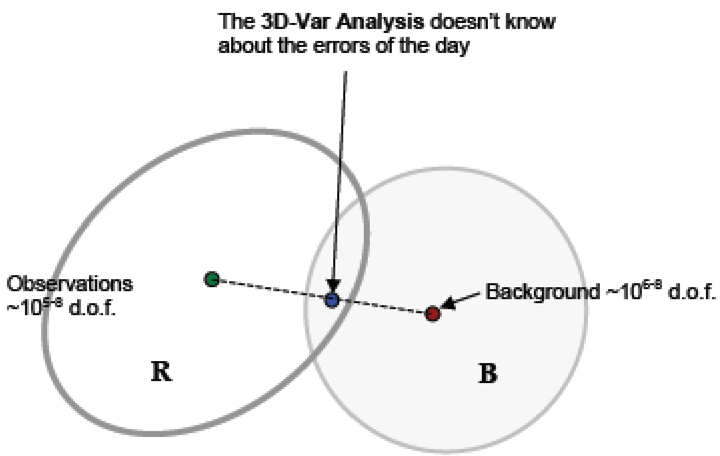
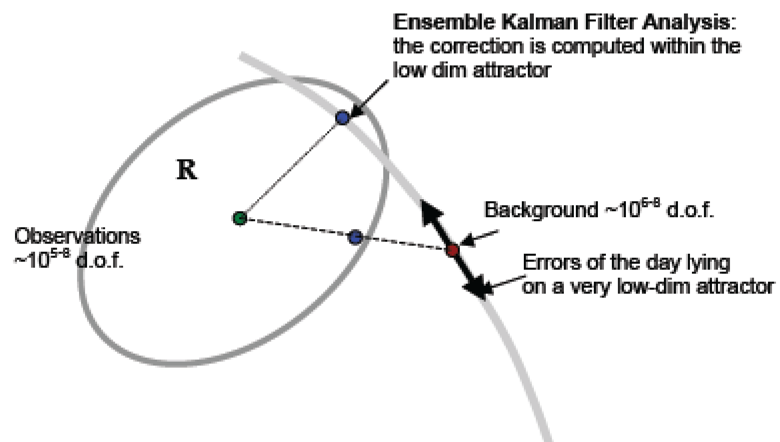
Ensemble data Assimilation technique, Hybrid data assimilation tech

Sumit Kumar
NCMRWF

Errors of the day: the variation in error due to the locations of recent instabilities and observations





in the case of 3DVar
is isotropic and constant in time

Here ensemble forecast members define a subspace within which the analysis lies

A key component of data assimilation systems is the background-error covariance matrix, which controls how the information from observations spreads into the model space.

Kalnay, Fisher

Table 20-2. Standard deviation σ_o of observation errors.	
Sensor Type	σ_o
Wind errors in the lower troposphere:	(m/s)
Surface stations and ship obs	3 to 4
Drifting buoy	5 to 6
Rawinsonde, wind profiler	0.5 to 2.7
Aircraft and satellite	3
Pressure errors:	(kPa)
Surface weather stations & Rawinsonde	0.1
Ship and drifting buoy	0.2
S. Hemisphere manual analysis	0.4
Geopotential height errors:	(m)
Surface weather stations	7
Ship and drifting buoy	14
S. Hemisphere manual analysis	32
Rawinsonde	13 to 26
Temperature errors:	(°C)
ASOS surface automatic weather stn.	0.5 to 1.0
Rawinsonde upper-air obs at z < 15 km	0.5
at altitudes near 30 km	< 1.5
Humidity errors:	
ASOS surface weather stations: T_d (°C)	0.6 to 4.4
Rawinsonde in lower troposph. RH (%)	5
near tropopause: RH (%)	15

A drifting buoy observes a wind of M = 10 m/s, while the first guess for the same location gives an 8 m/s wind with 2 m/s likely error. Find the analysis wind speed.

Solution

Given: M_O = 10 m/s, M_F = 8 m/s, σ_f = 2 m/s
Find: M_A = ? m/s

Use Table 20-2 for Wind Errors: Drifting buoy:
σ_o = 6 m/s

Use eq. (20.20) for wind speed M:

$$\boxed{M_A = M_F \cdot \frac{{\sigma_o}^2}{{\sigma_f}^2 + {\sigma_o}^2} + M_O \cdot \frac{{\sigma_f}^2}{{\sigma_f}^2 + {\sigma_o}^2}}$$

$$M_A = (8\text{m/s}) \cdot \frac{(6\text{m/s})^2}{(2\text{m/s})^2 + (6\text{m/s})^2} + (10\text{m/s}) \cdot \frac{(2\text{m/s})^2}{(2\text{m/s})^2 + (6\text{m/s})^2}$$

$$= (8 \text{ m/s})\cdot(36/40) + (10 \text{ m/s})\cdot(4/40) = \underline{8.2 \text{ m/s}}$$

Discussion: Because the drifting buoy has such a large error, it is given very little weight in producing the analysis. If it had been given equal weight as the first guess, then the average of the two would have been 9 m/s. It might seem disconcerting to devalue a real observation compared to the artificial value of the first guess, but it is needed to avoid startup problems.

Outline

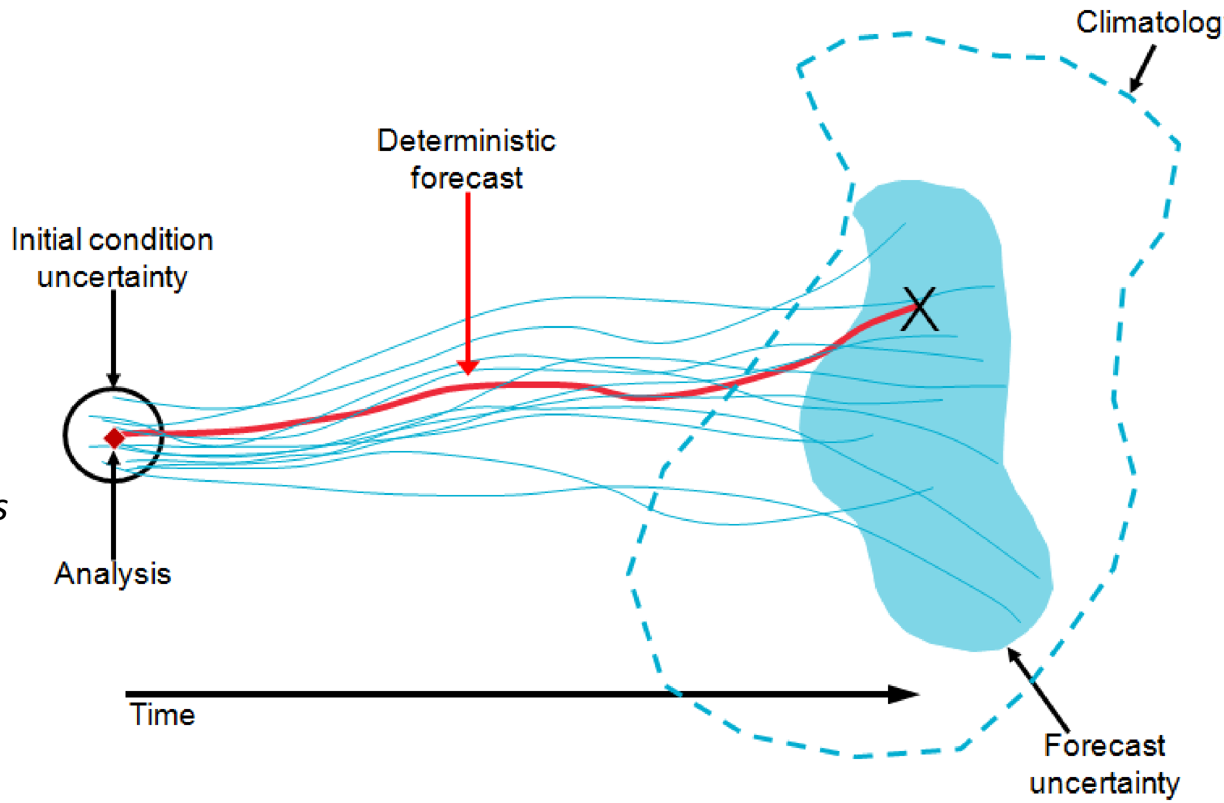
- Overview of hybrid DA methods
- NCMRWF hybrid 4DVAR system

We describe the hybrid ensemble/4D-Var global data assimilation system that is operationally running at NCMRWF.
The scheme uses the extended control variable technique to implement a hybrid background error covariance that combines the standard climatological covariance with a covariance derived from the 23-member operational ensemble NEPS.

