Atmospheric Data Assimilation

August 4, 2016

Tom Auligné (Co-Lead, JCSDA),

John Derber (Co-Lead, EMC),

Jeff Whitaker (ESRL),

Ricardo Todling (NASA GMAO),

Daryl Kleist (U. of Maryland),

Andrew Collard (EMC),

Nancy Baker (NRL Monterey),

Jeff Anderson (NCAR)

Objective

Create best possible (operational) atmospheric Data Assimilation system for the NGGPS era.

NGGPS Atmospheric Data Assimilation - Summary

- **Major Accomplishment in FY16:**
  - Operational implementation of 4d-hybrid EnVar system and use of cloudy radiances
- **Priority Foci for FY17 (all areas must be addressed)**
  - Operational implementations
  - Inclusion of GOES-R, JPSS-1 and other new data sources
  - Preparation for NGGPS dynamic core upgrade
  - Enhancements to 4d hybrid
    - Variable weighting of static and ensembles and localization weights
    - Ensemble resolution and number changes
  - Modifications necessary for changes in model resolution
  - Improvements to observation error specification (including correlated errors) and bias correction
  - Improved use of all-sky radiances (including model balance issues)
  - Enhanced quality control (variational and station history based)
  - Computational and structure optimization (including JEDI)
- **Key Issue**
  - Resources (computational and human) are not sufficient to meet expectations.
    - DA experiments require long term testing
    - Implementation testing requirements increasing rapidly
    - Low tolerance for problems
    - Increased resolution results in increased resource requirements
    - Insufficient long term storage
    - Relatively slow i/o – NEMS increases i/o requirements significantly
    - Expectation that all foci (above and others) will be addressed
    - Collaboration takes a lot of resources and often returns little

Operational implementations

- Goal of NGGPS is to improve operational forecasts – so transition to operations is a priority
- Operational atmospheric data assimilation system used for many systems
  - Global, RAP, HRRR, NMMB, HWRF, RTMA, URMA
  - All with annual (or greater) update schedule
  - Any operational implementation issues must be addressed immediately or implementation may be missed
  - Testing and transition requirements are not stable (generally becoming more burdensome with time)
  - Trying to coordinate between applications and get all changes into trunk

Inclusion of GOES-R, JPSS-1 and other new data sources

- Observing systems are expensive so expectation is that new data are used ASAP and that they will have significant positive impact on forecasts
- Data flow and formatting often complicated
- Huge volumes taxing compute, communication and storage infrastructure
- DA team attempts to get infrastructure in place prior to availability of data
- After data available, quality control, bias correction (if necessary), error specification and impact testing (must be non-negative and over extended period since impact of any new data will be small) must be performed prior to operationalization
- Inclusion in implementation package

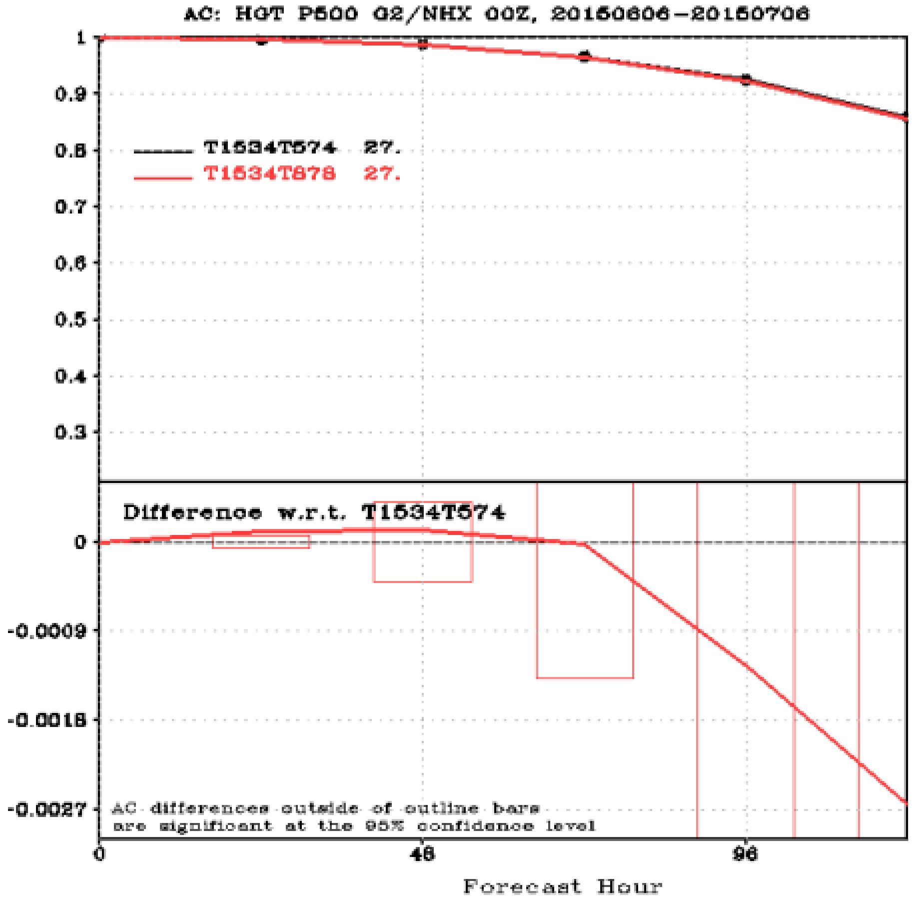
Preparation for NGGPS dynamical core upgrade

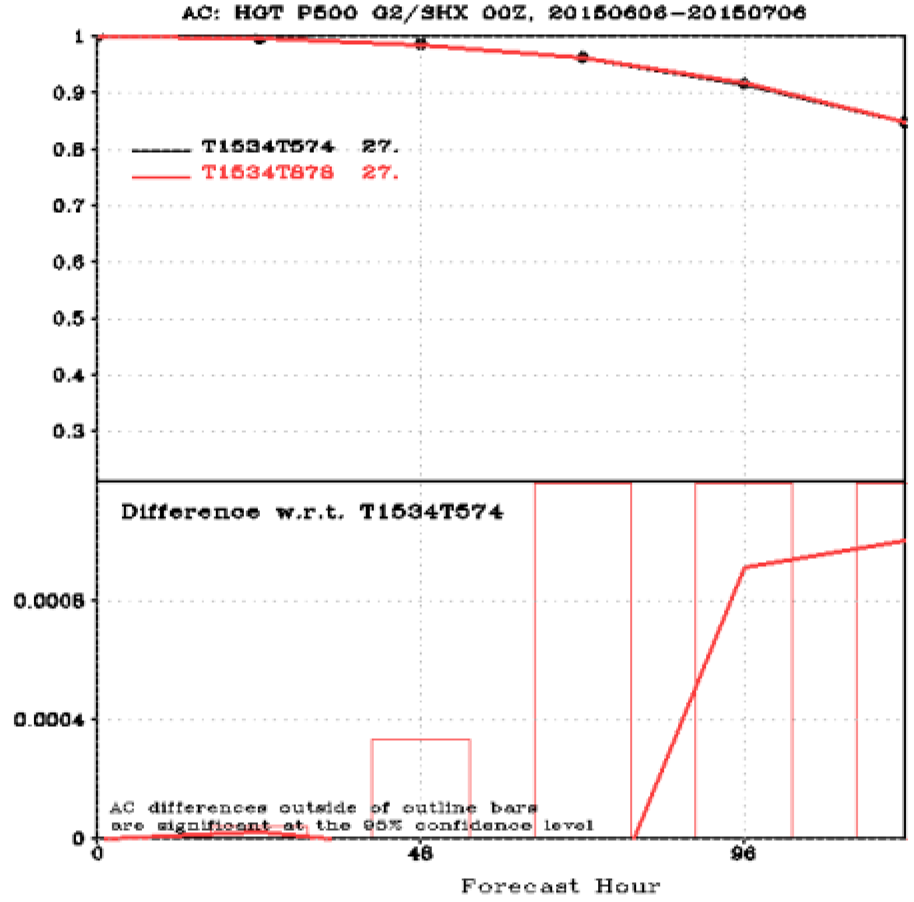
- Experiments performed by ESRL/PSD using modified sub-optimal GSI
- Not currently supported at EMC
- Main issues are analysis grids, interpolation to observations and non-hydrostatic
  - Sub-optimal implementation – first step
    - Interpolation to regular grid for analysis
    - Interpolation to observations from analysis grid
    - Non-hydrostatic analysis increment assumed zero
    - Interpolation of analysis increment from regular grid to model grid
  - First upgrade will be interpolation of model forecast to observations from model grid - JEDI

