Joint Effort for Data Assimilation Integration (JEDAI)

Relevance to post-processing

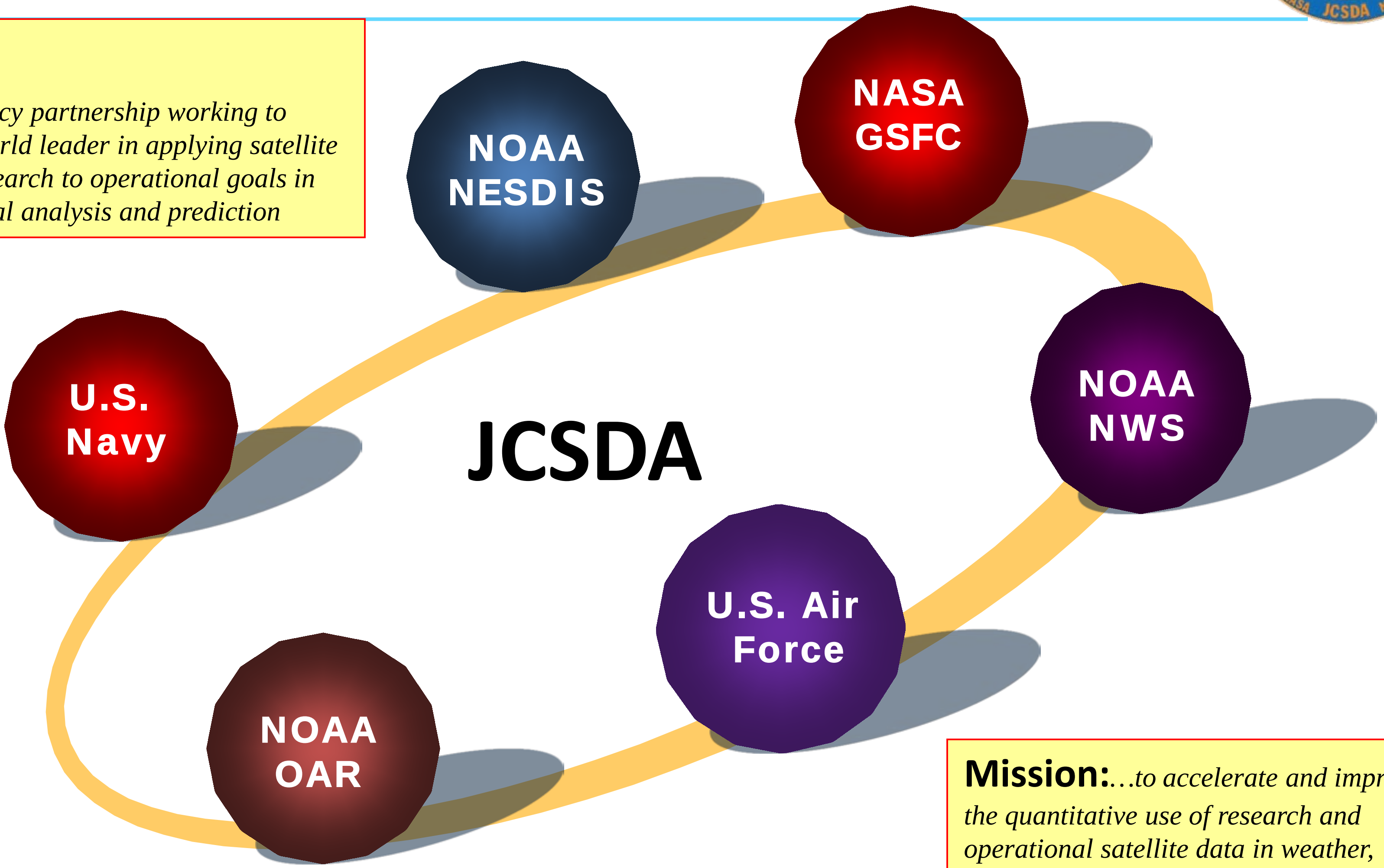
Tom Auligné

Director, JCSDA

Description of the JCSDA

Vision:

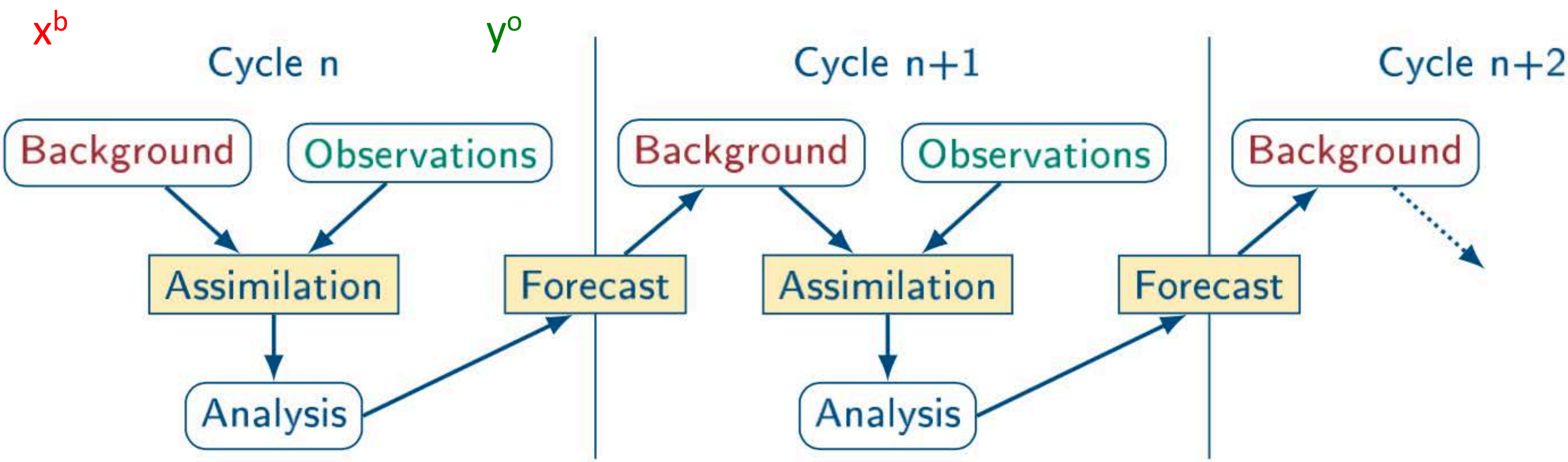
An interagency partnership working to become a world leader in applying satellite data and research to operational goals in environmental analysis and prediction



Mission:*…to accelerate and improve the quantitative use of research and operational satellite data in weather, ocean, climate and environmental analysis and prediction models.*

Introduction to Data Assimilation

Data assimilation systems usually combine together information from a set of observations, a short term forecast, and possibly other information to estimate the most probable state of atmosphere.