Instead of running just a single forecast, the computer model is run a number of times from *slightly different starting conditions*. The complete set of forecasts is referred to as the ensemble, and individual forecasts within it as ensemble members.

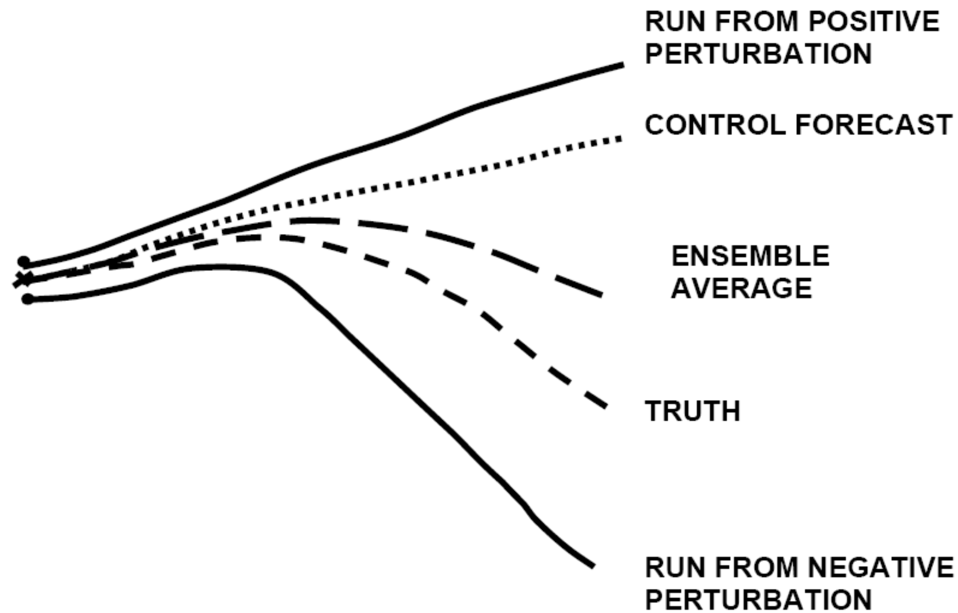
Ensemble forecast?

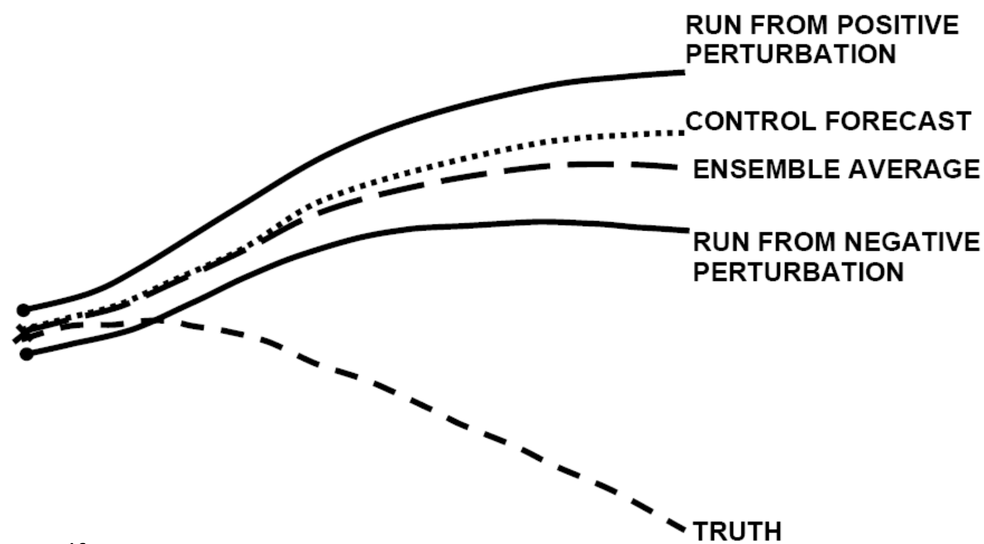
An ensemble samples the uncertainty of the forecast, assuming that the forecast model is perfect.



Ensemble forecast systems are designed so that *each member should be equally likely*. The *initial differences* between the ensemble members are *small,* and *consistent with uncertainties in the observations*. But when we look several days ahead the forecasts can be quite different.

Schematic of the essential components of an ensemble of forecasts: The analysis (denoted by a cross) constitutes the initial condition for the control forecast (dotted); two initial perturbations (dots around the analysis), chosen in this case to be equal and opposite; the perturbed forecasts (full line); the ensemble average (long dashes); and the verifying analysis or truth (dashed). The first schematic is a “good ensemble” in which the truth is a plausible member of the ensemble. The second is an example of a bad ensemble, quite different from the truth, pointing to the presence of deficiencies in the forecasting system (in the analysis, in the ensemble perturbations and/or in the model).





https://www2.atmos.umd.edu/~ekalnay/pubs/ECMWFPredictKalnay5.pdf

Data assimilation methods

Variational data assimilation (VAR) is the method of choice in many numerical weather prediction centres to estimate the state of the atmosphere for weather prediction

Can assimilate efficiently

Direct observations of meteorological fields from *in situ* instruments on sondes, aircraft and in weather stations,

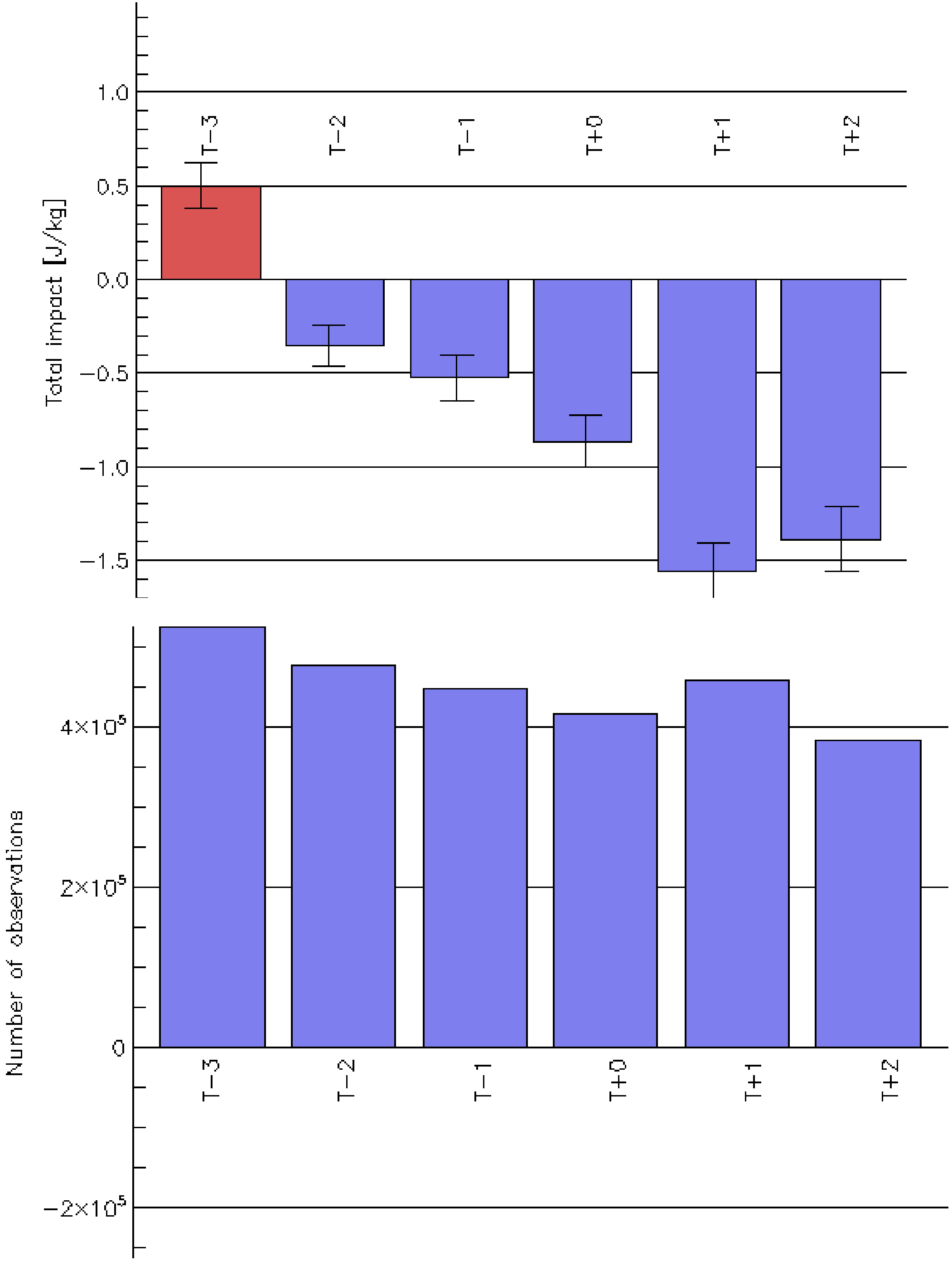
Indirect observations from satellites and from ground-based radar