Enhancements to 4d hybrid

- Variable Beta weights and localization scales
  - Beta weights – weights for ensemble and static background terms
    - Currently assumed to sum to 1.
    - In global constant in vertical. Regional varies in vertical (using global ensemble)
  - Localization scales
    - Necessary because of limited ensemble
    - Currently assumed same for all variables and no spatial variation (vertical variation)
  - Generalization of both of these terms underway
- Ensemble resolution and number changes
  - Lower resolution work by Whitaker and Lei appears to show increasing resolution results in greater improvement
  - Higher resolution tests underway (Mahajan)

T878 vs T574 ensembles



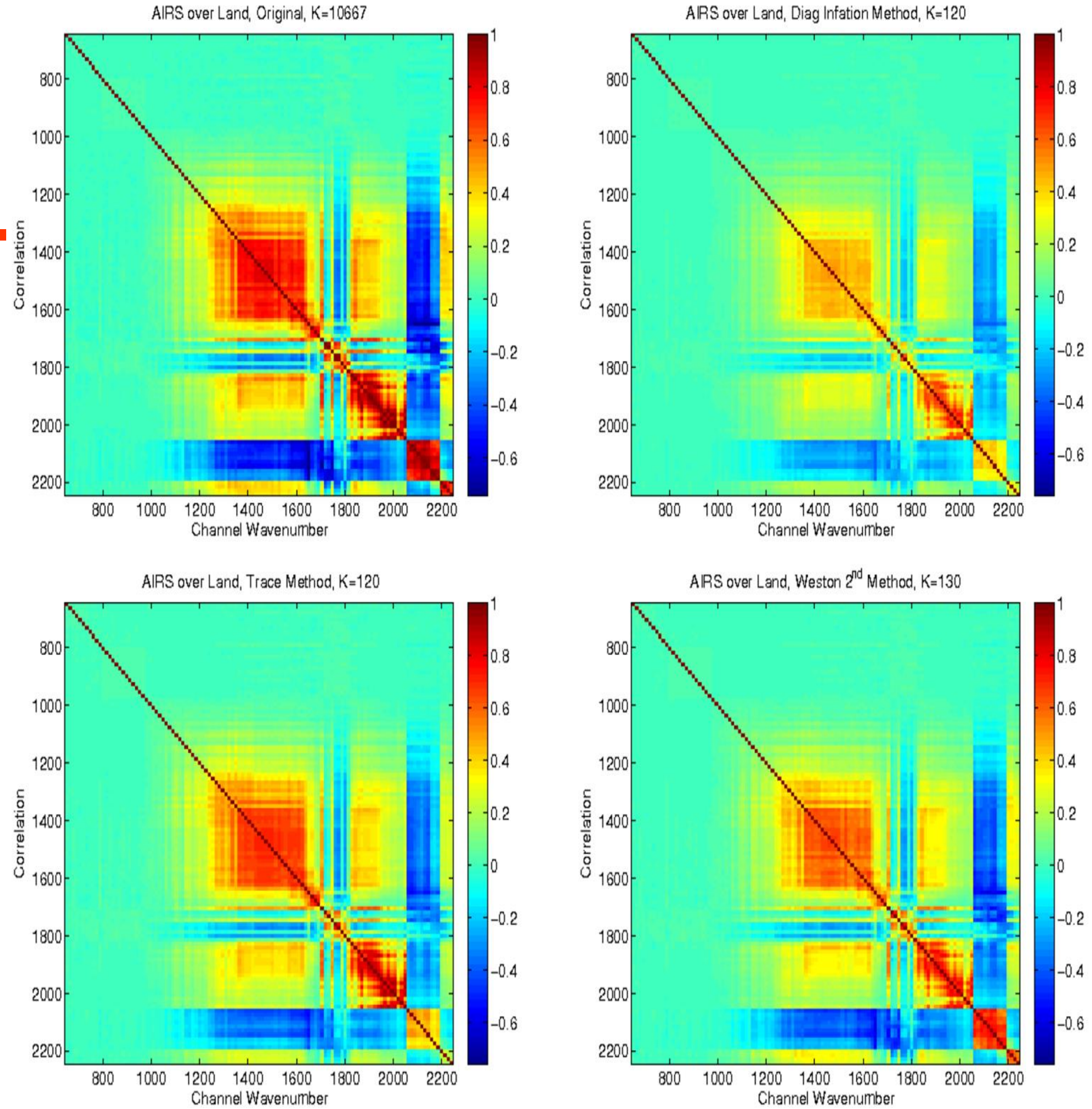


Modifications necessary for changes in model resolution

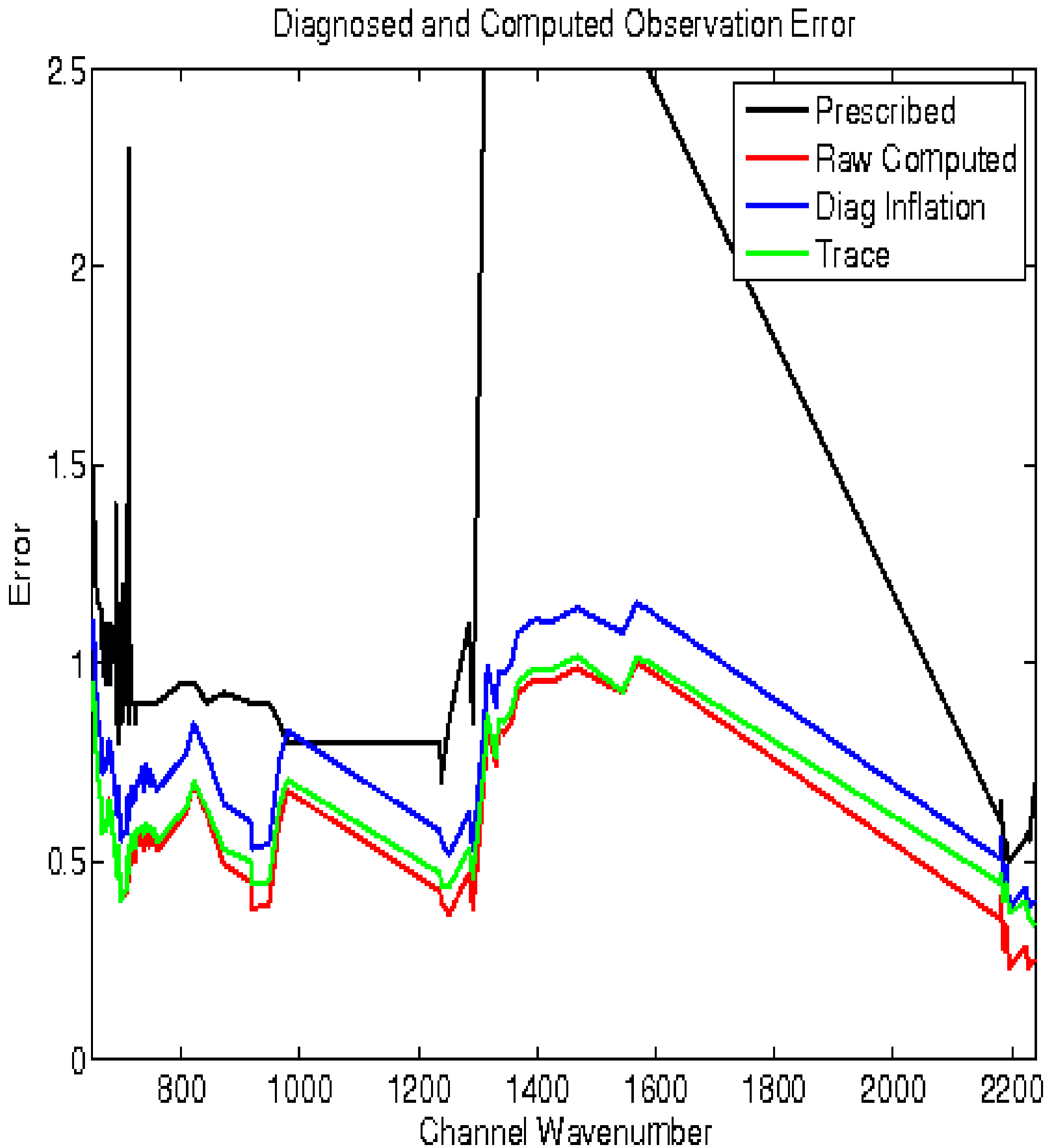
- Current plans for global model include increase in horizontal and vertical resolution in Q3FY18 (requiring system to be complete in Q2/3FY17)
  - Vertical resolution change results in most difficulties
  - As system set up requires 24/48 hour forecasts over ~1 year for definition of background error covariance (and other statistical relationships)
  - Retuning (requiring substantial rerunning of system) usually necessary
  - Data selection and radiance bias correction may need to be modified (as model top moves up)
  - Changes in model resolution will probably require enhancements in code optimization (especially i/o)