Courtesy: Ménétrier

Introduction to Data Assimilation

Hypotheses: observation and model background have errors that are uncorrelated, unbiased, normally distributed, with known covariances

Method: Bayesian statistical framework combined with dynamical constraints

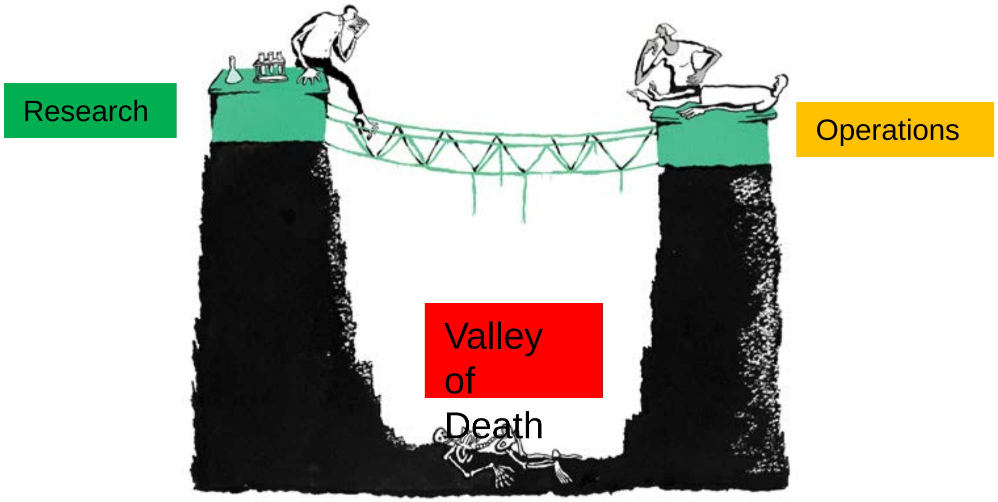
Outcome: "best estimate" of current state (maximum likelihood, minimum RMSE)

JEDAI: Motivation

- DA science has matured with many non-trivial components. Yet, codes = unnecessary complex.
- External collaborations at code level are very limited.
- Developments & transition to operations are slow
- Testing and debugging are even slower due to poorly understood code teleconnections
- Some potentially good ideas simply cannot be tested in operation-relevant environment.
- Step change needed to allow coupled DA, data fusion, new grids, scalability, etc.

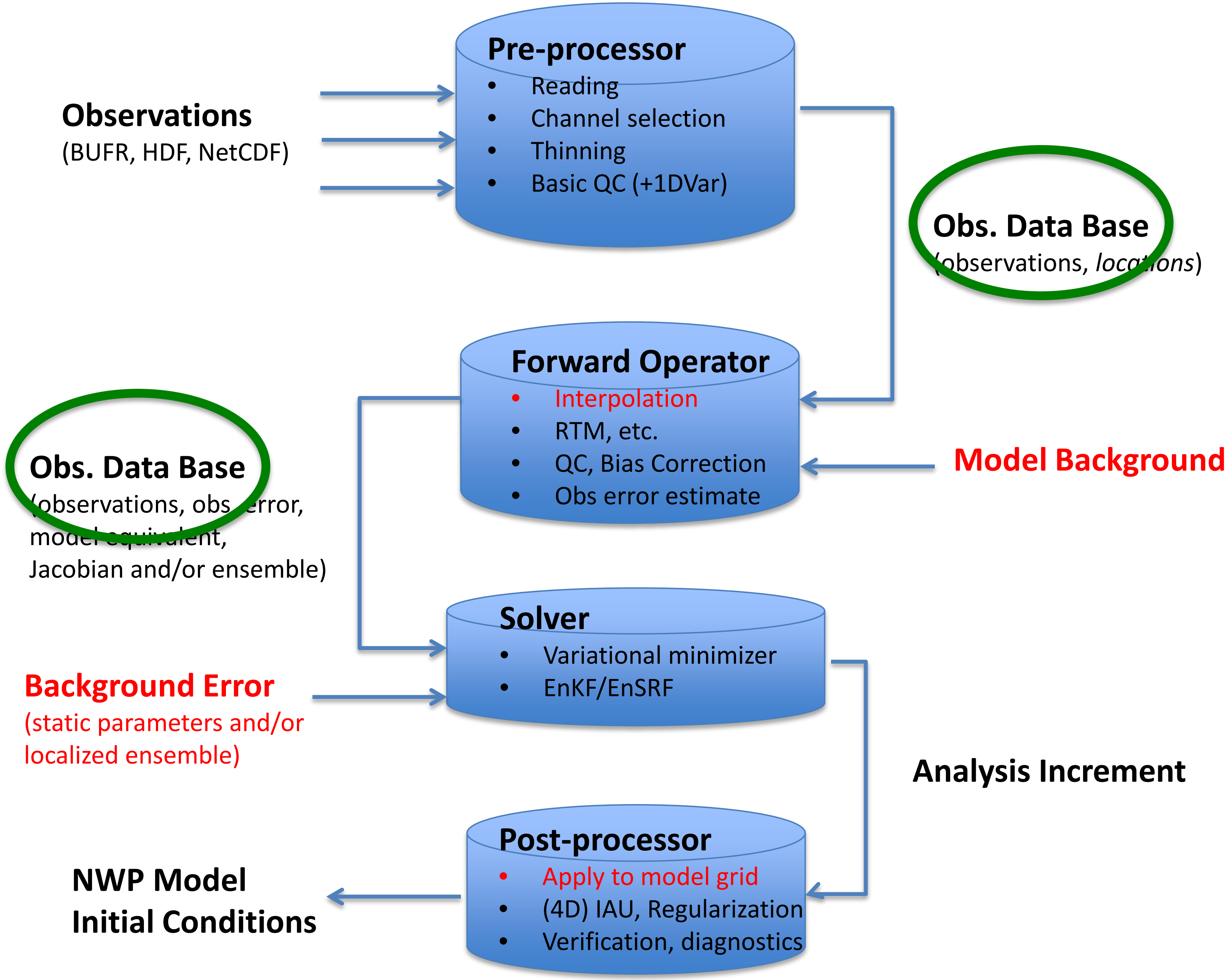
JEDAI: Stretch goals

- Unified nation-wide next-generation DA architecture
- Develop state-of-the-art collaboration across
  internal applications, JCSDA partners, the scientific community
- Bridge the valley of death



JEDAI: Strategy

- Mutualize redundant “plumbing” to focus on science
- Develop a flexible environment to quickly implement, evaluate, maintain and optimize scientific developments
- Build model-agnostic DA components



Pre-processor
• Reading
• Channel selection
• Thinning
• Basic QC (+1DVar)
Observations
(BUFR, HDF, NetCDF)
Obs. Data Base
(observations, *locations*)
Forward Operator
• Interpolation
• RTM, etc.
• QC, Bias Correction
• Obs error estimate
Model Background
Obs. Data Base
(observations, obs error,
model equivalent,
Jacobian and/or ensemble)
Solver
• Variational minimizer
• EnKF/EnSRF
Background Error
(static parameters and/or
localized ensemble)
Analysis Increment
Post-processor
• Apply to model grid
• (4D) IAU, Regularization
• Verification, diagnostics
NWP Model
Initial Conditions

Major Milestone

Unified Forward Operator (UFO)

1. Split Observer from Solver
2. Object-Oriented Design for Observation Operator
3. Connect with improved ODB
4. Define Observation “Locations”
5. Encapsulate Interpolation of model variables to observation locations
6. Replace grid-specific interpolation with ESMF regridding capabilities to develop a model-agnostic Observer