Despite this wealth of observational information available, VAR needs a priori (or background) state.

Importance of background state:

It provides information otherwise missing from observations, and provides a realistic reference state needed to form the nonlinear observation operators used to assimilate many of the indirect observations.

As with all information, the background state is also prone to error, and VAR must account for this.



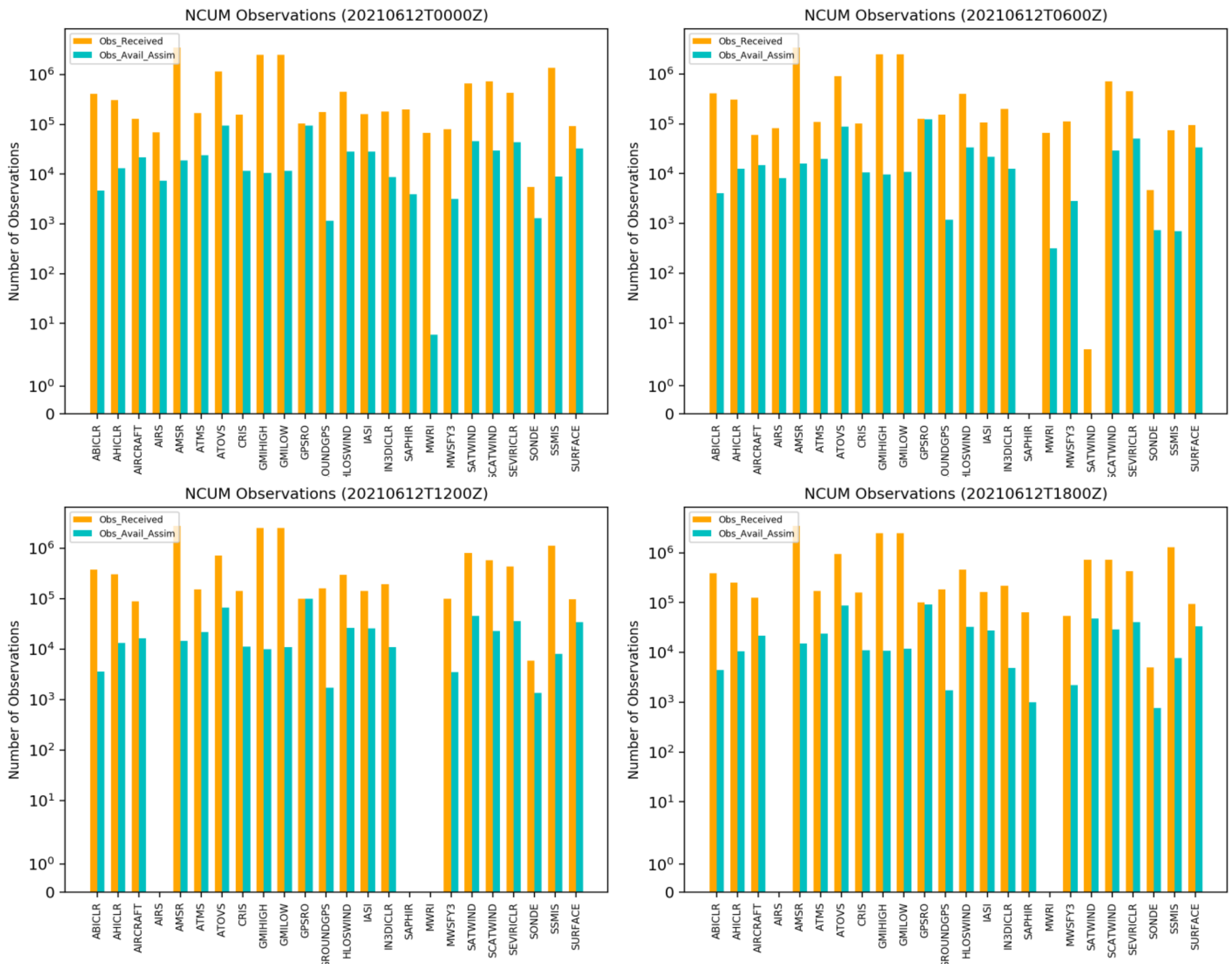
4DVAR:Assimilation of observation within 6 hour window

How 'large' is an operational DA system?

- $\boldsymbol{x}$ has typically $n \sim 10^6$-10^7 elements ($\therefore$ the **B**-matrix has 10^{12}-10^{14} matrix elements!).
- $\boldsymbol{y}$ has typically $p \sim 10^5 - 10^6$ observations.

Note that this is an order of magnitude smaller than the number of unknowns in x (hence the need to include the a-priori term).

Typical numbers of observations made by instruments (green) and assimilated in the NCMRWF global data assimilation system (magenta).



Data assimilation methods: Background error

This is most conveniently achieved through the so-called ***background error covariance statistics***, which are represented by the matrix **B**.

$$J[\delta \mathbf{x}, \mathbf{x}^g] = \frac{1}{2}\left(\delta \mathbf{x} - \delta \mathbf{x}^b\right)^T \mathbf{B}^{-1}\left(\delta \mathbf{x} - \delta \mathbf{x}^b\right) + \frac{1}{2}\left\{\mathbf{y}^o - H(\mathbf{x}^g + \delta \mathbf{x})\right\}^T \mathbf{R}^{-1}\left\{\mathbf{y}^o - H(\mathbf{x}^g + \delta \mathbf{x})\right\}$$

The cost, J, is minimized with respect to the increment $\delta\mathbf{x}$. At the minimum, $\delta\mathbf{x}$ describes the analysis, $\mathbf{x}^a$, specified with respect to a reference (or 'guess') state, $\mathbf{x}^g$, i.e. $\mathbf{x}^a = \mathbf{x}^g + \delta\mathbf{x}$.

Similarly $\delta\mathbf{x}^b$ is the incremental description of the background, $\mathbf{x}^b = \mathbf{x}^g + \delta\mathbf{x}^b$ (often the reference state is the background state and so $\delta\mathbf{x}^b = 0$)

covariance matrices uses:
weight the previous forecast, which contains information about past observations, and recent observations.

