

Stochastic array design with the sangoma_arm tool

Pierre De Mey, LEGOS, Toulouse

SANGOMA project meeting
Liège, 1-2 avril 2014

A simple problem

x augmented state vector $(n,1)$ over time interval of interest

(let me insist on the fact that this is an augmented state vector – everything that will be shown in this talk includes time as well as space in the definition of observations and prior state estimate)

y^o observations $(p,1)$ verifying $\mathbf{y}^o = H(\mathbf{x}^t) + \varepsilon$, with:

$H(\cdot)$ observation operator (not necessarily linear, but use linearized approximation)

$$\varepsilon \in N(0, \mathbf{R})$$

Q: how can we characterize the performance of an observational array (H, \mathbf{R}) ?

Assume we have a prior state estimate of **x** and associated error statistics (if not, any observational array will bring valuable information proportionately to its cost):

$\mathbf{x}^f = \mathbf{x}^t + \eta$, with:

$$\eta \in N(0, \mathbf{P}^f)$$

A qualitative/intuitive criterion of array performance

Incremental information brought in by the array (on top of prior):

Innovation vector $\mathbf{d} \equiv \mathbf{y}^o - \mathbf{y}^s = \mathbf{y}^o - H(\mathbf{x}^f) \approx \boldsymbol{\varepsilon} - \mathbf{H}\boldsymbol{\eta}$

The 2nd-order statistics of innovation can be used to characterize that incremental information:

$$\langle \mathbf{d}\mathbf{d}^T \rangle = \mathbf{R} + \mathbf{H}\mathbf{P}^f\mathbf{H}^T$$

Qualitative/intuitive **criterion** of array performance:

- \mathbf{R} “dominates”
 - most of the discrepancies are attributable to observational error
 - observations are not very useful
- $\mathbf{H}\mathbf{P}^f\mathbf{H}^T$ “dominates”
 - most of the discrepancies are attributable to prior state errors
 - observations can be used to identify and correct prior state errors

Towards a formal criterion of array performance

Two paths (among others) to formalize the intuitive order relationship...

Bennett's "array modes" (e.g. Bennett et al., 1997): these are orthonormal rotation vectors β obtained by diagonalizing the representer matrix:

$$\mathbf{H}\mathbf{P}^f\mathbf{H}^T = \beta\lambda\beta^T$$

β : observable degrees of freedom of the physical system for that configuration

λ : spectrum of RM, to be compared to the diagonal of \mathbf{R} (obs. noise floor)

Le Hénaff & De Mey (Le Hénaff et al., 2009): in the general case of non-homogeneous, non-diagonal \mathbf{R} , and observational samples scattered in time, space, and across variables, use spectrum σ and array modes μ of the scaled representer matrix χ :

$$\chi = \mathbf{R}^{-1/2}\mathbf{H}\mathbf{P}^f\mathbf{H}^T\mathbf{R}^{-1/2} = \mu\sigma\mu^T$$

μ : spectrum of SRM, to be compared to the diagonal of \mathbf{I} (obs. noise floor)

Representer Matrix Spectrum (RM Spectrum) method

Le Hénaff, De Mey and Marsaleix, *Ocean Dynamics*, 2009:

$$\text{Scaled Representer Matrix: } \boldsymbol{\chi} = \mathbf{R}^{-1/2} \mathbf{H} \mathbf{P}^f \mathbf{H}^T \mathbf{R}^{-1/2} = \boldsymbol{\mu} \boldsymbol{\sigma} \boldsymbol{\mu}^T \quad (\text{RMS1})$$

- $\boldsymbol{\mu}$: “array modes”
- $\boldsymbol{\sigma}$: spectrum of SRM, to be compared to the diagonal of \mathbf{I} (obs. noise floor)

From (RMS1) and the orthogonality of array modes:

$$\boldsymbol{\mu}^T \boldsymbol{\chi} \boldsymbol{\mu} = \boldsymbol{\mu}^T \mathbf{R}^{-1/2} \mathbf{H} \mathbf{P}^f \mathbf{H}^T \mathbf{R}^{-1/2} \boldsymbol{\mu} = \boldsymbol{\sigma} \quad (\text{RMS2})$$

- $\boldsymbol{\sigma}$ appears as a rotated scaled representer matrix in the new basis defined by $\boldsymbol{\mu}^T$
- $\boldsymbol{\rho}_\mu = \mathbf{P}^f \mathbf{H}^T \mathbf{R}^{-1/2} \boldsymbol{\mu}$ can be seen as a matrix of representer for the array modes = “modal representer”

Stochastic implementation of RM analysis (1/2) (ArM, De Mey, 2010)

Assume we have a way of generating m prior error samples e.g. from forecast Ensemble anomalies, or stochastic modelling.

