Adaptive observations, the Hessian metric and singular vectors

Martin Leutbecher

ECMWF

. . . with thanks to Alexis Doerenbecher

Adaptive observations for NWP

• place additional in situ obs. in sensitive regions in order to constrain forecast error (range/region).

Atlantic:	FASTEX (1997),	TOST (2003)
Pacific:	NORPEX (1998),	WSRP 99, 00,. . . 03

Several objective techniques have been developed that hold promise of identifying optimal sites for additional observations.

• How optimal are the optimal sites really?
Proper test with real obs is outstanding as it requires a large number of additional targeted and un-targeted obs, and a large number of cases.

Perhaps in THORPEX?

• a more general goal for adaptive observation techniques:

Prediction of how forecast uncertainty changes due to intermittent modifications of the observing network.

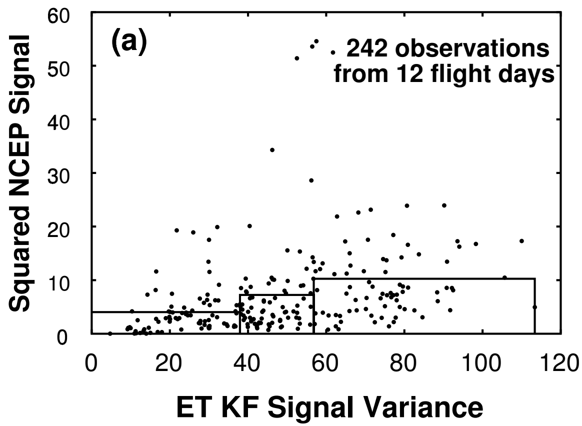
Change of uncertainty predictions can be verified with OSEs in a similar way as the predictions of ensemble spread can be verified.

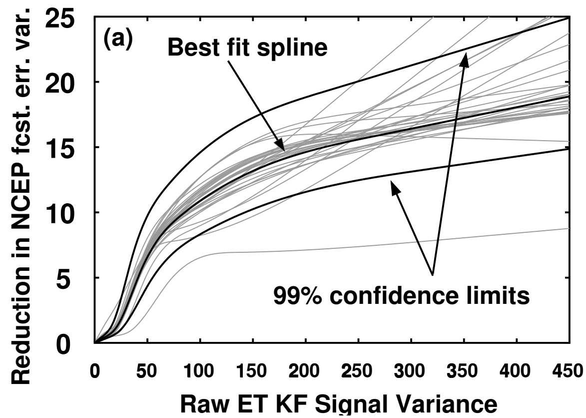
$$\frac{\text{ensemble fc}}{\text{spread-skill relationship}} \qquad \frac{\text{adaptive obs}}{\text{(change of fc uncertainty) - (change of skill) relationship}}$$

Ensemble Transform Kalman filter

- Ensemble Transformation: Bishop and Toth (1999)
- Ensemble Transform KF (ETKF): Bishop, Etherton and Majumdar (2001)

- Prediction of the reduction of forecast error variance: Majumdar et al. (2001)





signal= fc(routine + adaptive obs.) − fc(routine obs.)
y-axis: (sample mean over many) realizations using 3D-Var assimilation
x-axis: variance prediction assuming ETKF assimilation.

$\longrightarrow$ *towards consistency between targeting and data assimilation?*

Techniques to predict changes of fc uncertainty due to the assimilation of additional observations

ingredients: statistics of initial condition errors,
of how they change due to an assimilation of additonal obs.,
perturbation dynamics from the obs. time to the verification time.

dynamics: ensembles ↔ tangent-linear/adjoint techniques

data assimilation: Gaussian statistics

- Kalman filter
  - Ensemble KF
    - pert. obs.
    - sqrt-type (ETKF)
  - Var
  - OI

consistency issue: targeting method ↔ operational assimilation scheme

Adjoint-based methods to predict sensitive regions

Prediction of sensitive regions of the atmosphere using

- adjoint sensitivity with respect to the ICs, Rabier et al. (1996), Langland et al. (1996), Gelaro et al. (1998)
- total energy singular vectors: Buizza and Montani (1999), Gelaro et al. (1999)
- Hessian singular vectors: Leutbecher et al. (2002)

- TE-metric inconsistent with structure functions used in data assimilation, evidence from Cardinali and Buizza (2003):
  Diagnostic of targeting based on total energy SVs: On average, small projection of analysis differences (due to additional obs) on singular vectors, whereas forecast errors have a large projection on SV subspace. What would the same diagnostic yield for Hessian SVs?

But: How many obs? Which error characteristics? What spacing are required in sensitive region to effectively constrain fc error?

Adjoint-based methods to predict forecast error variance reductions

- Assimilation of additional obs. is accounted for.

- Techniques are consistent with the variance estimates of the underlying variational assimilation schemes.

  - prediction of fc error variance reduction due to additional obs in the direction of an adjoint sensitivity: Kalman filter sensitity (Bergot and Doerenbecher 2002)

    The Kalman filter sensitivity is closely related to the sensitivity with respect to observations (Baker and Daley 2000; Doerenbecher and Bergot 2001)

  - prediction of fc error variance reduction in a SV subspace, Hessian reduced rank estimate (Leutbecher 2003)

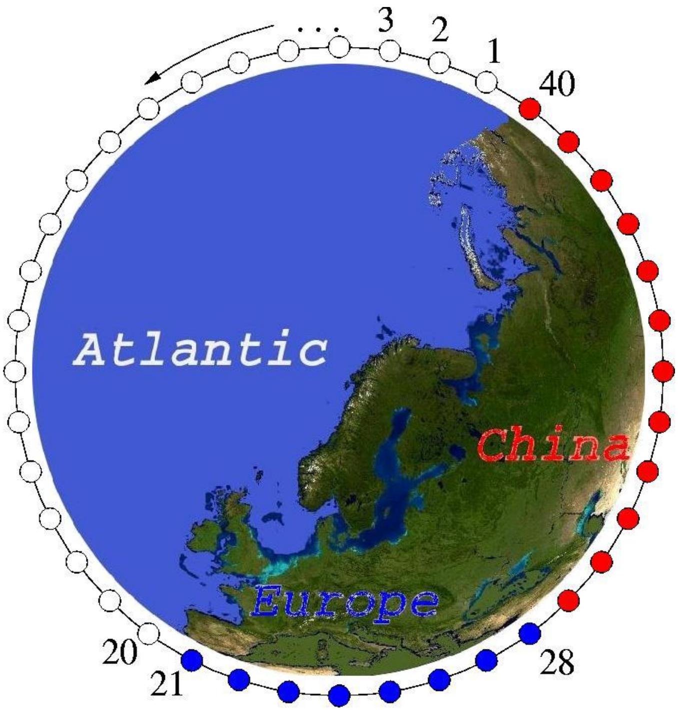
- Both techniques can be seen as reduced rank approximations of Extended Kalman filter predictions of forecast error variance reductions, where the KF is initialized with the static **B** of the variational assimilation scheme.

Outline

- evaluation of predictions of forecast error variance reductions in a simple model
  - Kalman filter
  - Kalman filter - reduced rank estimate
  - OI/3D-Var - reduced rank estimate
- the Hessian reduced rank estimate in an (almost) operational 4D-Var configuration
- limitations/ future directions

Planet L95

“Weather” on Planet L95



$$\frac{dx_i}{dt} = -x_{i-2}x_{i-1} + x_{i-1}x_{i+1} - x_i + F$$

with $i = 1, 2 \dots 40$,
$x_0 = x_{40}, \quad x_{-1} = x_{39}, \quad x_{41} = x_1$
and $F = 8$

(Lorenz, 1995, ECMWF Sem. on Predictability and Lorenz and Emanuel, 1998)

- time unit of 1 corresponds to 5 days
- chaotic system: 13 positive Lyapunov exponents, the largest corresponds to a doubling time of 2.1 d

“Weather” on Planet L95

(1998)

- variables fluctuate about mean in a non-periodic manner with a climatological standard deviation of $\sigma_{\text{clim}} = 3.6$
- perturbation of initial conditions grows with time and its leading edge propagates “eastward” at a speed of about 25 degrees/day
- homogeneous dynamics (invariant under transformation $i \rightarrow i+1$
- system of ODEs integrated with 4th order Runge-Kutta, $\Delta t = 3h$

LORENZ AND EMANUEL

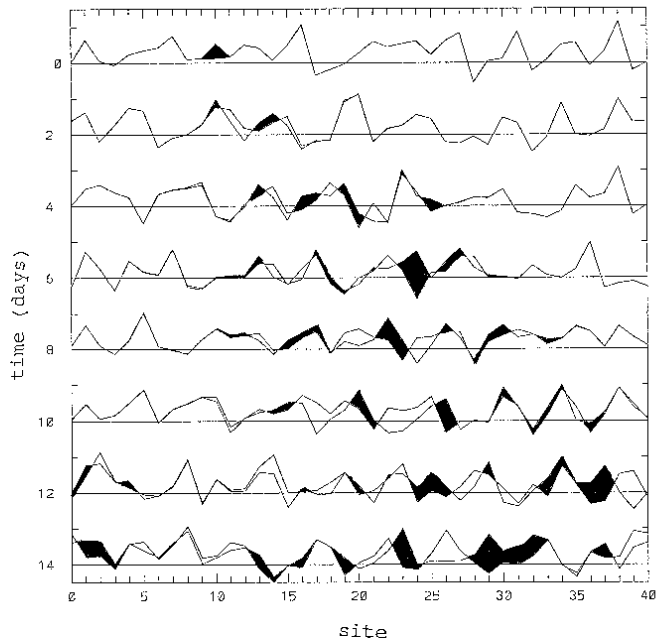


FIG. 3. Longitudinal profiles of X_j as in Fig. 2, but at 2-day intervals, with the initial profile of Fig. 2, and with a second set of profiles superposed. The superposed initial profile is formed by adding 4.0 units to X_{10}. Where the second profile lies above the original one, the area between the profiles is shaded.