Improvements to observation error specification and bias correction

- Inclusion of correlated errors for satellite radiances (GMAO, Bathmann and Collard)
  - Reduction in specified observational error variance
  - Regularization of covariance matrix
- Bias correction of aircraft data (Zhu, Purser and Yang)
  - Evaluating different predictors
- Radiation correction for Rawinsondes (Merkova)
- Update of observation error variance (e.g. AMV's) (Genkova, Su and UW)
  - correlated error considerations

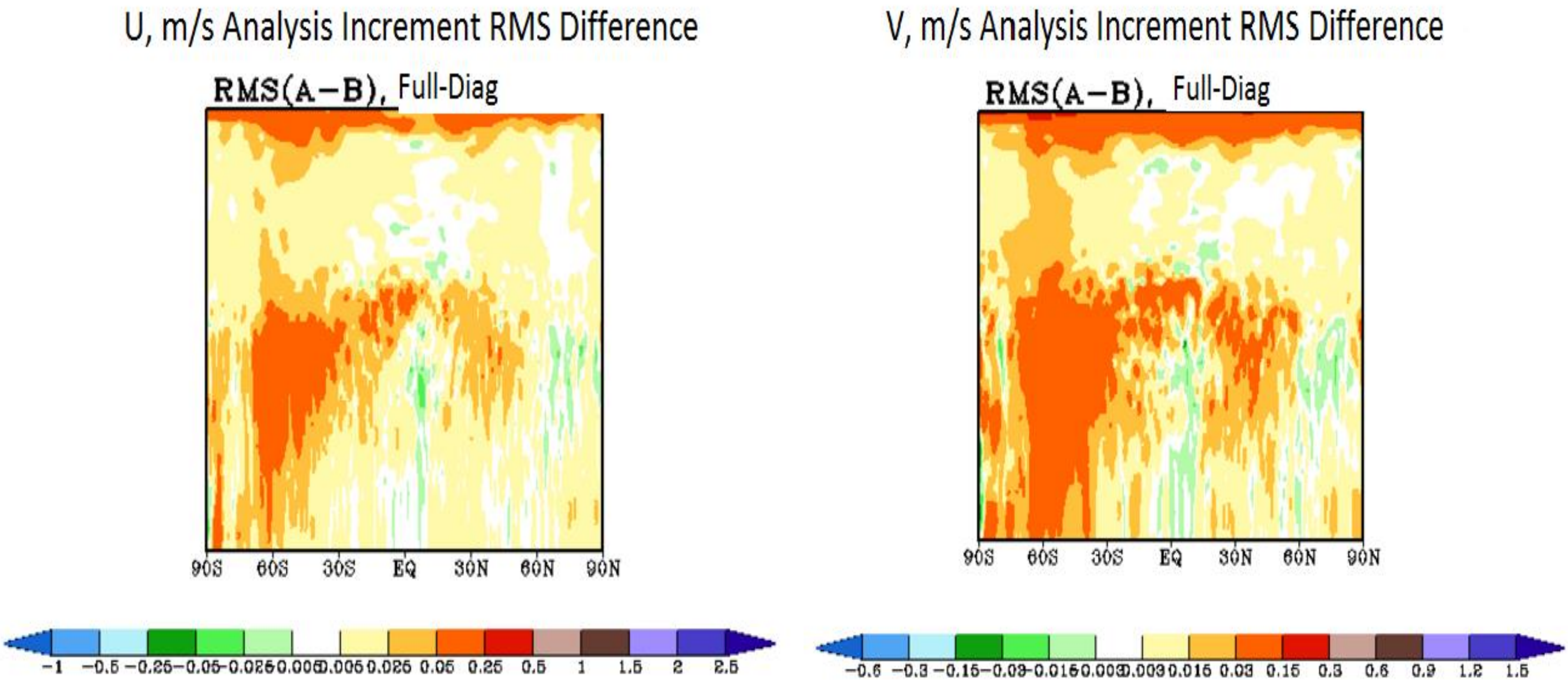


Comparison of the observation error correlation matrix for AIRS over land (top left), reconditioned with the diagonal inflation method (top right), trace method (bottom left) and Weston's method (bottom right). Here, K stands for condition number.



Observation errors that were prescribed to AIRS, compared the errors computed from Desroziers' method (over land), before and after reconditioning R.

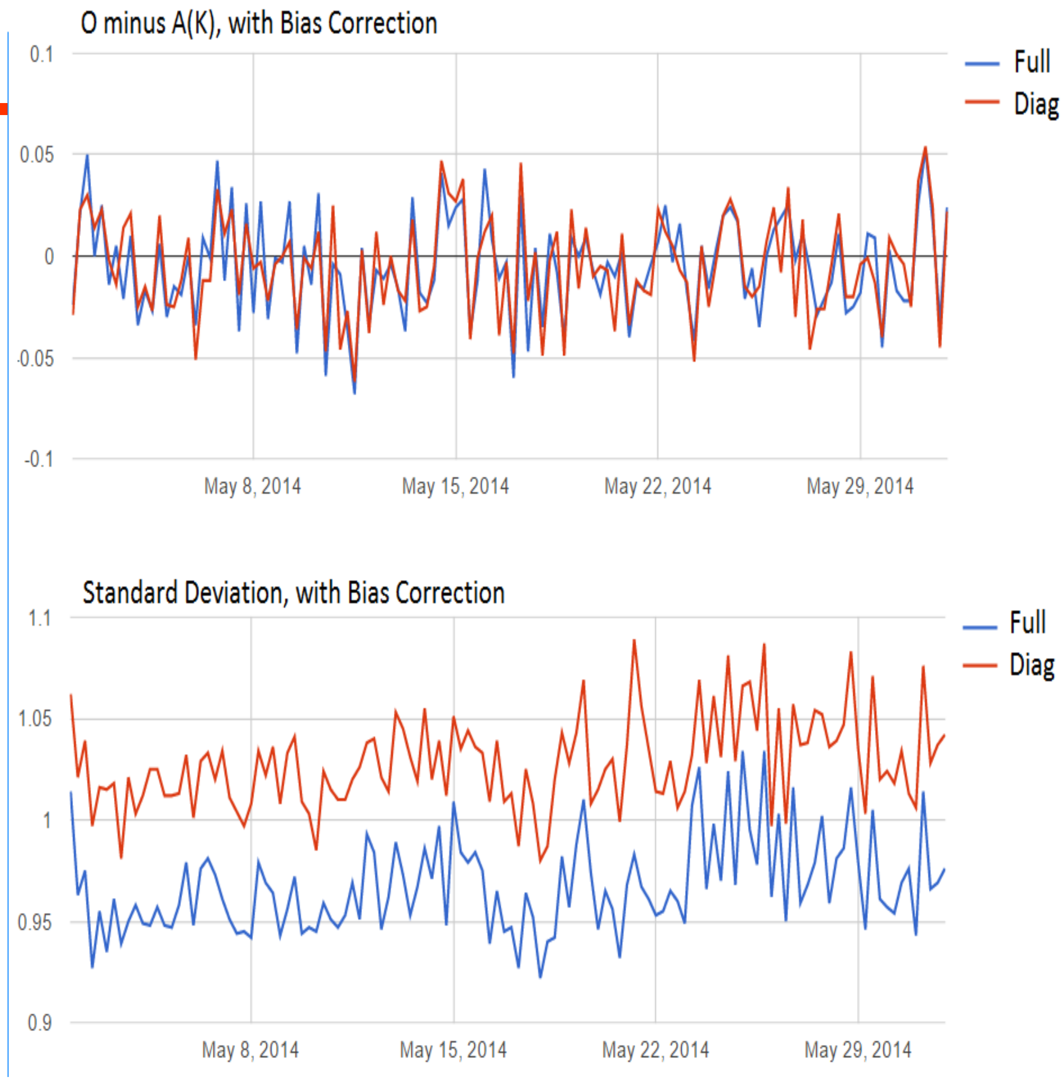
Improvements to observation error specification and bias correction



Wind analysis RMS increment after using a full R globally for AIRS and IASI in a 2 month parallel GFS experiment