Matrix of samples (centered): \mathbf{A}^f

We get stochastic estimates:

$$\bullet \hat{\mathbf{P}}^f = \frac{1}{m-1} \mathbf{A}^f \mathbf{A}^{fT} \quad (\text{ARM1})$$

$$\bullet \hat{\boldsymbol{\chi}} = \frac{1}{m-1} (\mathbf{R}^{-1/2} \mathbf{H} \mathbf{A}^f) (\mathbf{R}^{-1/2} \mathbf{H} \mathbf{A}^f)^T = \mathbf{S} \mathbf{S}^T \quad (\text{ARM2})$$

using $\mathbf{S} = \frac{1}{\sqrt{m-1}} \mathbf{R}^{-1/2} \mathbf{H} \mathbf{A}^f$ = scaled Ensemble observation anomalies (e.g. Sakov et al., *Ocean Dynamics*, 2009)

Stochastic implementation of RM analysis (2/2) (ArM, De Mey, 2010)

From (ARM2), the ev problem in the RM Spectrum method is now a sv problem in ArM.

We now have the following stochastic estimates:

- $\hat{\boldsymbol{\sigma}}$ = RM spectrum estimate = squares of the singular values of \mathbf{S}
- $\hat{\boldsymbol{\mu}}$ = Array Mode estimates = singular vectors of \mathbf{S}
- $\hat{\boldsymbol{\rho}}_{\mu} = \frac{1}{\sqrt{m-1}} \mathbf{A}\mathbf{S}^T \hat{\boldsymbol{\mu}}$ = Modal representer estimates

In practice there is no limitation in the choice of observation operator.

- It can operate in space, time, and across variables.
- In practice we calculate $\mathbf{H}\mathbf{A}^f$ as $H(\mathbf{A}^f)$ when calculating \mathbf{S} (e.g. Ensemble members made to generate their own observation proxies).

```

SUBROUTINE sangoma_arm( &
  nstate, nens, nobs, ndof, Af, Df, R, arm_spect, arm, arm_rep, status &
  ) BIND( C, name="sangoma_arm_" )

! PURPOSE
! Calculate array modes and associated quantities

! INPUT
! nstate          state size
! nens            Ensemble size
! nobs            number of observations
! ndof            =MIN(nens,nobs) number of d.o.f.s of problem
! Af              forecast ensemble anomalies, defined as Af(nstate,nens)
! Df              same as Af in data space, defined as Df(nobs,nens)
!                note: df can be directly generated by the model, or
!                linearly calculated as H*Af, H(nobs,nstate) = obs.op.
! R              observation error covariance matrix, def as R(nobs,nobs)

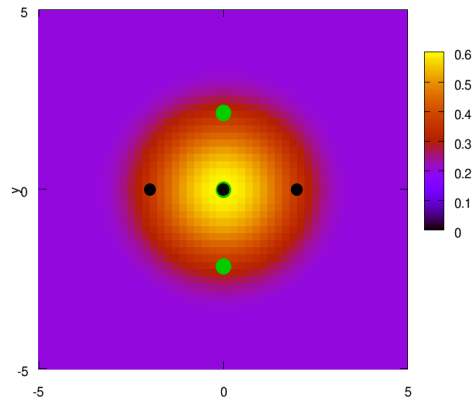
! OUTPUT
! arm_spect       array mode spectrum, defined as arm_spect(ndof)
! arm             array modes, defined as arm(nobs,ndof)
! arm_rep         modal representers, defined as arm_rep(nstate,ndof)
! status          status flag (0=success)

! NOTES
! 1. The actual precision of REAL is to be provided by the compiler. Only
!    KIND=8 will work with the current version (promotion of REAL to DOUBLE
!    PRECISION) because of the use of Dxxxxx BLAS/LAPACK calls. This can
!    be changed in a later version.
! 2. State space and data space are n-dimensional + (optionally) time.
!    - Each state-space sample and data-space sample can contain
!      information from several instants if desired.
!    - Modal representers can span space *and* time.

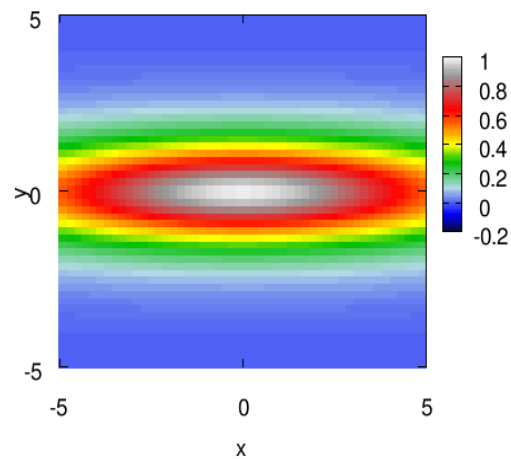
```