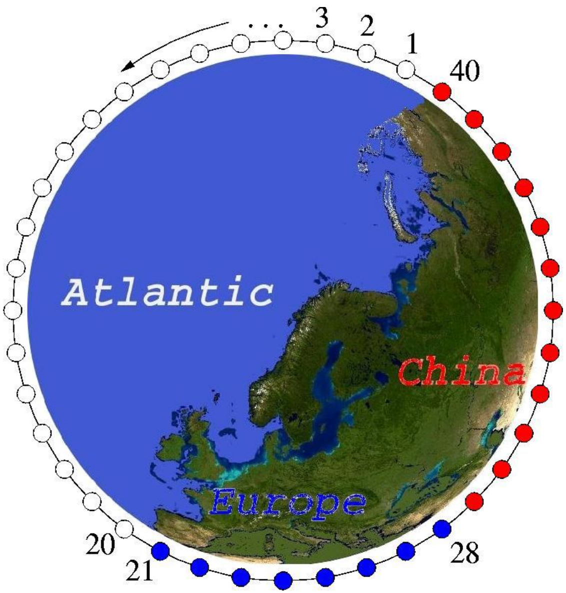
Previous studies of adaptive observing strategies

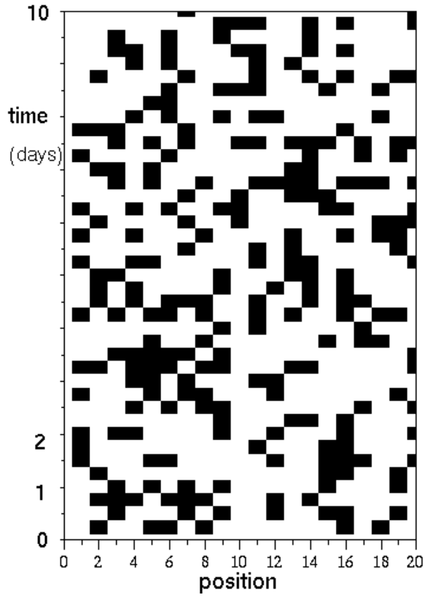
- Lorenz and Emanuel, 1998: obs assimilated with direct insertion; best strategy based on estimate of largest first guess error.
  $F_{\text{model}} = 0.95F$
- Berliner et al., 1999: Extended Kalman filter assimilation and targeting approach (no statistical verification, only 4 cases discussed); $F_{\text{model}} = F$
- Hansen and Smith, 2000: Direct insertion/ Ensemble KF assimilation; $F_{\text{model}} = 0.95F$

> "For analysis errors of sufficiently small magnitude, dynamically based selection schemes will outperform those based only upon uncertainty estimates; it is in this limit that singular vector-based adaptive observation strategies will be productive."

Routine observations used for NWP on Planet L95

- observations every 6 hours, unbiased normally distributed error
- over land (positions 21–40): obs. at every location, $\sigma_o = 0.05\,\sigma_{\text{clim}}$
- over ocean (positions 1–20) : $\sigma_o = 0.15\,\sigma_{\text{clim}}$ at "cloud-free locations". Clouds depend deterministically on **x** but in such a way that the space-time pattern looks like a random process. The probability for occurence of cloud is 0.7.

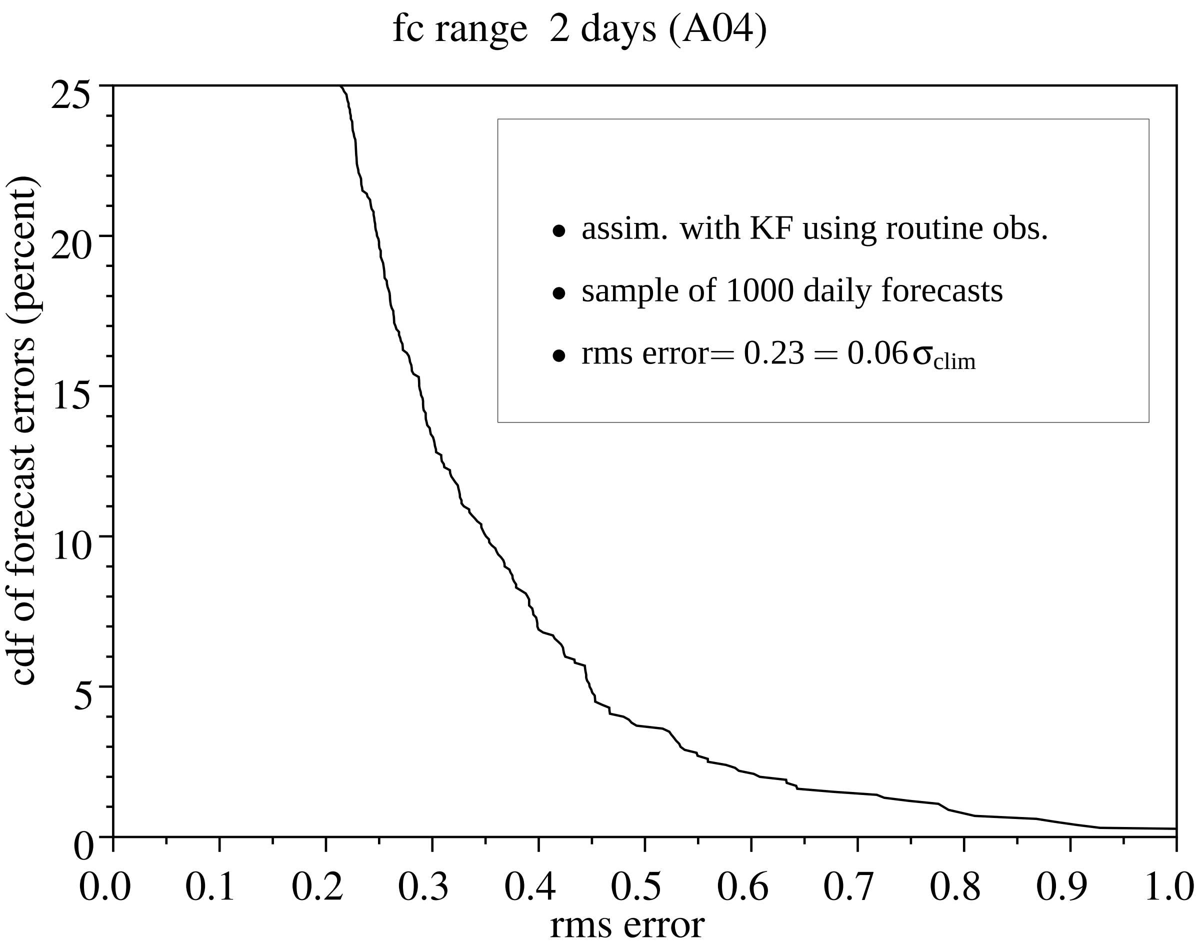




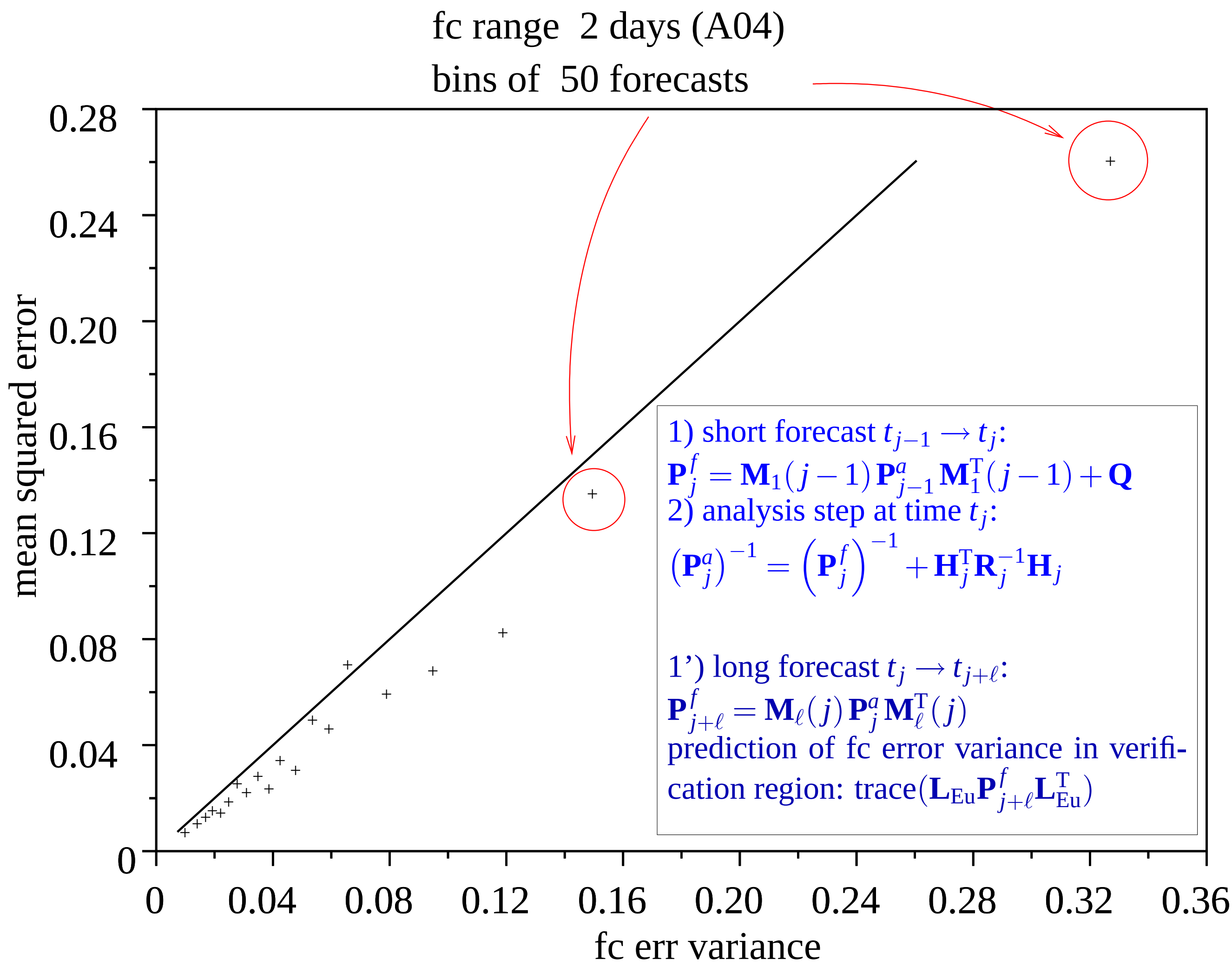
NWP Centres on L95

- EMO: European Meteorological Office, Extended Kalman filter
- MC: Meteo-China, Optimum-Interpolation = 3D-Var
- EMO and MC have very good (perfect) nonlinear forecast models.
- EMO and MC are interested in adaptive observations in order to improve the fc for Europe ($\mathbf{L}_{\mathrm{Eu}}$) at a range of up to 5 days. They both employ targeting schemes that use initial error statistics consistent with their respective assimilation schemes.
- Recently, they had a major experiment (**S**econd **H**emispheric **A**daptive observing, **P**redictability-**I**ntercomparison and **R**esearch **Ex**periment) in which additional observations had been taken every day over a 1000 day period at every ocean point with $\sigma_o = 0.05\sigma_{\mathrm{clim}}$. Routinely, both plan to implement an operational targeting scheme in which a targeted observation will be taken at one ocean site.
- EMO and MC have evaluated their targeting schemes using the data from their 1000 day observing experiment.