IASI Metop A, λ=6.71

Valid 2014060100



Fit to a passive IASI water vapor channel, in the 2 month GFS experiment

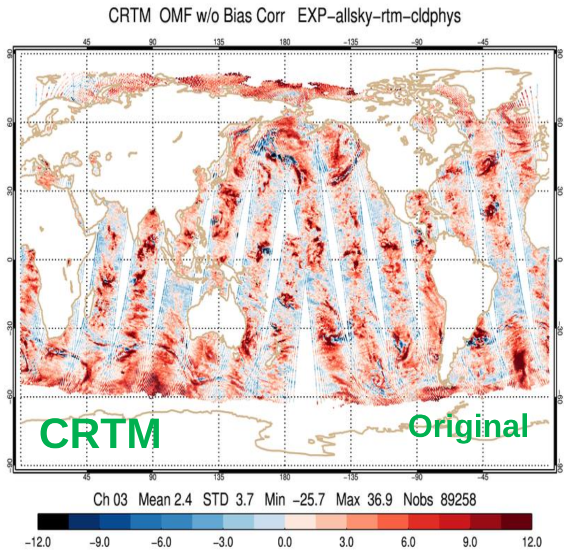
Improvements to observation error specification and bias correction

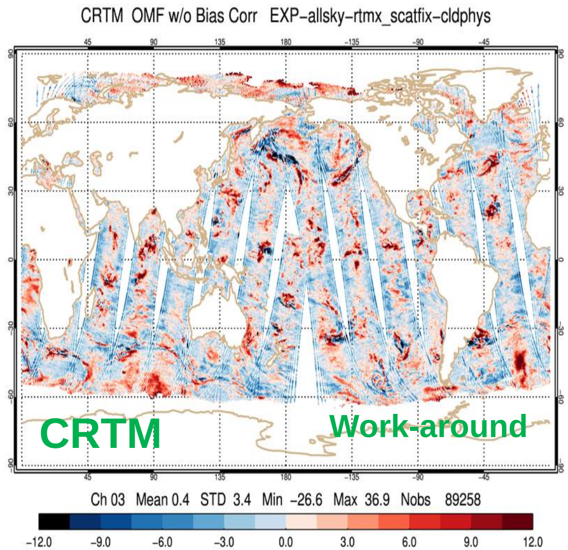
- Inclusion of correlated errors for satellite radiances (GMAO, Bathmann and Collard)
  - Reduction in specified observational error variance
  - Regularization of covariance matrix
- Bias correction of aircraft data (Zhu, Purser and Yang)
  - Evaluating different predictors
- Radiation correction for Rawinsondes (Merkova)
- Update of observation error variance (e.g. AMV's) (Genkova, Su and UW)
  - correlated error considerations

Improved use of all-sky radiances

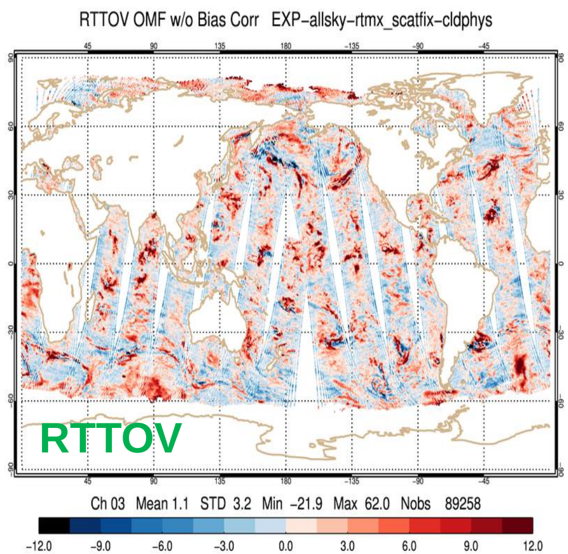
- Use of all-sky radiances still in infancy (Zhu, E. Liu, H. Liu, L. Bi, Collard).
- Improved simulation of (cloud-impacted) radiances from CRTM
- Extension of use from AMSU-A to ATMS
- Extension of use to IR (much more non-linear)
- **Model balance issues**: balance at beginning of forecast (decreasing spin-up/down of clouds)
- Choice of analysis variables
- Inclusion of convective clouds from/to model

Surface emissivity issues under scattering conditions – reflection of diffuse radiation and restricting to < 60 degrees





AMSU-A
Channel 3
Observation minus First-Guess



Enhanced quality control (QC)

- **Variational QC**
  - Current variational scheme can cause convergence problems and does not model error distribution as well as possible
  - New variational scheme (Purser, Su and Yang) partially incorporated in code and used in RTMA and URMA
  - Extension to all observation variables necessary. How to treat correlated errors?
- **Station history based QC**
  - Identifying bad observation based on history
    - Stuck instruments, frequent bad values, large biases, etc.
  - Must fit into operational environment and allow rehabilitation of observations

Computational and structure optimization

- Presentation by T. Auligné in second half on JEDI
  - Top down and bottom up strategy. Cannot pause GSI development for infrastructure upgrade.
- Computational efficiency vital
- Must fit into NCO requirements
- Must be usable and understandable by GSI community and partners
- I/O very important factor in current runtime
  - Ensembles, NEMS, higher resolutions, compute increasing faster than I/O capability
- Inclusion of all partners needs is difficult

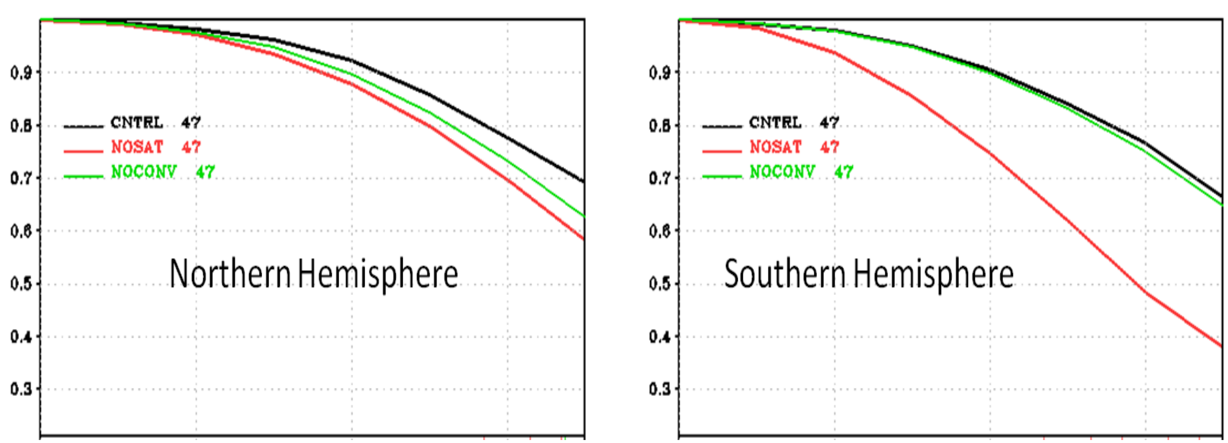
Joint Effort for Data assimilation Integration (JEDI)

Looking Forward (5-10 years)

World's Best Global Forecast is *impossible* without
World's Best Data Assimilation

Contributions to medium-range forecast error:
Initial Conditions = **Model error**
(Magnusson and Källen 2013)

Rule of thumb:
Data Assimilation ~50% effort in NWP (staff, HPC, software)



Source: Jung (2012)

