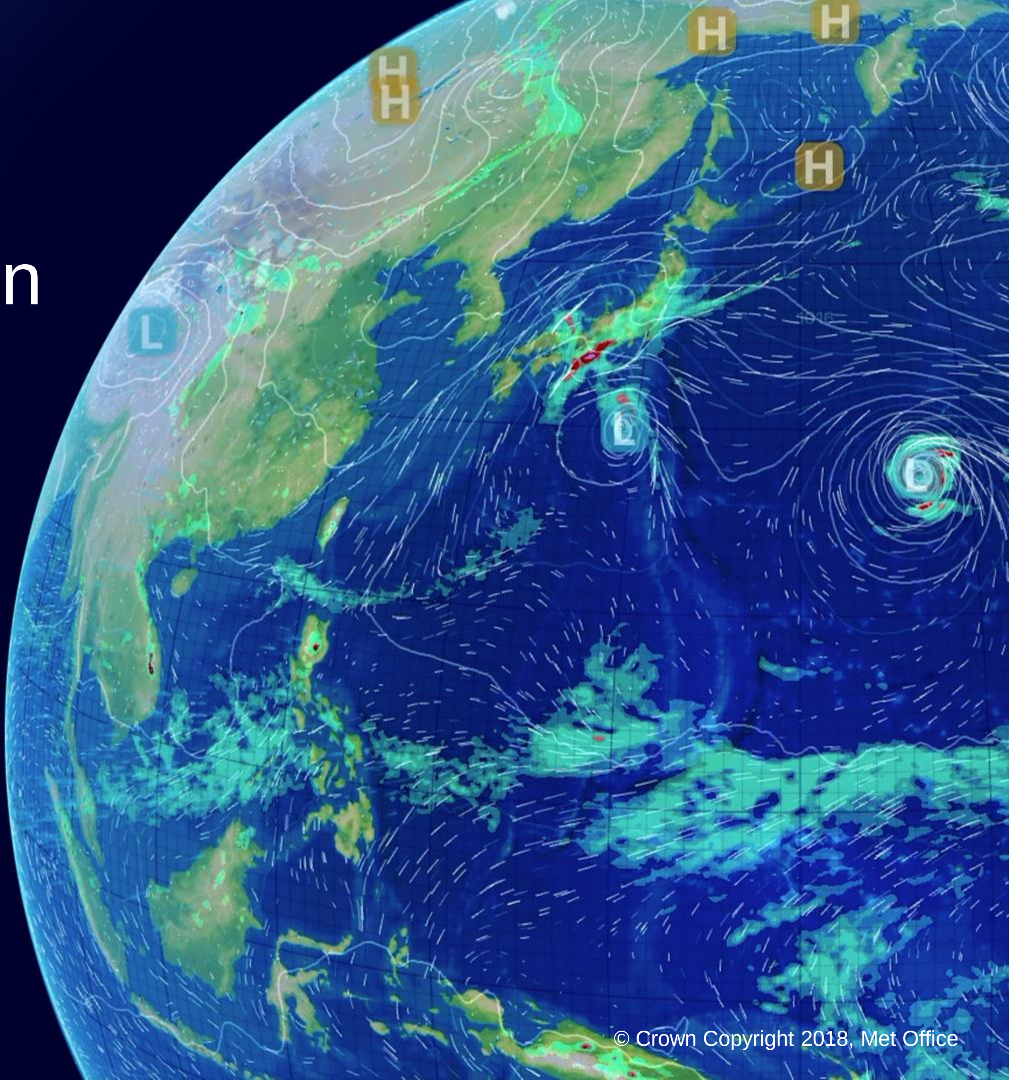
A comparison of hybrid variational data assimilation methods in the Met Office global NWP system



Andrew Lorenc

History of DA at the Met Office

Met Office

1970s Orthogonal polynomials in the 10-level model. Dixon 1972

Nudging

1982 FGGE scheme in global model and fine-mesh. Lyne *et al.* 1982

1988 Analysis Correction scheme. Lorenc *et al.* 1991

VAR

1993 Start of project

1999 3DVar in global & mesoscale models Lorenc *et al.* 2000

2004 4DVar in global model; 2006 NAE; 2017 UKV. Rawlins *et al.* 2007. Simonin *et al.* 2017

2011 Hybrid-4Var in global; 2018? UKV. Clayton *et al.* 2013

2014 Hybrid-4DEnVar trialled in global, En-4DEnVar for ensemble. Lorenc *et al.* 2015. Bowler *et al.* 2017a,b

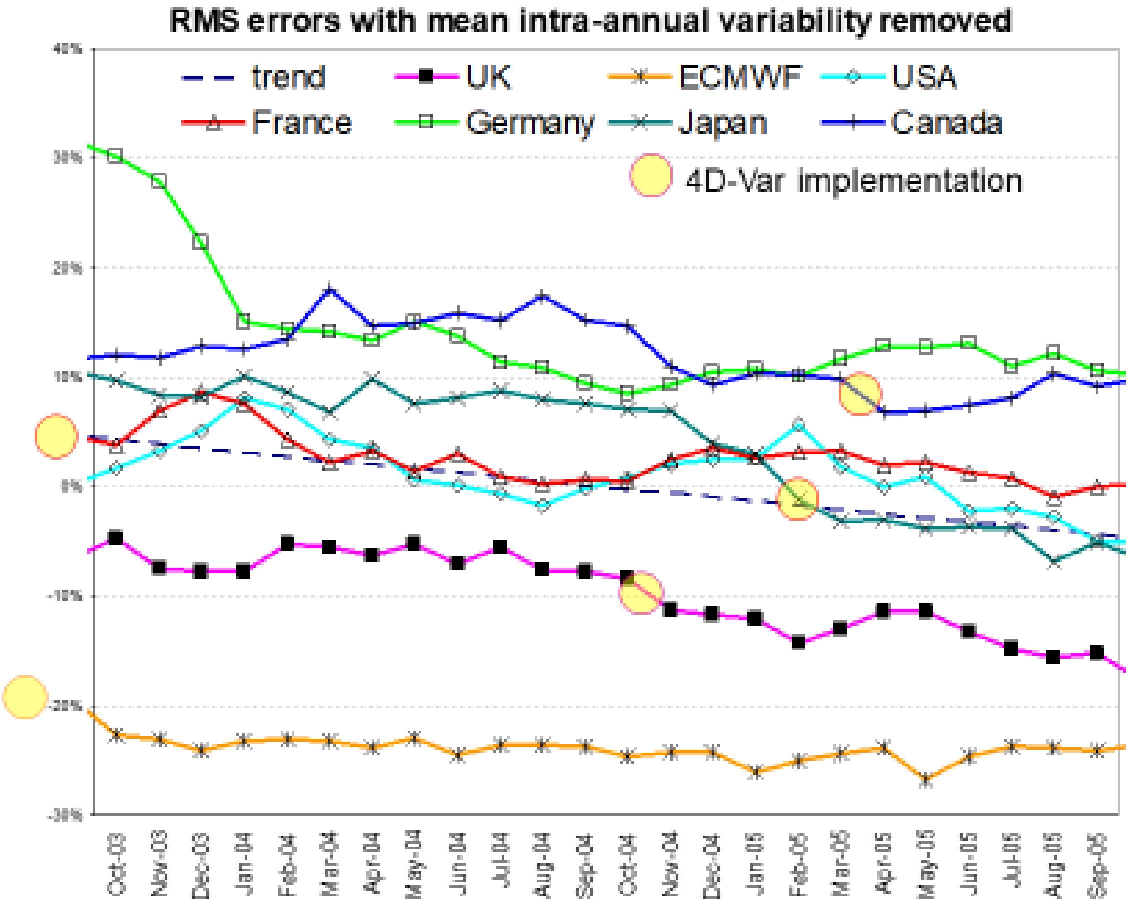
Exascale

2020 Decision to go ahead with new system

2023 Trialling.

Met Office

4DVar implementation at the leading global NWP centres.