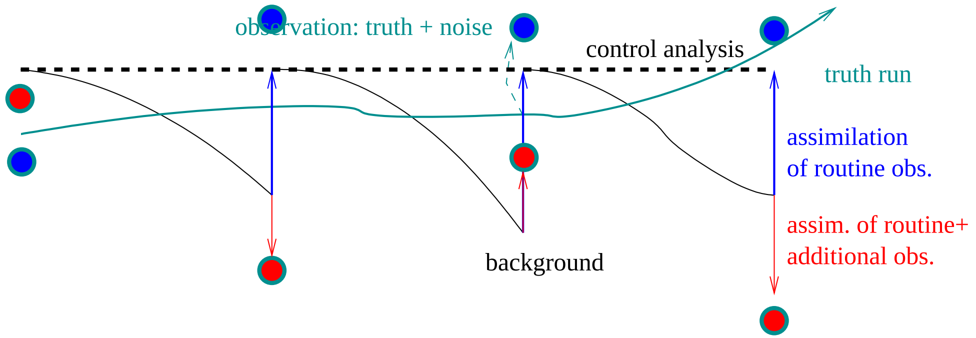
2-day fc errors for Europe (EMO)



(Co)variance prediction with the Kalman filter (routine obs)



Identifying the optimal position for an additional observation



Routine observations

analysis step at time t_j:

$$(\mathbf{P}_r^a)^{-1} = \left(\mathbf{P}_r^f\right)^{-1} + \mathbf{H}_r^{\mathrm{T}}\mathbf{R}_r^{-1}\mathbf{H}_r$$

long forecast $t_j \longrightarrow t_{j+\ell}$:

$$\mathbf{P}_r^f = \mathbf{M}\,\mathbf{P}_r^a\,\mathbf{M}^{\mathrm{T}}$$

Routine obs + additional obs. at position i

analysis step at time t_j:

$$(\mathbf{P}_i^a)^{-1} = \left(\mathbf{P}_r^f\right)^{-1} + \mathbf{H}_r^{\mathrm{T}}\mathbf{R}_r^{-1}\mathbf{H}_r + \mathbf{H}_i^{\mathrm{T}}\mathbf{R}_i^{-1}\mathbf{H}_i$$

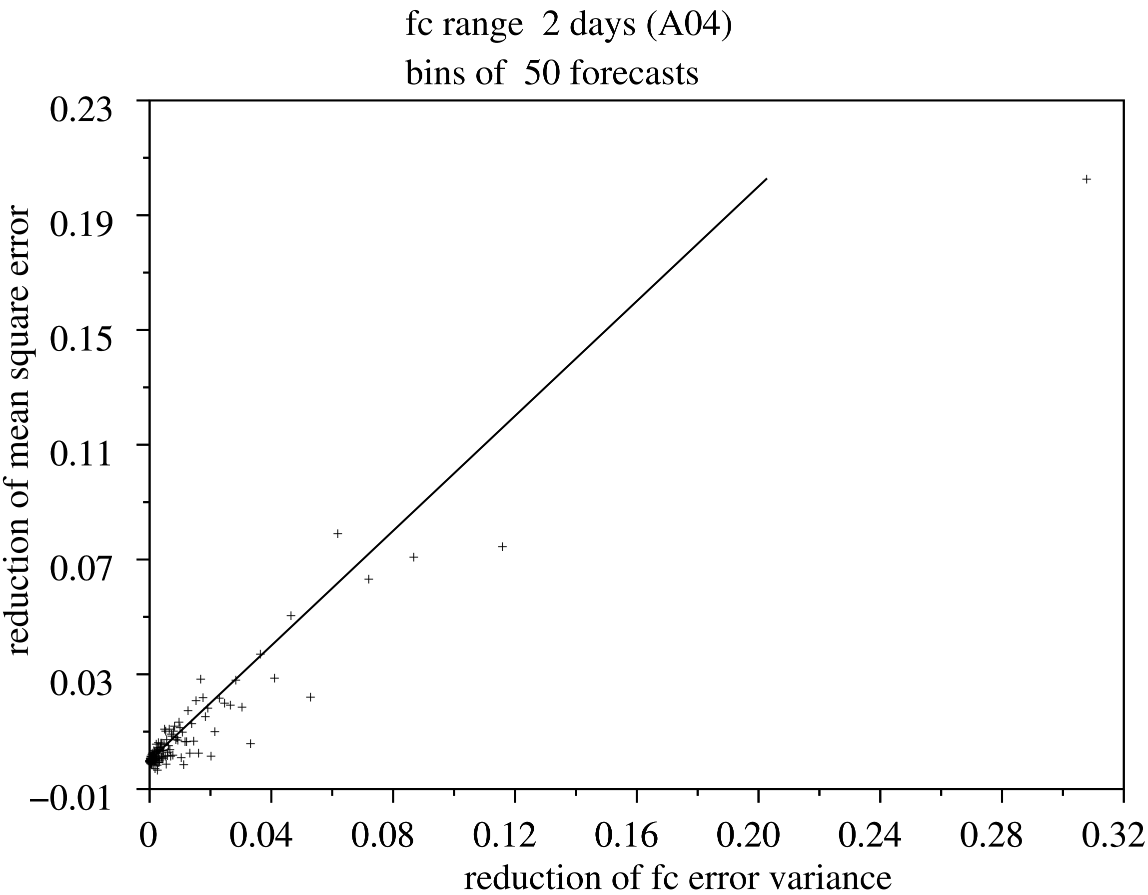
long forecast $t_j \longrightarrow t_{j+\ell}$:

$$\mathbf{P}_i^f = \mathbf{M}\,\mathbf{P}_i^a\,\mathbf{M}^{\mathrm{T}}, \qquad \text{for all } i = 1, \ldots 20$$

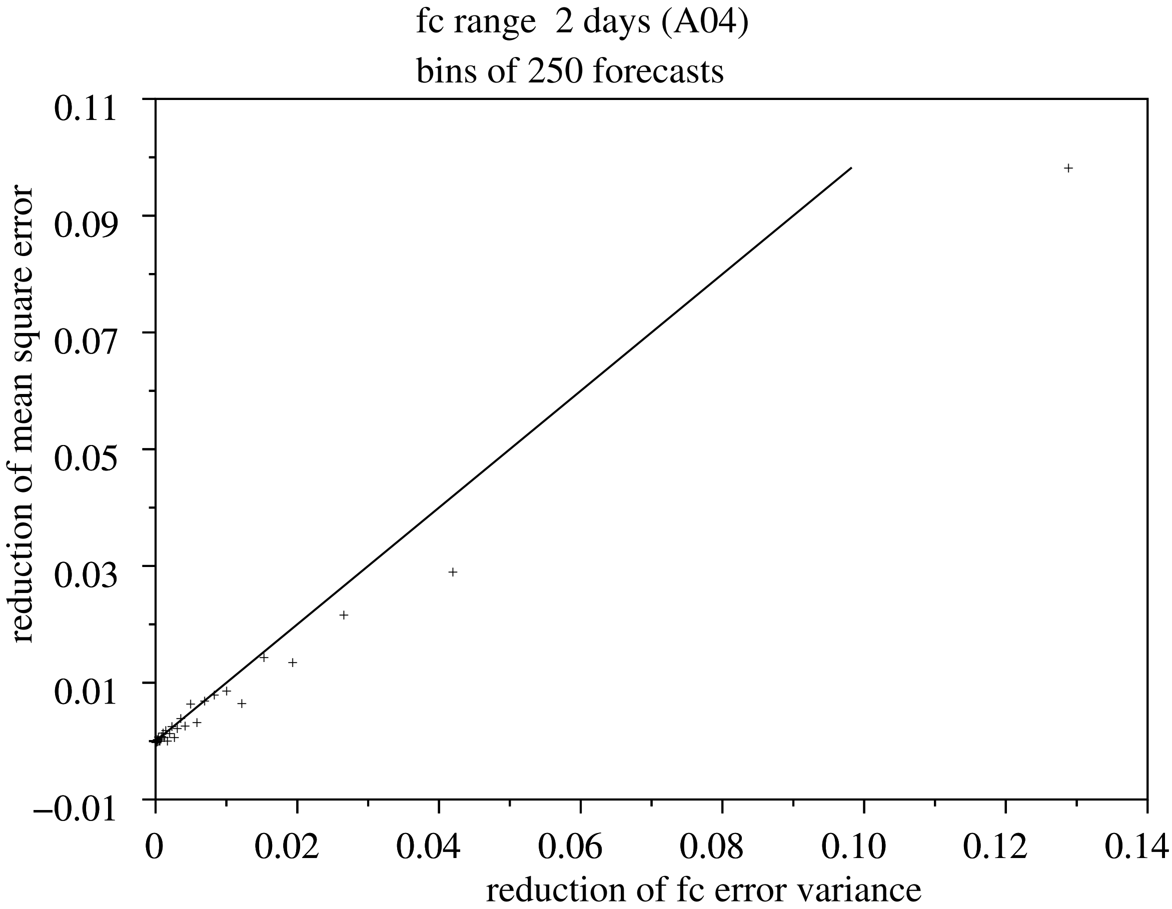
Optimal position i_*: select i that gives maximum reduction of fc error variance.

$$\max_{i=1\ldots20}\ \mathrm{trace}\left(\mathbf{L}_{\mathrm{Eu}}\left(\mathbf{P}_r^f - \mathbf{P}_i^f\right)\mathbf{L}_{\mathrm{Eu}}^{\mathrm{T}}\right)$$

Verification of the predictions of forecast error variance reductions due to additional observations (Kalman filter)



Verification of the predictions of forecast error variance reductions due to additional observations (Kalman filter)



Approximations to the Kalman filter approach

1. At EMO: fc error covariance estimate

$$\mathbf{M}\,\mathbf{P}^a\,\mathbf{M}^{\mathrm{T}}$$

reduced rank: evolve covariances only in a subspace

2. At MC: static background error covariances (OI≈ 3D-Var ≈ 4D-Var with obs. at beginning of assim. window)

replace $(\mathbf{P}^a)^{-1} = \left(\mathbf{P}^f\right)^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}$

by $\mathbf{A}^{-1} = \mathbf{B}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}$

Reduced rank estimate of forecast error variance

- Compute variance of forecast errors only in a subspace of leading singular vectors.

$$\mathrm{trace}(\widehat{\mathbf{\Pi}}_n \mathbf{L}_{\mathrm{Eu}} \mathbf{P}^f \mathbf{L}_{\mathrm{Eu}}^{\mathrm{T}} \widehat{\mathbf{\Pi}}_n^{\mathrm{T}}) \text{ instead of } \mathrm{trace}(\mathbf{L}_{\mathrm{Eu}} \mathbf{P}^f \mathbf{L}_{\mathrm{Eu}}^{\mathrm{T}})$$

- Here, $\widehat{\mathbf{\Pi}}_n$ denotes the projection on the subspace of the leading n (left) singular vectors of $\mathbf{L}_{\mathrm{Eu}}\mathbf{M}$.