Model-agnostic tool for verification, model development, data assimilation, inverse problems, statistical post-processing…

Challenges and Opportunities

- Leverage existing initiatives (ESMF, NEMS, ESPC, NGGPS, RUA, NFIDA-GSI, *etc*)
- Risk of conflicting requirements (e.g. research and operation). Trade-offs need to be identified.
- Maintain compatibility with existing operational systems
- Choice of algorithms and their implementation will be strongly influenced by future HPC machine architectures

- Collaborations on **architecture design & support**
  - JCSDA partners (EMC, OAR, NESDIS, NASA, Navy, USAF)
  - NCAR Data Assimilation Initiative (+ Academic community)
  - DTC code management & support
  - ECMWF OOPS project
  - Met Office Exascale DA project

JEDAI: Approach

- **Use Object-Oriented Design + modern coding standards to achieve better**
  - Flexibility Portability Exchangeability Maintainability
  - Scalability Readability Reusability Testability

- **Top-down approach**
  start with a clean slate
  Redesign entire architecture
  ECMWF OOPS project

- **Bottom-up approach**
  start from existing codes
  Incremental modularization
  GSI refactoring

- Approach: simultaneous **top-down and bottom-up**
  Roadmap gradually connecting the two approaches.

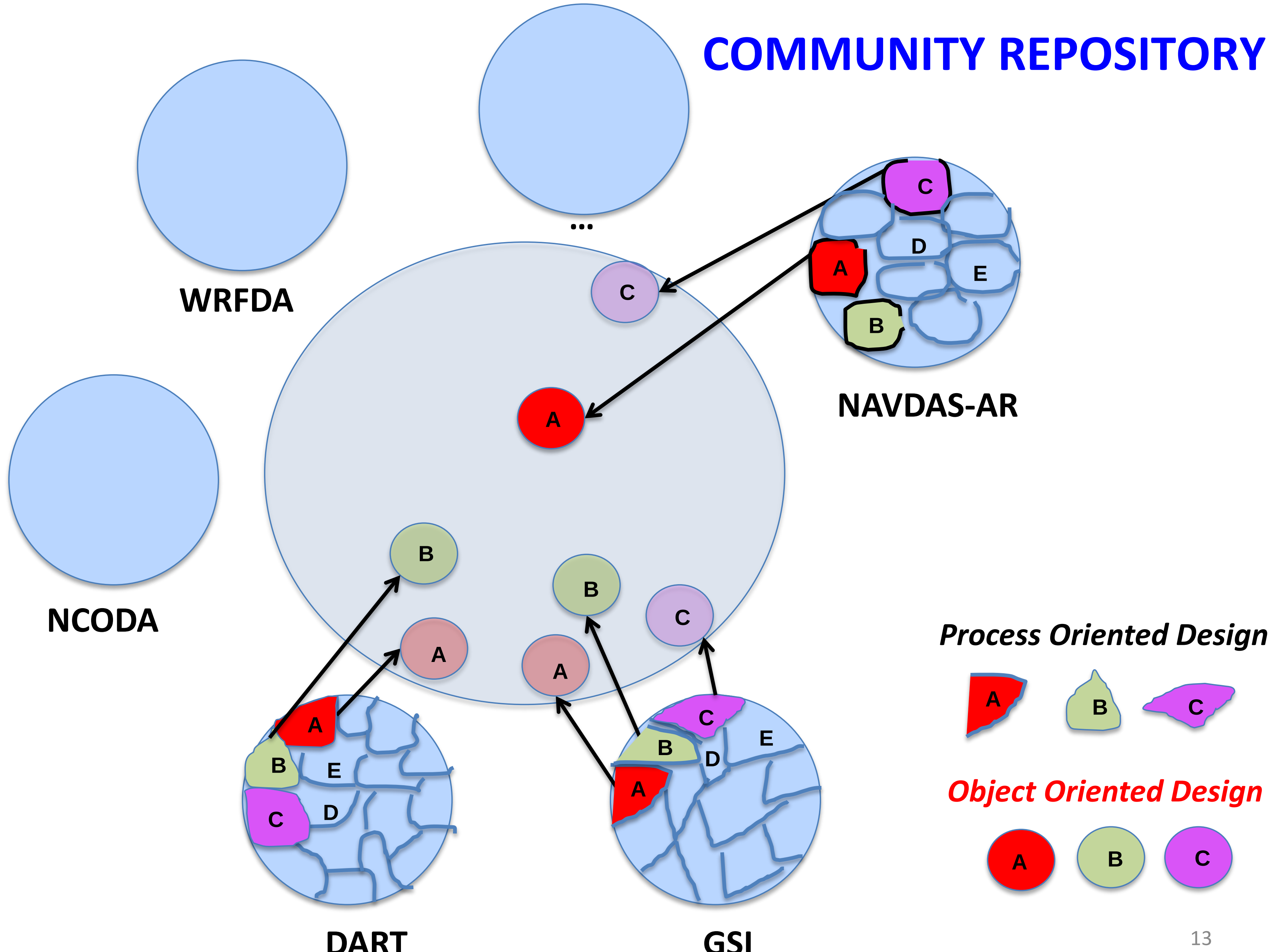
JEDAI: Modes of operation

- **Governance**
  - Allow for multiple levels of engagement
  - Coordination at the level of Object Design (UML)
  - Collegial decisions

- (Single GIT) **Community Repository**

- **Entropy Management Team** (EMT)
  - Both police & support team for scientists
  - Coding standards: ensure readability, avoid redundancy
  - Regression testing