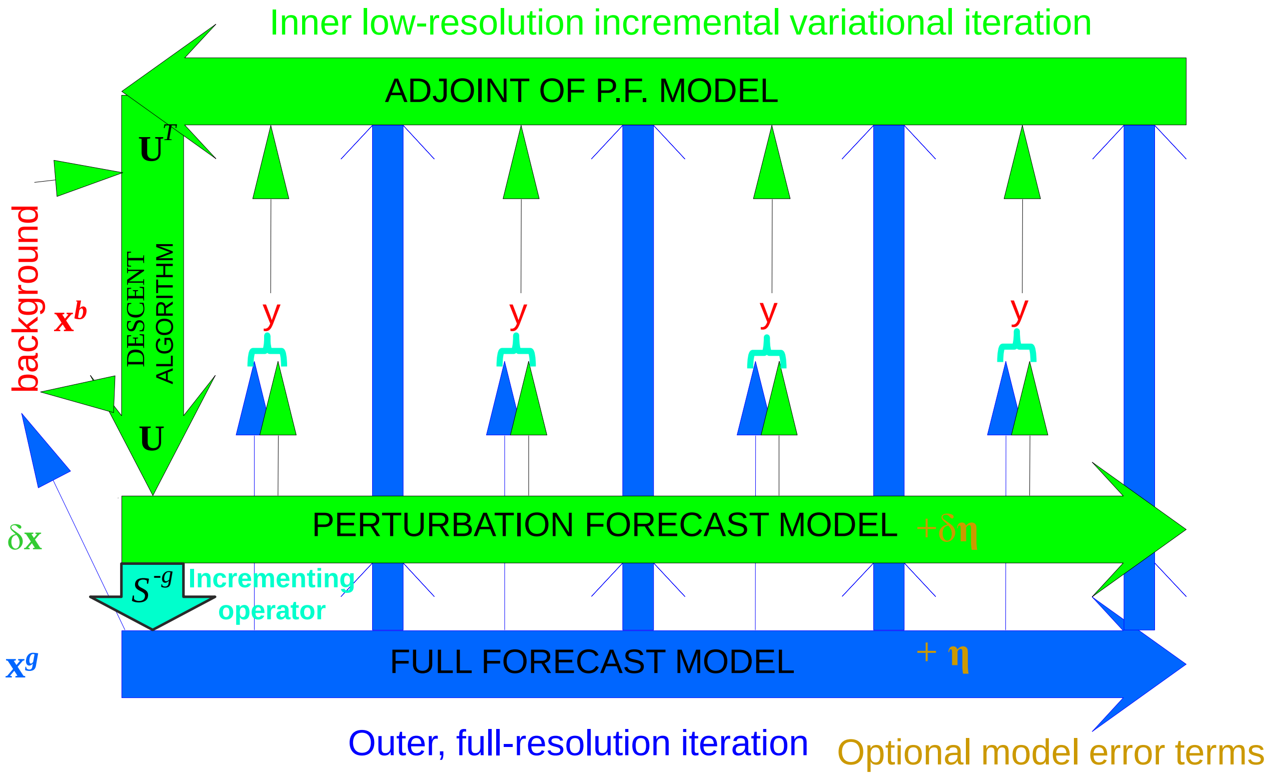
Rawlins *et al.* 2007.

The Met Office VAR system

Characteristics

- Global and LAM configurations.
- Copy file formats & IO and MPP methods from UM, but separate F90 code.
- Incremental for fields, but full nonlinear observation operators.
- Separate OPS, to interpolate full (outer-loop) model fields to observations; and do obs. selection, preliminary 1DVar of radiances, quality control, etc.
- From 3DVar to 4DVar using simplified PerturbationForecast model; simplifications include using a single total moisture increment.
- Hybrid-4DVar option uses hybrid static & ensemble covariances.
- Hybrid-4DEnVar option uses 4D ensemble covariances instead of PF model.
- Options for bias correction, quality control, impact assessment of obs (VarBC, VarQC, FSOI).

Incremental 4DVar with Outer Loop



An outer-loop reruns the FULL FORECAST MODEL. The guess $\mathbf{x}^g$ is updated; the background $\mathbf{x}^b$ is not.

Incremental Approach

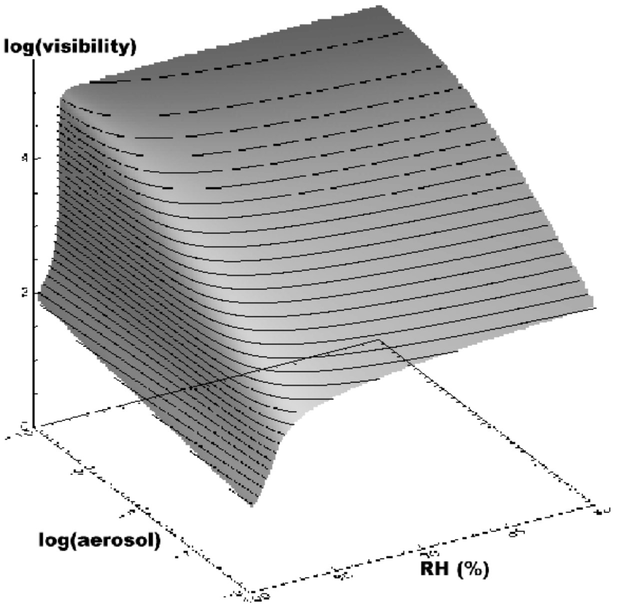
Essential for reducing relative cost 4DVar v Forecast (with best available model)

Suggested by John Derber and elaborated by Courtier *et al*. (1994).

Usual fully-incremental approach fits $\mathbf{Hx}'$ to $\mathbf{d}=\mathbf{y}^o-H(\mathbf{x}^b)$ – the variational penalty uses linear $\mathbf{H}$

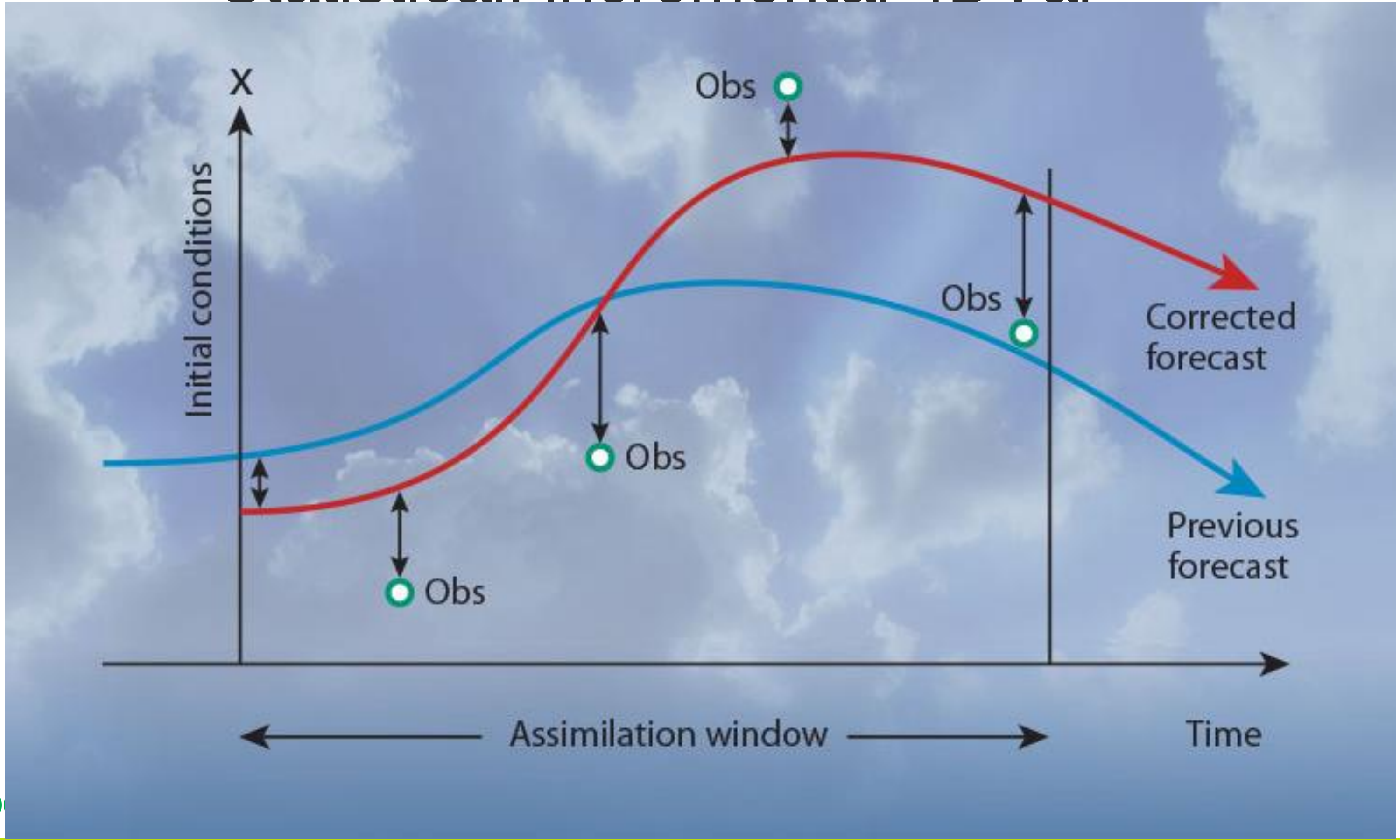
We wanted to use highly nonlinear observations such as visibility in MES 3DVar, so H is split into horizontal- and time-interpolation (in the OPS) to columns $\mathbf{c}_x$. VAR interpolates and adds an increment, to give $\mathbf{c}_x^+$, so it can fit a nonlinear $\mathbf{y}=H(\mathbf{c}_x^+)$ to $\mathbf{y}^o$

This makes the penalty function non-quadratic, which rules out some minimisation algorithms.