- The inverse of the routine analysis error covariance matrix $(\mathbf{P}_r^a)^{-1}$ is an appropriate choice for the initial time metric because the SVs computed with this metric evolve into the leading eigenvectors of the routine forecast error covariance matrix $\mathrm{trace}(\mathbf{L}_{\mathrm{Eu}} \mathbf{P}_r^f \mathbf{L}_{\mathrm{Eu}}^{\mathrm{T}})$.

- Notation: matrix containing the leading n initial SVs as column vectors $\mathbf{V}_n = (\mathbf{v}_1 \ldots \mathbf{v}_n)$.

Variance reductions in the singular vector subspace

- Representation of analysis error covariances in the subspace spanned by the leading n initial SVs:

$$\begin{array}{rll} \text{routine network:} & \mathbf{V}_n\mathbf{V}_n^{\mathrm{T}}, \text{ where} & \mathbf{V}_n^{\mathrm{T}}\left(\mathbf{P}_r^a\right)^{-1}\mathbf{V}_n = \mathbf{I} \\ \text{modified network } i^{\dagger}\text{:} & \mathbf{V}_n\boldsymbol{\Gamma}_i\boldsymbol{\Gamma}_i^{\mathrm{T}}\mathbf{V}_n^{\mathrm{T}}, \text{ where} & \left(\mathbf{V}_n\boldsymbol{\Gamma}_i\right)^{\mathrm{T}}\left(\mathbf{P}_i^a\right)^{-1}\left(\mathbf{V}_n\boldsymbol{\Gamma}_i\right) = \mathbf{I} \end{array}$$

†) involves approximating $\mathbf{P}_i^a$ by a block-diagonal matrix in the space in which $\mathbf{P}_r^a$ is the identity.

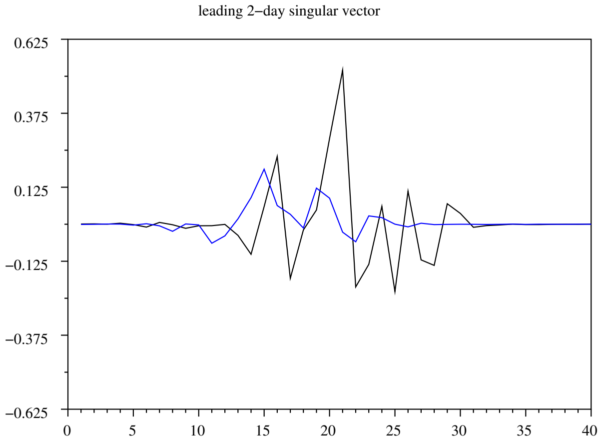
- the transformation matrix $\boldsymbol{\Gamma}_i$ is an inverse square root of the $n \times n$ matrix $\mathbf{C}_i = \mathbf{V}_n^{\mathrm{T}}\left(\mathbf{P}_i^a\right)^{-1}\mathbf{V}_n = \mathbf{I}_n + \mathbf{V}_n^{\mathrm{T}}\mathbf{H}_i^{\mathrm{T}}\mathbf{R}_i^{-1}\mathbf{H}_i\mathbf{V}_n$. Matrix $\mathbf{C}_i$ expresses the modified analysis error covariance metric in the basis of the singular vectors.

- forecast error variance:

$$\mathrm{trace}\left(\widehat{\boldsymbol{\Pi}}_n\mathbf{L}_{\mathrm{Eu}}\mathbf{P}^f\mathbf{L}_{\mathrm{Eu}}^{\mathrm{T}}\widehat{\boldsymbol{\Pi}}_n^{\mathrm{T}}\right) = \begin{cases} \sum_{j=1}^{n}\sigma_j^2 & \text{routine network} \\ \mathrm{trace}\left(\boldsymbol{\Gamma}^T\mathrm{diag}(\sigma_1^2,\ldots,\sigma_n^2)\boldsymbol{\Gamma}\right) & \text{modified network,} \end{cases}$$

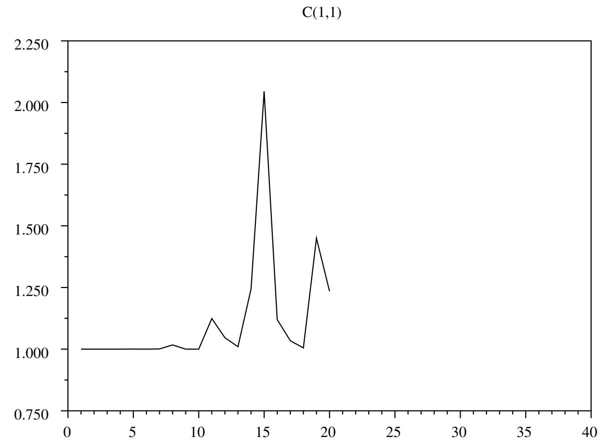
where σ_j denotes the singular value SV $\mathbf{v}_j$.

A rank-1 example



leading initial singular vector $\mathbf{v}_1$
leading evolved singular vector $\mathbf{M}\mathbf{v}_1$

2-day optimization period, projection operator for "Europe" (positions 21–28).

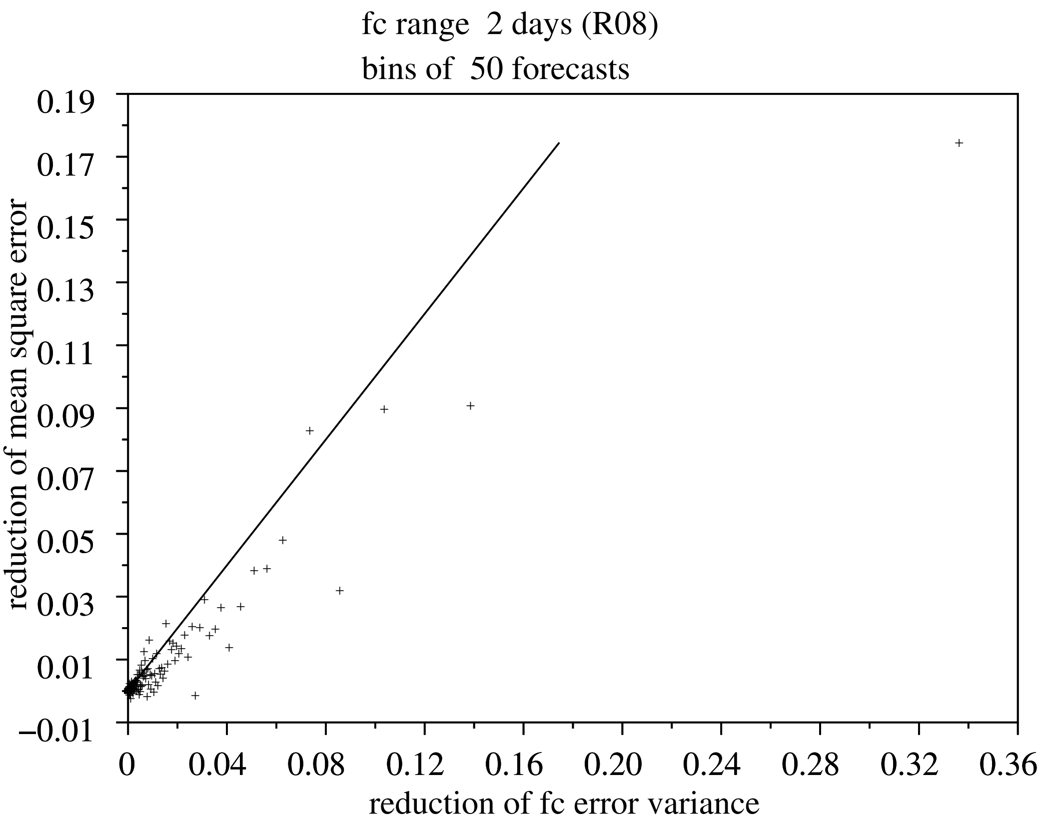


The modified analysis error covariance metric evaluated in the direction of the leading SV

$$C_i = \mathbf{v}_1^{\mathrm{T}} \left(\mathbf{P}_i^a\right)^{-1} \mathbf{v}_1 = 1 + v_1^2(i)/\sigma_o^{-2}$$

- The predicted forecast error variance for an additional observation at i is σ_1^2/C_i.
- An observation at position 15 would result in a halfing of the 2-day forecast error variance in the direction of the leading evolved SV.

Evaluation of the KF rank-1 predictions of forecast error variance reductions due to additional observations



Approximation of the background error covariances

Replace the routine background error covariance matrix $\mathbf{P}_r^f$ by a static background error covariance matrix $\mathbf{B}$ (OI at MC)

- in the variance prediction (targeting) and
- in the assimilation algorithm.

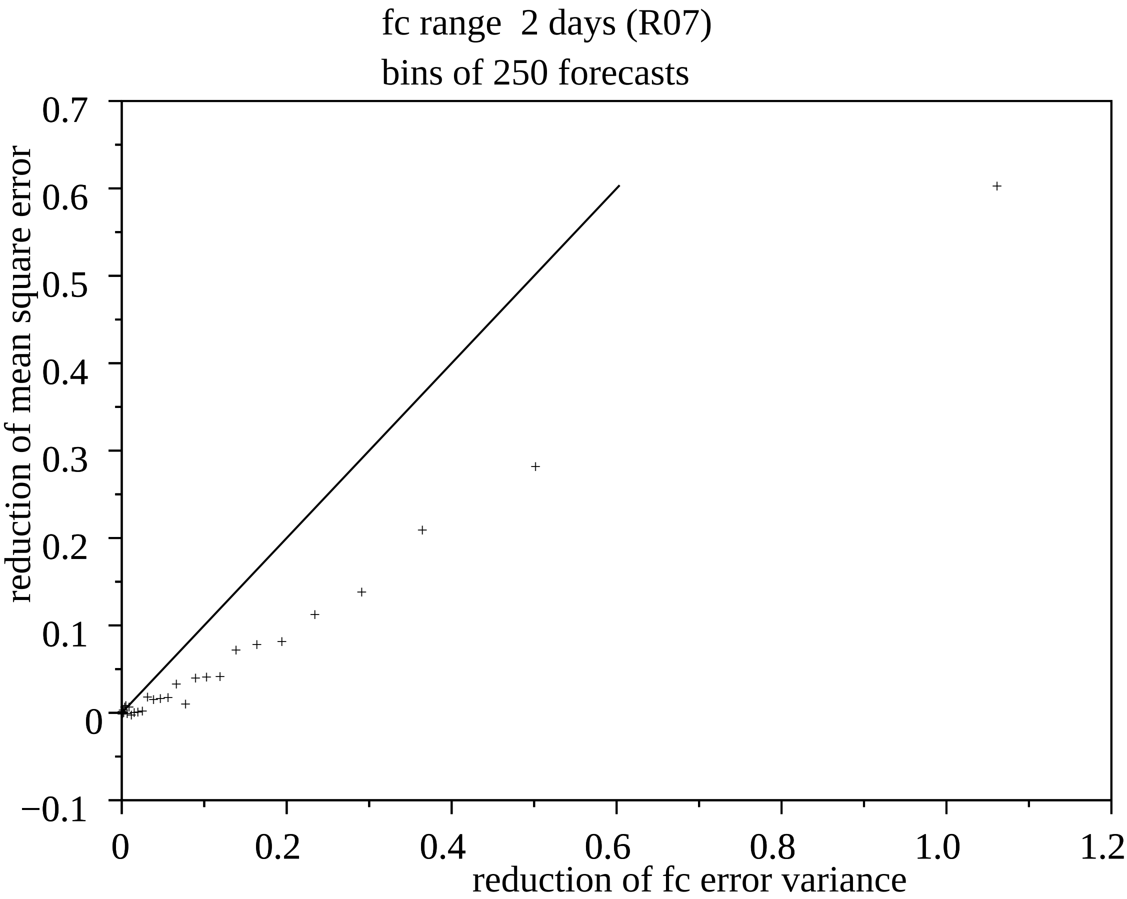
MC has the *magic* **B**: The static background error covariance matrix is a sample covariance matrix computed from forecast − truth differences.
Several iterations of updating $\mathbf{B} = \langle(\mathbf{x}^f - \mathbf{x}^t)(\mathbf{x}^f - \mathbf{x}^t)^{\mathrm{T}}\rangle$ over the 1000 day sample.

