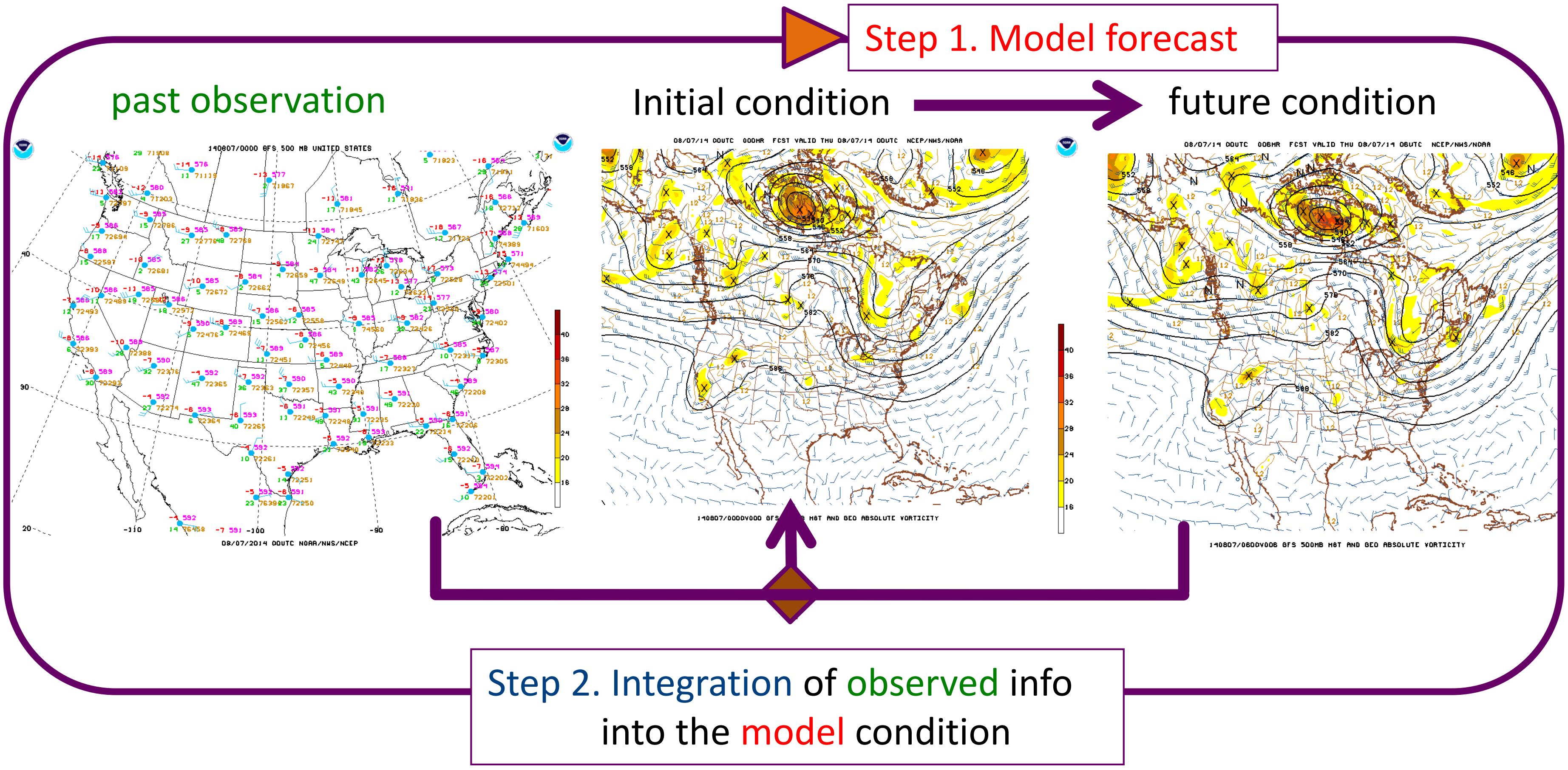
Data Assimilation Plans at NOAA/NWS/NCEP