Covariance matrices describe:

- how the uncertainties in different quantities are ***correlated***, allowing us to give ***greater weight to more-accurate data*** and also to ***spread observational information*** between different atmospheric variables.

For example, a temperature observation can also be used to adjust our estimate of the wind. Operational forecast models require too many pieces of information to explicitly account for all the inter-relationships. Instead, 4DVar models the correlations using physics principles in the form of an unchanging covariance matrix.

Bannister, 2008

Covariance Matrix

A covariance matrix (also known as auto-covariance matrix or variance–covariance matrix) is **a square matrix giving the covariance between each pair of elements of a given random vector.**

A **useful tool for separating the structured relationships in a matrix of random variables**.

In a scalar system, the background error covariance is simply the variance, or the average squared departure from the mean, where

$$B = \overline{(e_b - \overline{e_b})^2}$$

In multi-dimensional system

$$B = \overline{(e_b - \overline{e_b})(e_b - \overline{e_b})^T}$$

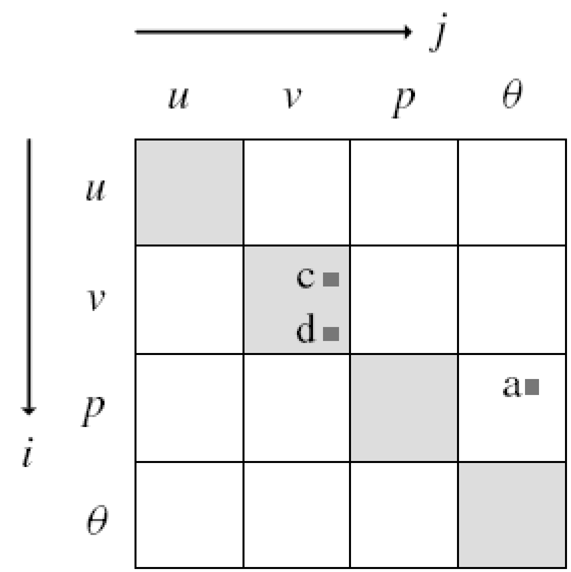
Which is a square, symmetric matrix with variance along the diagonal. E.g. for a very simple three dimensional system:

$$\mathbf{B} = \begin{bmatrix} \mathrm{var}(e_1) & \mathrm{cov}(e_1, e_2) & \mathrm{cov}(e_1, e_3) \\ \mathrm{cov}(e_1, e_2) & \mathrm{var}(e_2) & \mathrm{cov}(e_2, e_3) \\ \mathrm{cov}(e_1, e_3) & \mathrm{cov}(e_2, e_3) & \mathrm{var}(e_3) \end{bmatrix}$$

The off-diagonal terms are cross-covariances between each pair of “variables” in the model, the term variable here corresponds to the value of each physical dependent variable at each grid point.

The no of variables, and the dimension of the matrix, is the product of the no of physical variables and no of grid points.

Basic structure of the **B**-matrix for a system with four variables: zonal wind *(u)*, meridional wind *(v)*, pressure *(p)* and potential temperature *(ϑ)*. Each variable is a discrete three-dimensional field whose values at each position are represented as a vector and whose covariances are represented as a submatrix in the above. Submatrices that are the autocovariances of a single field between pairs of positions in space are the shaded block diagonal matrices, and submatrices that are the cross covariances between different variables and between pairs of positions in space (multivariate covariances) are unshaded.



Mathematical properties of the B-matrix

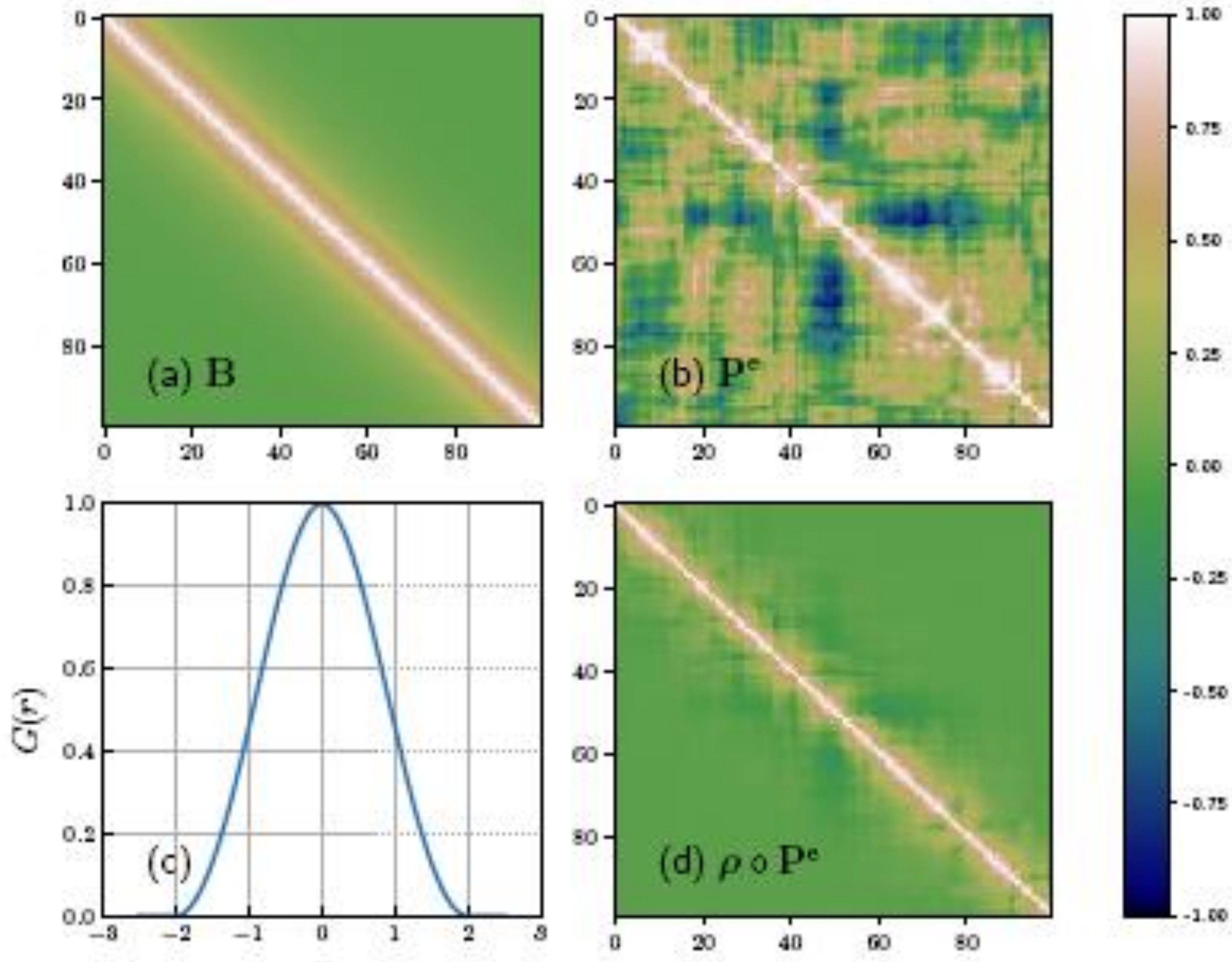
The observation error covariance matrix (right) shown against the observation vector (left). Often observation errors are taken to be uncorrelated with each other and so is diagonal. The diagonal matrix elements are the respective observation variances (equal to the square of the standard deviations) and the off-diagonal elements are zero. There are observations.

$$\vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \\ \\ \\ \\ y_p \end{bmatrix} \qquad \mathbf{R} = \begin{bmatrix} \sigma_1^2 & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \sigma_p^2 \end{bmatrix}$$