KF versus OI performance, global rms errors

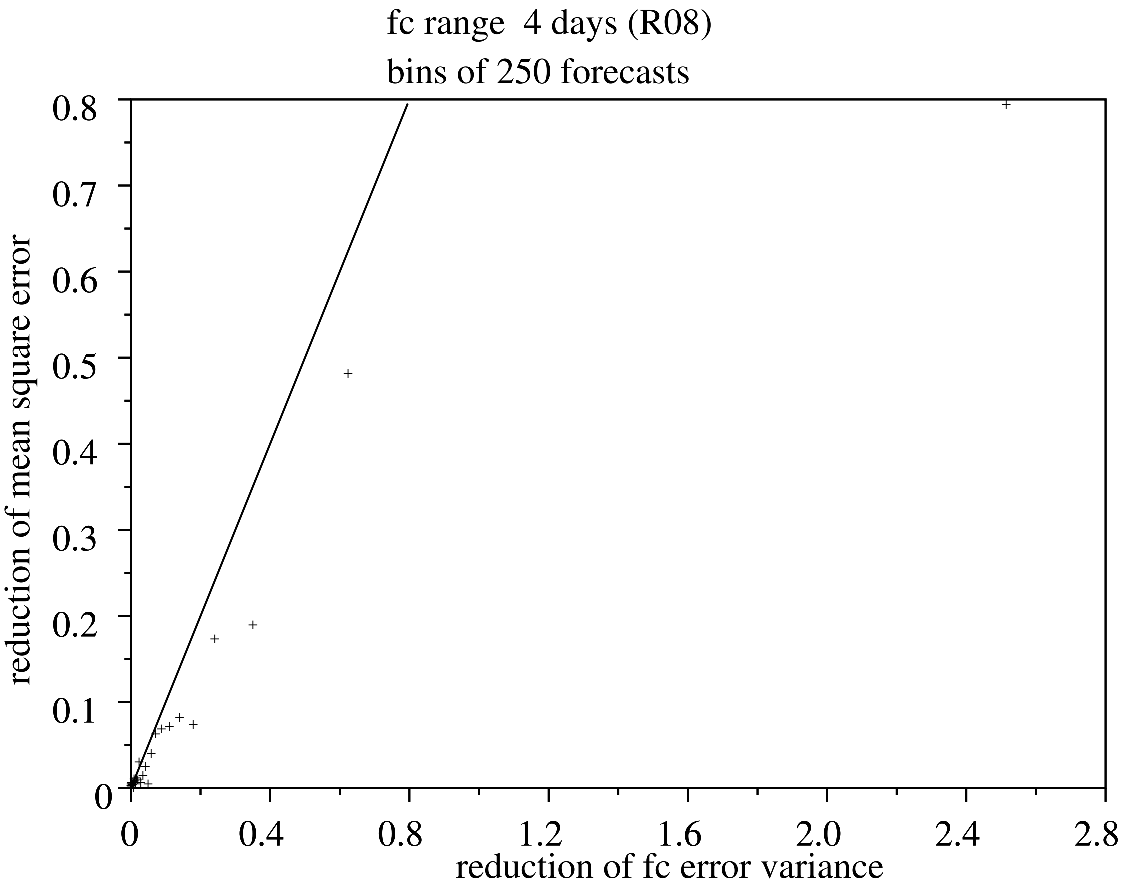
range (d)	0	1	2	3	4	5
KF	0.11	0.17	0.25	0.38	0.57	0.80
OI	0.33	0.49	0.74	1.04	1.38	1.74

$\sigma_{\text{clim}} = 3.6$, $\sigma_{o,\text{land}} = 0.18$, $\sigma_{o,ocean} = 0.54$

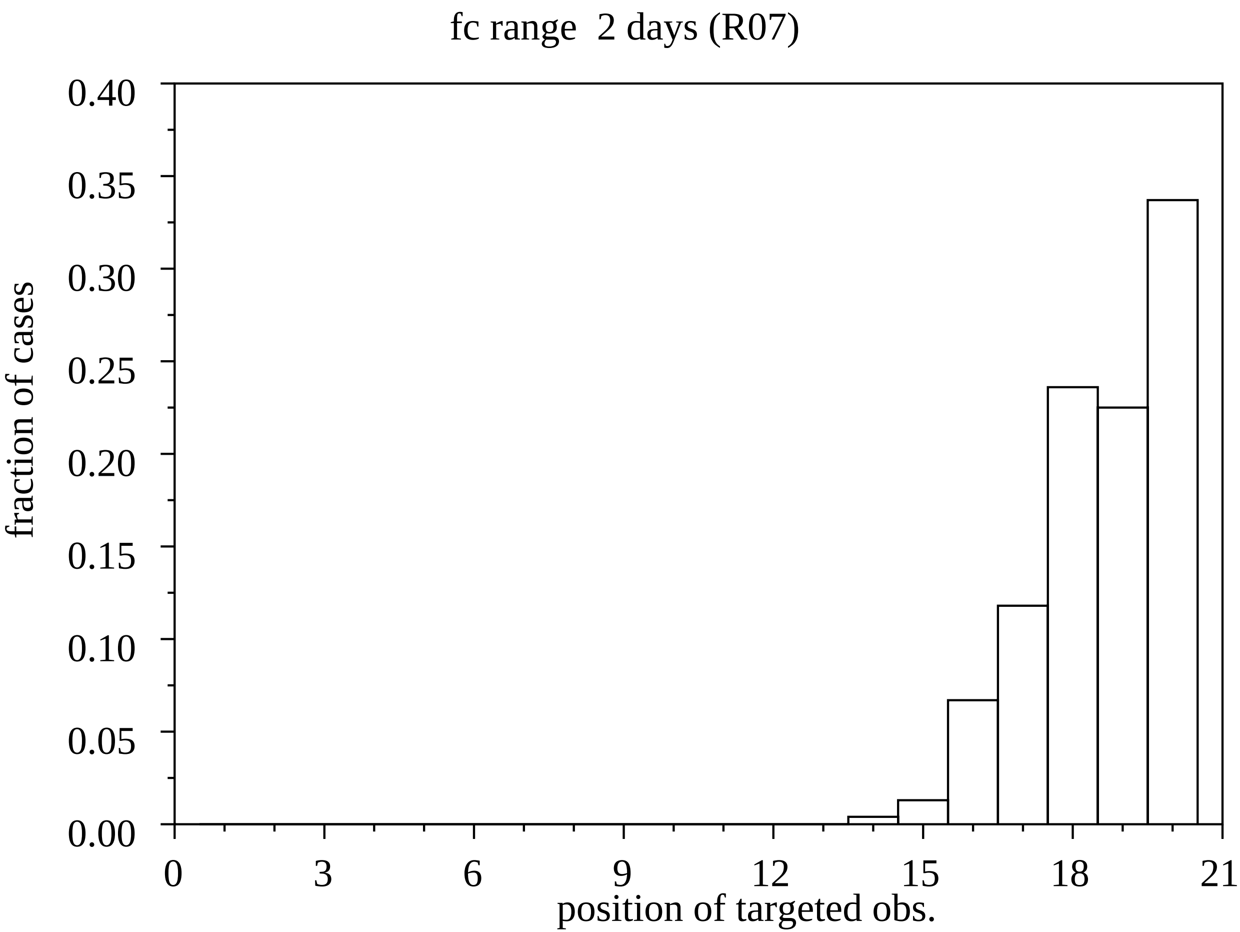
Verification of 2-day forecast error variance reductions: OI-rank 1



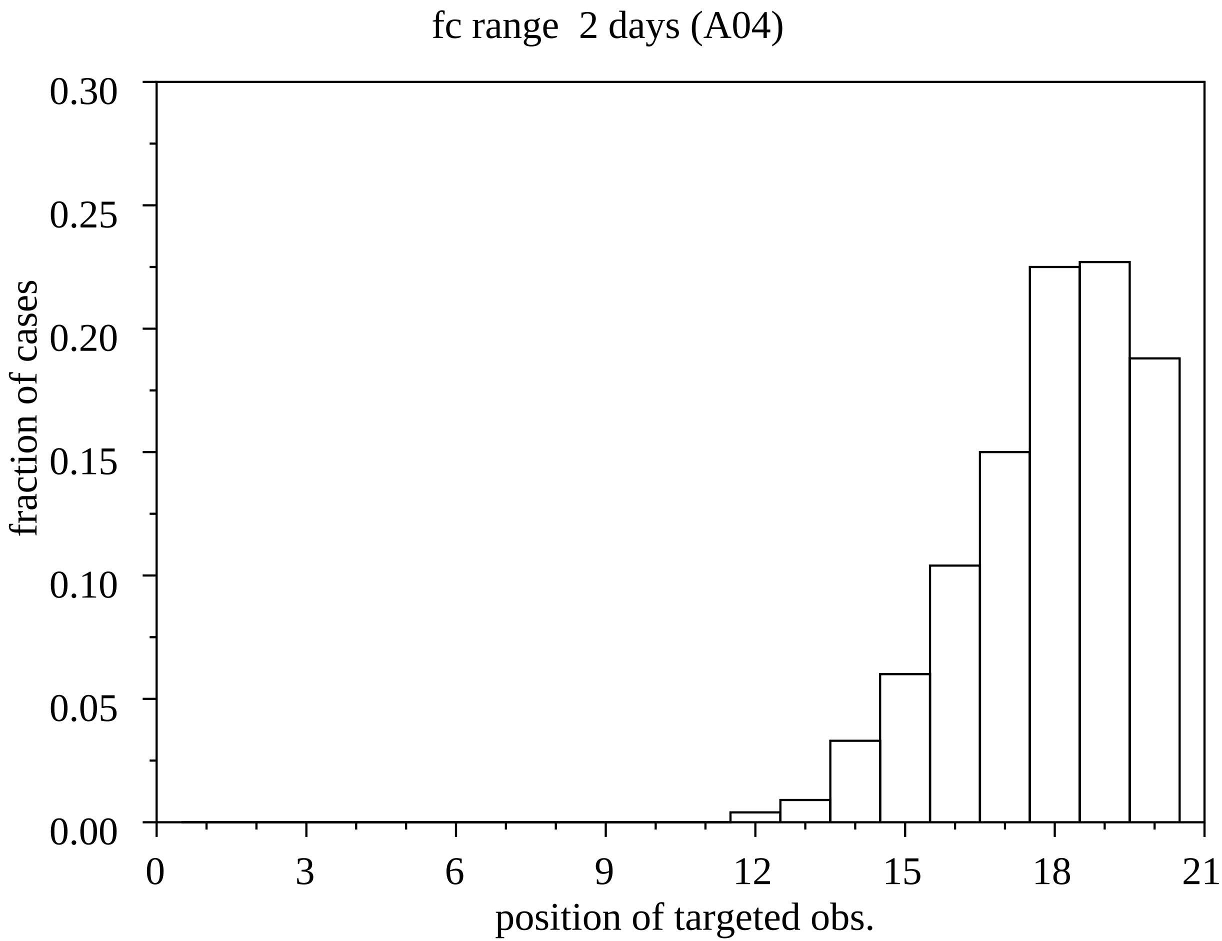
Verification of 4-day forecast error variance reductions: KF-rank 1