Daryl Kleist
NOAA/NWS/NCEP/EMC

Yannick Trémolet
JCSDA

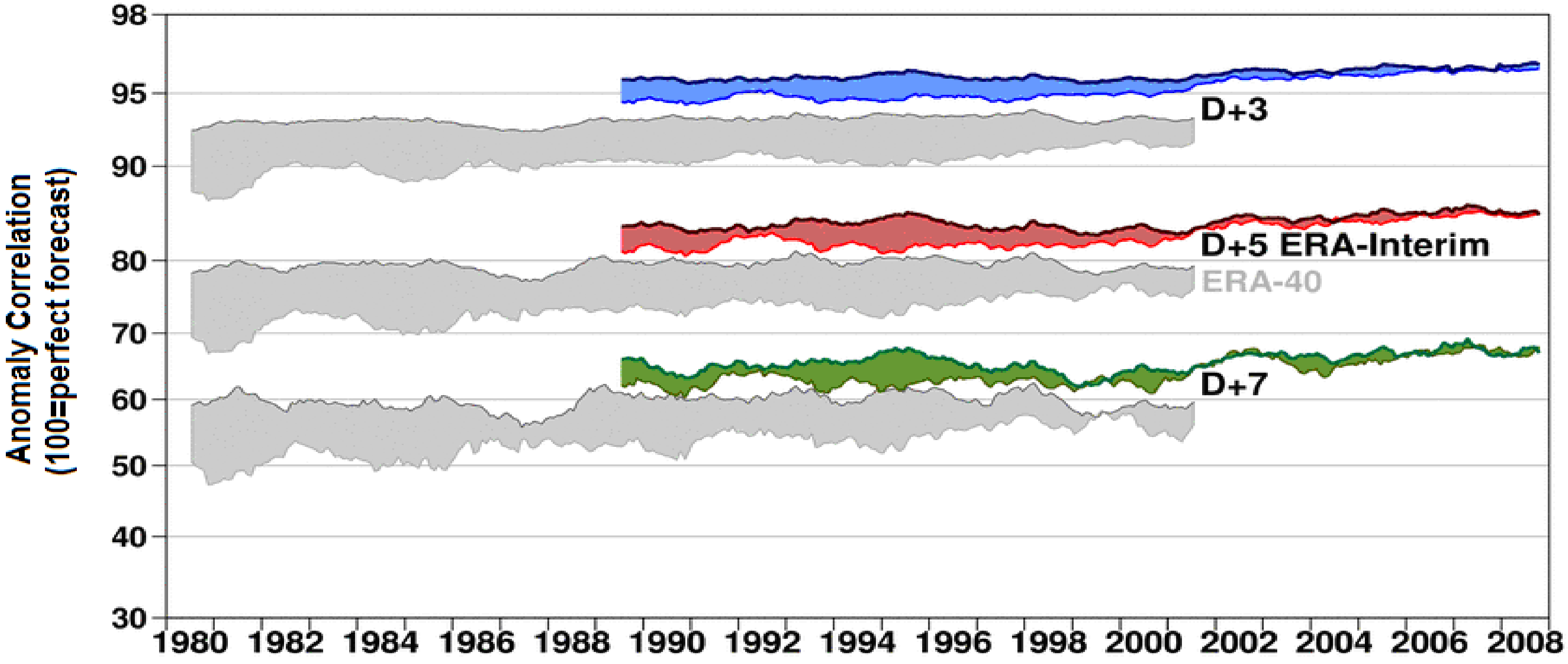
8th NCAR MultiCore Workshop (MC8)
Boulder, CO 18-19 September 2018

Data Assimilation



Courtesy: Kayo Ide

Importance of DA Algorithm: Ability to extract more information



ECMWF 500-hPa geopotential height anomaly correlations from two different reanalysis systems. Gray: ERA-40 (Uppala et al., 2005) with 3D-Var (ca. 1998); Colors: ERA-Interim (Dee et al., 2011) which uses 4D-Var (ca. 2005). “D+3” corresponds to the 3-day forecast; “D+5” the 5-day forecast; and, “D+7” the 7-day forecast. In each case the top line is the anomaly correlation of the forecasts started from the reanalysis for the Northern Hemisphere, and the bottom line is the corresponding forecast for the Southern Hemisphere. Note the improvement brought about by the improvement of the data assimilation system, which is especially important in the Southern Hemisphere. SOURCE: NRC, 2010a and ECMWF.