A simple 2D, univariate example (1/3)

Prior state error variance



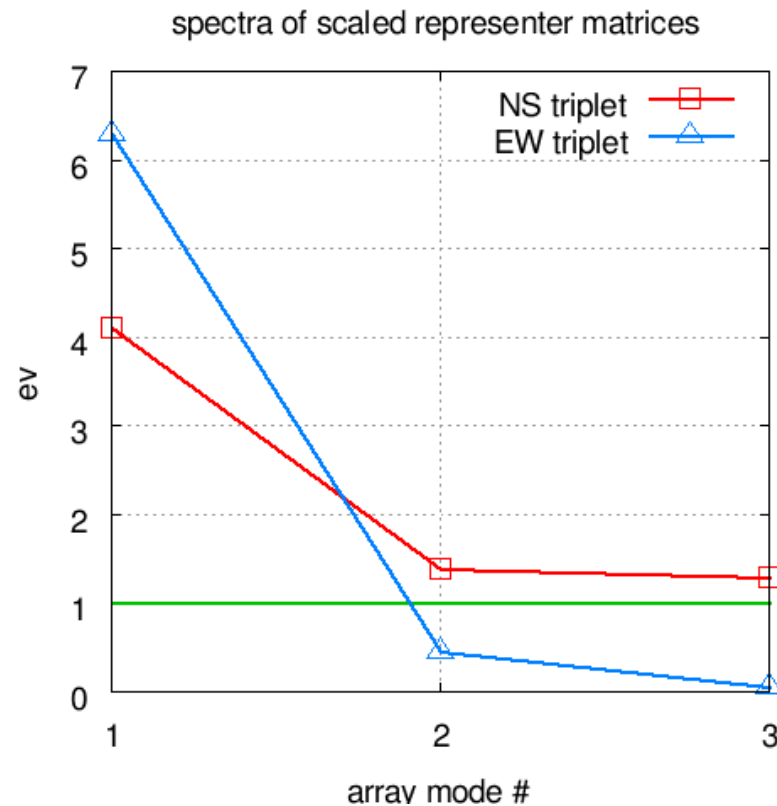
Prior state error correlation



A simple 2D, univariate example (2/3)

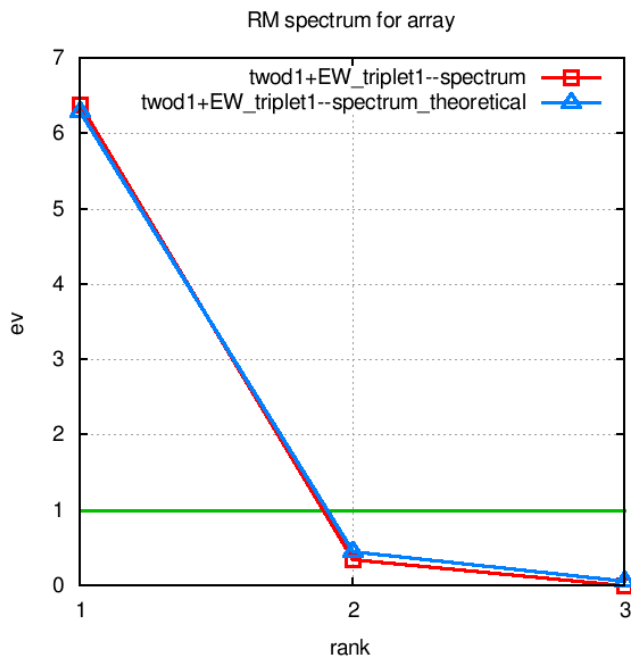
Two observational strategies:

- **N-S triplet** detects more d.o.f.s (3) amidst observational noise
- E-W triplet more redundant (1 d.o.f.)



A simple 2D, univariate example (3/3)

- 20 random samples with \mathbf{P}^f statistics (Gaussian generator)



EW triplet
Theoretical vs.

sangoma_arm

RM spectrum estimates

Detectable modal representers

(Theor.) RM Spectrum

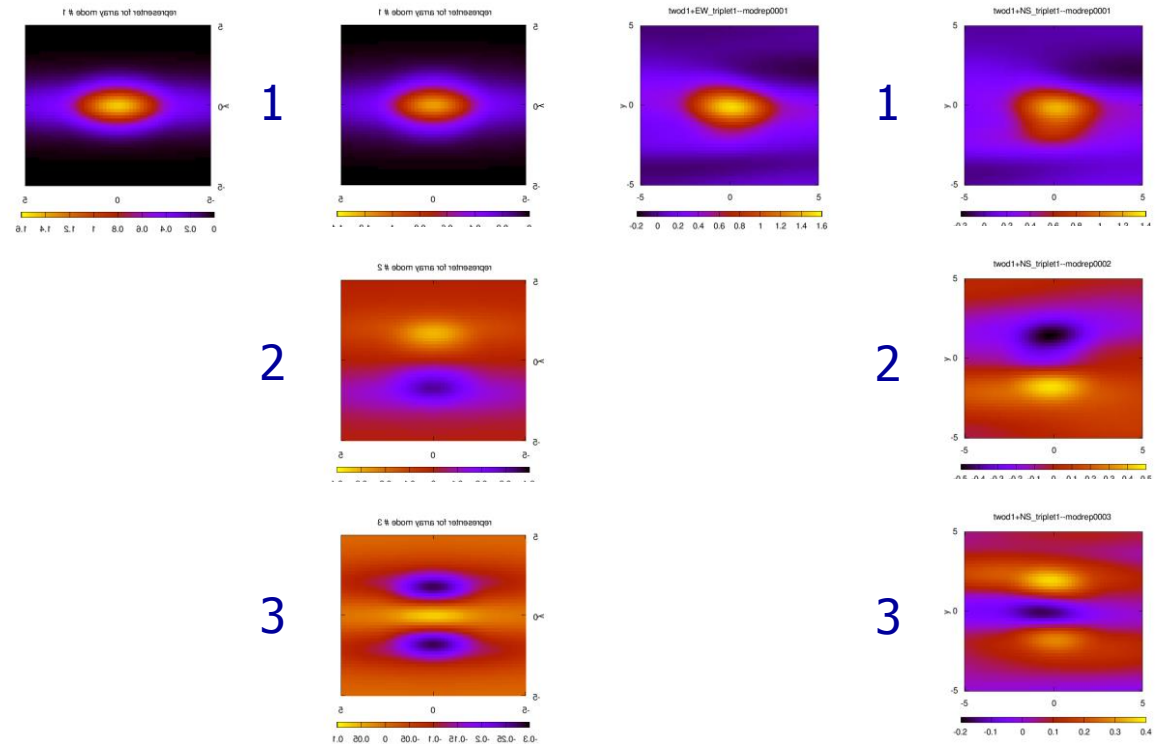
sangoma_arm

EW triplet

NS triplet

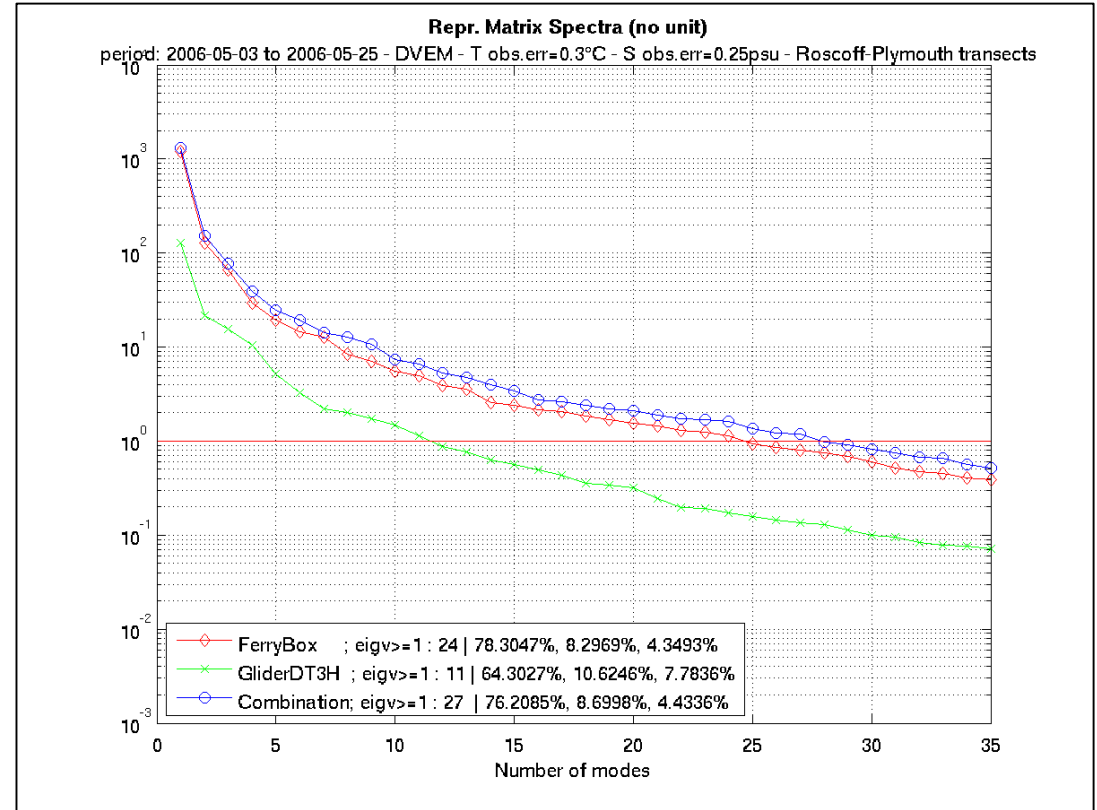
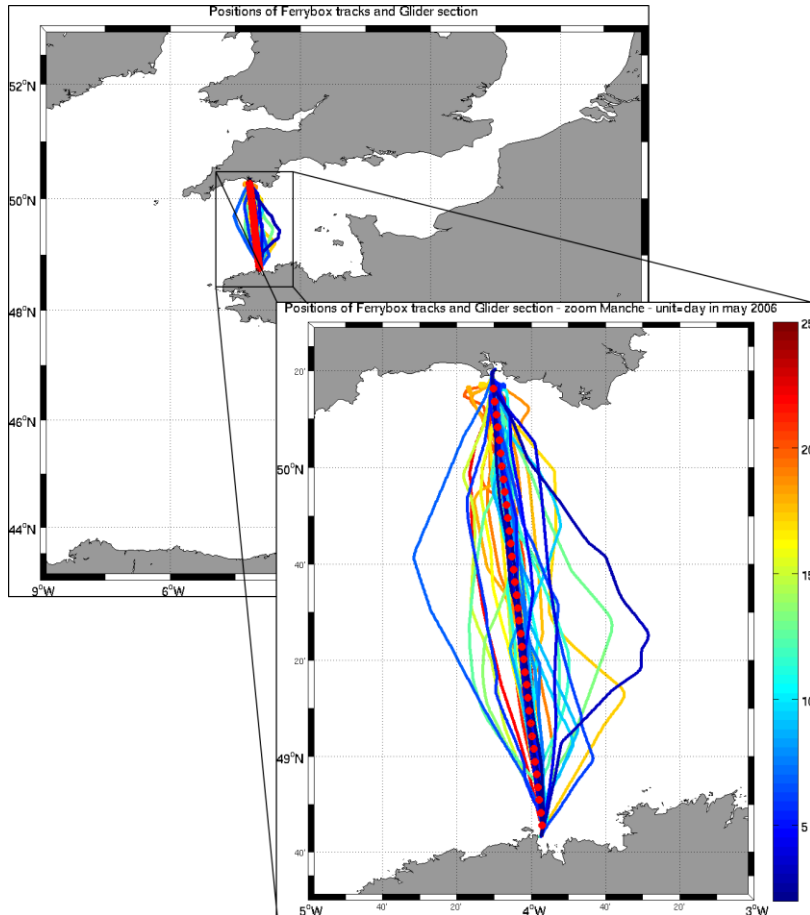
EW triplet