Clark *et al.* 2008

Statistical, incremental 4DVar



Lorenc (20

Idealised General Bayesian 4D DA

$\underline{\mathbf{x}}^b$	background trajectory
$\underline{\mathbf{P}}$	4D error covariance of $\underline{\mathbf{x}}^b$
$\delta\underline{\mathbf{x}}$	4D analysis increment
$\underline{\mathbf{y}} = \underline{H}\left(\underline{\mathbf{x}}^b + \delta\underline{\mathbf{x}}\right)$	model estimate of obs
$J\left(\delta\underline{\mathbf{x}}\right) = \frac{1}{2}\delta\underline{\mathbf{x}}^T\underline{\mathbf{P}}^{-1}\delta\underline{\mathbf{x}} + \frac{1}{2}\left(\underline{\mathbf{y}} - \underline{\mathbf{y}}^o\right)^T\underline{\mathbf{R}}^{-1}\left(\underline{\mathbf{y}} - \underline{\mathbf{y}}^o\right)$	penalty function

$\underline{\mathbf{P}}$ is **big!** We cannot even estimate it fully, let alone compute $\frac{1}{2}\delta\underline{\mathbf{x}}^T\underline{\mathbf{P}}^{-1}\delta\underline{\mathbf{x}}$.

The solution is to model $\underline{\mathbf{P}}$ using a sequence of operations we can compute, then use these to transform $\delta\underline{\mathbf{x}}$ so that $\frac{1}{2}\delta\underline{\mathbf{x}}^T\underline{\mathbf{P}}^{-1}\delta\underline{\mathbf{x}}$ simplifies.

4DVar: using static covariance $\mathbf{B}$

Model 3D covariance using transforms	$\mathbf{B} = \mathbf{U}\mathbf{U}^T$
3D analysis increment	$\delta\mathbf{x}_0 = \mathbf{U}\mathbf{v}^c$
made 4D using linear forecast model $\underline{\mathbf{M}}$	$\delta\underline{\mathbf{x}} = \underline{\mathbf{M}}\delta\mathbf{x}_0$
Implicit 4D prior covariance	$\underline{\mathbf{P}} = \mathbf{S}^{-I}\underline{\mathbf{M}}\mathbf{B}\underline{\mathbf{M}}^T\mathbf{S}^{-T}$
Transformed penalty function	$J\left(\mathbf{v}^c\right) = \frac{1}{2}\mathbf{v}^{cT}\mathbf{v}^c + \frac{1}{2}\left(\underline{\mathbf{y}} - \underline{\mathbf{y}}^o\right)^T\underline{\mathbf{R}}^{-1}\left(\underline{\mathbf{y}} - \underline{\mathbf{y}}^o\right)$

hybrid-4DVar

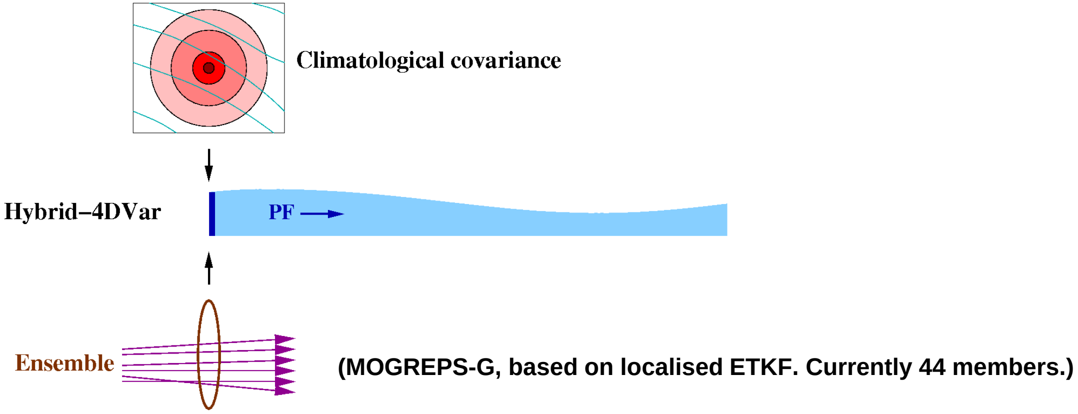
4D analysis increment $$\underline{\delta \mathbf{x}} = \underline{\mathbf{M}}\left(\beta_c \mathbf{U}\mathbf{v}^c + \beta_e \Sigma_{k=1}^{N} \mathbf{U}^\alpha \mathbf{v}_k^\alpha \circ \mathbf{x}'_k\right)$$

Localized 4D covariance $$\underline{\mathbf{P}} = \mathbf{S}^{-I}\underline{\mathbf{M}}\left(\beta_c^2 \mathbf{B} + \beta_e^2 \left(\mathbf{C} \circ \mathbf{X}\mathbf{X}^T\right)\right)\underline{\mathbf{M}}^T \mathbf{S}^{-T}$$

concatenated control vectors $$\mathbf{v}^T = \left[\mathbf{v}^{cT}, \mathbf{v}_1^{\alpha T} \cdots \mathbf{v}_N^{\alpha T}\right]$$

1% improvement in rms errors when implemented at Met Office (Clayton *et al*. 2013)

Hybrid-4DVar



B implicitly propagated by a linear “Perturbation Forecast” (PF) model:

- ~100 PF + adjoint forecasts run serially.
- But PF model doesn’t scale well.
- And difficult to keep PF model in line with forecast model.

We need an alternative scheme for future supercomputers that excludes the PF model…

4DEnVar: using an ensemble of 4D trajectories

which samples background errors

Ensemble trajectory matrix $\underline{\mathbf{X}} = \left[\underline{\mathbf{x}}'_1 \cdots \underline{\mathbf{x}}'_N \right]$

where $\underline{\mathbf{x}}'_k = \frac{1}{\sqrt{N-1}} \left(S\left(\underline{\mathbf{x}}_k\right) - \overline{S\left(\underline{\mathbf{x}}\right)} \right)$

Model 4D $\underline{\mathbf{P}}$ directly, as localised ensemble covariance,

$$\underline{\mathbf{P}} = \mathbf{S}^{-I} \left(\underline{\mathbf{C}} \circ \underline{\mathbf{X}}\,\underline{\mathbf{X}}^T \right) \mathbf{S}^{-T}$$

then assume persistence for $\underline{\mathbf{C}}$

$$\underline{\mathbf{C}} = \underline{\mathbf{I}}\mathbf{C}\underline{\mathbf{I}}^T$$

4D localised linear combination of ensemble trajectories

$$\underline{\boldsymbol{\alpha}}_k = \underline{\mathbf{I}}\mathbf{U}^{\alpha}\mathbf{v}_k^{\alpha}$$

$$\delta\underline{\mathbf{x}} = \Sigma_{k=1}^{N} \underline{\boldsymbol{\alpha}}_k \circ \underline{\mathbf{x}}'_k$$