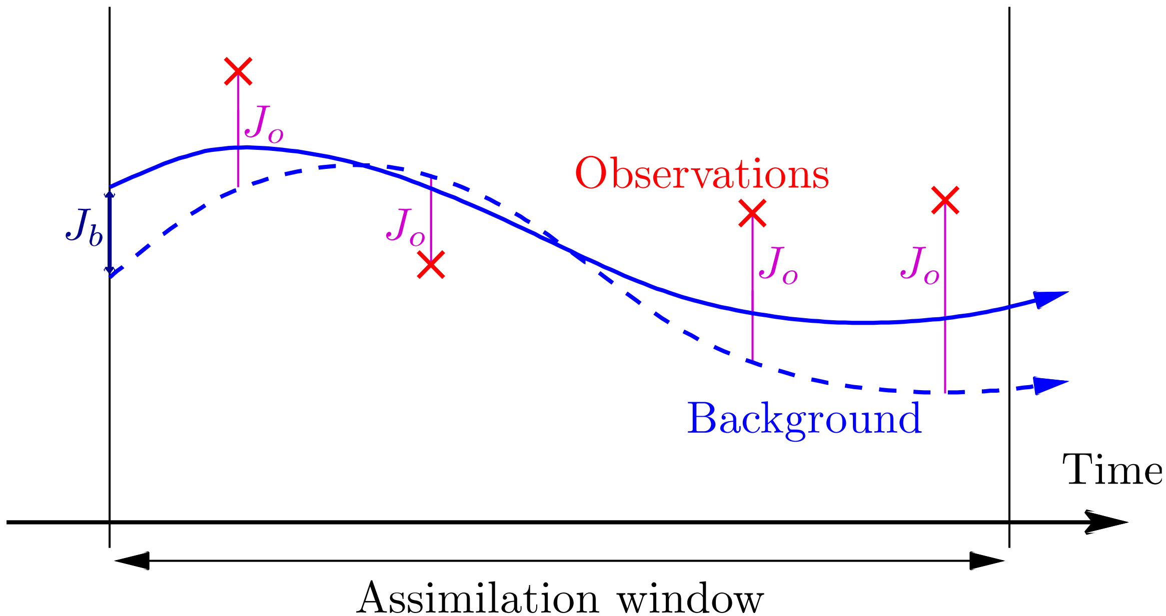
Variational Data Assimilation

Courtesy: Yannick Trémolet

Variational Data Assimilation is used by operational centers for NWP (GSI, NAVDAS, IFS, VAR, ...)

Principle: minimize the distance between the analysis and all available observations over the assimilation window. ***Solved for iteratively.***

- $\mathbf{x}_b$ Background state
- $\mathbf{y}$ Observations
- $\mathcal{H}$ Observation operator
- $\mathbf{B}$ Background error covariance
- $\mathbf{R}$ Observation error covariance



$$J(x) = \frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) + \frac{1}{2}[\mathcal{H}(\mathbf{x}) - \mathbf{y}]^T \mathbf{R}^{-1}[\mathcal{H}(\mathbf{x}) - \mathbf{y}]$$

Some Considerations

Courtesy: Yannick Trémolet

Computational issue: the size of the data assimilation problem

- The size of **x** is $O(10^9)$
- The size of **y** is $O(10^7)$
- The observation error covariance matrix is diagonal (or nearly)
- The background error covariance matrix cannot be stored or inverted
  - it must be modeled and coded as a series of linear operators: Spectral, Wavelets, Recursive filters, Diffusion operator…
  - Change of variable to avoid inversion
- Even vectors (**x** and **y**) have to be distributed across many processors to fit in memory
  - Adds complexity, especially for non local observations

Use of ensembles and "hybrids"

- Operational NWP has been utilizing variational DA for decades
  - At least partly driven by direct assimilation of satellite data
  - Requires TL/AD version of model in solver (more on this next slide)

- Ensemble techniques provide clear benefit from flow-dependent, multivariate background error estimates (**B**)
  - Finite-sized ensembles introduce serious sampling issues

- Blended covariance estimates have been demonstrated to be superior for many circumstances (including simple toy applications)