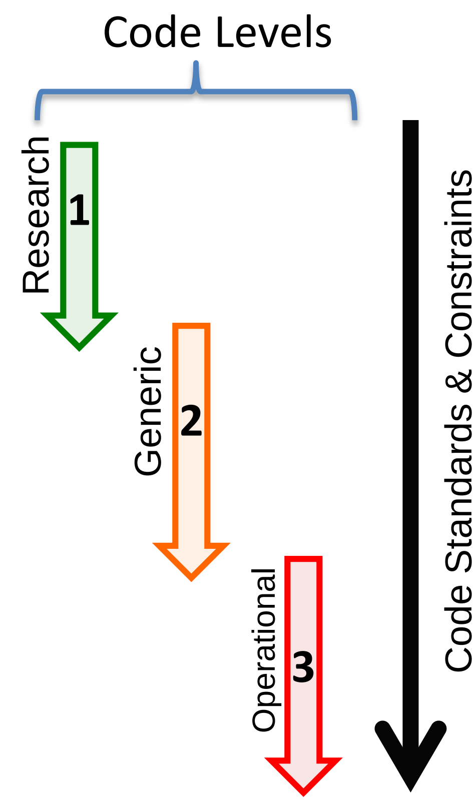
COMMUNITY REPOSITORY



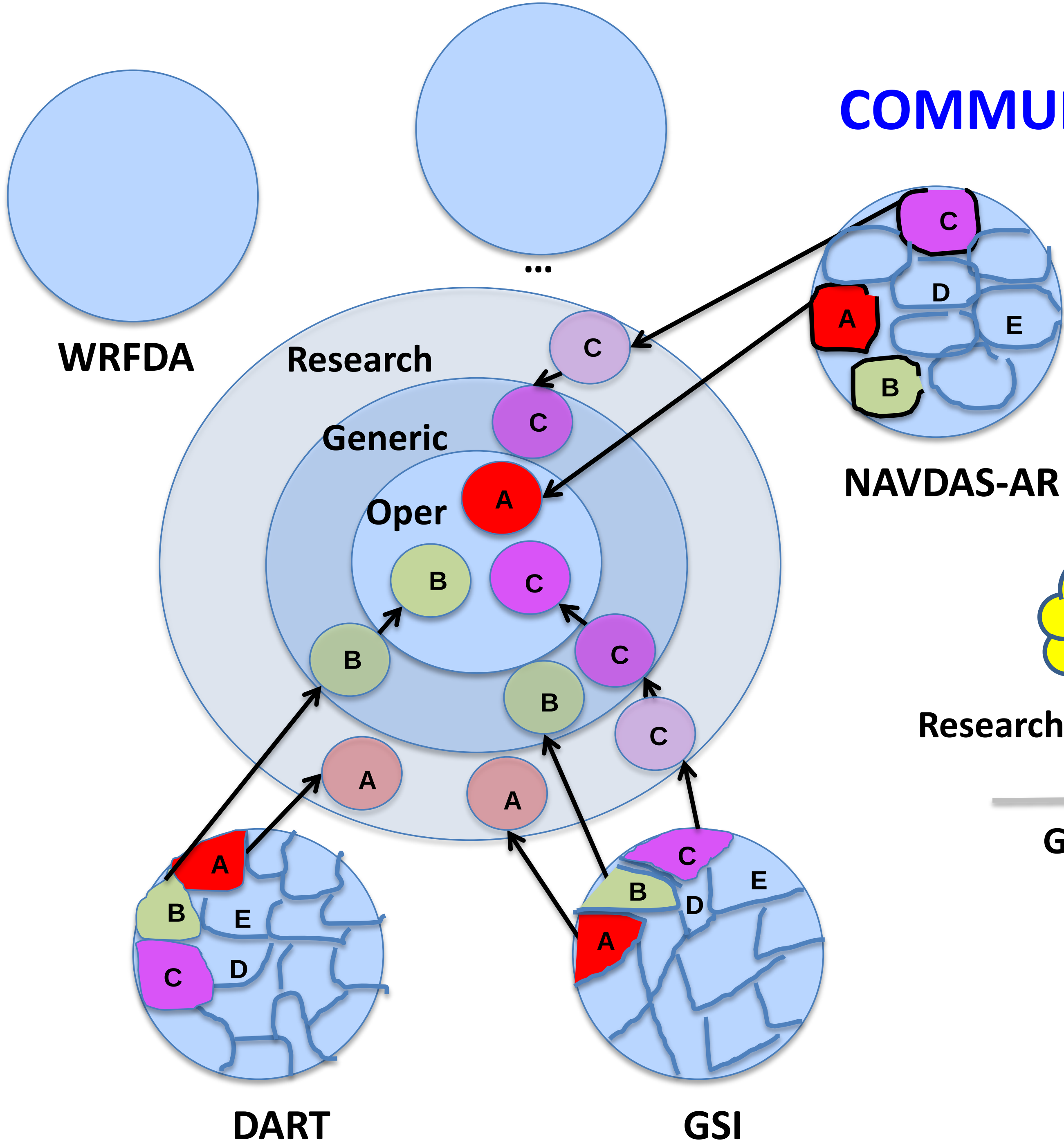
Adapted from Toth and Xie

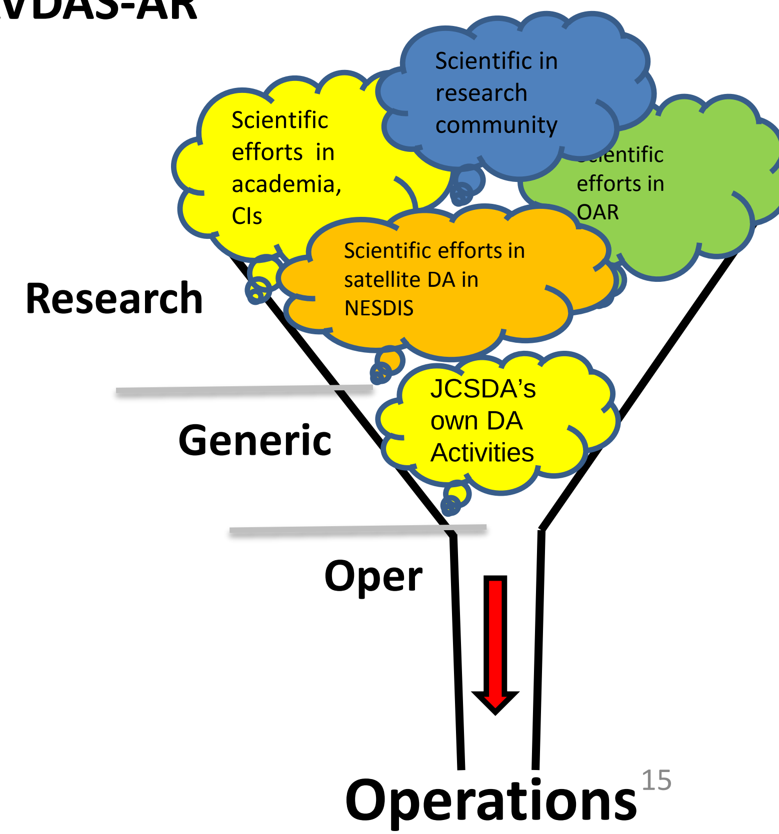
Code Maturity Assessment

NOAA Technical Readiness Level (TRL)	
TRL 1	Basic principles observed and reported
TRL 2	Technology concept and/or application formulated
TRL 3	Analytical and experimental critical function and/or characteristic proof-of-concept
TRL 4	Component/subsystem validation in laboratory environment
TRL 5	System/subsystem/component validation in relevant environment
TRL 6	System/subsystem model or prototyping demonstration in a relevant end-to-end environment
TRL 7	System prototyping demonstration in an operational environment
TRL 8	Actual system completed and "mission qualified" through test and demonstration in an operational environment
TRL 9	Actual system "mission proven" through successful mission operations
TRLs 1-2 are nominally considered Research, TRLs 3-5 are Development, TRLs 6-8 are Demonstration, and TRL 9 is Deployment, Implementation, or Operational Transition.	

