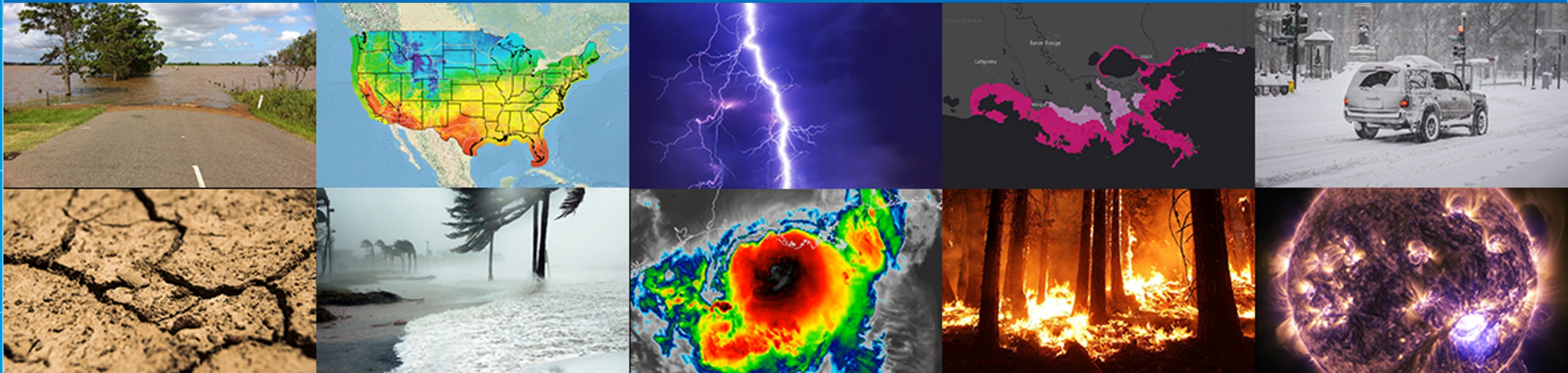
Data Assimilation –
Introduction for Breakout Session

NATIONAL WEATHER SERVICE

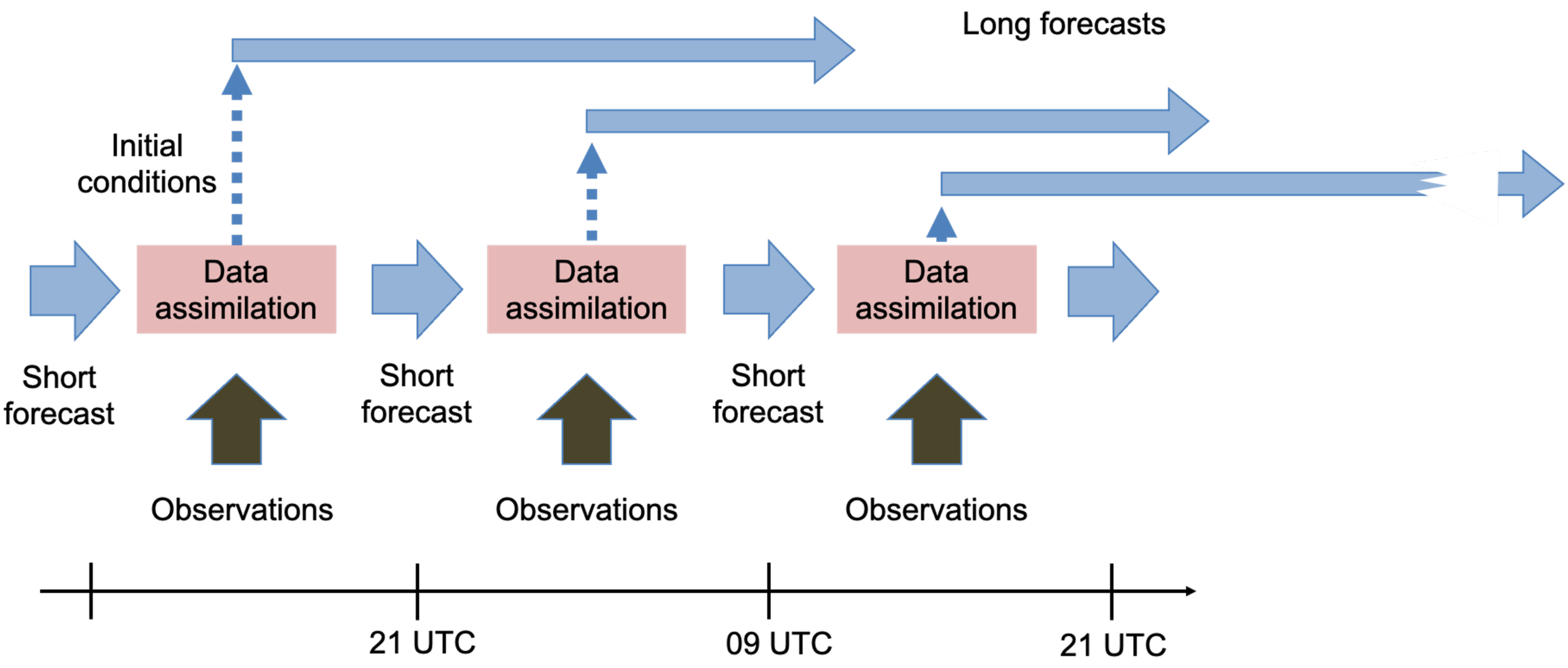
Daryl Kleist - Chief, Data Assimilation & Quality Control Group
Environmental Modeling Center
NOAA/NWS/NCEP