Hybrid Variational DA Solvers

- Hybrid 4DVar

$$J(\mathbf{x}_c',\alpha)=\beta_c\frac{1}{2}(\mathbf{x}')^{\mathrm{T}}\mathbf{B}_c^{-1}(\mathbf{x}')+\beta_e\frac{1}{2}\alpha^{\mathrm{T}}\mathbf{L}^{-1}\alpha+\frac{1}{2}\sum_{k=1}^{K}(\mathbf{y}_k'-\mathbf{H}_k\mathbf{M}_k\mathbf{x}_0')^{\mathrm{T}}\mathbf{R}_k^{-1}(\mathbf{y}_k'-\mathbf{H}_k\mathbf{M}_k\mathbf{x}_0')$$

$$\mathbf{x}_0'=[\mathbf{x}_c'+\sum_{m=1}^{M}(\alpha^m\circ(\mathbf{x}_e)^m)]$$

- Key Points:
  - As in 4DVar, linearized model (**M**) and adjoint (**M**$^{\mathrm{T}}$) are part of the iterative solver
  - Ensemble is used to help prescribe background error covariance at beginning of assimilation window

Hybrid Variational DA Solvers

- Hybrid 4D EnVar

$$J(\mathbf{x}'_{\mathrm{c}},\alpha)=\beta_{\mathrm{c}}\frac{1}{2}(\mathbf{x}')^{\mathrm{T}}\mathbf{B}_{\mathrm{c}}^{-1}(\mathbf{x}')+\beta_{\mathrm{e}}\frac{1}{2}\alpha^{\mathrm{T}}\mathbf{L}^{-1}\alpha+\frac{1}{2}\sum_{k=1}^{K}(\mathbf{y}'_k-\mathbf{H}_k\mathbf{x}'_k)^{\mathrm{T}}\mathbf{R}_k^{-1}(\mathbf{y}'_k-\mathbf{H}_k\mathbf{x}'_k)$$

$$\mathbf{x}'_k=\mathbf{C}_k\,[\mathbf{x}'_{\mathrm{c}}+\sum_{m=1}^{M}(\alpha^m\circ(\mathbf{x}_{\mathrm{e}})_k^m)]$$

- Key Points:
  - Unlike hybrid 4DVar, linearized model (**M**) and adjoint (**M**$^{\mathrm{T}}$) are not part of solver (this may or may not be good….).
  - 4D-ness is extracted from pre-computed ensemble trajectories (IO and other issues induced.

Use of Ensembles

- Most of computational cost is not in critical path. Easy to parallelize.

- Rank Deficiency: We can only afford to run a small sized ensemble relative to size of problem (even locally). We are currently using O(100) members.
  - Localization

- Ensembles mean lots of data to read in/out
  - IO is a HUGE challenge

- Need to maintain/update the ensemble to be representative of background and analysis error covariances
  - Currently running a second, ensemble data assimilation

Observations

Animation Courtesy Will McCarty (NASA GMAO)

Observations

- Number of observations growing exponentially
  - Especially with hyperspectral IR from satellite, dual-polarization radar, etc.
  - More observations coming from non-traditional (and private) sources

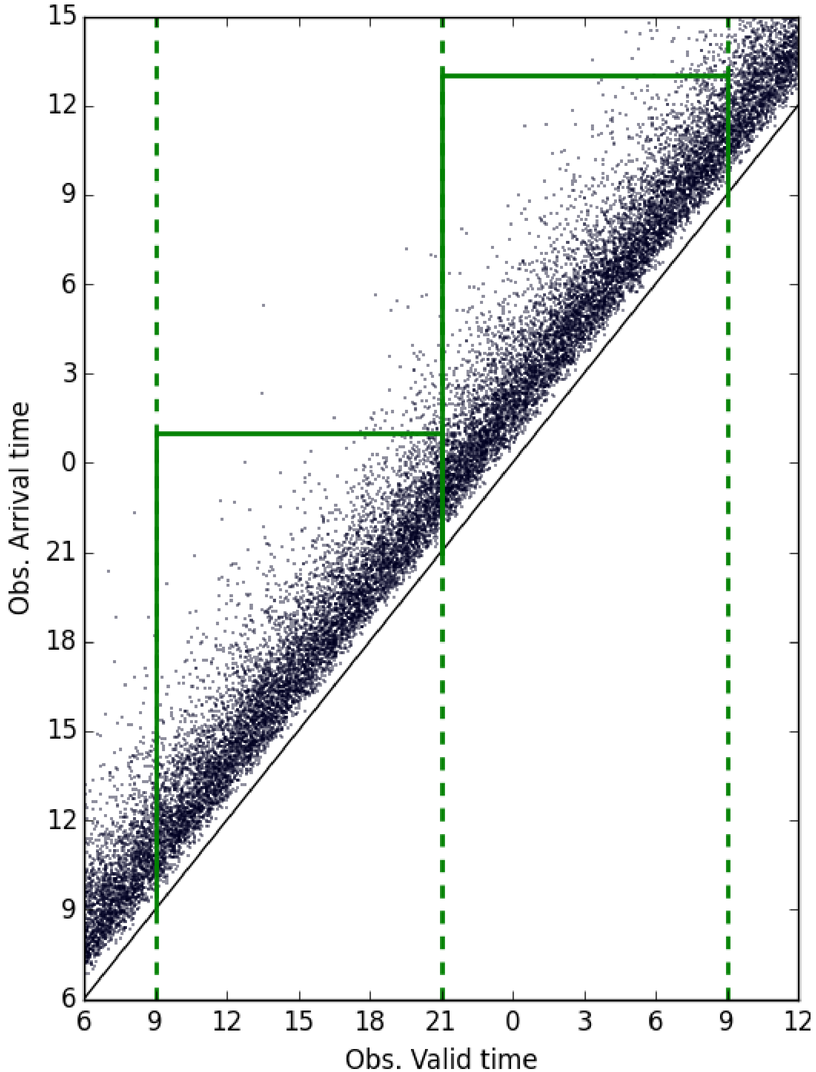
- Given asynchronous nature of observations, can be challenge from HPC perspective
  - Load balancing, communication, quality control