$\vec{u}$ E.ward wind field
$\vec{v}$ N.ward wind field
$\vec{\theta}$ pot. temp. field
$\vec{p}$ pressure field
$\vec{q}$ humidity field

$$\vec{x}_B = \begin{bmatrix} \\ \\ \\ \\ \\ \end{bmatrix} \qquad \mathbf{B} = \begin{array}{c|c} & \vec{u} \quad \vec{v} \quad \vec{\theta} \quad \vec{p} \quad \vec{q} \\ \hline \begin{array}{c} \vec{u} \\ \vec{v} \\ \vec{\theta} \\ \vec{p} \\ \vec{q} \end{array} & \begin{bmatrix} \mathbf{B}_{uu} & \mathbf{B}_{uv} & \mathbf{B}_{u\theta} & \mathbf{B}_{up} & \mathbf{B}_{uq} \\ \mathbf{B}_{uv} & \mathbf{B}_{vv} & \mathbf{B}_{v\theta} & \mathbf{B}_{vp} & \mathbf{B}_{vq} \\ \mathbf{B}_{u\theta} & \mathbf{B}_{v\theta} & \mathbf{B}_{\theta\theta} & \mathbf{B}_{\theta p} & \mathbf{B}_{\theta q} \\ \mathbf{B}_{up} & \mathbf{B}_{vp} & \mathbf{B}_{\theta p} & \mathbf{B}_{pp} & \mathbf{B}_{pq} \\ \mathbf{B}_{uq} & \mathbf{B}_{vq} & \mathbf{B}_{\theta q} & \mathbf{B}_{pq} & \mathbf{B}_{qq} \end{bmatrix} \end{array}$$

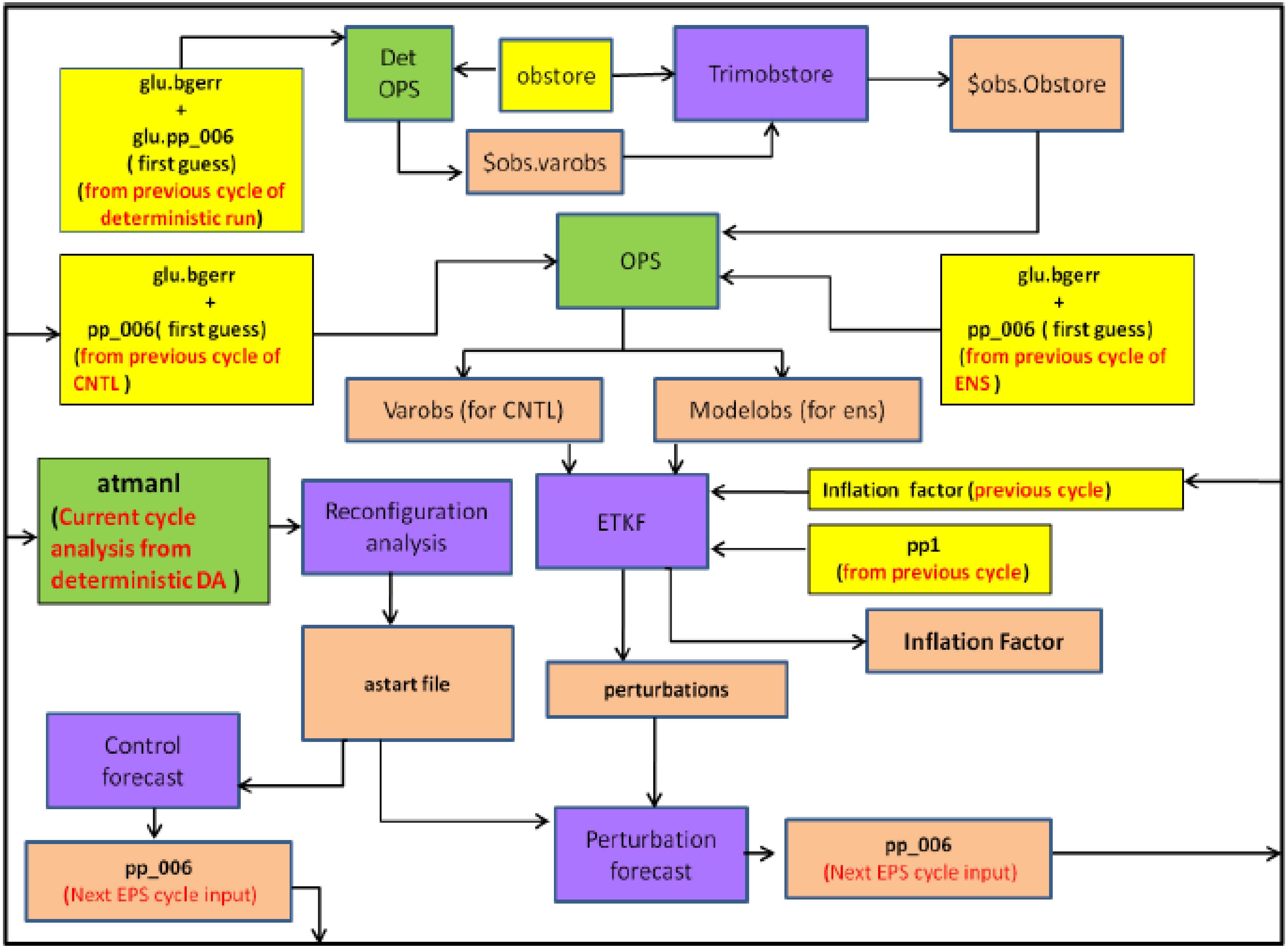
The background error covariance matrix for a forecast given in the state space. Each square is itself a matrix here. Sub-matrices along the diagonal (deep yellow) are called 'self covariances' and off-diagonal sub-matrices are called 'multivariate covariances



Panel a: True covariance matrix. Panel b: Sample covariance matrix. Panel c: Gaspari-Cohn correlation matrix used for covariance localization. Panel d: Regularized covariance matrix obtained from a Schur product.

