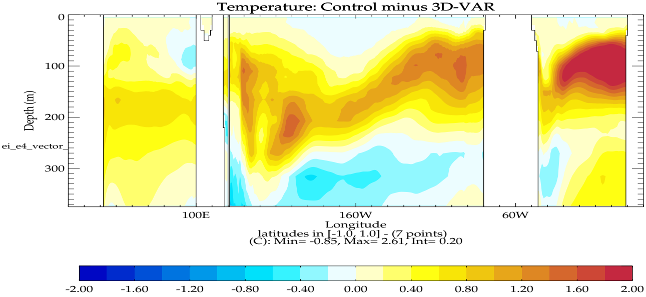
Histograms of T and S (model-obs) misfits



Warm bias

Zonal mean errors at the Equator





The relationship between the increments and model error is not obvious!

1. The assimilation system
2. Diagnostics from a global 2^o reanalysis with an ensemble 3D-Var
3. Diagnostics from a global 1^o 3D-Var reanalysis
4. Conclusions

Conclusions

- Ensemble 3D-Var
  - Difficulties in obtaining an adequate spread (what's new?). Possible remedies: inflation factor, model-error perturbations, more ensemble members,....
  - Nevertheless, positive results obtained: better balanced analyses and improved fit to independent (currents, SLA) data.
  - Expensive, but may be justified if can be used simultaneously for probabilistic forecasting as well as background-error estimation.
- Desroziers *et al.* diagnostics
  - Useful for testing consistency of prescribed and diagnosed **B** and **R**.
  - Can be used for tuning **B** and **R**.
- Model error
  - A major problem that can significantly degrade assimilation results, particularly near the equator.
  - Important to account for it in the assimilation method: model bias correction in 3D-Var or a weak-constraint term in 4D-Var.