- Complexity continues to grow
  - Nonlinear, nonlocal, non-Gaussian

- Not all observations arrive in real-time

Observations

Courtesy: Yannick Trémolet



To perform analysis, observations are used within time window

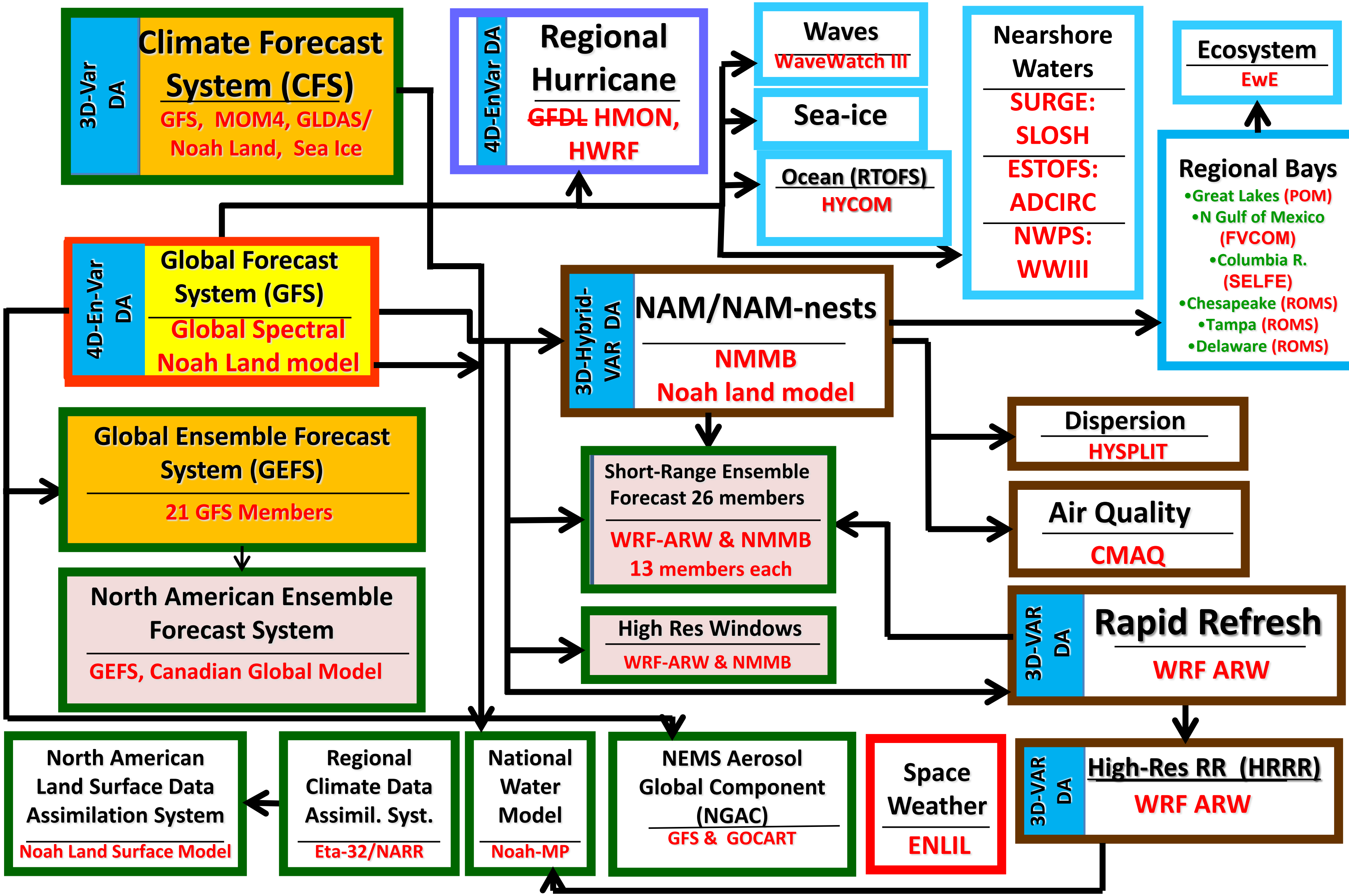
Computational issue: Observations do not instantaneously appear at operational centers:

- Communications delays
- Ground stations locations
- Pre-processing…

Some observations are lost

Some computational effort is lost

NOAA's Operational Numerical Model Guidance Suite



Status of Current Data Assimilation at NCEP

- For atmosphere, NCEP has leveraged single data assimilation code: GSI (Gridpoint Statistical Interpolation)
  - Operational for RAP/HRRR, NAM, GFS/GDAS, CFS/CDAS, HWRF, RTMA/URMA

- GSI includes (some) community engagement and support (largely through DTC)
  - Formal review process & committee

- GSI has also proven successful community tool, collaboratively developed between NCEP, GMAO, ESRL, and some academic partners
  - This includes EnKF (EnSRF & LETKF using GSI observation operators), now under the GSI “umbrella”

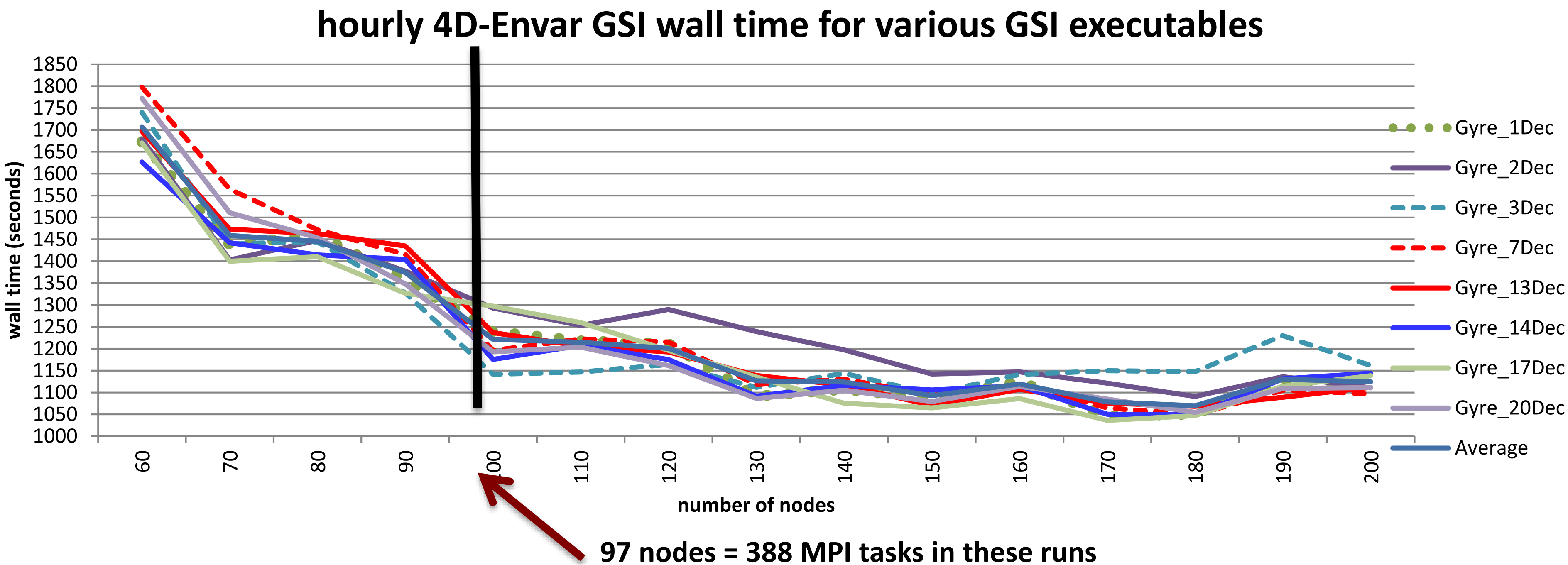
Status of Current Data Assimilation

- Ensembles and hybrids are now state-of-the science, operational for global NWP at most centers
  - Also critical for regional systems, sometimes leveraging global ensemble information
  - Direct connection to ensemble prediction systems
  - Variants also applicable to non-atmospheric component applications

- NCEP has largely pursued adjoint-free developments
  - For 4D, implementation of ***hybrid 4DEnVar*** for GDAS/GFS
  - This is the starting point for FV3-GFS

- Need for consolidation across most (all?) applications including non-atmospheric

Technical Challenge Example: Scalability



Current software has bottlenecks related to MPI decompositions to allow for 2D global operations (horizontal recursive filter here). Multiple domain decompositions also comes with significant communications overhead.

IFS Computational Cost Breakdown

Courtesy Yannick Trémolet

	Time (sec.)	%
$\mathcal{M}$	500	19.5
M	1143	44.5
$\mathcal{H}$	71	2.8
H	228	8.9
B	120	4.7
R	59	2.3
Lin. algebra	53	2.0
I/O	≈305	≈11.8
Other	92	3.6

IFS 4D-Var (12h window)

T1279/T255/T319/T399

≈30M active observations

528 MPI tasks, 18 OpenMP threads

264 nodes, 6336 cores

≈43 min. wall clock time

Scientific Plan for Data Assimilation

- Algorithm: Inter-comparison study of Hybrid 4DEnVar and Hybrid 4DVar (with FV3 TL/AD)
  - Consider implication of choices on coupled data assimilation
    - Is TL/AD available for coupled model, etc.