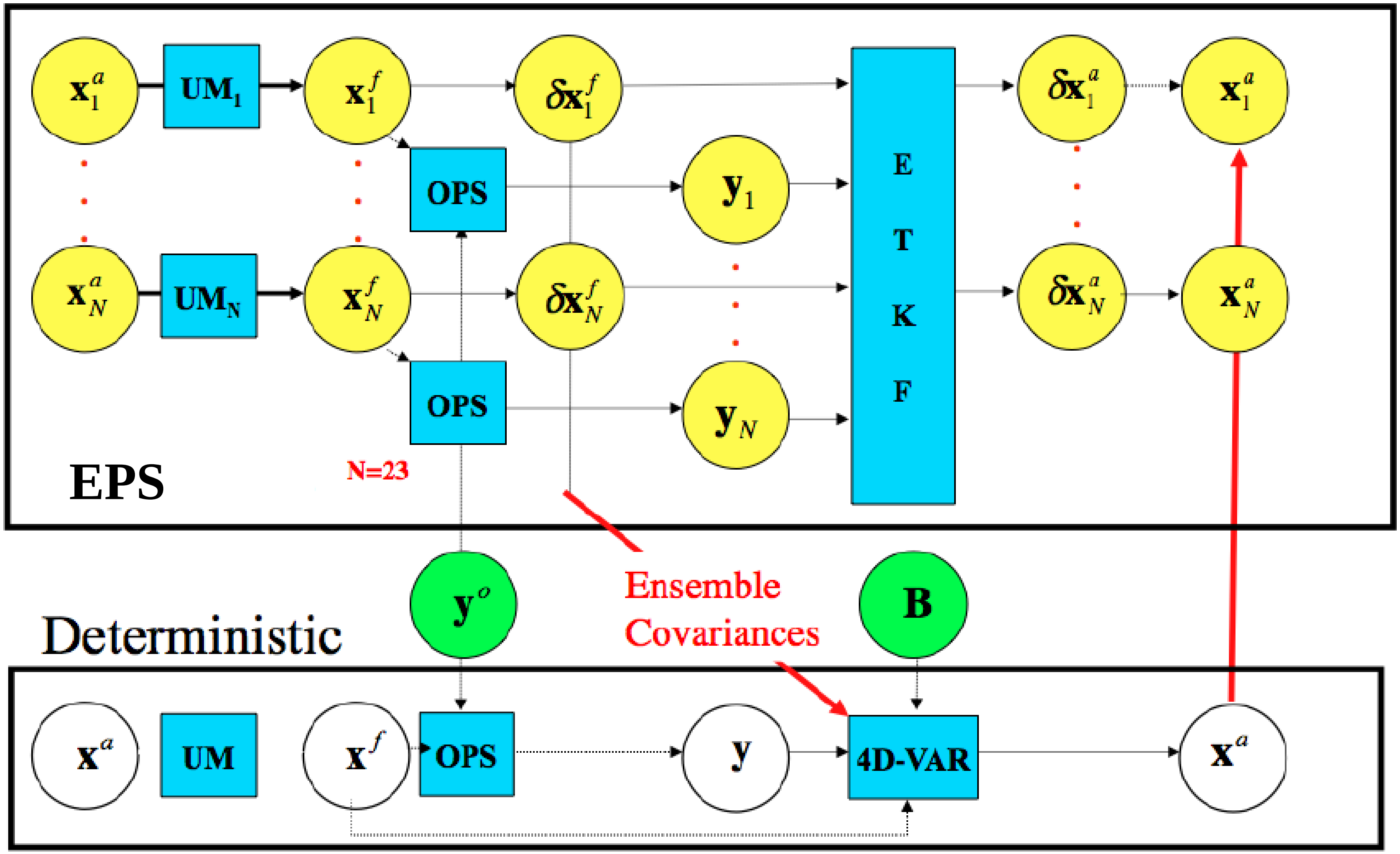
Schur (Hadamard) product of two vectors is very similar to <u>matrix addition</u>, elements corresponding to same row and columns of given vectors/matrices are multiplied together to form a new vector/matrix.
It is named after French Mathematician, Jacques Hadamard.

$$\overset{G}{\begin{bmatrix} 3 & 5 & 7 \\ 4 & 9 & 8 \end{bmatrix}} \circ \overset{H}{\begin{bmatrix} 1 & 6 & 3 \\ 0 & 2 & 9 \end{bmatrix}} = \overset{N}{\begin{bmatrix} 3\times 1 & 5\times 6 & 7\times 3 \\ 4\times 0 & 9\times 2 & 8\times 9 \end{bmatrix}}$$

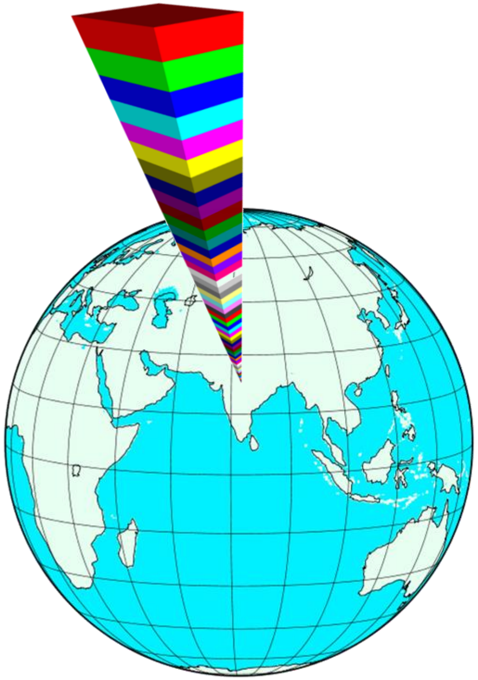

glu.bgerr
+
glu.pp_006
(first guess)
(from previous cycle of
deterministic run)
Det
OPS
obstore
Trimobstore
$obs.Obstore
$obs.varobs
OPS
glu.bgerr
+
pp_006(first guess)
(from previous cycle of
CNTL)
glu.bgerr
+
pp_006 (first guess)
(from previous cycle of
ENS)
Varobs (for CNTL)
Modelobs (for ens)
atmanl
(Current cycle
analysis from
deterministic DA)
Reconfiguration
analysis
ETKF
Inflation factor (previous cycle)
pp1
(from previous cycle)
Inflation Factor
astart file
perturbations
Control
forecast
pp_006
(Next EPS cycle input)
Perturbation
forecast
pp_006
(Next EPS cycle input)

Hybrid 4DVar

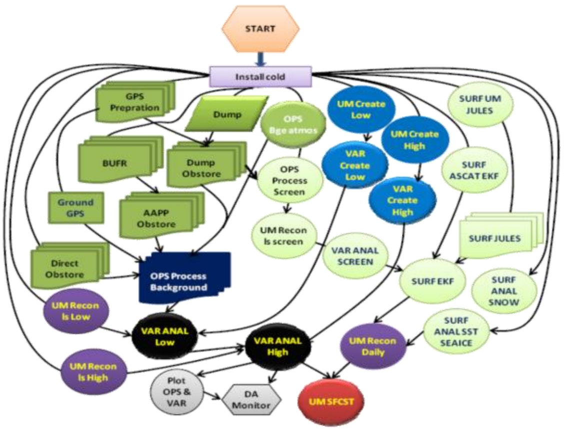
A drawback of 4DVar is that the forecast covariance matrix does not take account of the day-to-day weather characteristics. For example, we would expect correlations to be stretched along a weather front. One way to include this flow-dependence is to use ensemble forecasting to represent the forecast uncertainty. We randomly perturb the previous forecast several times to produce a set of possible initial atmospheric states, each of which is evolved in time using the forecast model. The ensemble is then used to construct a flow-dependent covariance matrix. In practice, the ensemble size is limited by our supercomputing capacity so a hybrid approach is used that blends the unchanging covariance matrix of traditional 4DVar with the ensemble covariance matrix. Hybrid 4DVar is used operationally for global forecasting at the Met Office.



Sketch of the interactions between EPS (upper box) and high-resolution deterministic NWP (lower box) systems. UM=Unified model, OPS=Observation Preprocessing System, ETKF=Ensemble Transform Kalman Filter. The red arrow denotes the coupling supplying ensemble perturbations as estimates of flow-dependent forecast error to the data assimilation, and 4D-Var analysis to which ETKF-updated ensemble perturbations are added for the next cycle of ensemble forecasts.