NS triplet



(Color scales not exactly the same – signs not important – only shapes are to be trusted)

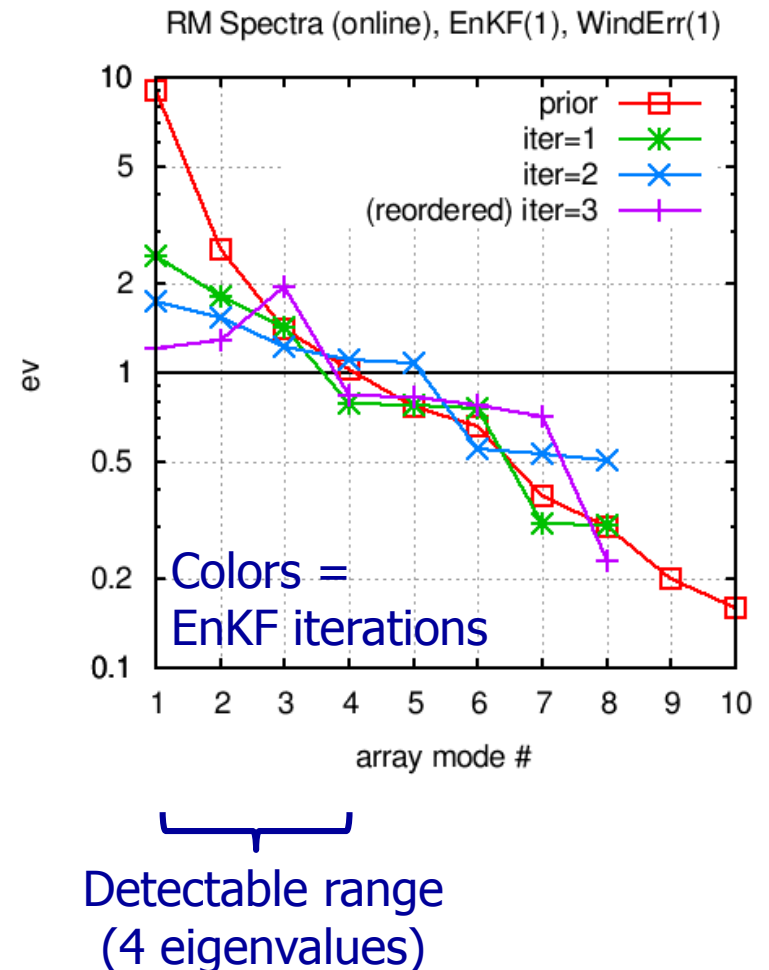
Roscoff-Plymouth Ferrybox vs. Glider, [sangoma_]arm analysis



- Higher repeat cycle of ferrybox critical here, despite being surface only, because of HF model errors (linked to tidal front displacements)