High-level requirements

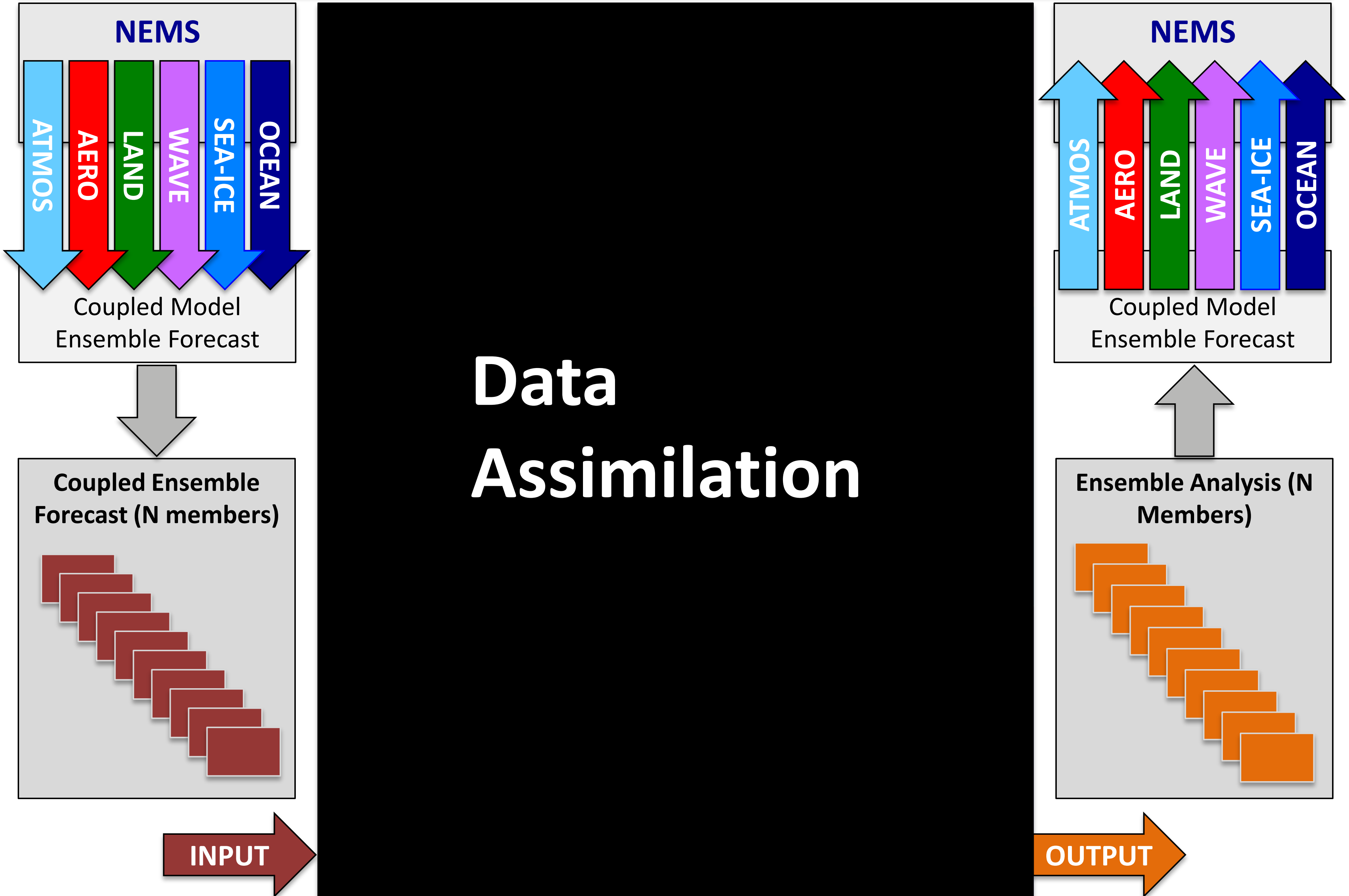
1. Major refactoring of existing DA systems

2. Coordinated community effort

1. Avoid piecemeal approach

   - across initiatives, domains and applications

NCEP Coupled Hybrid Data Assimilation and Forecast System



Adapted from Saha

JEDI

+ Academia

Nation Unified Next-generation Data Assimilation

STRATEGY

1. Collective path to unification, while allowing multiple levels of engagement
2. Modular, Object-Oriented code for flexibility, robustness and optimization
3. Mutualize **model-agnostic** components across
   - Applications (atmosphere, ocean, land, aerosols, strongly coupled, etc.)
   - Models & Grids (operational/research, regional/global models)
   - Observations (past, current and future)

OBJECTIVES

1. Facilitate innovative developments to address DA grand challenges
2. Increase **R2O** transition rate from community
3. Increase **science productivity** and **code performance**

JEDI: Project Schedule

- **Phase 0**: **FY16** Spin-up with JCSDA funds
- **Phase 1→Q2 FY18** Technical specs + Prototyping
- **Phase 2→Q4 FY19** Mature system demonstration
- **Phase 3→Q2 FY21** Transition to operations

Management Process: **Phase 0**

- **8-10 March 2016**: Joint NCAR/JCSDA Workshop: "Blueprints of Next-Generation DA Systems"

- **Draft White Paper:** define JEDI project, objectives, and approach

- **Q4FY16**: JCSDA new hires
  - "JEDI master" (design and planning, project management)
  - 2 scientists (GSI Forward Operator, and sea-ice DA)

Management Process: Phase 1

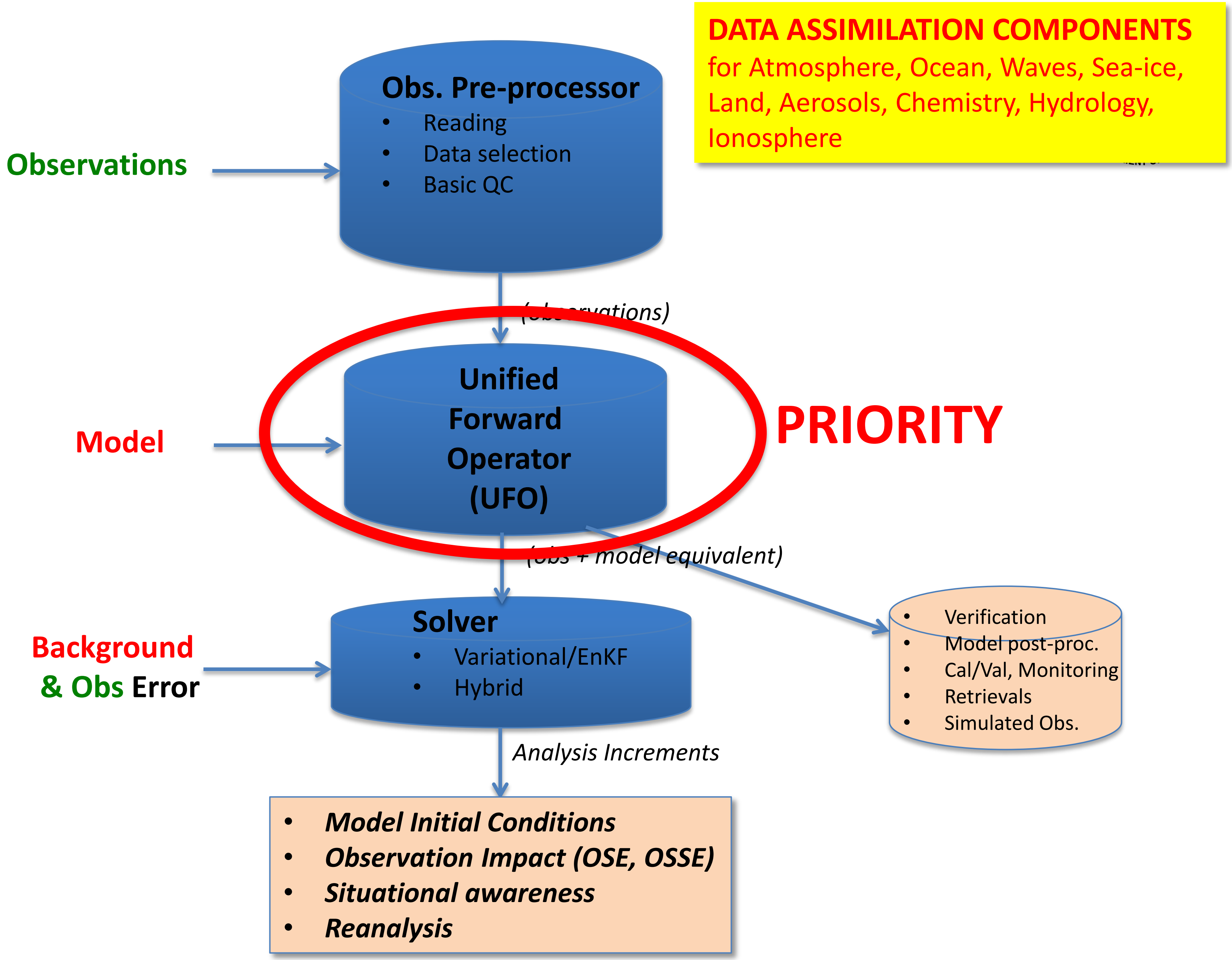
- **Q1FY17**: create "JEDI Council" (= Nation Unified DA Project Team)
  - JEDI core developers + project management
  - ~3 independent senior consultants/advisors
  - Representatives from partners with direct involvement
  - Liaisons with NGGPS Task Teams
    (GMTB, Aerosols and atmospheric composition, Post-processing, Verification, Marine, Land, Overarching system, Software architecture)
- **Q2FY17**: NGGPS DA workshop (series of focused sessions)
- **Q4FY17**: Planning documents:
  - JEDI system requirements (*What*)
  - NGGPS DA roadmap (*When*)
  - Community code governance (*How*)