AI4NWP Workshop, 28-29 November 2023, Boulder, CO



Data assimilation for numerical weather prediction: (NWP)

Blending short forecasts and new observational data in a statistically optimal way



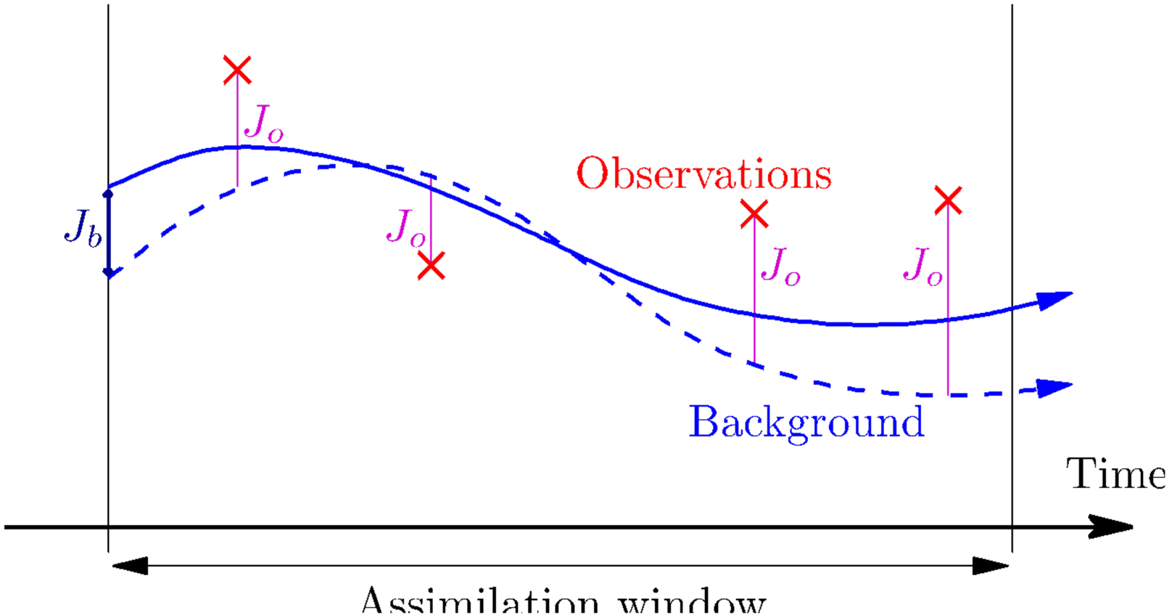
From Alan Geer (ECWMF, Presentation at 2nd NOAA AI Workshop)

Variational DA (adapted from Y. Tremolet)