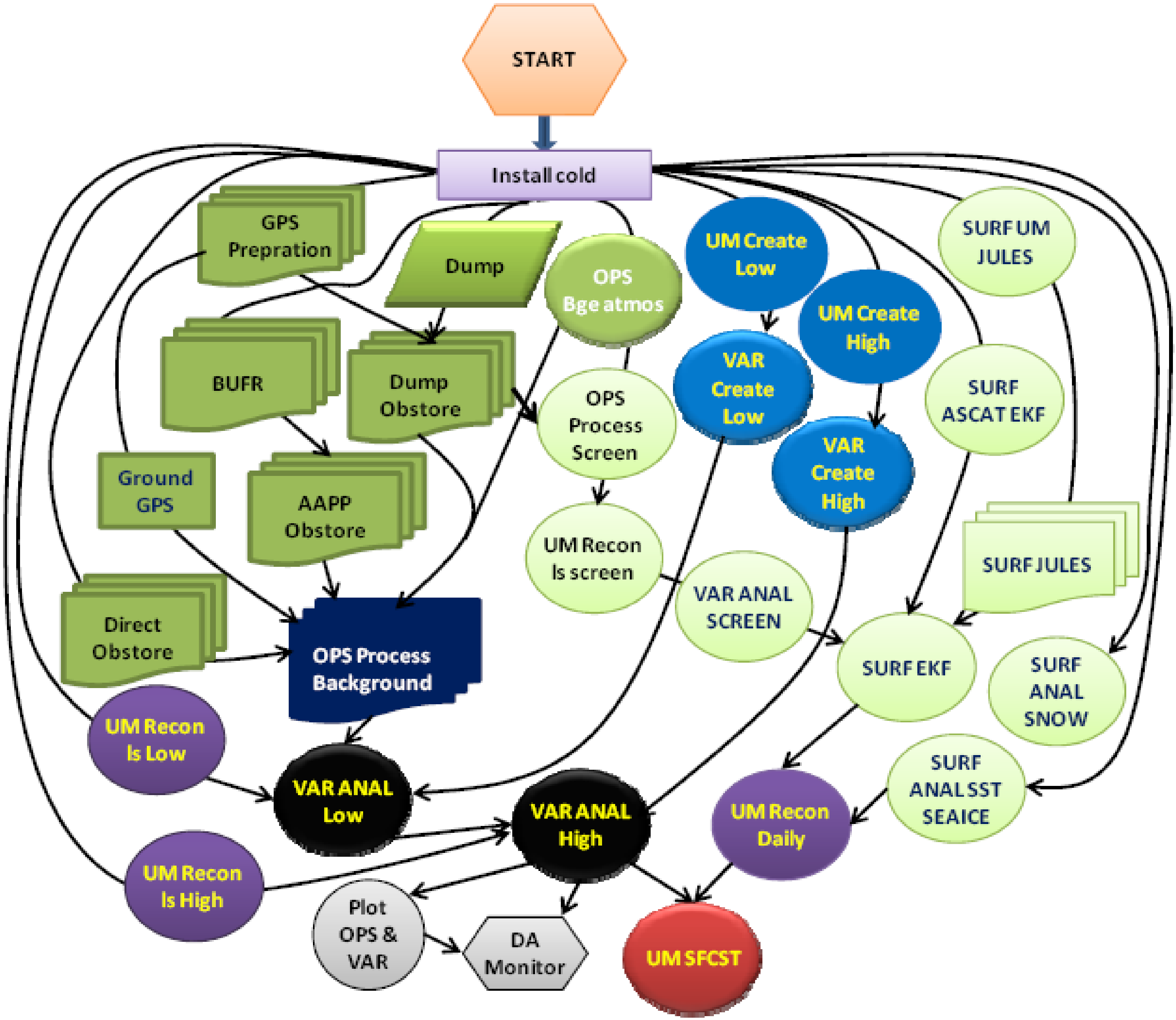
Global NCUM System (12 km resolution)



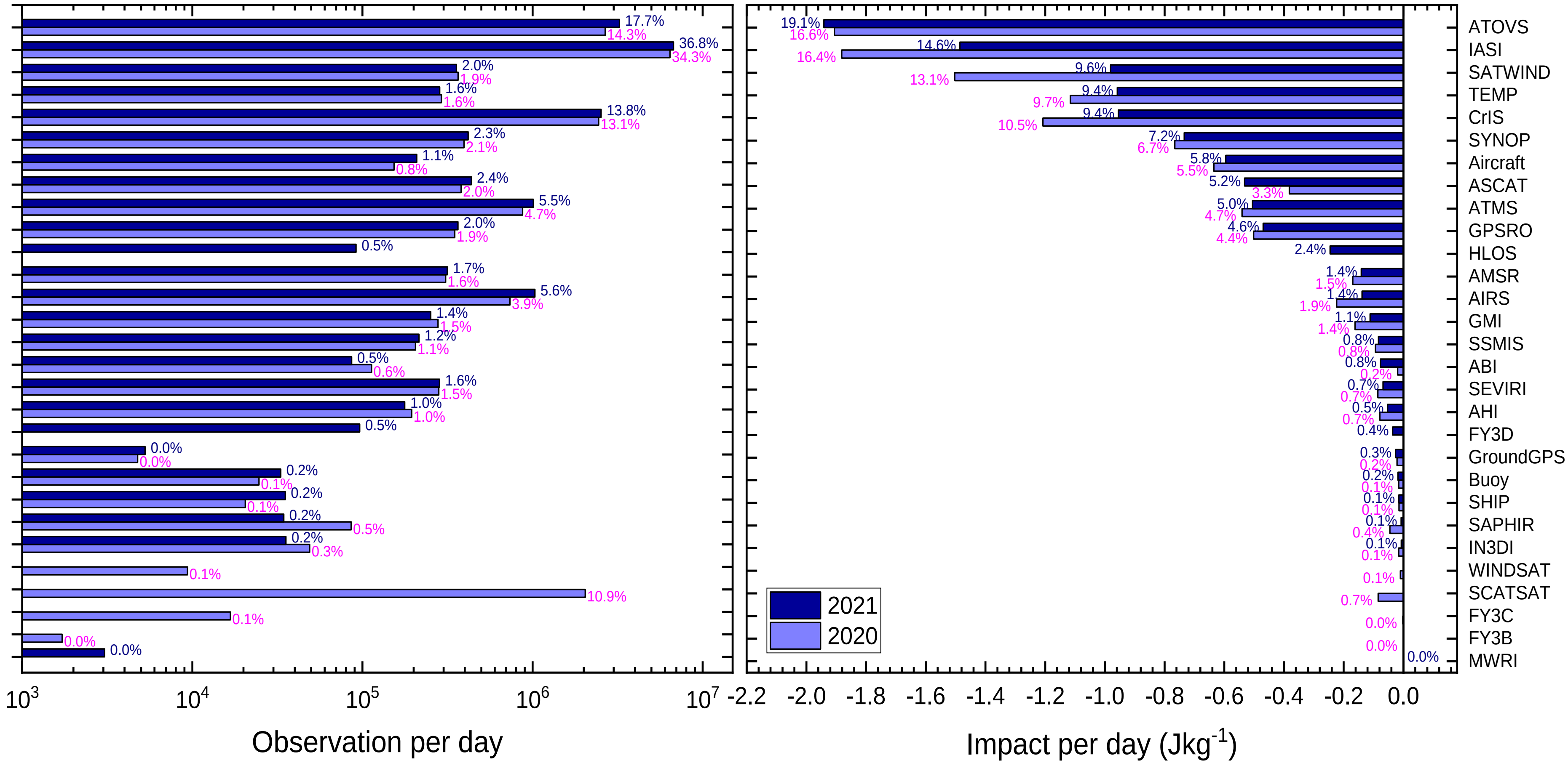
Flow chart of NCUM system (in Mihir HPCS)