  - Further exploitation of information from ensembles
    - Scale dependent hybrids (weights, localization), shifting/lagging, multi-resolution

- How to deal with differing temporal / spatial scales of components for coupled system?
  - Alternate cycling strategy
    - Various overlapping windows with differing lengths?

- Range of applications creates significant challenges

- ***Choice of algorithms may be application dependent***! This has implications for coupled assimilation and rapidly updating, convective scale DA.

Scientific Plan for Data Assimilation: "Longer Term"

- Currently in process of drafting research & strategic plan for improving operational data assimilation capabilities in the 5-10 year timeframe. Beyond some of the aforementioned:
  - ***Careful consideration to computing aspects***
  - Leveraging machine/deep learning
  - Alternate cycling strategies including overlapping windows
  - "In-core" data assimilation
  - Non-Gaussian, nonlinear errors
  - Bridging very short timescale (WoF) to S2S and beyond

Joint Effort for Data assimilation Integration (JEDI)

STRATEGY

1. Collective path toward National Unified Next-Generation Data Assimilation
2. Modular, Object-Oriented code for flexibility, robustness and optimization
3. Mutualize **model-agnostic** components across
   - Applications (atmosphere, ocean, land, aerosols, etc.)
   - Models & Grids (regional/global, FV3)
   - Observations (past, current and future)

OBJECTIVES

1. Facilitate **innovation** to address next scientific grand challenges
2. Increase **R2O** transition rate
3. Increase **science productivity** and code **performance**

JEDI

- Leveraging the Object Oriented Prediction System (OOPS)
  - Partially designed to investigate scalability. Flexibility to explore various algorithms for different architectures
  - Designed to do more in memory (observation equivalents, in core solver)
    - Reduce IO
  - Latest tests show OOPS-based DA 20% faster for IFS for same algorithm
    - Maintainability and flexibility do not necessarily compromise performance

- More generally, bringing modern software development methods to our community

- Plan to replace current operational DA for global NWP with JEDI-based system in less than 4 years
  - Incremental components to be implemented when ready