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

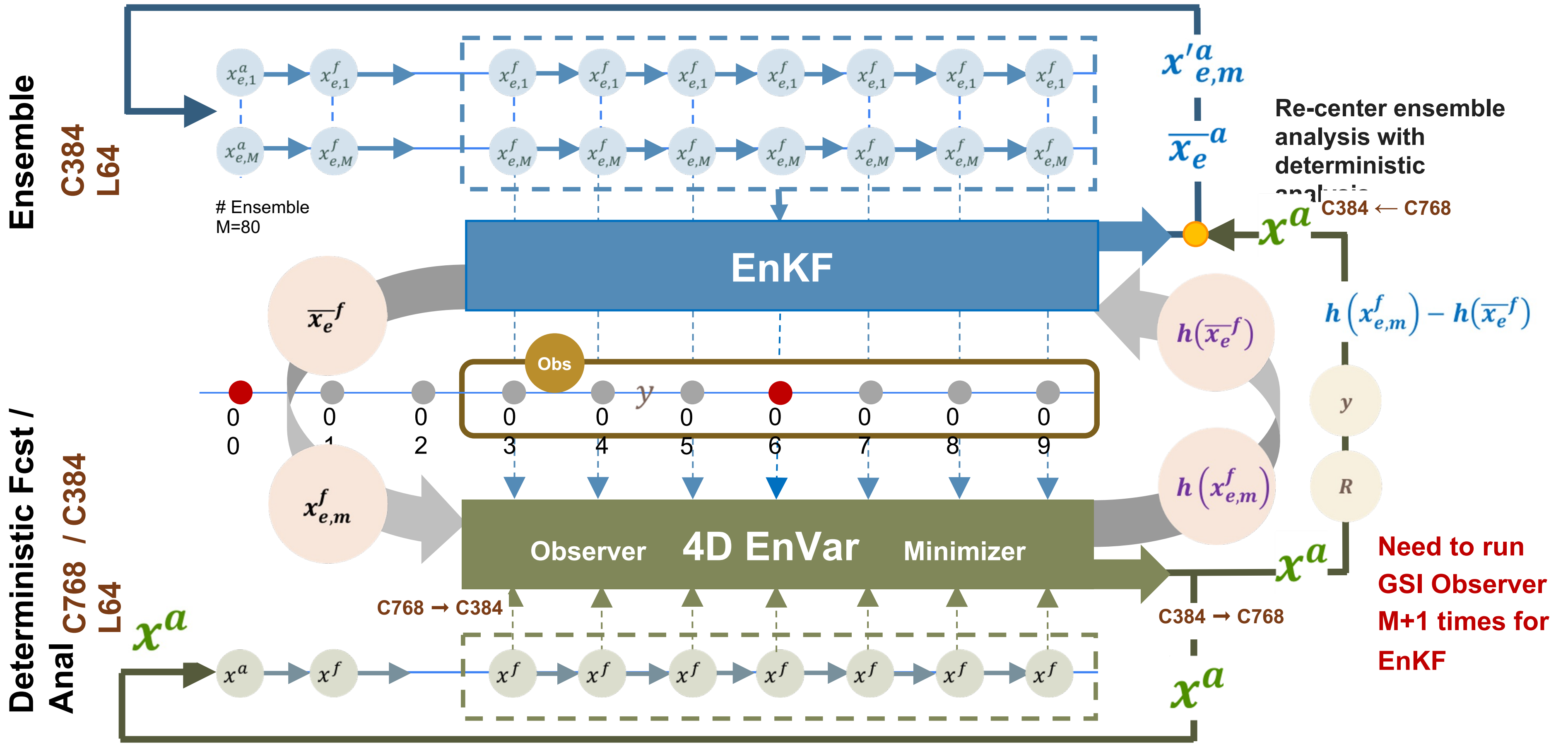
Principle: minimize the distance between the analysis and all available observations over the assimilation window. ***Solved for iteratively.***