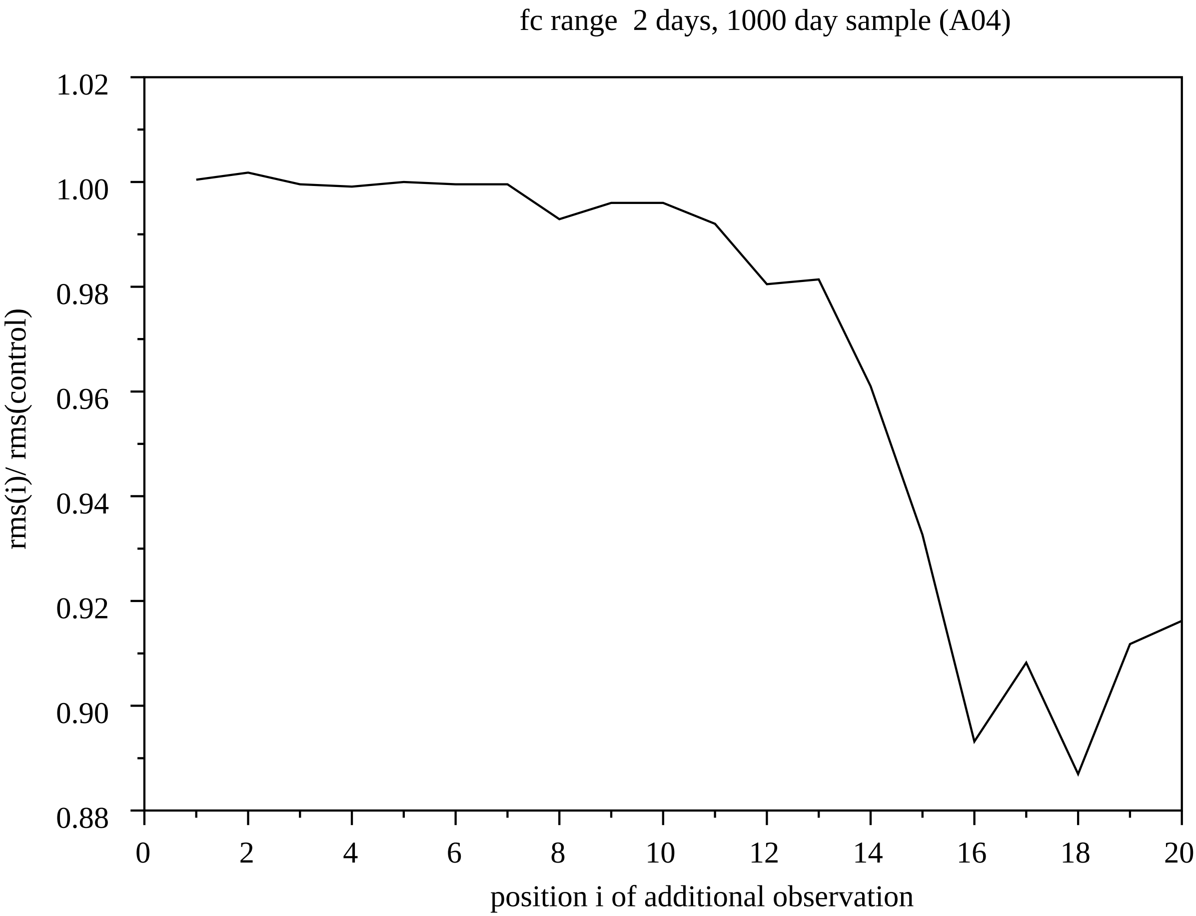
Distribution of the optimal position (OI-rank 1) for an additional obs in the Atlantic to improve the 2-day fc over Europe



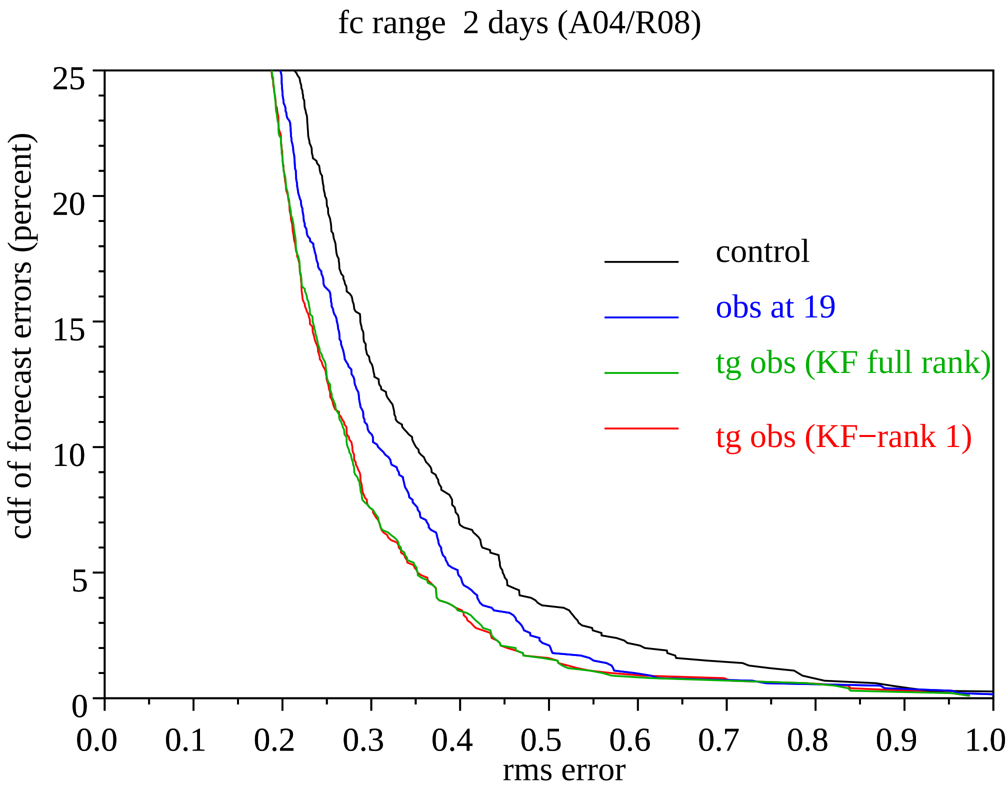
Distribution of the optimal position for an additional obs in the Atlantic to improve the 2-day fc over Europe (KF full rank)



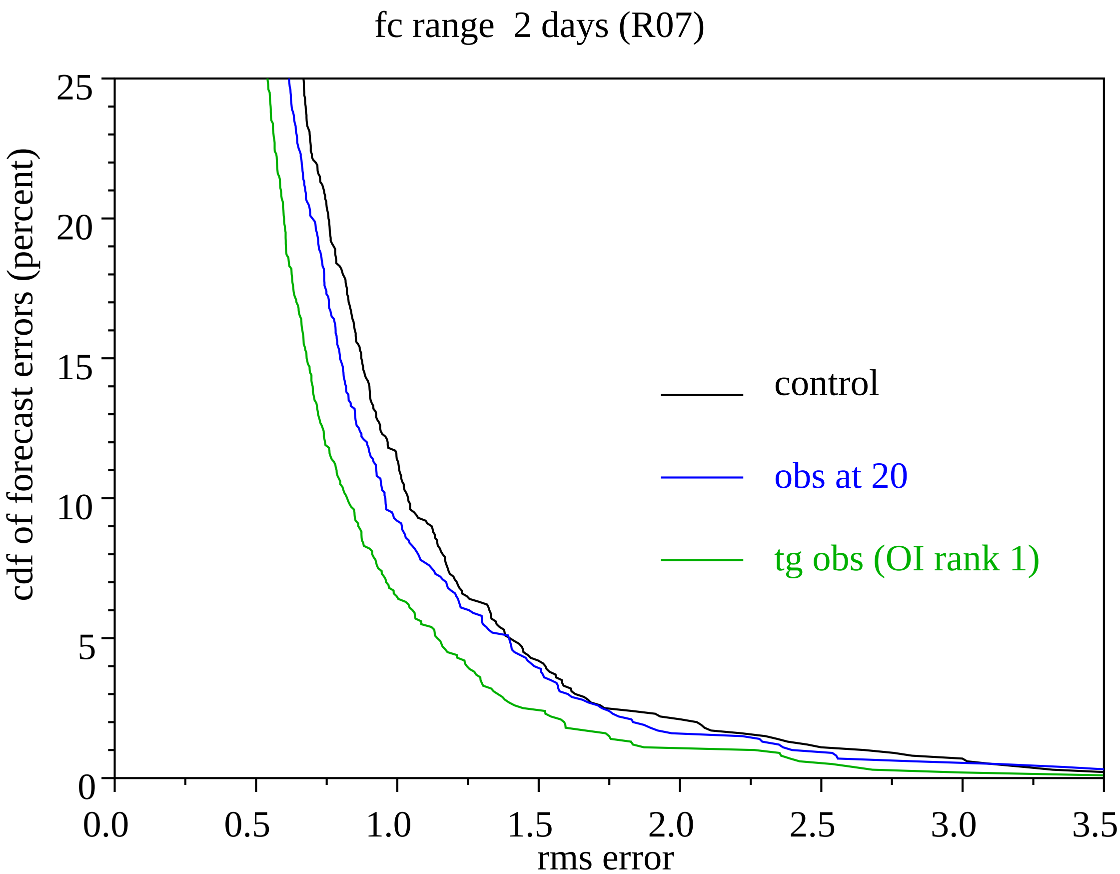
Reduction of 2-day forecast error over Europe due to additional observation at position *i* (KF)