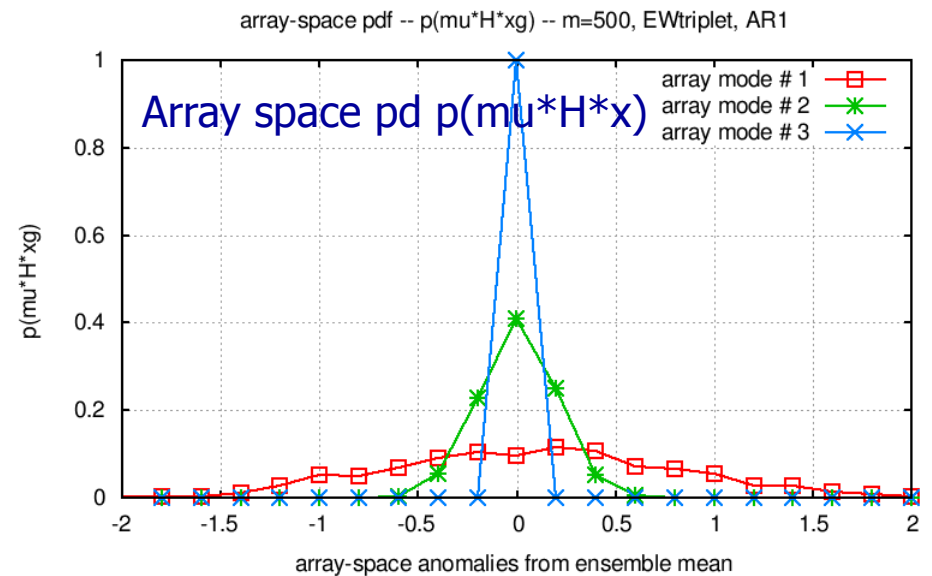
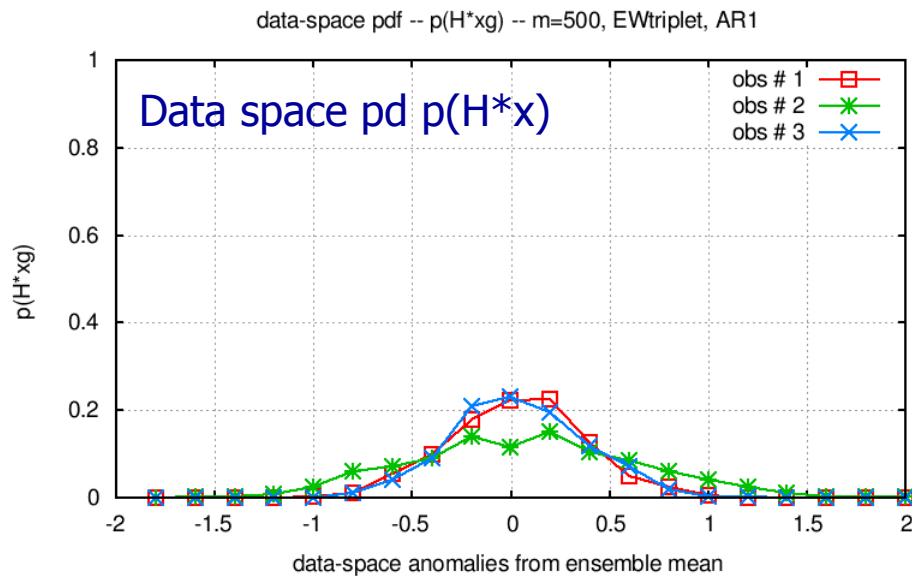
Online [sangoma_]arm diagnostic analysis with 4-D local EnKF

- Assimilate simulated SWOT wide-swath altimeter on 10-day orbit for 2 months in summer 2004 in Bay of Biscay
- Carry out ArM analysis online at each 10-day assim cycle (invariant **H**)
- Localized EnKF (BELUGA)
- Rank is approximately conserved through assimilation
- Spectra whiten in detectable range
 - Array info is being extracted
 - Mostly large-scale and mesoscale error processes constrained
 - No eigenvalue decrease for high-frequency shelf processes → need for sustained observations of such processes



Can array modes help Ensemble consistency analyses?

- Problem: check whether model forecast *pdf* (from Ensemble) and observations (innovation *pdf*) are consistent with each other
- Low-order array-space forecast pd's have broadest base (by design)
 - Hierarchize ensemble consistency checks from easiest to hardest to pass
 - Perhaps use some form of Brier score
- Sangoma_arm_CA tool in preparation



EW triplet, AR1 process, 500 members