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

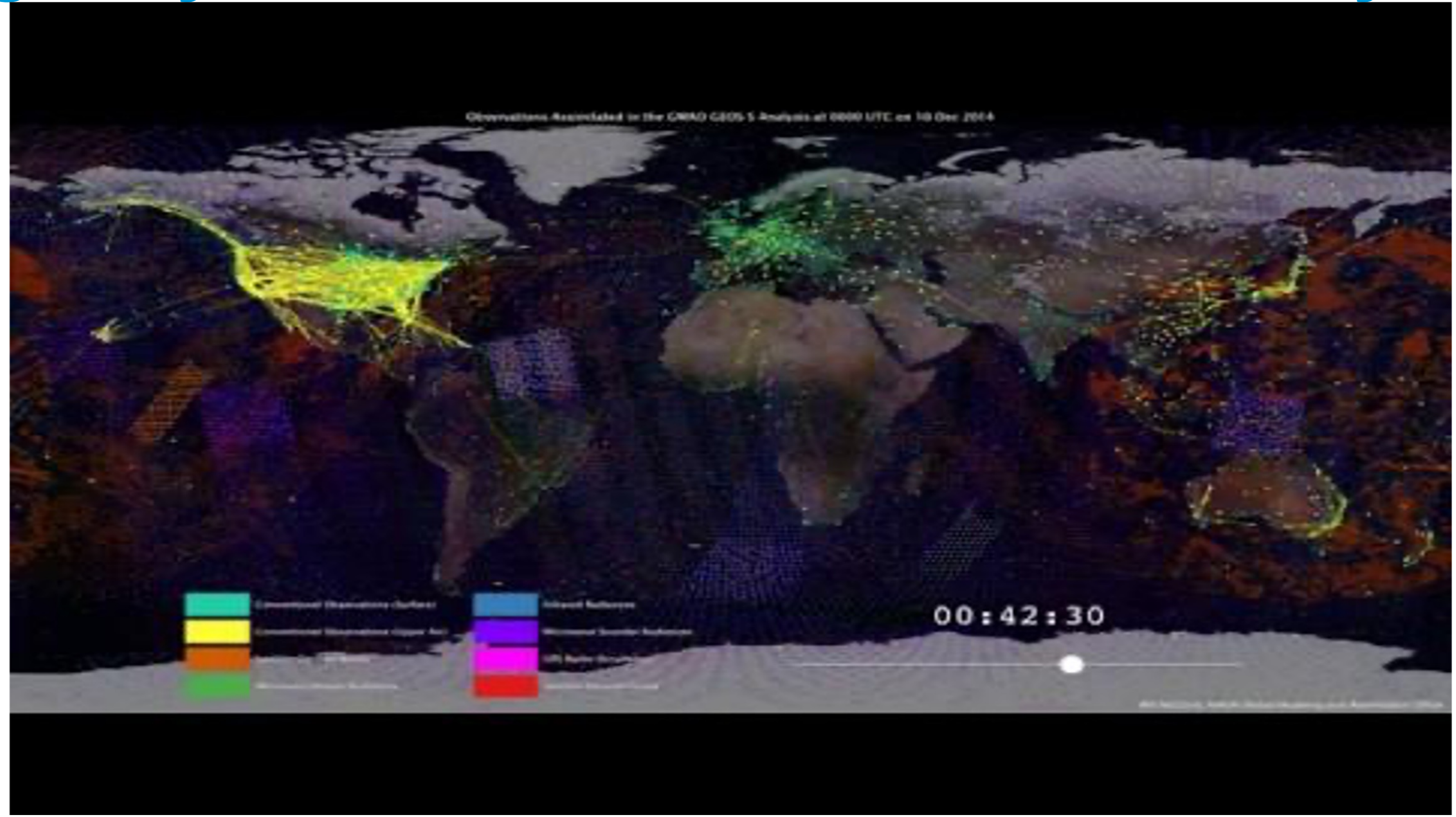


$$J(x) = \frac{1}{2}(\mathbf{x}-\mathbf{x}_b)^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b) + \frac{1}{2}[\mathcal{H}(\mathbf{x})-\mathbf{y}]^T\mathbf{R}^{-1}[\mathcal{H}(\mathbf{x})-\mathbf{y}]$$

Coupled Deterministic / Ensemble Systems (courtesy Emily Liu)

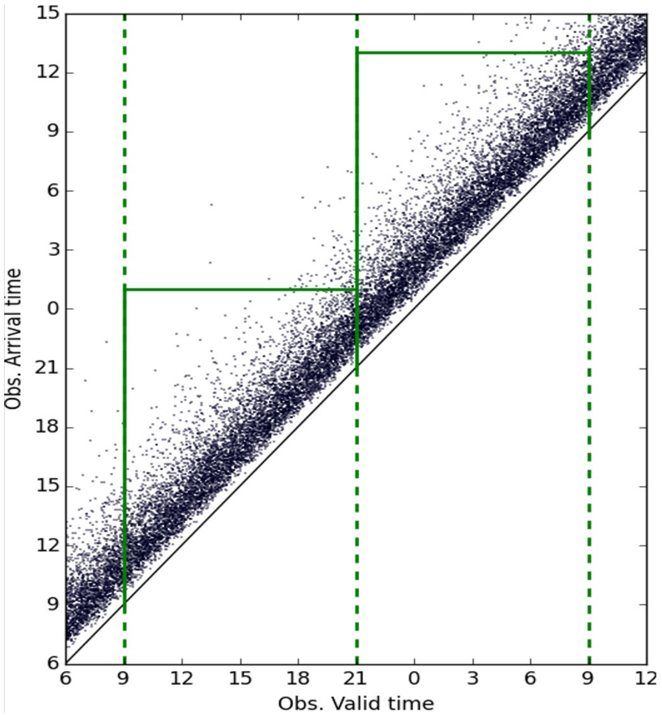


Single Cycle Global Data Assimilation System



Example of observations assimilated into single six-hour GDAS updated. Animation courtesy of Will McCarty (NASA) – circa 2014

Observations (from Y. Tremolet/JCSDA)