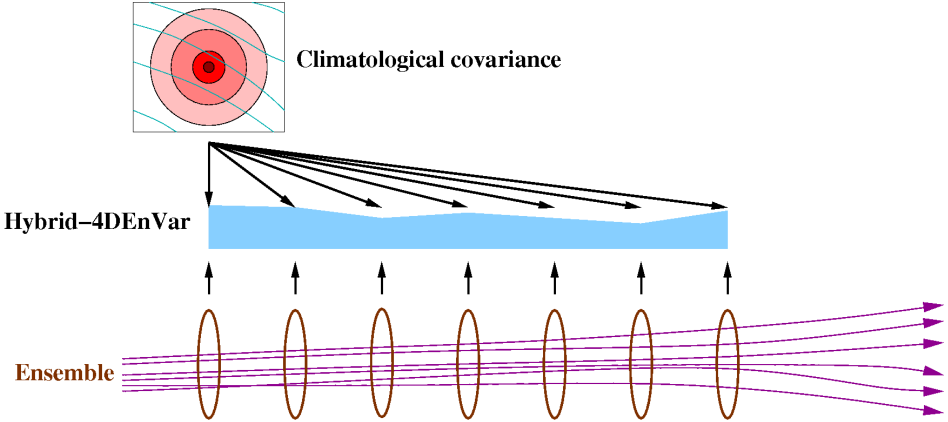
concatenated control vectors

$$\mathbf{v}^T = \left[\mathbf{v}_1^{\alpha T} \cdots \mathbf{v}_N^{\alpha T} \right]$$

Transformed penalty function

$$J\left(\mathbf{v}\right) = \tfrac{1}{2}\mathbf{v}^T\mathbf{v} + \tfrac{1}{2}\left(\underline{\mathbf{y}} - \underline{\mathbf{y}}^o\right)^T \underline{\mathbf{R}}^{-1} \left(\underline{\mathbf{y}} - \underline{\mathbf{y}}^o\right)$$

Hybrid-4DEnVar



No PF model, but much more IO required to read ensemble data:

- 11 times faster with 22 N216 members and 384 PEs. (IO around 30% of cost)

Analysis consists of two parts:

- A 3DVar-like analysis based on the climatological covariance $\mathbf{B}_c$
- A 4D analysis consisting of a linear combination of the ensemble perturbations.

Localisation is currently in space only: same linear combination of ensemble perturbations at all times.

Summary comparison

hybrid-4DEnVar

4D analysis increment $\underline{\delta\mathbf{x}} = \beta_c \underline{\mathbf{I}} \delta\mathbf{x}_0 + \beta_e \Sigma_{k=1}^{N} \underline{\boldsymbol{\alpha}}_k \circ \underline{\mathbf{x}}_k'$

Localized 4D covariance $\underline{\mathbf{P}} = \beta_c^2 \underline{\mathbf{I}}\mathbf{B}\underline{\mathbf{I}}^T + \beta_e^2 \left(\underline{\mathbf{C}} \circ \underline{\mathbf{X}}\,\underline{\mathbf{X}}^T\right)$

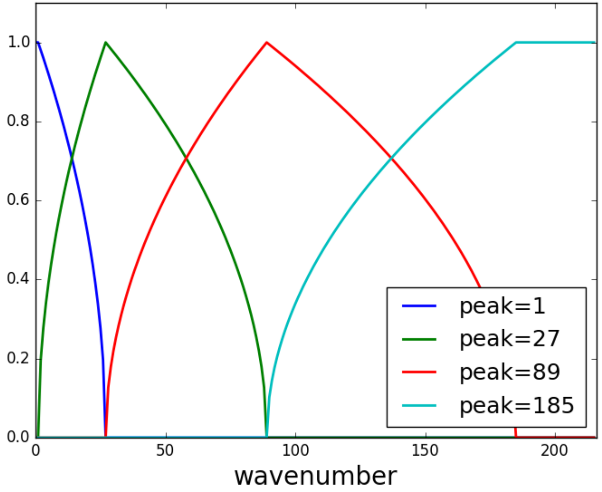
hybrid-4DVar

4D analysis increment $\underline{\delta\mathbf{x}} = \underline{\tilde{\mathbf{M}}} \left(\beta_c \delta\mathbf{x}_0 + \beta_e \Sigma_{k=1}^{N} \boldsymbol{\alpha}_k \circ \mathbf{x}_k'\right)$

Localized 4D covariance $\underline{\mathbf{P}} = \underline{\tilde{\mathbf{M}}} \left(\beta_c^2 \mathbf{B} + \beta_e^2 \left(\mathbf{C} \circ \mathbf{X}\mathbf{X}^T\right)\right) \underline{\tilde{\mathbf{M}}}^T$

Met Office

Scale-dependent localisation & waveband filtering Lorenc (2017)



Filtered perturbation matrices for each waveband $\mathbf{X}_e = \left[\mathbf{x}'_{1,e} \cdots \mathbf{x}'_{N,e} \right]$ where

$$\mathbf{x}'_{k,e} = \frac{1}{\sqrt{N-1}} \mathbf{F}_e \left(\left(S\left(\mathbf{x}_k\right) - \overline{S\left(\mathbf{x}\right)} \right) \right)$$

and the $\mathbf{F}_e$ for each band are such that the covariances sum to the unfiltered

$$\mathbf{X}\mathbf{X}^T = \sum_{e=1}^{N_{bands}} \mathbf{X}_e \mathbf{X}_e^T$$

Apply a different scale-dependent localisation $\mathbf{C}_e = \mathbf{U}_e^{\alpha} \mathbf{U}_e^{\alpha T}$ to each band. 6241, 919, 389, 256km

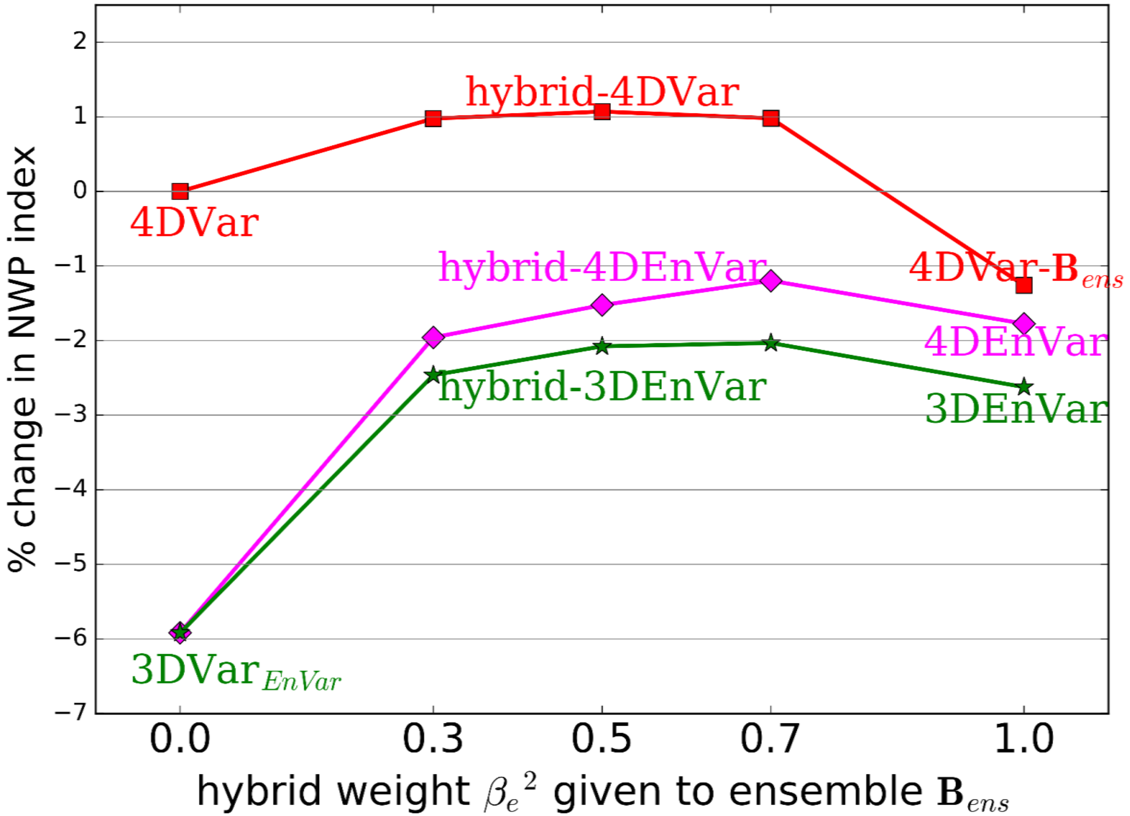
Each band has separate control variables $\boldsymbol{\alpha}_{k,e} = \mathbf{U}_e^{\alpha} \mathbf{v}_{k,e}^{\alpha}$

The analysis increment is summed over bands $\delta\mathbf{x} = \sum_{e=1}^{N_{bands}} \sum_{k=1}^{N} \boldsymbol{\alpha}_{k,e} \circ \mathbf{x}'_{k,e}$

Comparison of hybrid-Var methods

All experiments used an Ne=44 ensemble from our current MOGREPS local ETKF system