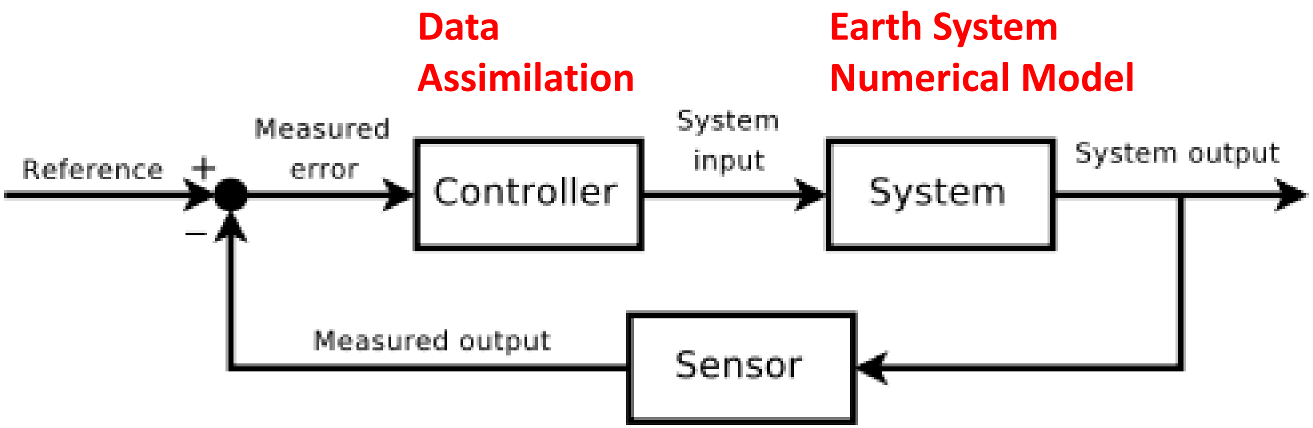


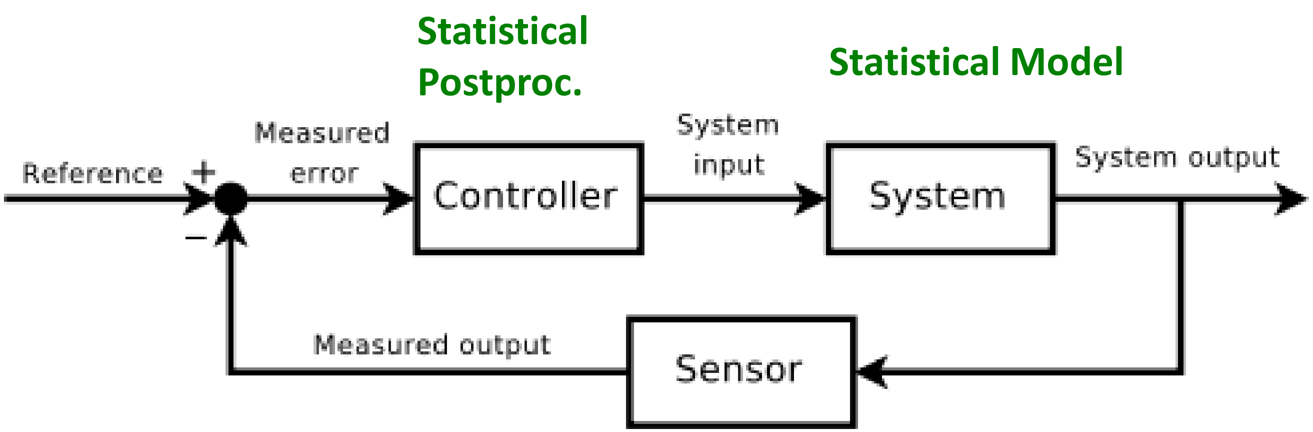
MULTI-LEVEL COMMUNITY REPOSITORY



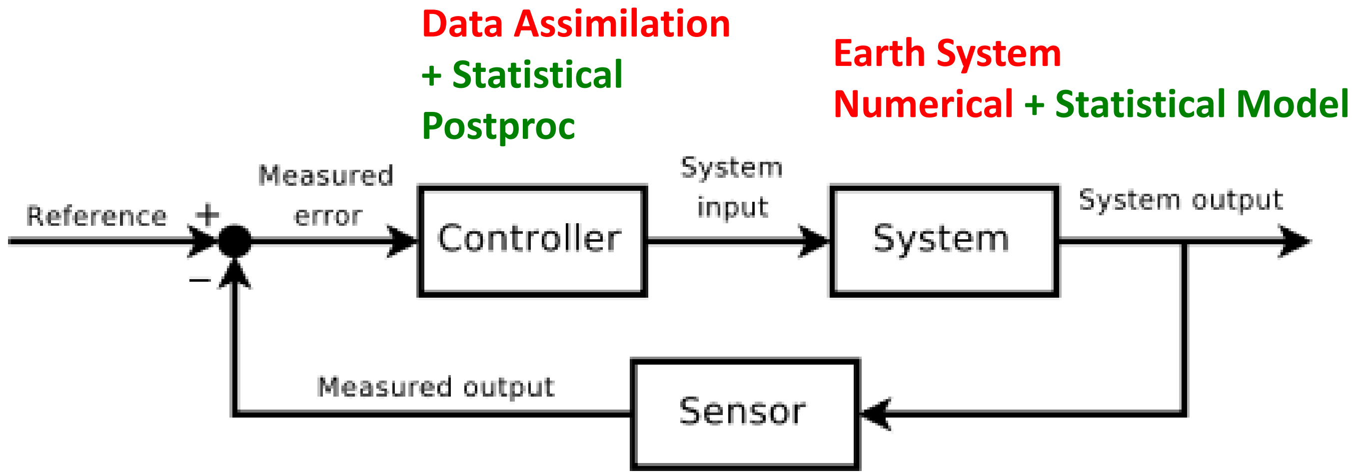


Control System Theory





Control System Theory



Example: Variational Bias Correction

$$\tilde{H}(\mathbf{x}, \boldsymbol{\beta}) = H(\mathbf{x}) + \sum_{i=0}^{N} \beta_i p_i(\mathbf{x})$$

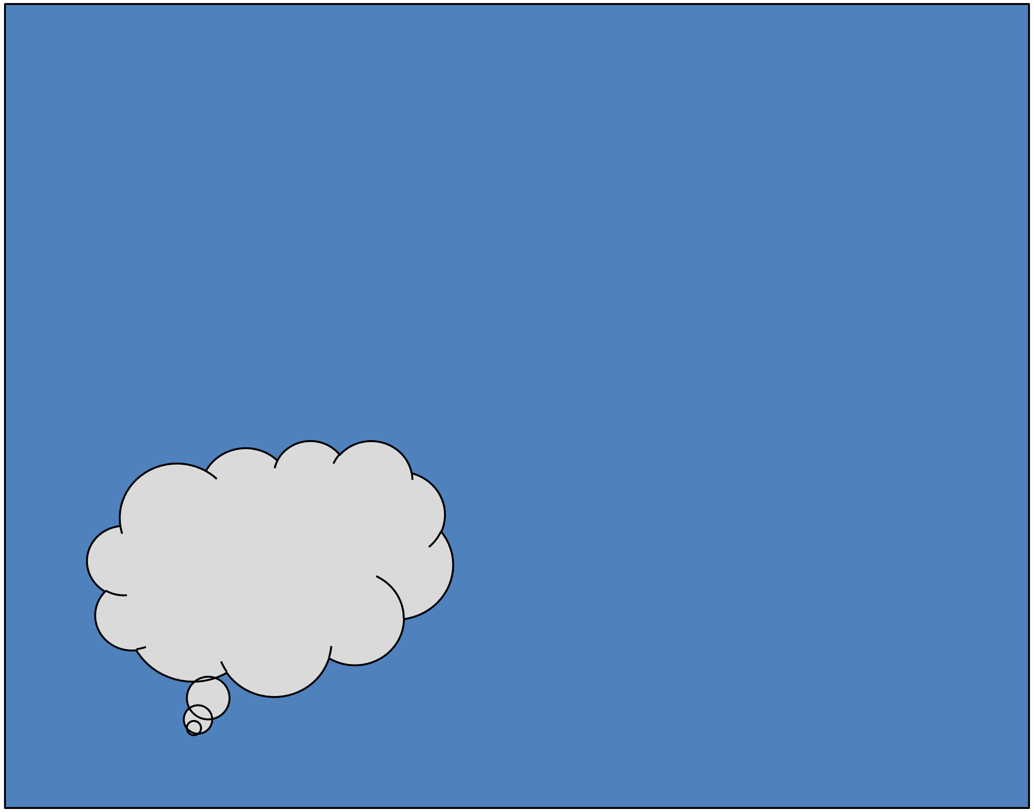
$$J(\mathbf{x}, \boldsymbol{\beta}) = \frac{1}{2}[\mathbf{y} - \tilde{H}(\mathbf{x}, \boldsymbol{\beta})]^{\mathrm{T}}\mathbf{R}^{-1}[\mathbf{y} - \tilde{H}(\mathbf{x}, \boldsymbol{\beta})] + \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_{\mathrm{b}})^{\mathrm{T}}\mathbf{B}_{\boldsymbol{\beta}}^{-1}(\boldsymbol{\beta} - \boldsymbol{\beta}_{\mathrm{b}}) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_{\mathrm{b}})^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_{\mathrm{b}}),$$