To perform analysis, observations are used within time window

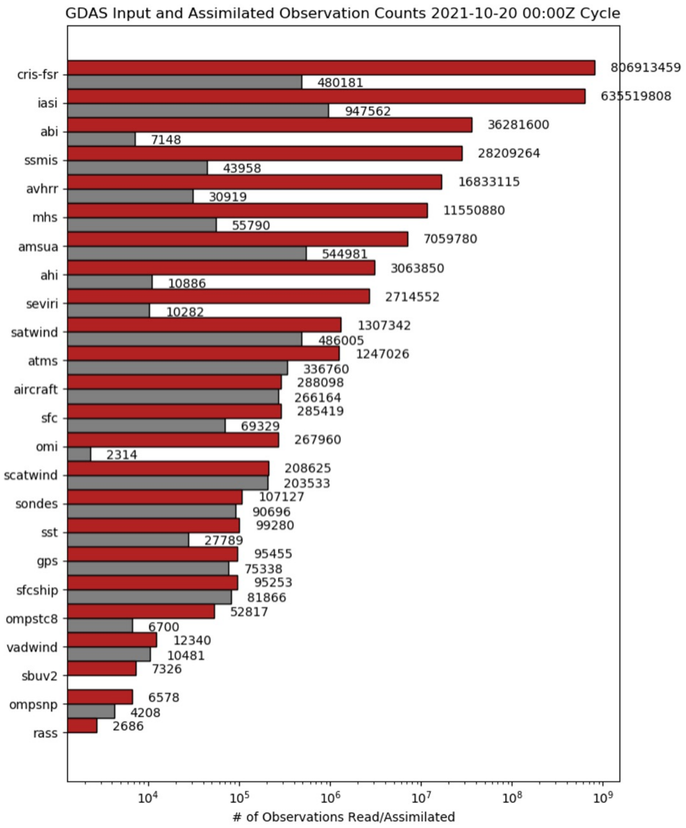
Computational issue: Observations do not instantaneously appear at operational centers:

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

Some observations are lost

Some computational effort is lost

Underutilized Observing System



- The current global observing system for NWP is diverse, robust and rapidly evolving
- We are still under-utilizing what we already have
  - Representativeness issues
  - Scientific challenges (e.g., all-sky/all-surface radiance assimilation)
  - Redundancy and risk reduction
  - Complexity / computational limitations
- Potential for use of AI/ML for data (including channel) selection, superobbing (maximizing information content), and quality control

From NCEP 10 Year Development Strategy (to be published Dec. 2023)

ML for DA priorities

1. **Observations** – *quality control, data selection, bias correction, super-observations, extraction of maximal information content, anomaly detection and operational monitoring*;
2. **Forward operator emulation** – *computational efficiencies, replacement for complex operators*;
3. **Background error** – *computational efficiencies, multivariate aspects and coupled assimilation, parameter estimation for error models*;
4. **Background** – *dynamic downscaling, bias correction*;
5. **Model error** – *estimation and correction;*
6. **Emulator exploitation** – *replacement for TL/AD in 4DVar, efficient creation of huge ensembles to avoid localization*.

Joint Effort for Data assimilation Integration (JEDI) Infrastructure for Unified Data Assimilation

JEDI is a project within the Joint Center for Satellite Data Assimilation (JCSDA)

JEDI provides a **software infrastructure for data assimilation** that

- ❑ is model agnostic
- ❑ is generic and portable, from toy models running on laptops to operational Earth system coupled models running in the cloud.
- ❑ enables DA on the model native grid
- ❑ does not impose one specific DA methodology or algorithm
- ❑ provides a framework for rapid uptake of new observations into operations with generic observation handling and modeling
- ❑ encourages implementation of model-independent observation operators
- ❑ provides a unified Interface for Observation Data Access (IODA)

NOAA is committed to JEDI – and could be enabler for integration of ML to help with DA problems….

Or…

Do we even need DA?

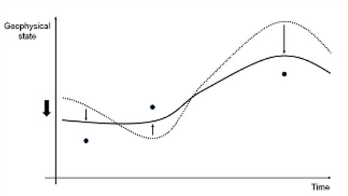
Backup Slides

Cost / loss function equivalence of ML and variational DA

Assume Gaussian errors (error standard deviation σ)
and for clarity here simplify to scalar variables
and ignore any covariance between observation, model or state error