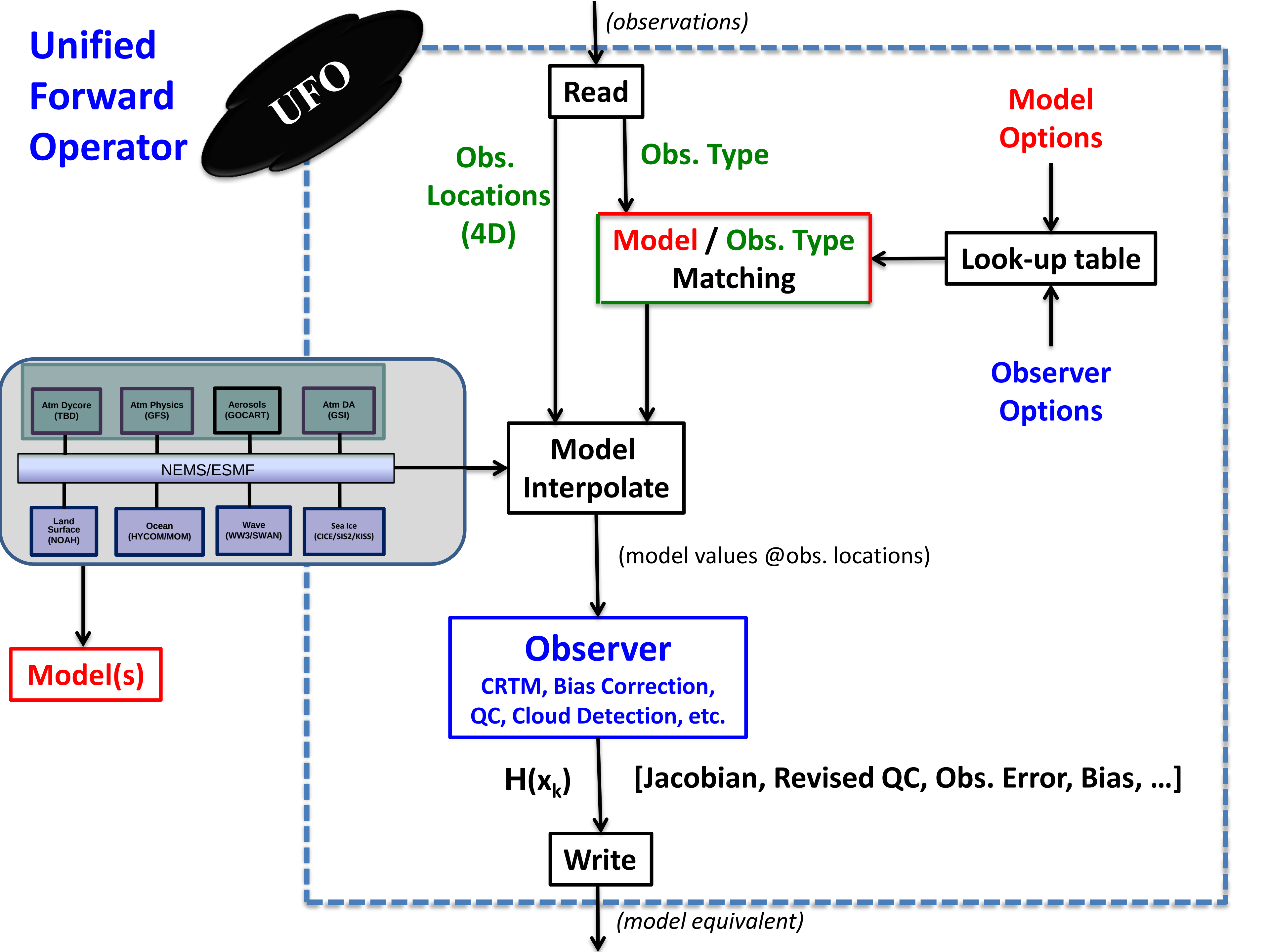
NGGPS DA Roadmap

Roadmap connecting two simultaneous approaches

- **Top-down approach:** *start with clean slate…*
  - Gather users' needs and produce requirements document.
  - Evaluate ECMWF OOPS beta version.
  - *Key partners: JCSDA, ESRL, NGGPS Teams*

- **Bottom-up approach:** *start from existing codes…*
  - Incremental refactoring of GSI code
  - High-level modularization and polymorphism (OO code)
  - *Key partners: JCSDA, EMC, NASA/GMAO, NGGPS Teams*


DATA ASSIMILATION COMPONENTS
for Atmosphere, Ocean, Waves, Sea-ice, Land, Aerosols, Chemistry, Hydrology, Ionosphere
Observations
Obs. Pre-processor
• Reading
• Data selection
• Basic QC
(observations)
Model
Unified Forward Operator (UFO)
PRIORITY
(obs + model equivalent)
• Verification
• Model post-proc.
• Cal/Val, Monitoring
• Retrievals
• Simulated Obs.
Background & Obs Error
Solver
• Variational/EnKF
• Hybrid
Analysis Increments
• Model Initial Conditions
• Observation Impact (OSE, OSSE)
• Situational awareness
• Reanalysis

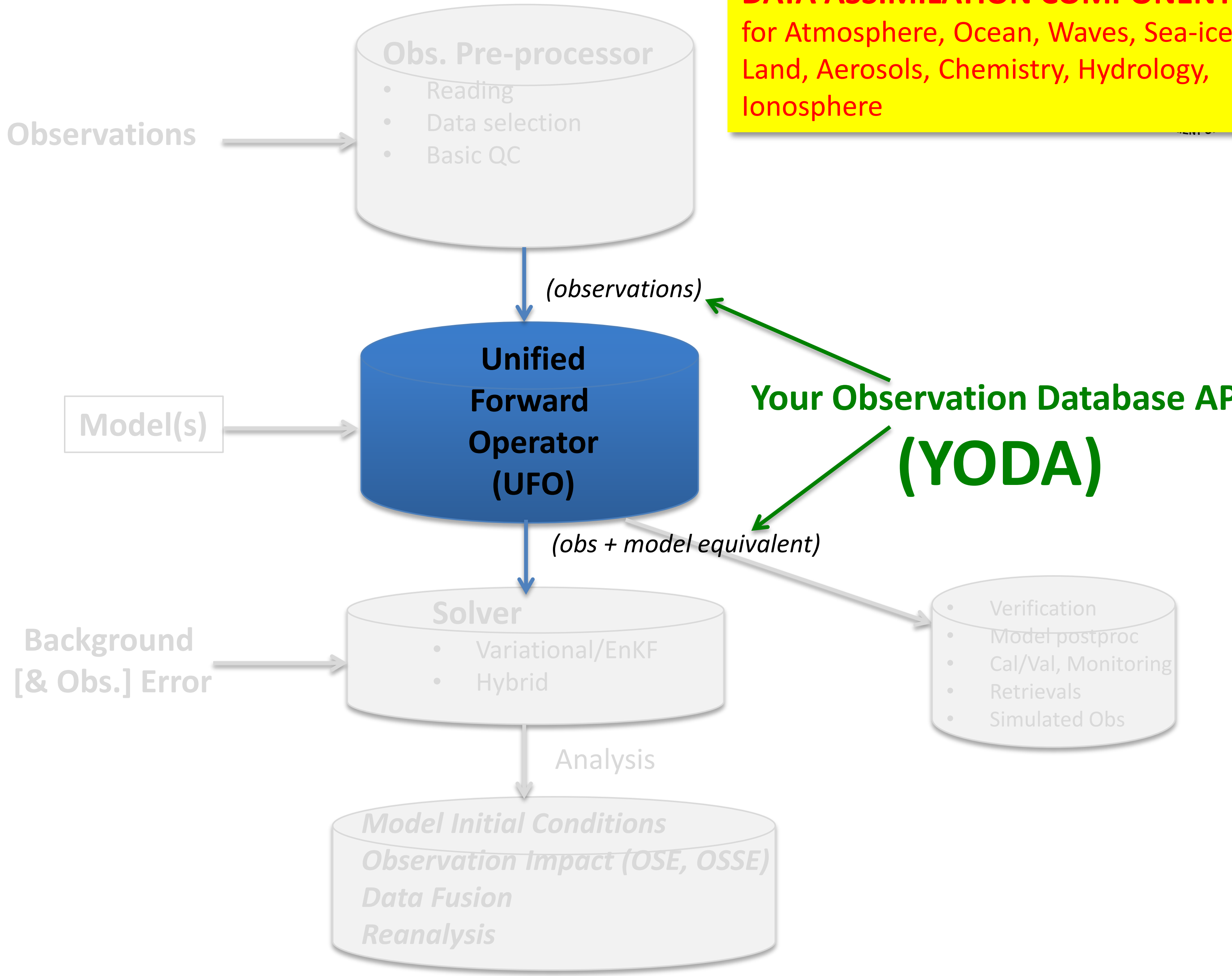

Unified
Forward
Operator
UFO
(observations)
Read
Obs. Locations (4D)
Obs. Type
Model Options
Model / Obs. Type Matching
Look-up table
Observer Options
Atm Dycore (TBD)
Atm Physics (GFS)
Aerosols (GOCART)
Atm DA (GSI)
NEMS/ESMF
Land Surface (NOAH)
Ocean (HYCOM/MOM)
Wave (WW3/SWAN)
Sea Ice (CICE/SIS2/KISS)
Model Interpolate
(model values @obs. locations)
Model(s)
Observer
CRTM, Bias Correction, QC, Cloud Detection, etc.
H(x_k)
[Jacobian, Revised QC, Obs. Error, Bias, ...]
Write
(model equivalent)