- 4DVar improves static $\mathbf{B}_c$ by 6% but doesn't much improve $\mathbf{B}_{ens}$ – with β_e^2=1 4DEnVar is as good
- Allowing for the obs-time in 4DEnVar gains 1% over 3DEnVar
- Hybrid-4DVar is 1% better than 4DVar and 2% better than best hybrid-4DEnVar



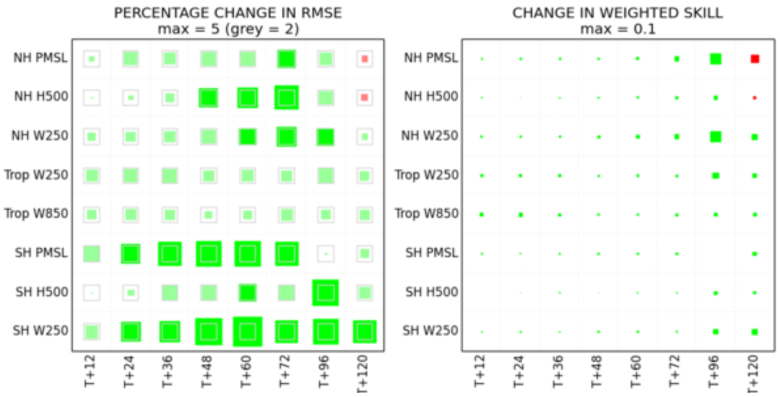
Lorenc & Jardak 2018

Scorecards for best method

Hybrid-4DVar with β_e^2=0.5

beats 4DVar with β_e^2=0.0

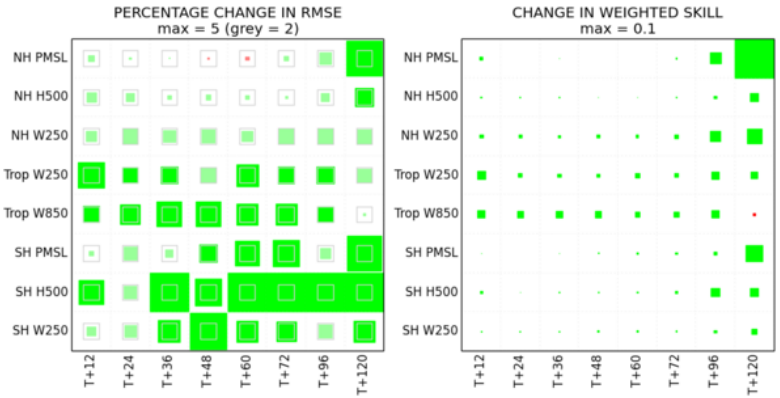
+1.1% on Index



Hybrid-4DVar with β_e^2=0.5

beats hybrid-4DEnVar with β_e^2=0.7

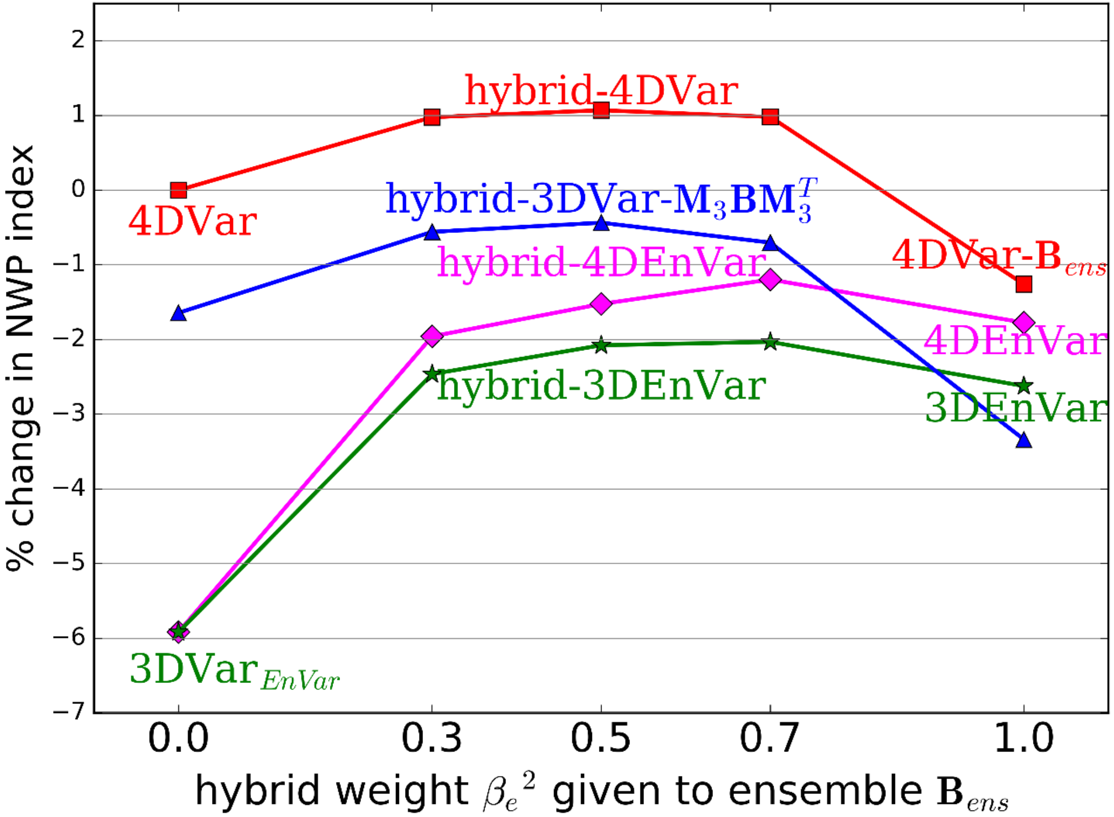
+2.2% on Index



Comparison of hybrid-Var methods

Used 4DVar software to run 3DVar with covariances improved by 3hrs evolution (as in Lorenc & Rawlins 2005)

- This improves static $\mathbf{B}_c$, explaining most of the benefit of 4DVar
- Allowing for the obs-time in 4DVar gains ~1% (like 4DEnVar over 3DEnVar)
- Time-evolved $\mathbf{M}_3\mathbf{B}_{ens}\mathbf{M}_3{}^T$ does not improve on $\mathbf{B}_{ens}$



Lorenc & Jardak 2018

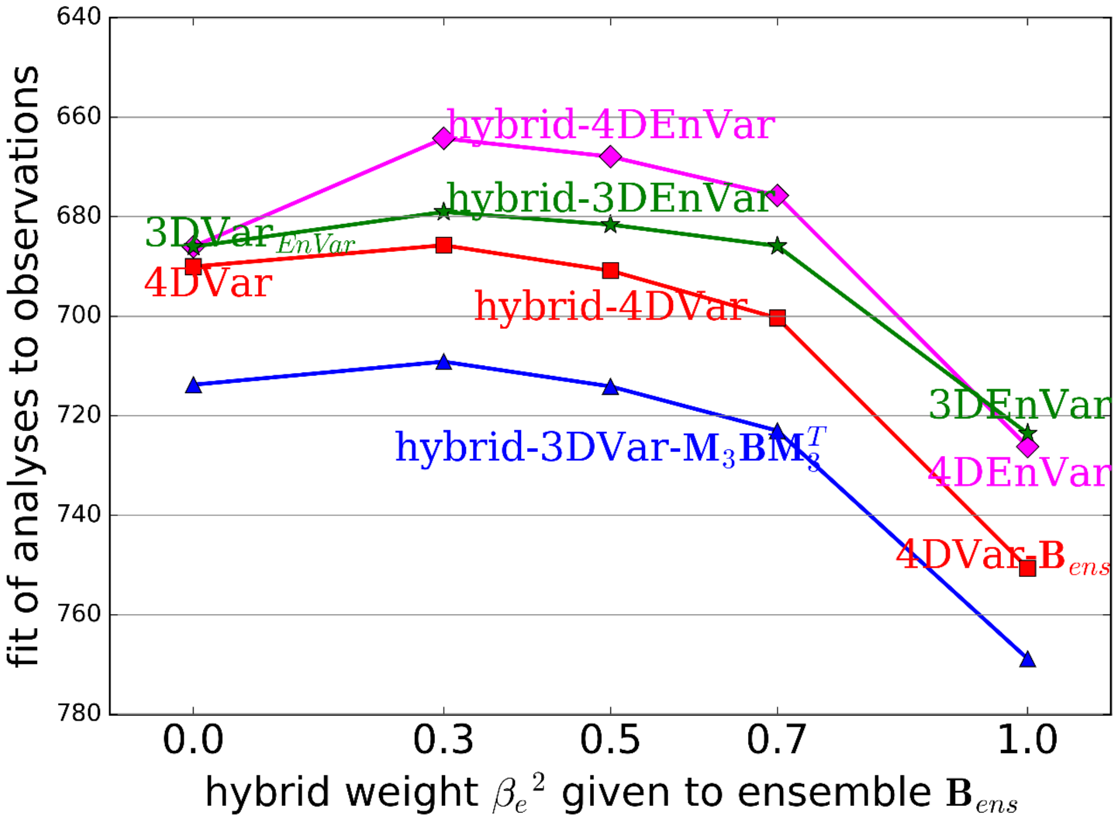
Explanation using ideas from nonlinear dynamics