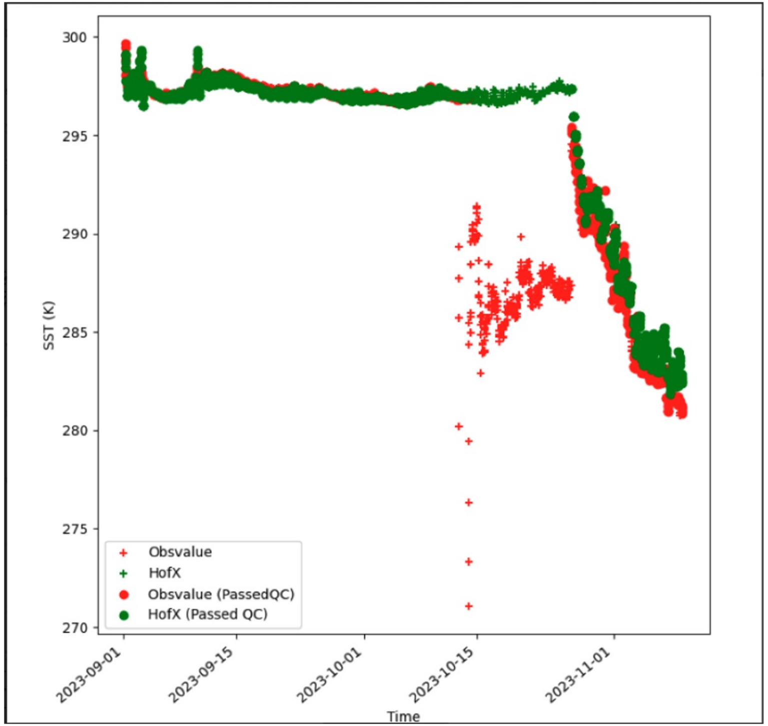
ML	Loss function	Basic loss function	Feature error?	Weights regularisation
DA	Cost function	Observation term	Prior knowledge of state	Prior knowledge of model

$$J(x,w) = \underbrace{\frac{(y-h(x,w))^2}{(\sigma^y)^2}}_{J^y} + \underbrace{\frac{(x^b-x)^2}{(\sigma^x)^2}}_{J^x} + \underbrace{\frac{(w^b-w)^2}{(\sigma^w)^2}}_{J^w}$$

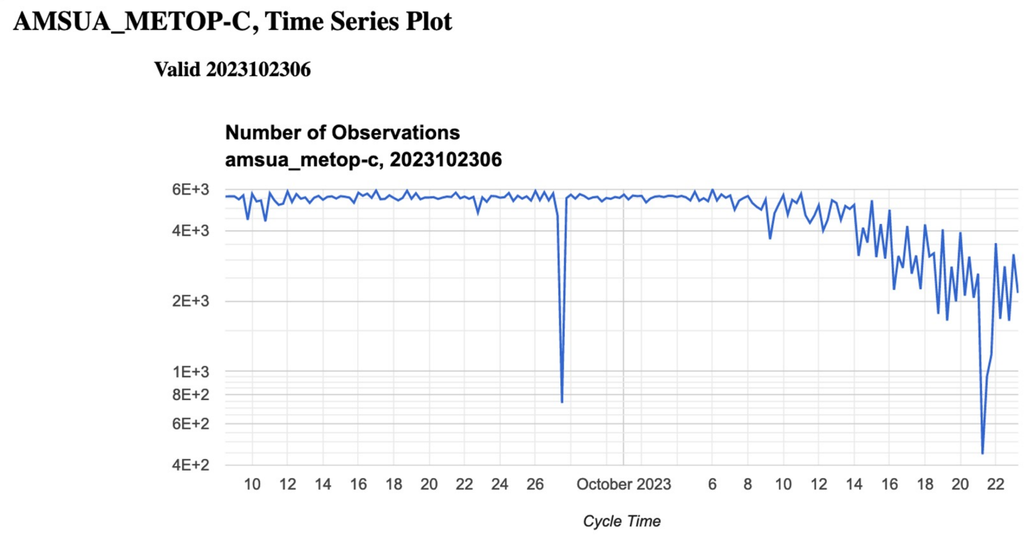


From Alan Geer (ECWMF, Presentation at 2nd NOAA AI Workshop)