Roadmap

UFO: Unified Forward Operator

1. Split GSI into Pre-processor, UFO, Solver, (+Utils) independent libraries
2. Develop **flexible** UFO infrastructure
    - Read/write observations and model equivalent
    - Match observation types with model variables
    - Encapsulate+generalize+optimize interpolation of model fields to observation locations
3. Polymorphic version of GSI CRTM interface, bias correction, QC, error estimation, etc.
4. Expand UFO to sea-ice, ocean (NCODA?), …

DATA ASSIMILATION COMPONENTS
for Atmosphere, Ocean, Waves, Sea-ice, Land, Aerosols, Chemistry, Hydrology, Ionosphere



Roadmap

YODA: Your Observation Database API

- **Objectives**: Standardized input/output API
  - Data Assimilation (atmos., ocean, reanalysis, …)
  - Verification/Validation, Model Post-processing,
  - Cal/Val, Retrievals, OSSEs

- **Q2 FY17**: hire "Database Guru"
- **Q3 FY17**: Placeholder in GSI = API to flat NetCDF files
- **Q4 FY17**: Requirements document & tech. specs
  - *Metadata* for variety of sensors (past, current, future)
  - Flexible data manipulation, yet fast; low cost
  - Parallel distribution; archiving; [data on the Cloud]

JEDI: Modes of operation

- **Governance**
  - Collegial decisions ("JEDI Council”)
  - Coordination at Object Design level
  - Allow for multiple levels of engagement

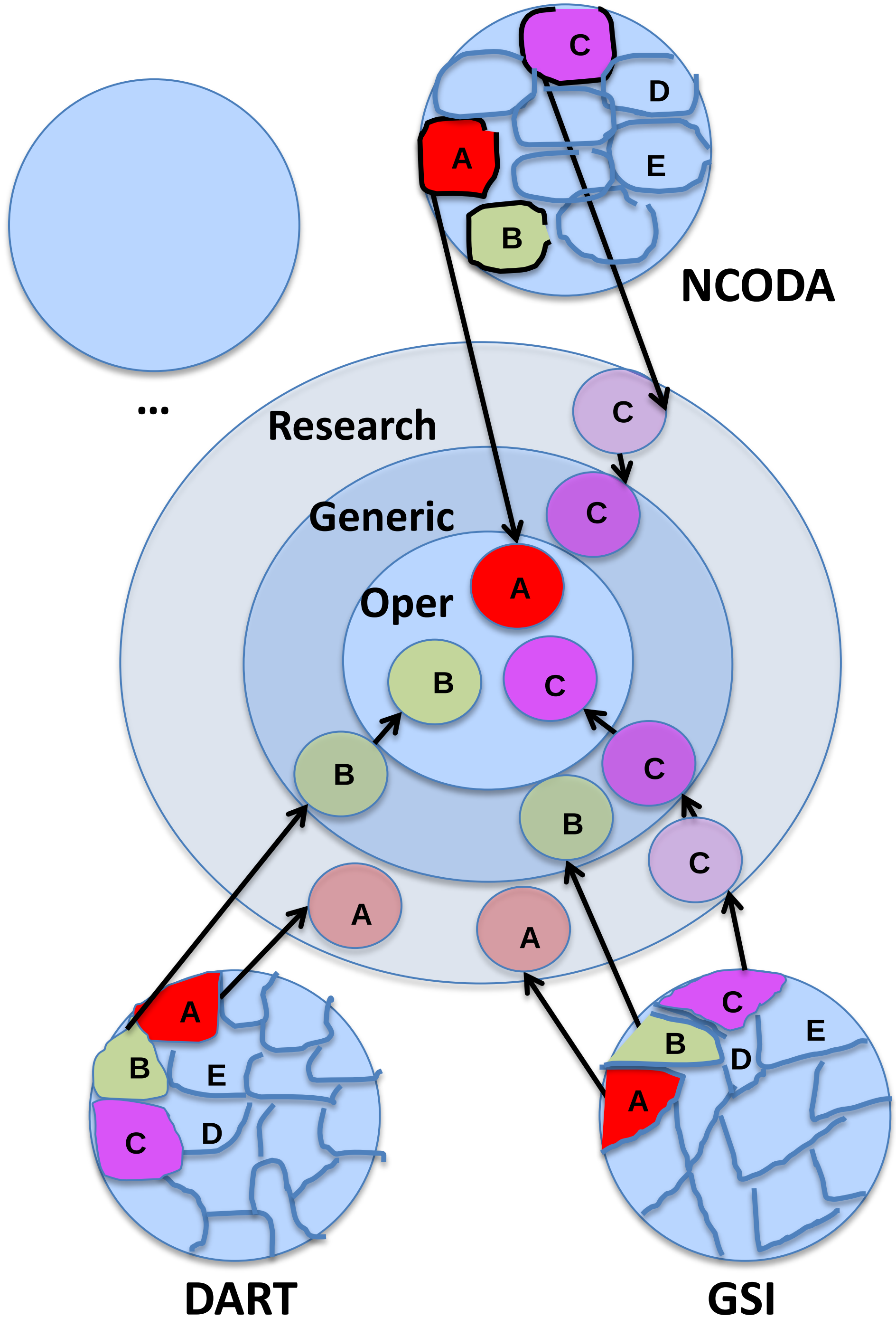
- (Single GIT) **Community Repository**

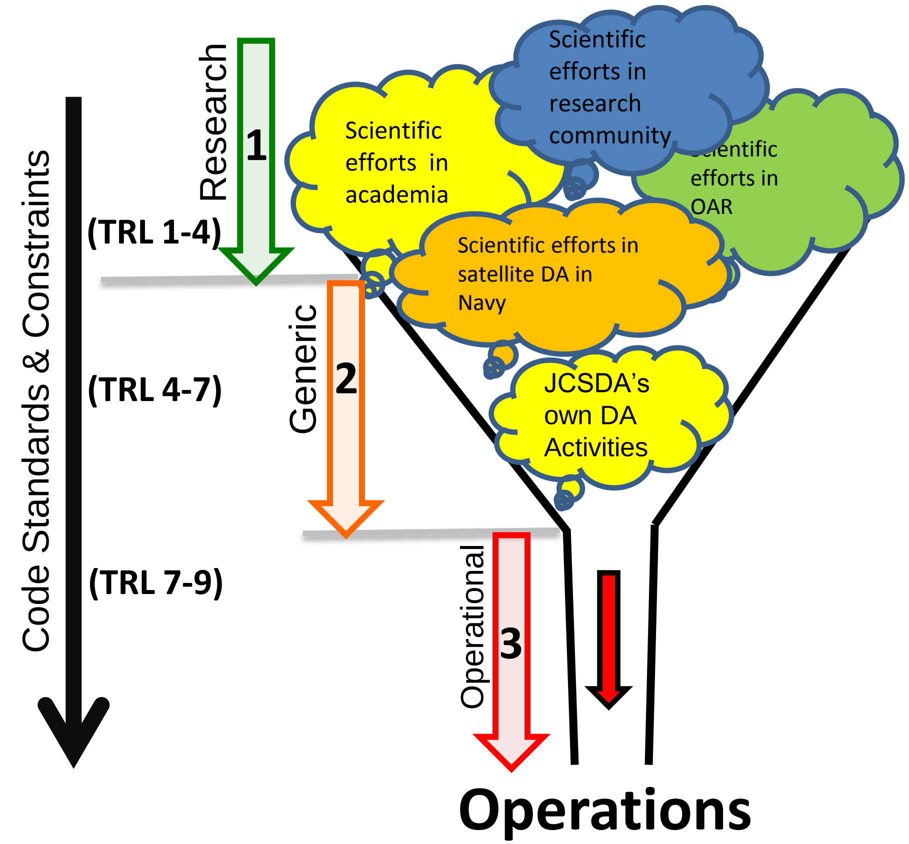
- Define **requirements and metrics** for accepting developments

- **Entropy Management Team** (EMT)
  - Support for scientists: promote generality and avoid redundancy
  - Enforce coding standards: ensure readability
  - Support documentation and regression testing

Key partners: GSI/EnKF Review Committee, NCAR, DTC, GMTB, Community Model Infrastructure Team