- For a perfect chaotic model, the errors from a Kalman filter based DA system asymptote to the **unstable**–neutral **subspace**
- Ensemble DA only works if N_e > dimensions of unstable-neutral sub-space (Bocquet & Carrassi 2017). Our Ne=44 was too small, so needs augmenting by $\mathbf{B}$.
- 4DVar methods work best when increments are in the unstable sub-space (Trevisan *et al.* 2010)
- Perturbations which grow typically slope – our $\mathbf{B}$ model is isotropic.
  So $\mathbf{B}$ has only a small projection on the unstable sub-space – $\mathbf{MBM}^T$ does better.
- The ideas above are only an idealised limiting case, based on small perfect models.
  NWP models cannot be perfect because of the butterfly effect.
  The dimension of the unstable sub-space is uncertain due to localisation.

Analysis fit to observations

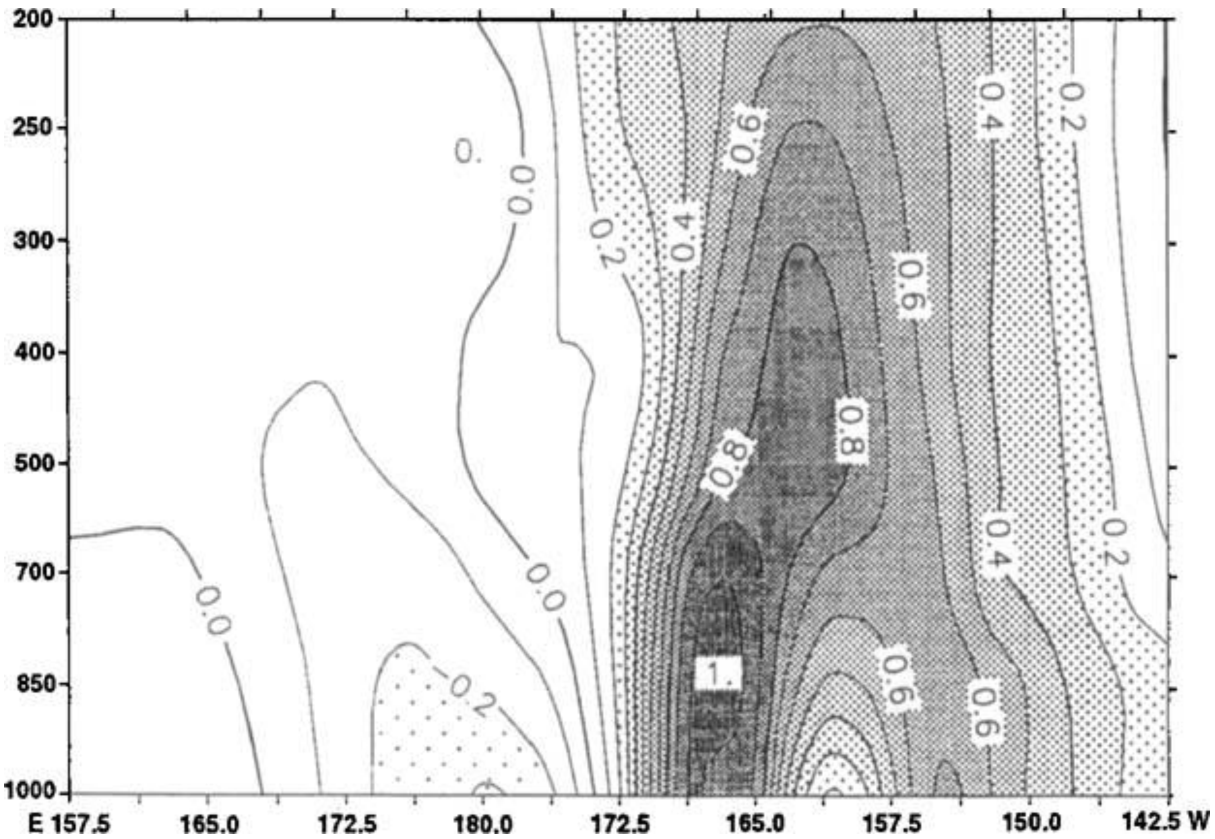
The ability to fit the observations (to within their standard errors) is a necessary (but not sufficient) measure of a good analysis.

Even with waveband and scale-dependent localisation, and augmentation by lagging & shifting, our 44 member ensemble covariance was not good at fitting the observations.



Dynamical structure functions in 4DVar

Thépaut *et al.* (1996)



Time evolution (in this case for 24hrs) produces structures that are tilted and look like singular vectors.

Our isotropic **B**-model (in which scale is independent of direction) cannot give preference to such growing structures.

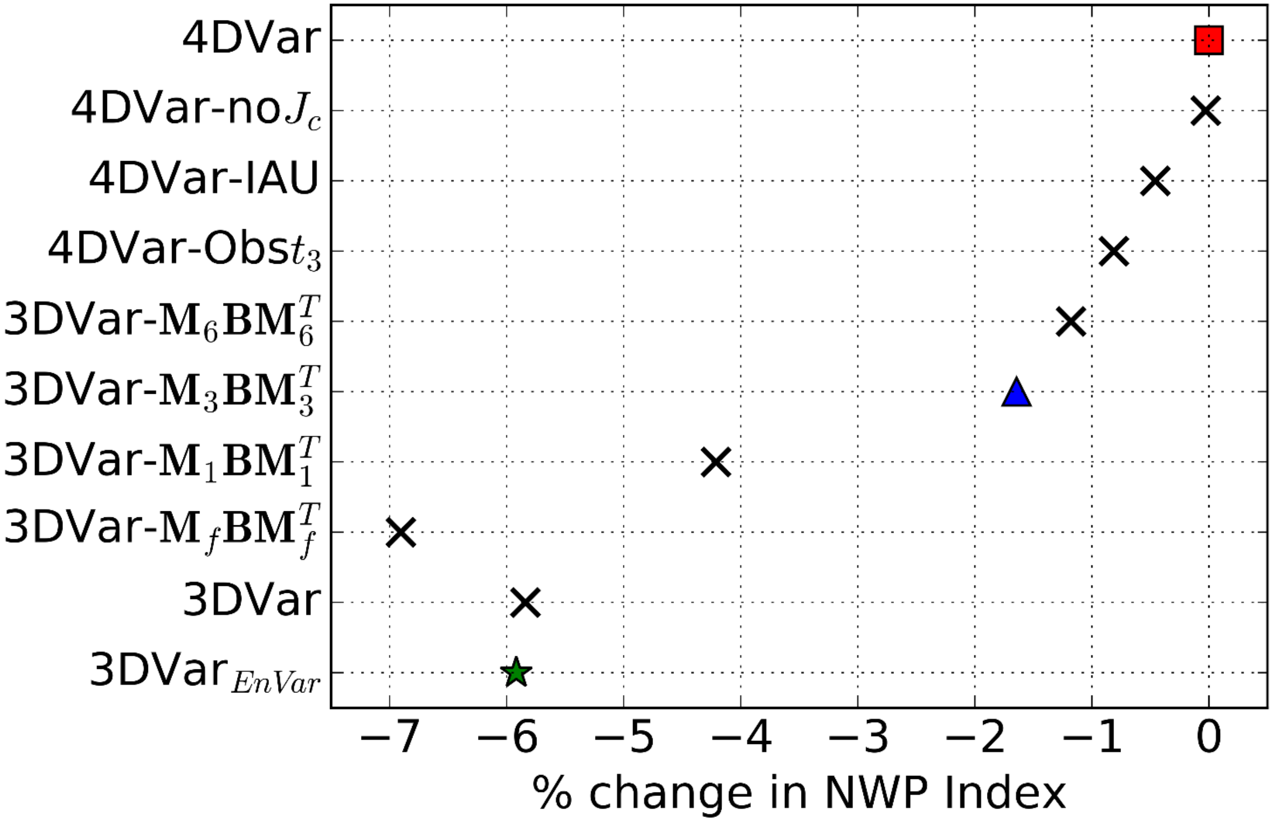
Extra slides

- Showing NWP Index from extra non-hybrid experiments
  - Exploring the effect of changing the length of evolution of **B**
  - Comparing detailed differences in our implementation.