AAARGH: Adaptive Advection Adjustment via Realigned Grid mesH

00UTC

01UTC

02UTC



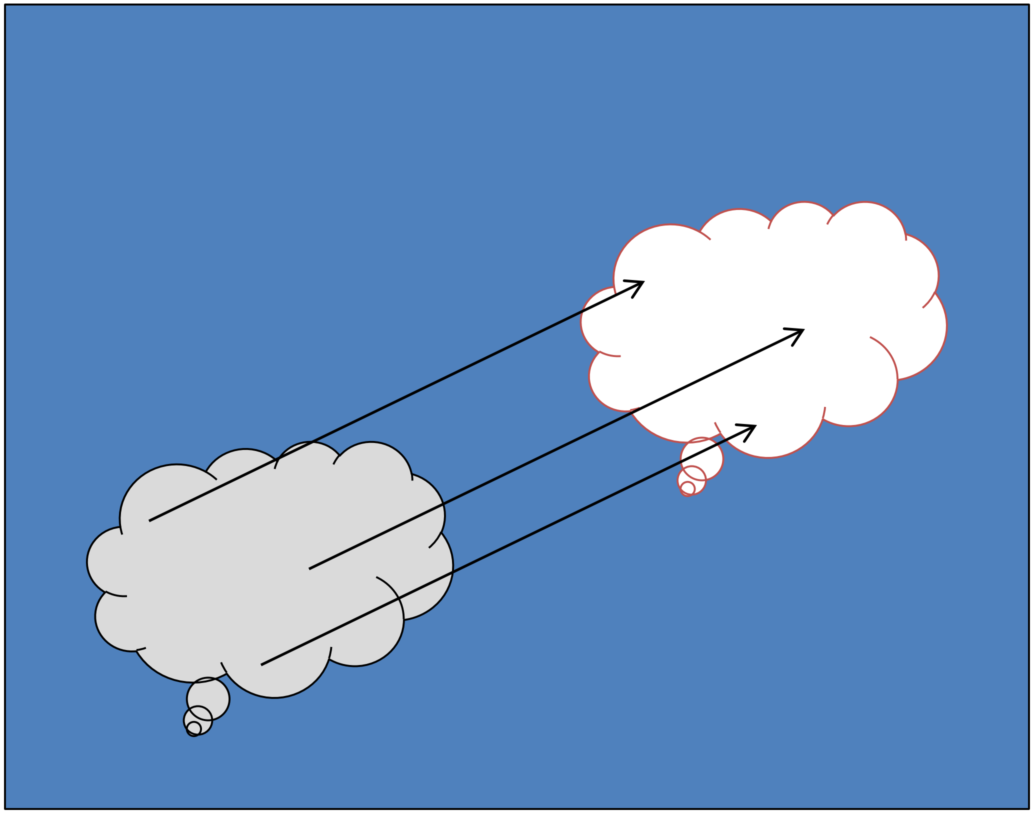
MADCast

AAARGH: Adaptive Advection Adjustment via Realigned Grid mesH

00UTC

01UTC

02UTC



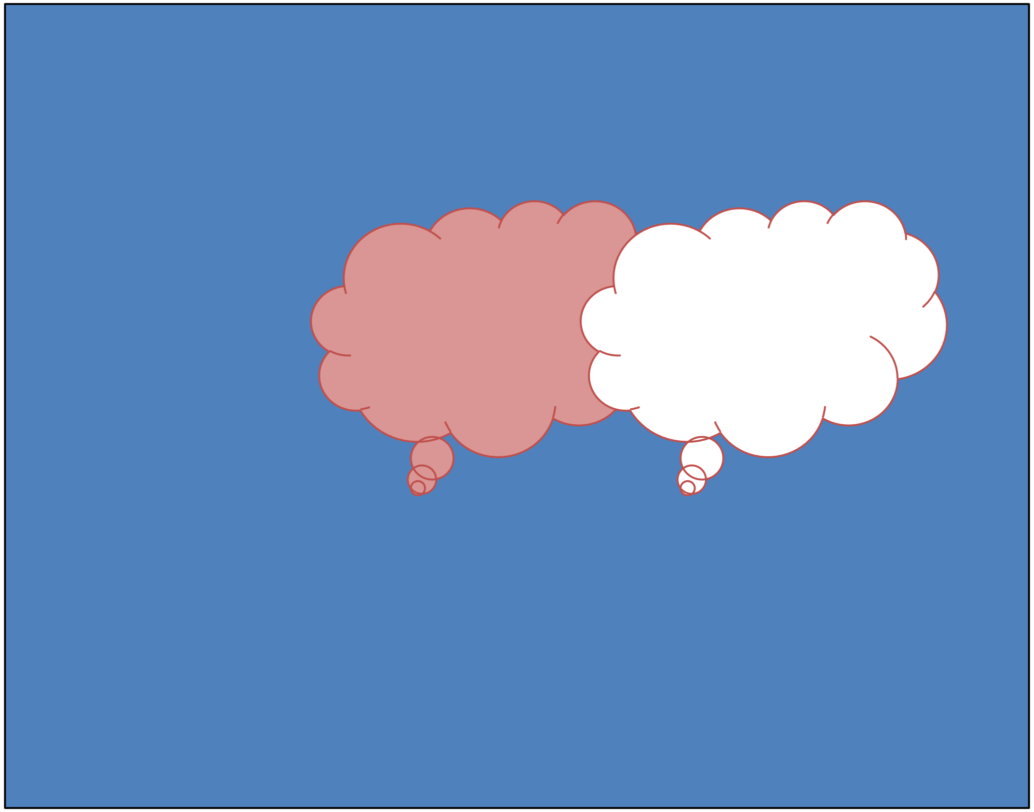
MADCast (WRF dynamical transport)