Dynamic Quality Control

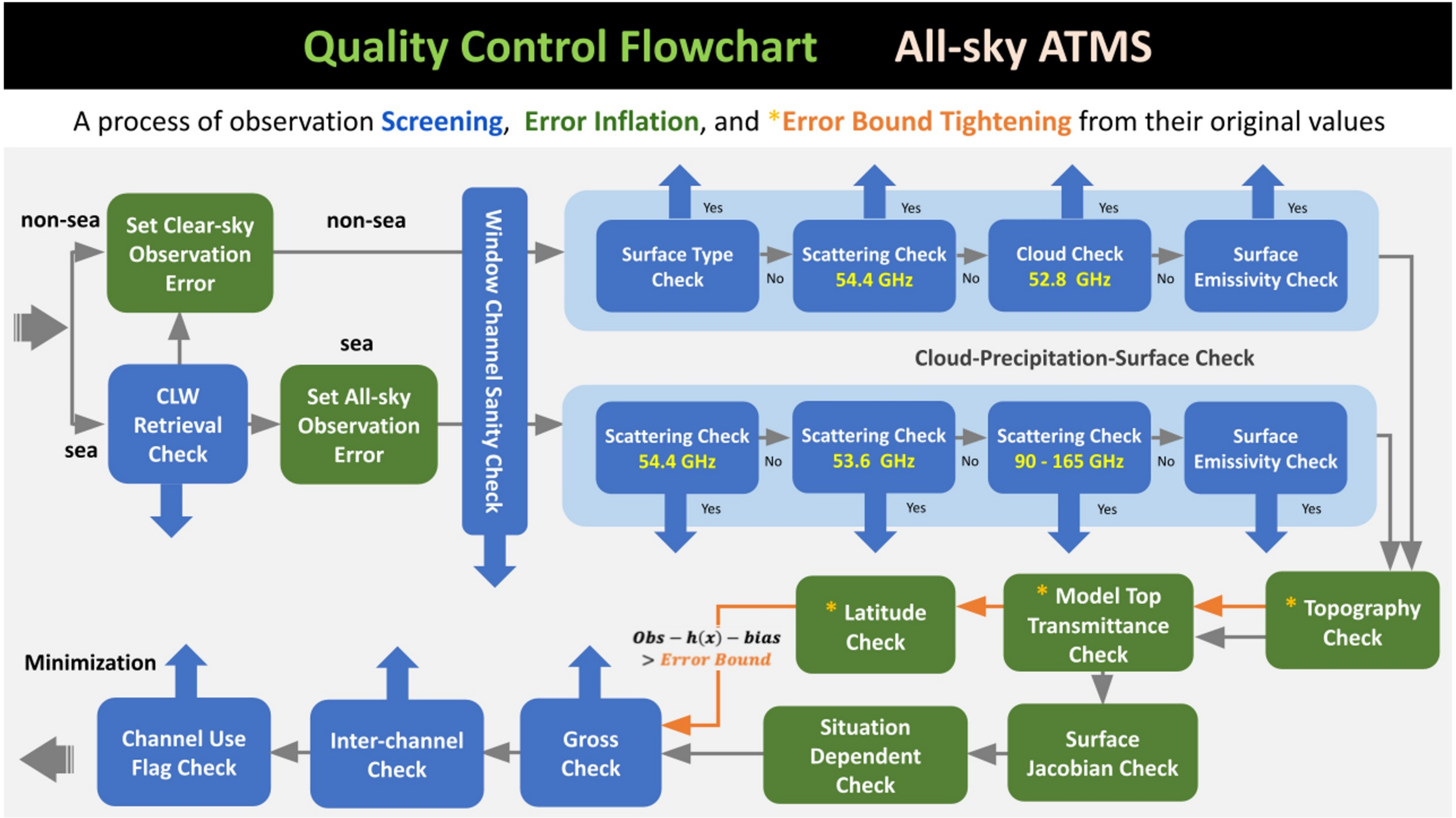


Example of buoy observation gone awry – Obs values (red), Hx (green) before (cross) or having passed QC (dots). In this case, the buoy went bad in such a way in early November that it was able to pass QC while getting worse, dragging the SST analysis with it.



Time series of # observations accept for assimilation from a particular channel from amsu-a on metop-c. In this case, there was an encoding issue with a single channel that was causing erroneous quality control to throw out data.

Quality control can be quite complex

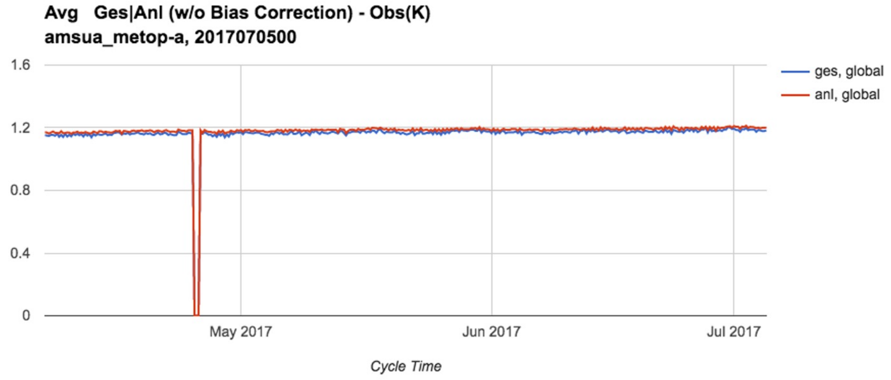


Bias Correction

- Many assumptions made to perform data assimilation problem – bias in observations (or simulated observations) and/or background violates some assumptions.
- Bias in simulated observation and observations themselves can come from many sources:
    - Inadequacies in characterization of instrument
    - Deficiencies in forward models (operators) - mapping model/background to observations space
    - Errors in processing data
    - Bias in background
- Generally need to remove biases somehow to perform assimilation of observation
- For satellite radiances and aircraft temperatures, currently use a procedure within the assimilation itself called Variational Bias Correction (VarBC)