- Showing the average fit of backgrounds to observations, for the hybrid expts.
- This is a good measure of quality, correlating well with NWP Index.

- Showing the regional variation in the split into wavebands.
- This means that waveband-specific tuning caters for regional variation in coeffs.

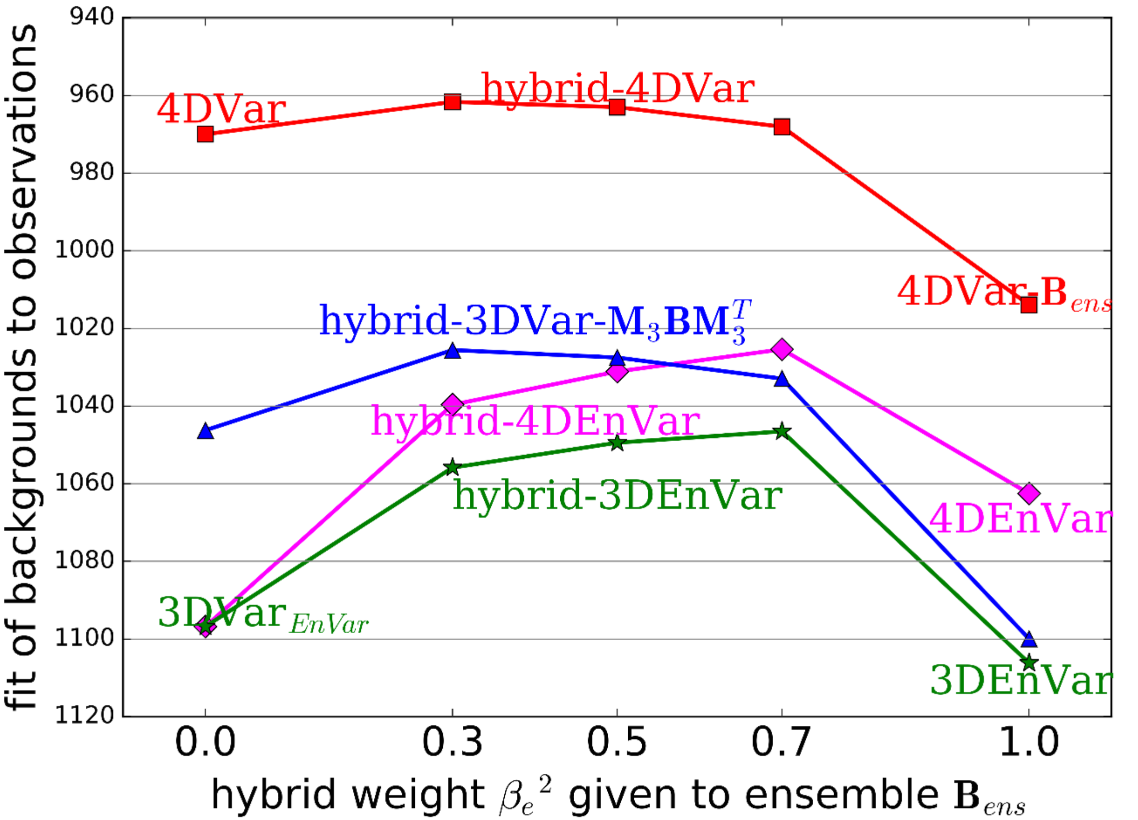
Extra non-hybrid experiments



Lorenc & Jardak 2018

Background fit to observations

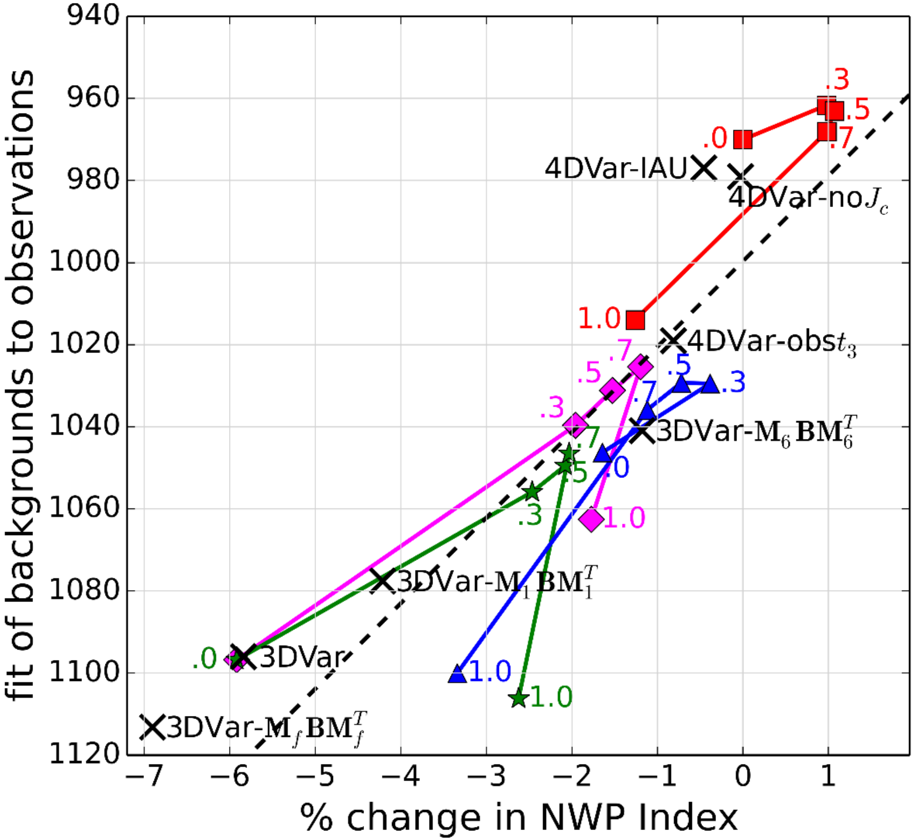
The ability to fit the observations (to within their standard errors) is *usually a reliable* measure of a good analysis.



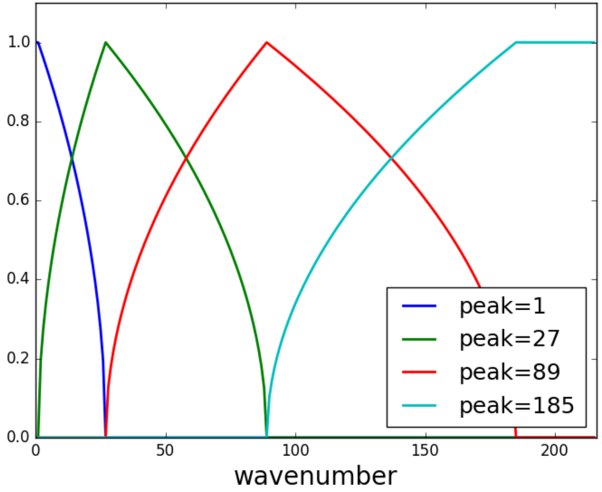
Background fit to observations

The ability to fit the observations (to within their standard errors) is *usually a reliable* measure of a good analysis.

This is shown by the high correlation with the NWP Index scores.

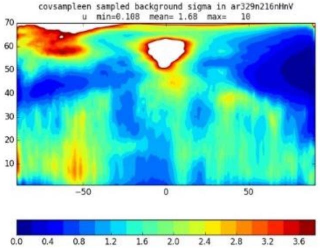


Wavebands copied from Lorenc (2017).