AAARGH: Adaptive Advection Adjustment via Realigned Grid mesH

00UTC

01UTC

02UTC



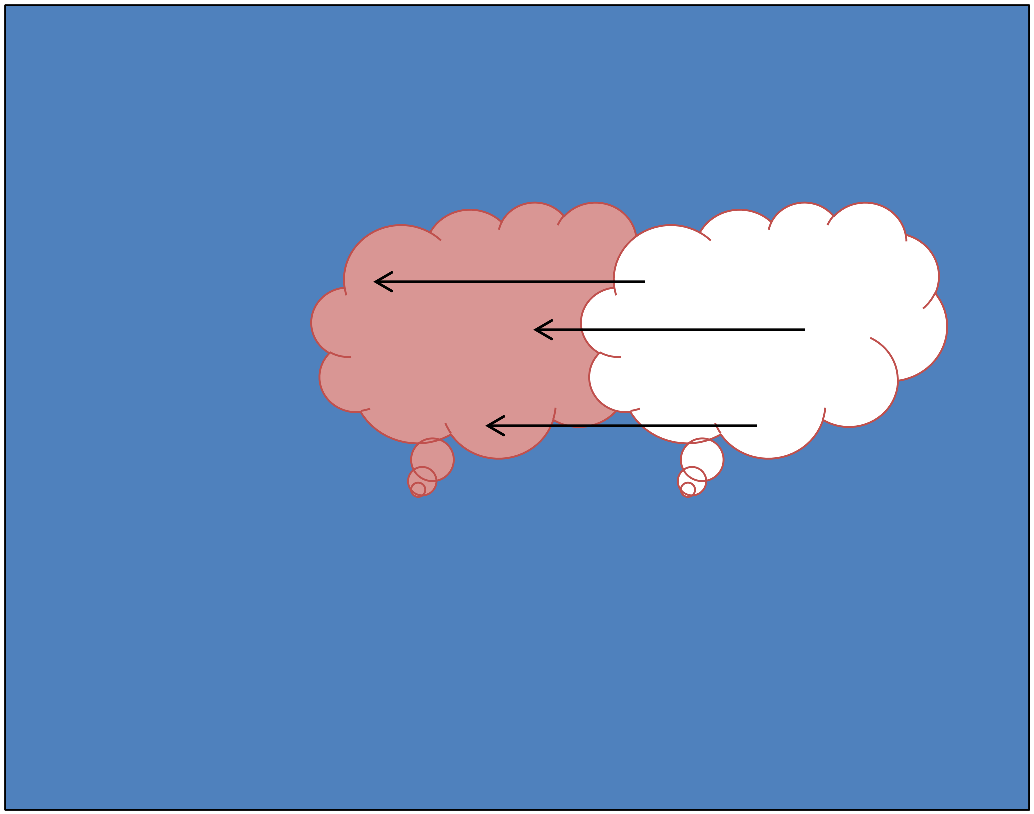
Observation

AAARGH: Adaptive Advection Adjustment via Realigned Grid mesH

00UTC

01UTC

02UTC



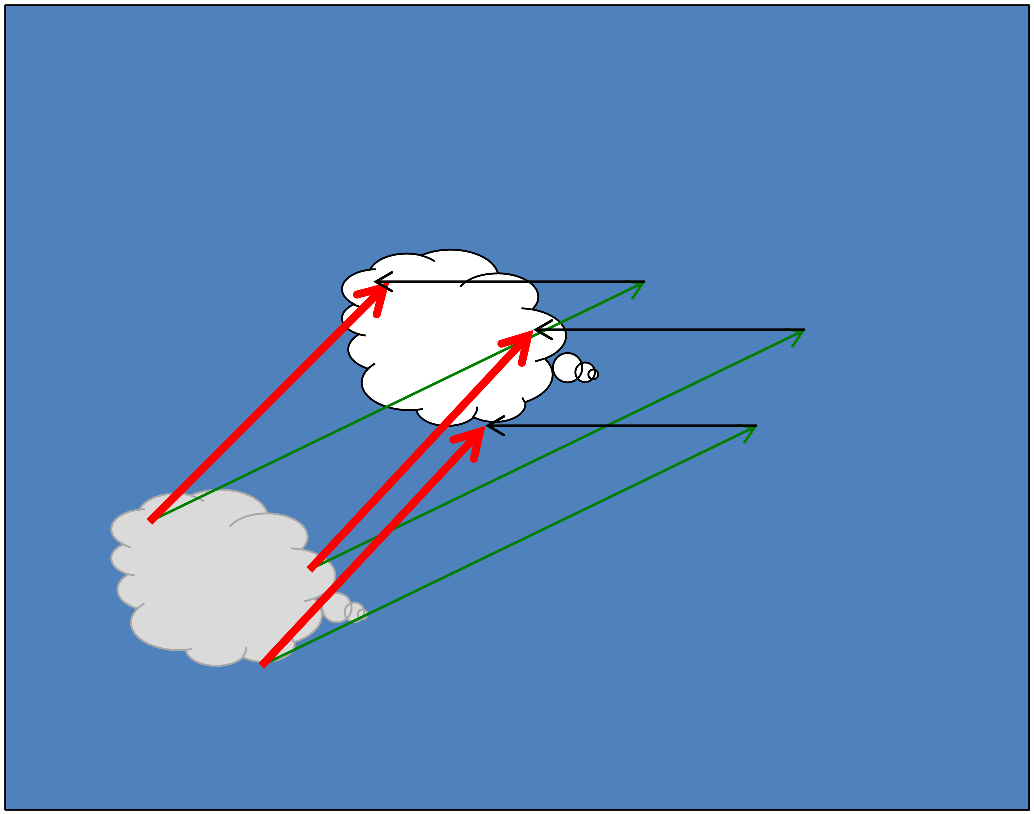
dWRF

AAARGH: Adaptive Advection Adjustment via Realigned Grid mesH

00UTC

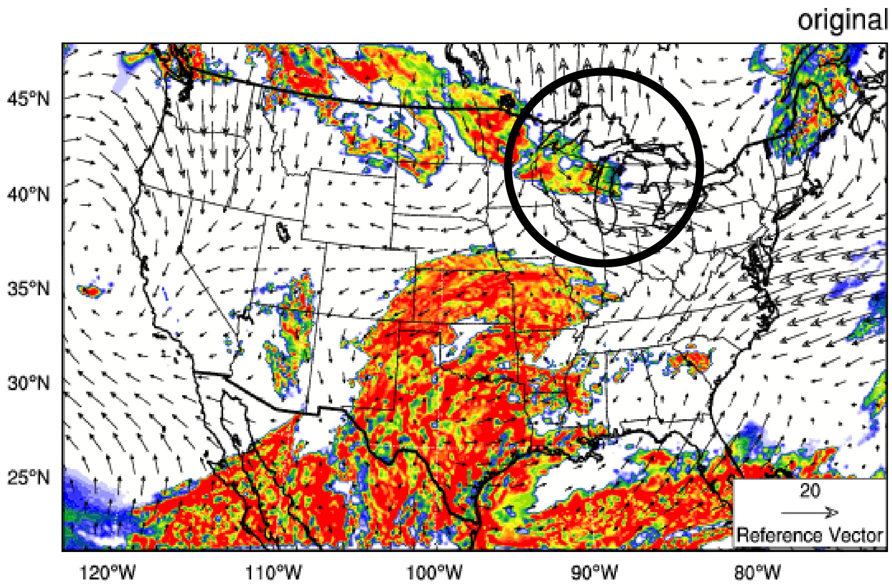
01UTC

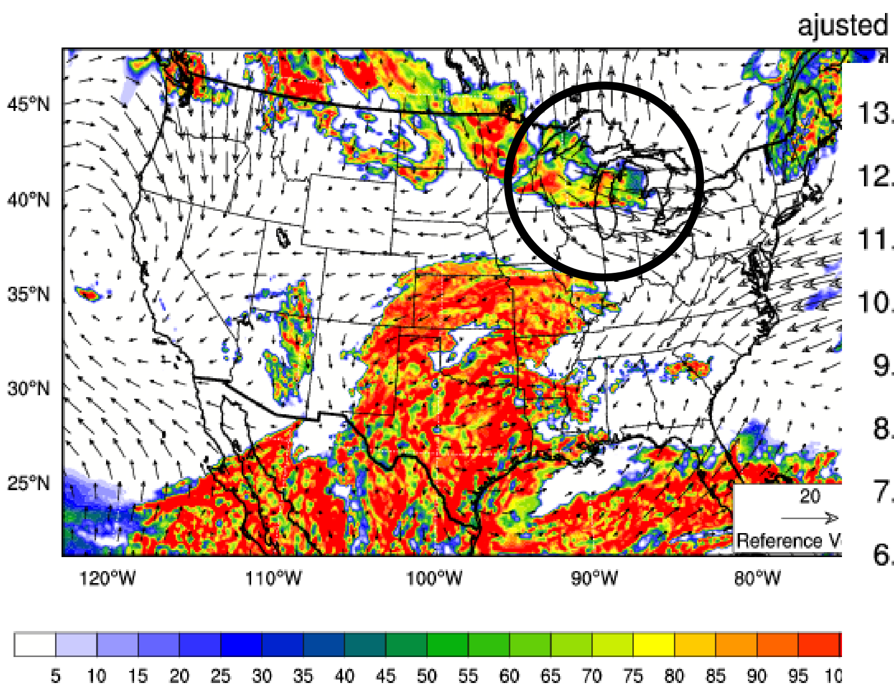
02UTC



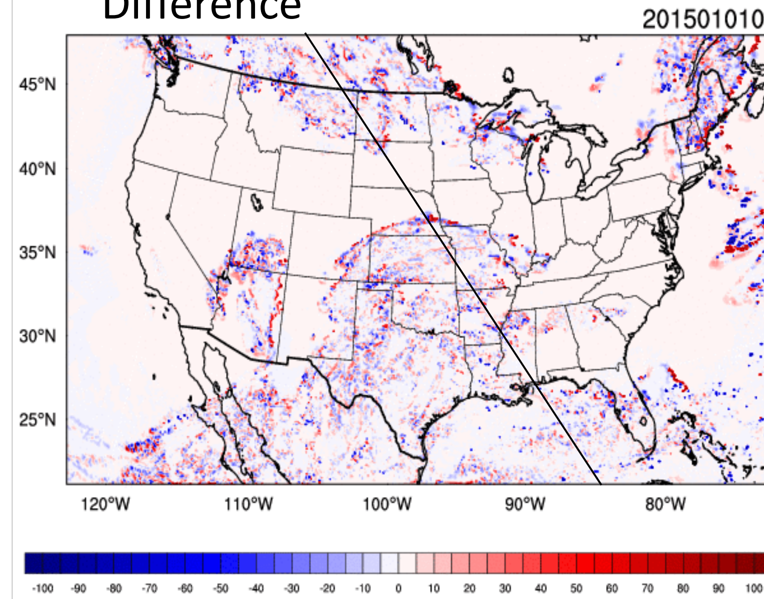
Assuming slow AAARGH (compared to rapid refresh)

MADCast + dWRF

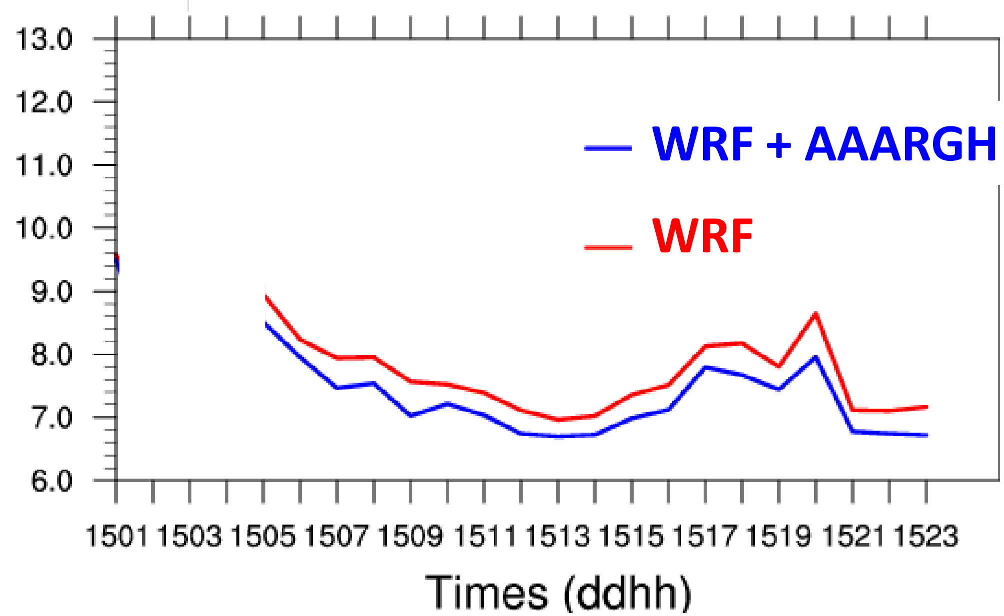




Difference



RMSE of forecast vs. GOES radiances



Modern DA on a Paper Napkin

Hypotheses: Background and observation errors are uncorrelated, unbiased, normally distributed, with known covariances **B** and **R**

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T+\mathbf{R})^{-1}$$

- Kalman Filter analysis:

$$x^a = x^b + \mathbf{K}d \qquad \mathbf{A} = (\mathbf{I}-\mathbf{K}\mathbf{H})\mathbf{B}$$

- Model forecast:

$$x^b \leftarrow M(x^a) \qquad \mathbf{B} \leftarrow \mathbf{M}\mathbf{A}\mathbf{M}^T + \mathbf{Q}$$

Conclusions

- **GSI Refactoring = component of the Joint Effort for Data Assimilation Integration (JEDAI)**. Inter-agency effort spearheaded by JCSDA
  - Modular, flexible, object-oriented code
  - Improved readability, maintenance, testing and optimization
  - Collaborative applications for operation & research
  - Model-agnostic DA components

- **Data Assimilation and Post-Processing are tied by the hip and can greatly benefit each other**