Salient Features of NCUM Assimilation–Forecast System

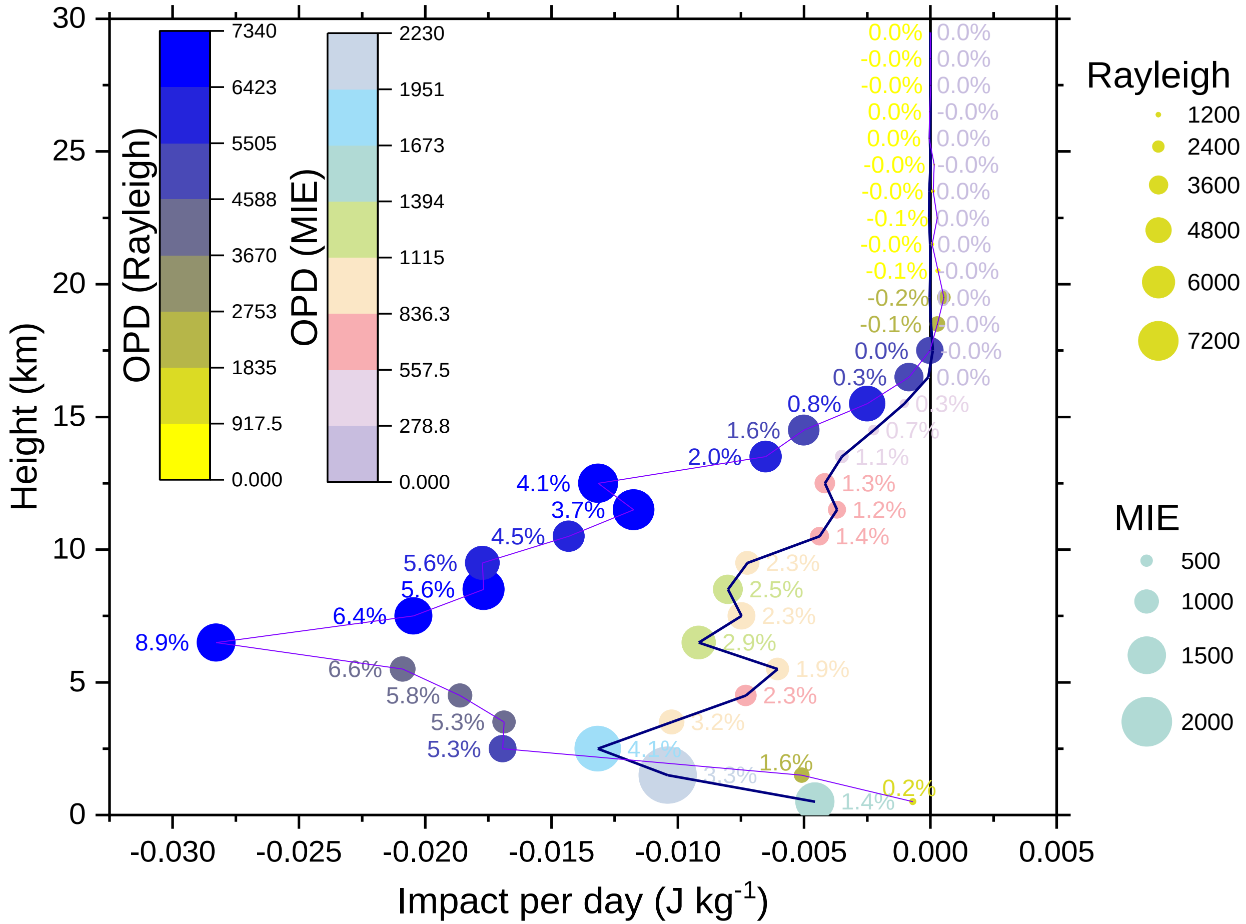
Model	Atmospheric Data Assimilation	Surface Analysis
Model: Unified Model; Version 10.8 **Domain:** Global **Resolution:** 12 km, Levels 70 **No. of Grids:** 2048x1536 **Time Step:** 5 minutes **Physical Parameterizations:** based on GA6.1 (Walters et al, Geosci. Model Dev., 10: 1487-1520, 2017) **Dynamical Core:** ENDGame **Forecast length:** 10 days (based on 00 UTC and 12 UTC initial conditions)	**Resolution:** N320L70 (~40 km) with N144L70 Hessian based pre-conditioning **Method:** Hybrid incremental 4D-Var. Information on "errors of the day" is provided by NEPS forecast at every data assimilation cycle **Data Assimilation Cycles:** 4 analyses per day at 00, 06, 12 and 18 UTC. Observations within +/- 3 hrs from the cycle time is assimilated in respective DA cycle **Observations:** Observation Processing System does the quality control of observations. Variational bias correction is applied to satellite radiance observations.	**Soil Moisture Analysis:** *Method:* Extended Kalman Filter *Analysis time:* 00, 06, 12 and 18 UTC *Observations assimilated:* ASCAT soil wetness observations, Screen level Temperature and Humidity (pseudo observations from 3D-Var screen analysis) **Sea Surface Temperature:** Updated at 12 UTC DA cycle with OSTIA based SST and sea-ice analysis **Snow Analysis:** Satellite-derived snow analysis. Updated at 12 UTC DA cycle


START
Install cold
GPS Prepration
Dump
OPS Bge atmos
UM Create Low
SURF UM JULES
UM Create High
BUFR
Dump Obstore
VAR Create Low
SURF ASCAT EKF
OPS Process Screen
VAR Create High
Ground GPS
AAPP Obstore
UM Recon ls screen
SURF JULES
VAR ANAL SCREEN
Direct Obstore
OPS Process Background
SURF EKF
SURF ANAL SNOW
UM Recon ls Low
VAR ANAL Low
VAR ANAL High
SURF ANAL SST SEAICE
UM Recon Daily
UM Recon ls High
Plot OPS & VAR
DA Monitor
UM SFCST

Observation Type	Observation Description	Assimilated Variables
Surface	Surface observations over Land and Ocean, TC bogus (Surface Pressure)	Wind, Temp, Humidity, Surface Pressure
Sonde	Radiosonde (TAC & BUFR), Pilot balloons, Wind profiles &Radar VAD winds	Wind, Temp, Humidity
Aircraft	Upper-air wind and temperature from aircraft (AMDAR & AIREP)	Wind, Temp
GroundGPS	Ground based GPS observations	Zenith Total Delay
Satellite:GPSRO	Global Positioning System Radio Occultation observations from various satellites (Terra-Sar X, COSMIC, FY3D, KOMPSAT, MetOp (A, B & C))	Bending Angle
Satellite:Satwind	Atmospheric Motion Vectors from geostationary and polar orbiting satellites (MSG, JMA, GOES,MetOp,INSAT-3D & INSAT-3DR, MODIS, NOAA)	Wind
Satellite:Scatwind	Advanced Scatterometer in MetOp-A & B, ScatSat-1, WindSat	Wind
Satellite:MicroWave Sounder/Imager	Microwave sounders / imagers ATMS, AMSU, GMI, MWHS, AMSR2, SAPHIR, SSMIS	Brightness Temperature
Satellite:Hyperspectral IR	Hyperspectral infrared sounders IASI,CrIS,AIRS	Brightness Temperature
Satellite: Geostationary Sounder/Imager	Sounder/Imagers from MSG,GOES,Himawari,INSAT	Brightness Temperature
Satellite:HLOS Wind	Mie-scattering and Rayleigh-scattering Horizontal Line-Of-Sight (HLOS) winds from AEOLUS satellite	HLOS wind
Satellite:Aerosol Optical Depth (AOD)	Dust aerosol optical depth from MODIS (Terra & Aqua) satellite	AOD



ATOVS
IASI
SATWIND
TEMP
CrIS
SYNOP
Aircraft
ASCAT
ATMS
GPSRO
HLOS
AMSR
AIRS
GMI
SSMIS
ABI
SEVIRI
AHI
FY3D
GroundGPS
Buoy
SHIP
SAPHIR
IN3DI
WINDSAT
SCATSAT
FY3C
FY3B
MWRI
2021
2020
10^3 10^4 10^5 10^6 10^7
Observation per day
-2.2 -2.0 -1.8 -1.6 -1.4 -1.2 -1.0 -0.8 -0.6 -0.4 -0.2 0.0
Impact per day (Jkg^{-1})



OPD (Rayleigh)
7340
6423
5505
4588
3670
2753
1835
917.5
0.000
OPD (MIE)
2230
1951
1673
1394
1115
836.3
557.5
278.8
0.000
Rayleigh
1200
2400
3600
4800
6000
7200
MIE
500
1000
1500
2000
Height (km)
Impact per day (J kg-1)
8.9%
6.4%
5.6%
5.6%
4.5%
3.7%
4.1%
2.0%
1.6%
0.8%
0.3%
0.0%
6.6%
5.8%
5.3%
5.3%

Summary

- There is a natural linkage between data assimilation and ensemble forecasting: ensemble forecasts can estimate the flow-dependent uncertainty of the forecast;
- Data assimilation techniques require accurate estimates of forecast uncertainty in order to optimally blend the prior forecast(s) with new observations.
- Thus these two endeavors are united and this union certainly improve the quality of both initial conditions and subsequent forecasts.

Hybrid (variational/ensemble) data assimilation approaches attempt to combine the best of both variational and ensemble frameworks.

Thank you for patience listening