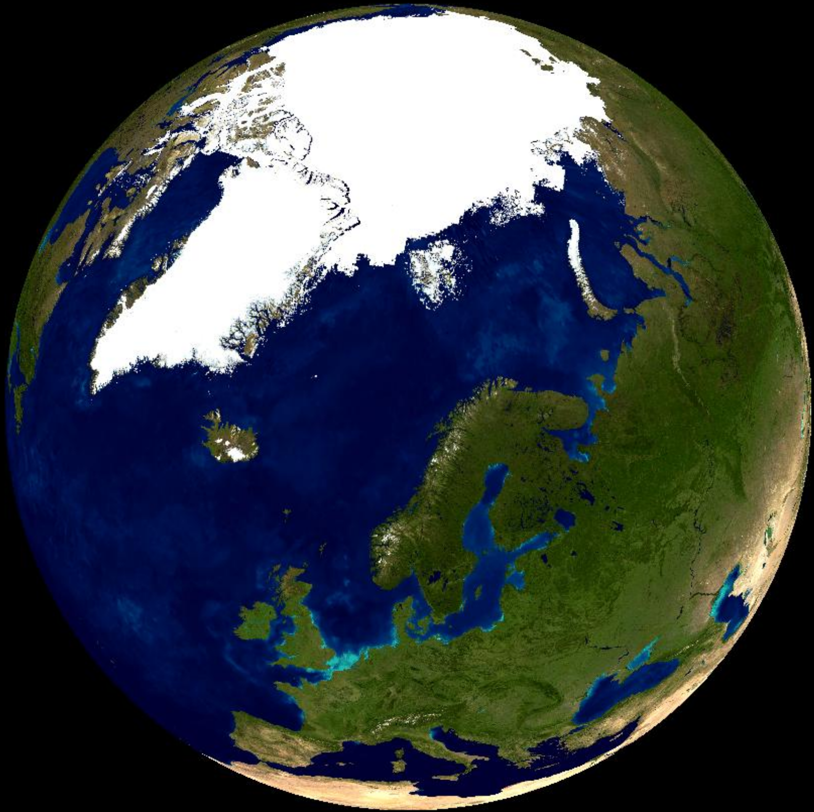


Distribution of 2-day forecast errors: KF-full rank versus KF-rank 1



Distribution of 2-day fc errors over Europe (OI-rank 1)





- prediction of forecast error variance reductions due to additional observations

- rank reduction using a subspace of Hessian singular vectors using the routine observations in J_o

- initial error estimates are consistent with the covariance estimates of an (almost) operational variational data assimilation scheme (6 h 4D-Var, T42 inner loop only)

The Hessian $\quad \dfrac{\partial\partial J}{\partial x_i \partial x_j}$

Ludwig Otto Hesse
(1811–1874, Königsberg, Heidelberg, München)

- Variational data assimilation, incremental formulation, “inner loop”

$$J(\mathbf{x}) = \frac{1}{2}\mathbf{x}^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{x} + \frac{1}{2}(\mathbf{H}\mathbf{x} - \mathbf{d})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{H}\mathbf{x} - \mathbf{d}),$$

where **x** denotes the departure from a background state and **d** the departure of the observed values from the background interpolated to observation locations.

- Currently, the background error covariances **B** are estimated from an ensemble of analyses.
- An estimate of the analysis error covariances is obtained by the Hessian

$$\nabla\nabla J = \mathbf{B}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}$$

A reduced rank estimate of forecast error variance reductions based on the Hessian

Similar to KF/OI-reduced rank estimates but uses an analysis error covariance estimates based on the Hessian.

- subspace: Hessian singular vectors $\mathbf{v}_i$ computed with the metric $\nabla\nabla J_{\text{routine}}$ based on (an estimate of) the routine observing network
- efficient computation of the analysis error covariance metric for modified obs. network (routine + additional) in the subspace:

$$C_{ij} = \mathbf{v}_i^{\mathrm{T}} \nabla\nabla J_{\text{mod}} \mathbf{v}_j = \mathbf{v}_i^{\mathrm{T}} \left(\nabla\nabla J_{\text{routine}} + \mathbf{H}_a^{\mathrm{T}} \mathbf{R}_a^{-1} \mathbf{H}_a \right) \mathbf{v}_j$$
$$= \delta_{ij} + \left(\mathbf{H}_a \mathbf{v}_i \right)^{\mathrm{T}} \mathbf{R}_a^{-1} \mathbf{H}_a \mathbf{v}_j$$

- estimate of forecast error variance reduction:

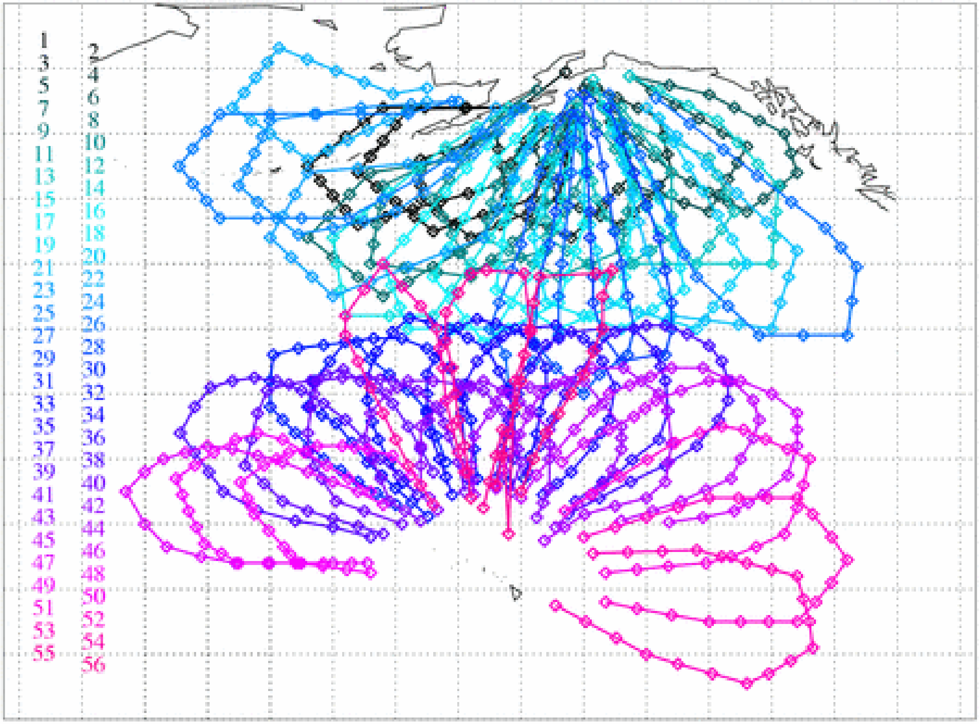
$$\text{trace}\left(\left[\mathbf{I} - \mathbf{C}^{-1} \right] \text{diag}(\sigma_1^2 \ldots \sigma_{\text{n}}^2) \right),$$

where σ_j denotes the singular value of the routine Hessian SV $\mathbf{v}_j$.

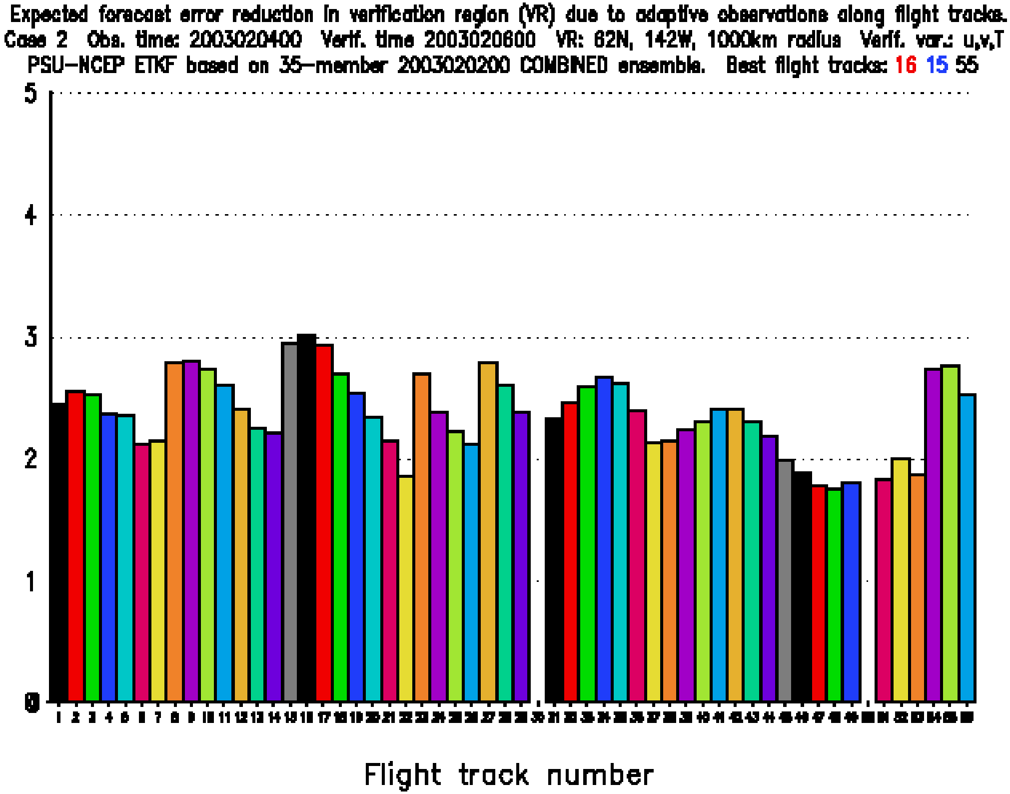
Hessian reduced rank estimate

more equations . . .