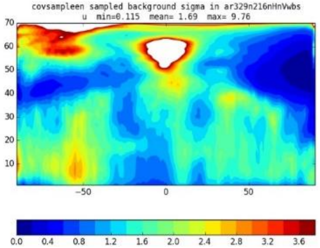


RMS zonal mean X-sections of **u′**

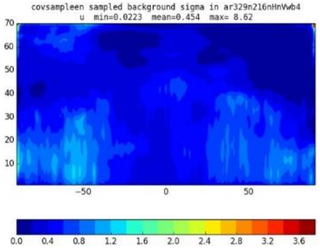
For a randomly chosen date in June



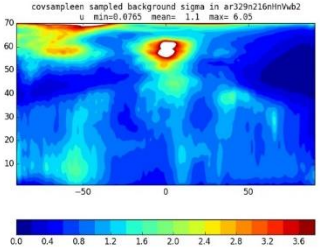
Raw ensemble



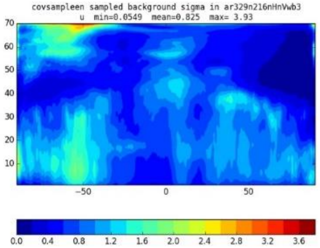
Sum of wavebands



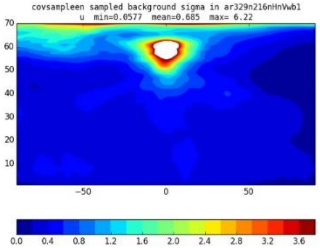
waveband 1



waveband 2

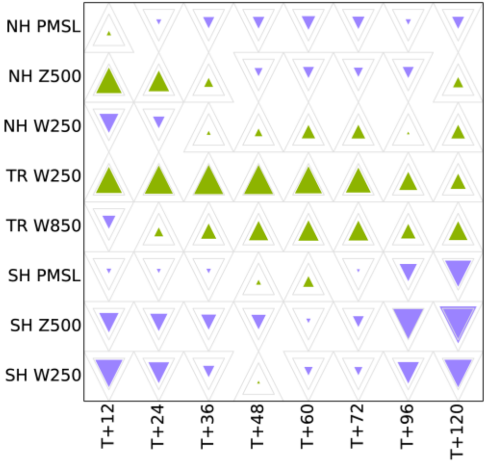


waveband 3

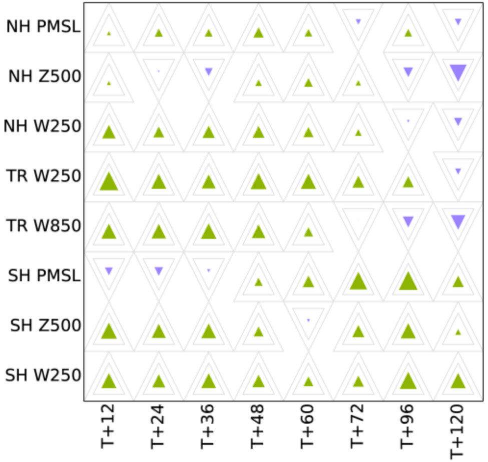


waveband 4

Wavebands & scale-dependent localisation help with some tuning!



Changing localisation scale (from 600km to 800km) has mixed benefit, depending on region, when not using wavebands.



Changing all localisation scales (by factor 5/6) has consistent benefit when using wavebands with scale dependent localisation.

$\Rightarrow$ The regional variations in the split between wavebands take care of many regional variations in optimal tuning. (Lorenc 2017)

Some details of the Met Office VAR equations were simplified to save time in this presentation; see Lorenc and Jardak (2018) for more precise versions.

References

Bocquet M, Carrassi A. 2017. Four-dimensional ensemble variational data assimilation and the unstable subspace. *Tellus A* **69**(1): 1304 504, doi:10.1080/16000870.2017.1304504, URL http://dx.doi.org/10.1080/16000870.2017.1304504.

Bowler NE, Clayton AM, Jardak M, Jermey PM, Lorenc AC, Wlasak MA, Barker DM, Inverarity GW, Swinbank R. 2017a. The effect of improved ensemble covariances on hybrid variational data assimilation. *Q. J. R. Meteorol. Soc.* **143**: 785–797, doi:10.1002/qj.2964.

Bowler NE, Clayton AM, Jardak M, Lee E, Lorenc AC, Piccolo C, Pring SR, Wlasak MA, Barker DM, Inverarity GW, Swinbank R. 2017b. Inflation and localisation tests in the development of an ensemble of 4d-ensemble variational assimilations. *Q. J. R. Meteorol. Soc.* **143**: 1280–1302, doi:10.1002/qj.3004.

Clark PA, Harcourt SA, Macpherson B, Mathison CT, Cusack S, Naylor M. 2008. Prediction of visibility and aerosol within the operational Met Office Unified Model. i: Model formulation and variational assimilation. *Q. J. R. Meteorol. Soc.* **134**: 1801–1816.

Clayton AM, Lorenc AC, Barker DM. 2013. Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office. *Q. J. R. Meteorol. Soc.* **139**: 1445–1461, doi:10.1002/qj.2054.

Courtier P, Thépaut JN, Hollingsworth A. 1994. A strategy for operational implementation of 4D-Var, using an incremental approach. *Q. J. R. Meteorol. Soc.* **120**: 1367–1387, doi:10.1002/qj.49712051912.

Dixon R, Spackman EA, Jones I, Francis A. 1972. The global analysis of meteorological data using orthogonal polynomial base functions. *J. Atmos. Sci.* **29**(4): 609–622, doi:10.1175/1520-0469(1972)029<0609:TGAOMD>2.0.CO;2, URL https://doi.org/10.1175/1520-0469(1972)029<0609:TGAOMD>2.0.CO;2.

Lorenc AC. 1981. A global three-dimensional multivariate statistical analysis scheme. *Mon. Weather Rev.* **109**: 701–721.

Lorenc AC. 2003. Modelling of error covariances by four-dimensional variational data assimilation. *Q. J. R. Meteorol. Soc.* **129**: 3167–3182, doi:10.1256/qj.02.131.

Lorenc AC. 2017. Improving ensemble covariances in hybrid-variational data assimilation, without increasing ensemble size. *Q. J. R. Meteorol. Soc.* **143**: 1062–1072, doi:10.1002/qj.2990.

Lorenc AC, Ballard SP, Bell RS, Ingleby NB, Andrews PLF, Barker DM, Bray JR, Clayton AM, Dalby T, Li D, Payne TJ, Saunders FW. 2000. The Met. Office global three-dimensional variational data assimilation scheme. *Q. J. R. Meteorol. Soc.* **126**: 2991–3012, doi:10.1002/qj.49712657002.

Lorenc AC, Bowler NE, Clayton AM, Fairbairn D, Pring SR. 2015. Comparison of hybrid-4DEnVar and hybrid-4DVar data assimilation methods for global NWP. *Mon. Weather Rev.* **143**: 212–229, doi:10.1175/MWR-D-14-00195.1.

Lorenc AC, Jardak M. 2018. A comparison of hybrid variational data assimilation methods for global NWP. *Q. J. R. Meteorol. Soc.* **in preparation**.

Lorenc AC, Payne T. 2007. 4D-Var and the butterfly effect: Statistical four-dimensional data assimilation for a wide range of scales. *Q. J. R. Meteorol. Soc.* **133**: 607–614, doi:10.1002/qj.36.

Lorenc AC, Rawlins F. 2005. Why does 4D-Var beat 3D-Var? *Q. J. R. Meteorol. Soc.* **131**: 3247–3257, doi:10.1002/qj.49711146703.

Lyne WH, Swinbank R, Birch NT. 1982. A data assimilation experiment and the global circulation during the FGGE special observing periods. *Q. J. R. Meteorol. Soc.* **108**: 575–594.

Rawlins F, Ballard SP, Bovis KJ, Clayton AM, Li D, Inverarity GW, Lorenc AC, Payne TJ. 2007. The Met Office global 4-dimensional variational data assimilation system. *Q. J. R. Meteorol. Soc.* **133**: 347–362, doi:10.1002/qj.32.

Simonin D, Pierce C, Roberts N, Ballard SP, Li Z. 2017. Performance of Met Office hourly cycling NWP-based nowcasting for precipitation forecasts. *Q. J. R. Meteorol. Soc.* doi:10.1002/qj.3136, URL `http://dx.doi.org/10.1002/qj.3136`.

Thépaut JN, Courtier P, Belaud G, Lemaître G. 1996. Dynamical structure functions in a four-dimensional variational assimilation: A case study. *Q. J. R. Meteorol. Soc.* **122**: 535–561.

Trevisan A, D'Isidoro M, Talagrand O. 2010. Four-dimensional variational assimilation in the unstable subspace and the optimal subspace dimension. *Q. J. R. Meteorol. Soc.* **136**(647): 487–496, doi:10.1002/qj.571.