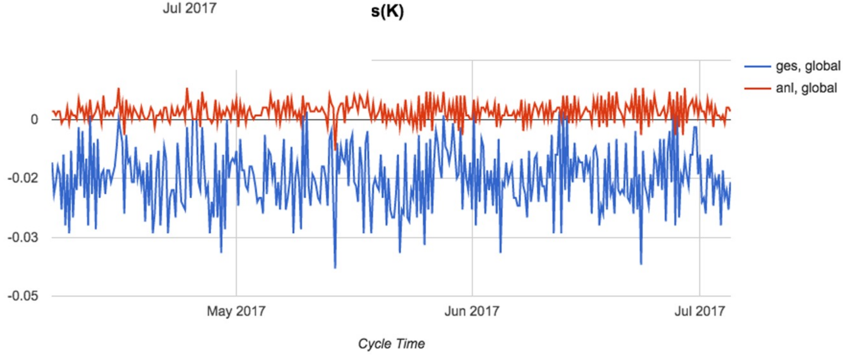
MetOp-A AMSU-A Ch 6
Mean Departures

Observed-Guess
Observed-Analysis



Before

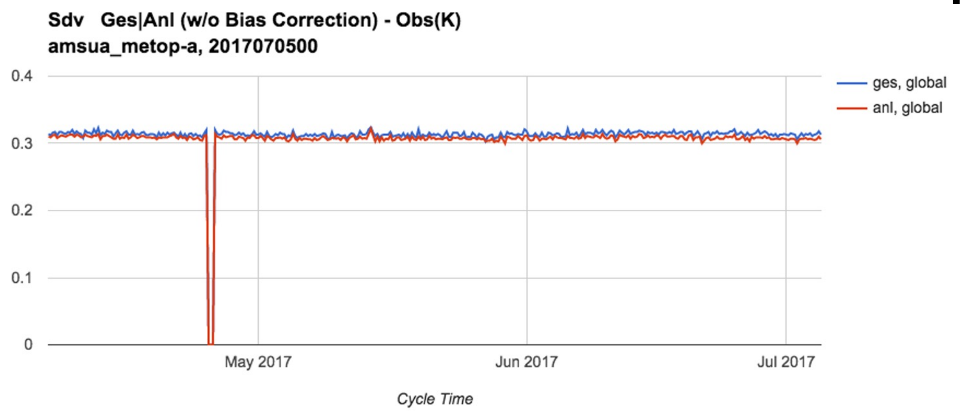
Bias Correction



After

Bias Correction

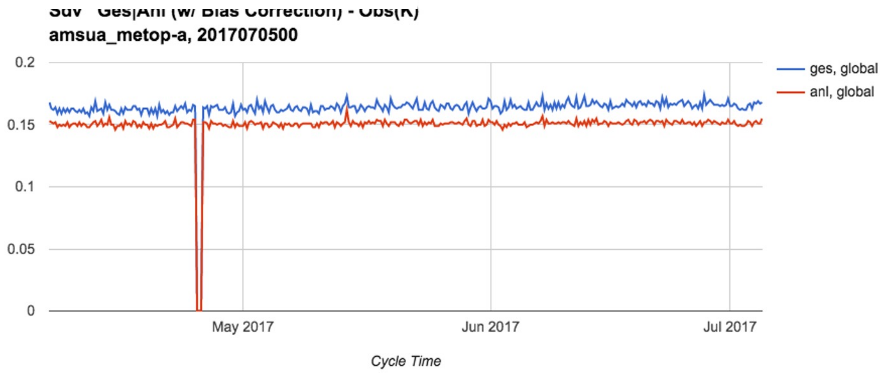
MetOp-A AMSU-A Ch 6
Std Dev of Departures



Observed-Guess
Observed-Analysis

Before
Bias Correction

After
Bias Correction



Thinning or Superobbing

- Thinning
  - Reducing spatial or spectral resolution by selecting a reduced set of locations or channels.
  - Can include “intelligent thinning” to use better observation.
- Superobbing
  - Reducing spatial or spectral resolution by combining locations or channels.
  - Can reduce noise.
  - Includes reconstructed radiances.
  - Can include higher moments contained in data Purser et al., 2010.
  - Can be done with obs or departures, but should be done after QC.
- Both can be used to address 3 problems:
  - Redundancy in data.
  - Reduce correlated error.
  - Reduce computational expense.