Operational observation targeting in WSRP: Predefined flight tracks



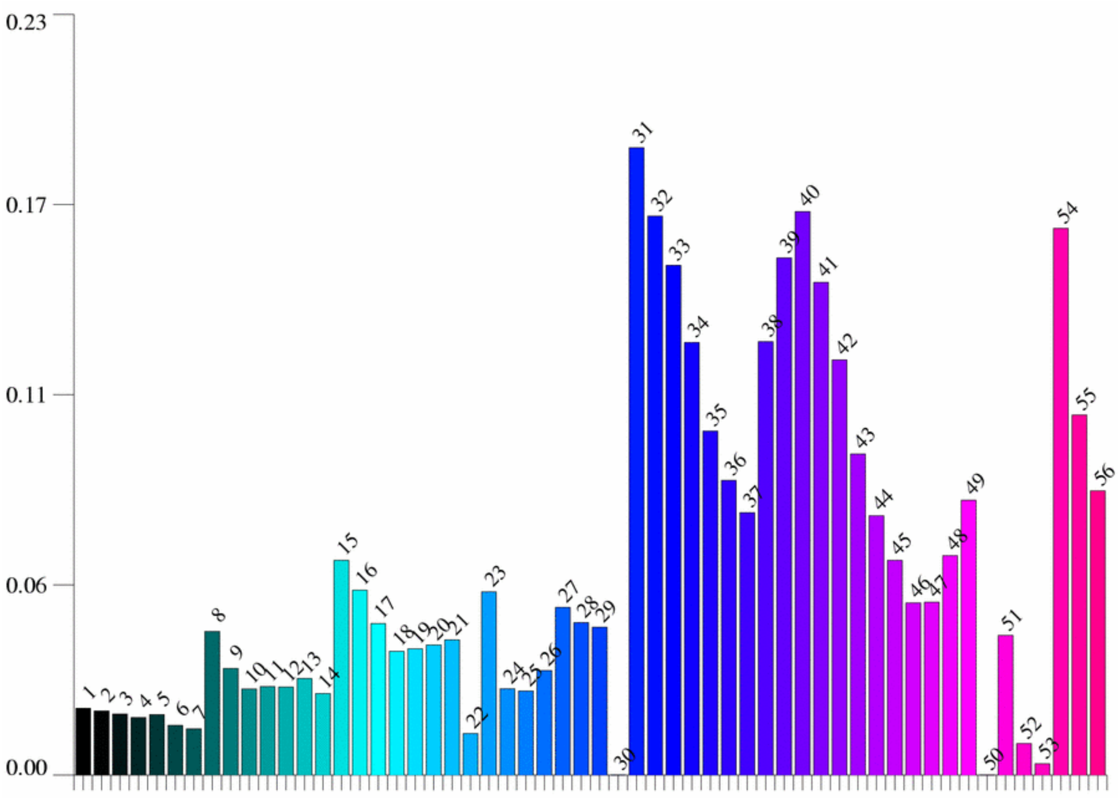
Ranking of flight tracks based on ETKF for add. obs. around 4 Feb, 00 UT, to improve the fc valid on 6 Feb, 00 UT over Alaska



flights from Anchorage | flights from Honolulu

Ranking of flight tracks based on Hessian reduced rank estimate for add. obs. around 4 Feb, 00 UT, to improve the fc valid on 6 Feb, 00 UT over Alaska

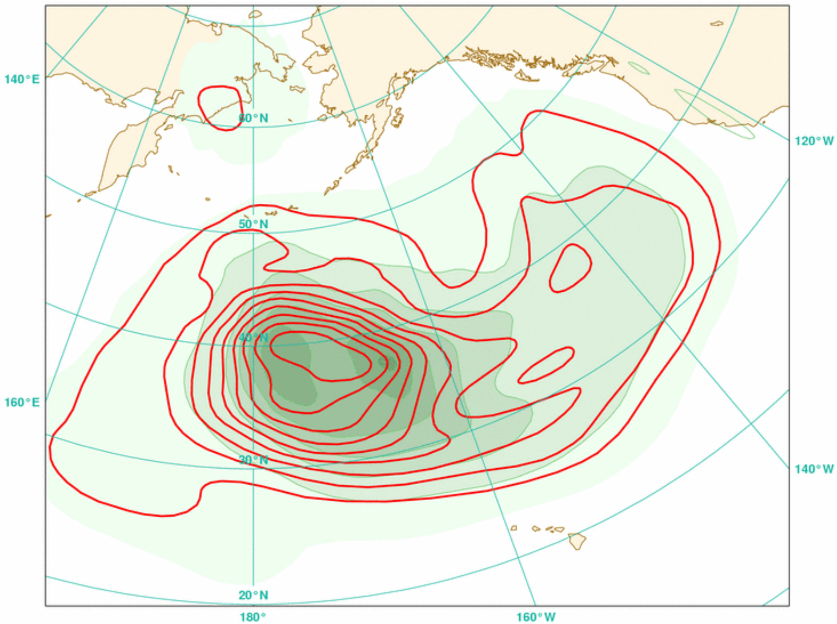
expected reduction of forecast error variance (TE) in SV subspace



flight track number

(courtesy of Alex Doerenbecher)

Maps of sensitive areas and associated reductions of forecast error variance



contours: SV summary map used to identify a sensitive region. The field is a singular value weighted average of the vertically integrated total energy of the leading 10 routine Hessian SVs.

shading: reduction of forecast error variance due to an additional sounding

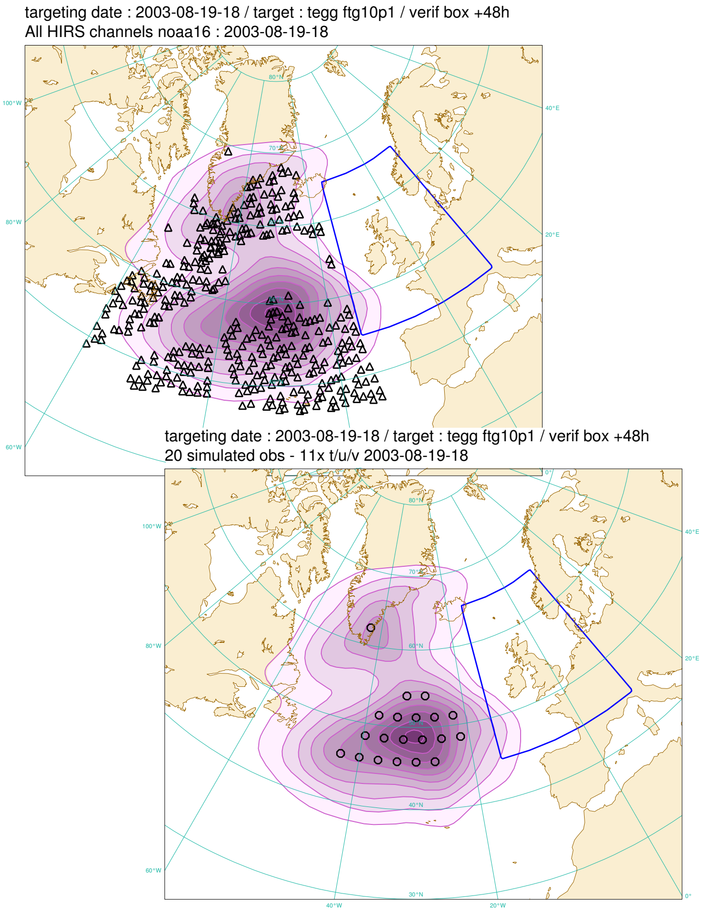
verification region: over Alaska

forecast range 48h

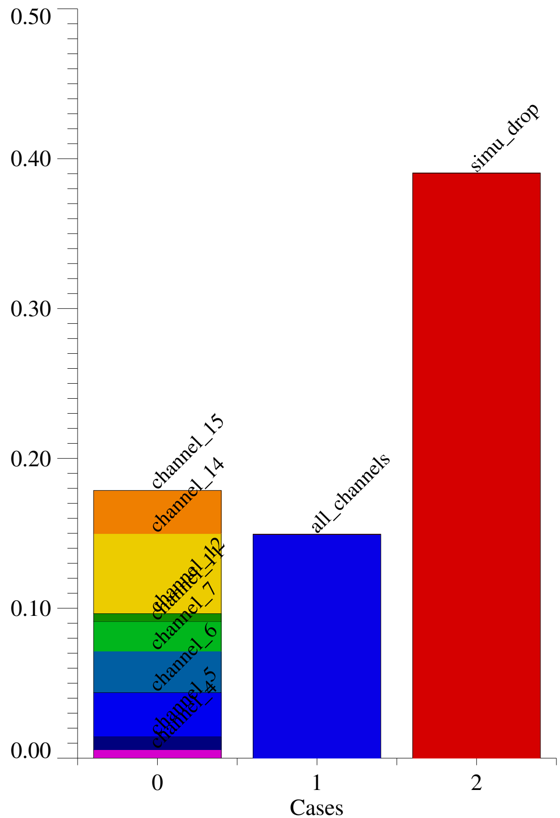
targeting time: 4 Feb 2003, 00UT

(courtesy of Alex Doerenbecher)

Comparing the expected impact of different observation types



expected reduction of forecast error variance (TE) in SV subspace



(courtesy of Alex Doerenbecher)

Atlantic TOST 2003

- SOP starts on 13th of October
- obs targeting guidance based on

  UKMO: Ensemble transform Kalman filter based on EC-ens.

  MF: total energy SVs run on a (possibly perturbed) trajectory

  ECMWF: 2 flavours of total energy SVs and Hessian SVs

- sampling of alternative targets unlikely.
- starting point for verification exercise of Hessian reduced rank estimate. But probably cases from many TOSTs need to be accumulated to arrive at a reliable verification. (tail of fc error distribution!)

Limitations/future directions - subspace

- How many SVs are required to reliably predict fc error variance reductions?
  - rank($\mathbf{L}_{\mathrm{Eu}}\mathbf{M}$) $\leq$ number of gridpoints in verification region.
  - L95: 1 SV is sufficient, leading SV explains $\sim$ 0.75 of total fc error variance in verification region
  - Cardinali and Buizza for 10 NORPEX cases: leading 10 TE-SVs explain $\approx$ 40% of total fc error (measured with total energy TE) in verification region.

- necessity for updates of the singular vector subspace?
  Depends on subspace size and the number and accuracy of added obs.

Limitations/future directions - perturbation dynamics

- simplifications of TL/AD: resolution, physical processes ($\rightarrow$ Marta Janiskova's presentation)
- validity of tangent-linear assumption
  Gilmour et al. 2001: probably not useful beyond 24 h; but measure of nonlinearity dominated by small scales

  Reynolds and Rosmond 2003: SVs useful up to 72 h (diagnostic in SV-space and scale-dependent diagnostic)
- perhaps combination of SV and ensemble based approach? Could handle non-linear saturation aspects more gracefully. But does validity of linear transformation in ETKF technique goes much beyond the validity of the TL-approximation?
- contribution of model error ???

Limitations/future directions - error statistics

shared with data assimilation:

- background error covariances…wavelet-approach (Mike Fisher), reduced-order evolved $\mathbf{P}^f$ (Brian Farrell and Petros J. Ioannou), ensemble based approaches (Peter Houtekamer)
- unaccounted correlation of obs error. Important for satellite data with obs. error correlations in space and between channels $\rightarrow$ optimal thinning (Liu and Rabier 2003)

additional challenge for adaptive obs methods:

- prediction of the observational coverage of *assimilated* obs. at a future time: Nontrivial task for satellite data affected by cloud. The importance of this issue will grow with the number and accuracy of used satellite data. $\rightarrow$ data coverage fc based on model cloud?

The End

slides