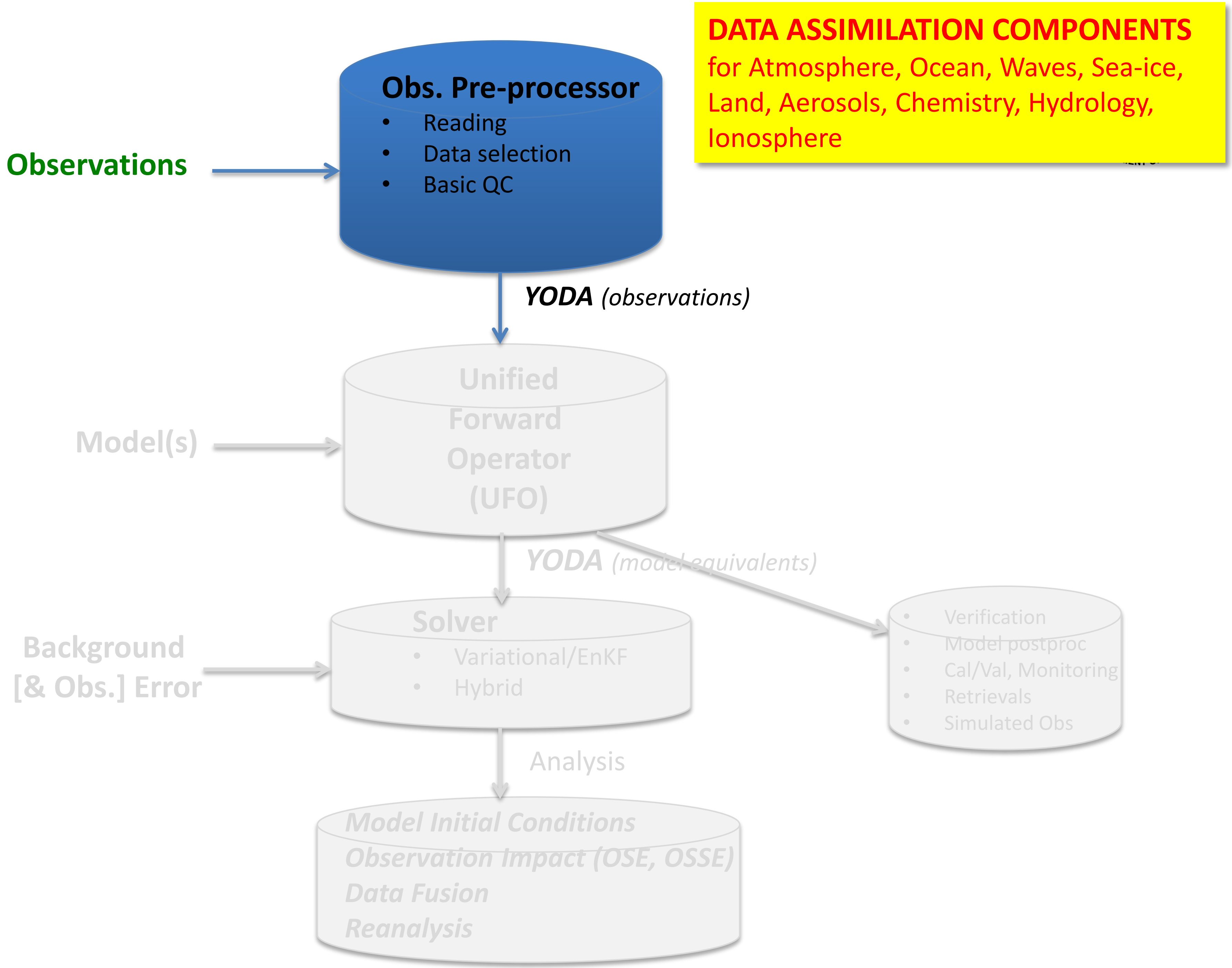
MULTI-LEVEL COMMUNITY REPOSITORY



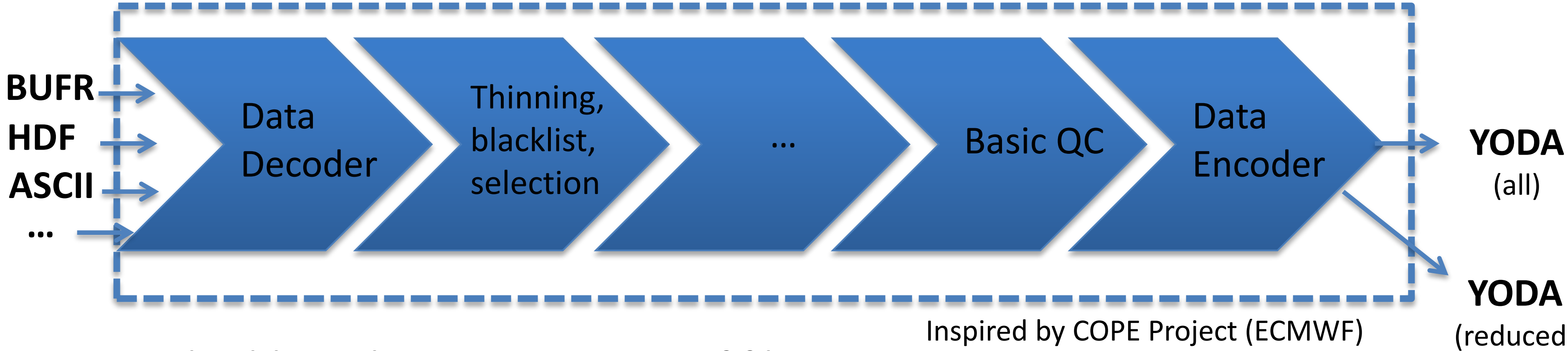


JEDI: Project Schedule

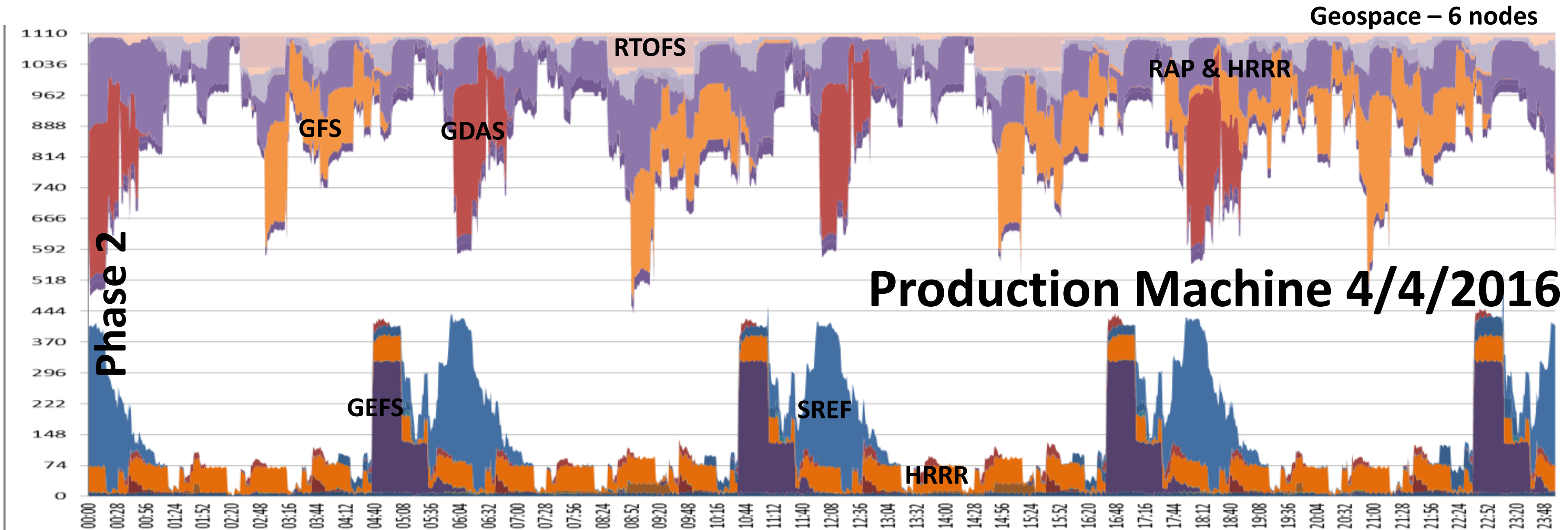
- **Phase 0: FY16** Spin-up with JCSDA funds
- **Phase 1: Q2 FY18** Technical specs + Prototyping
- **Phase 2: Q4 FY19** Mature system demonstration
- **Phase 3: Q2 FY21** Transition to operations

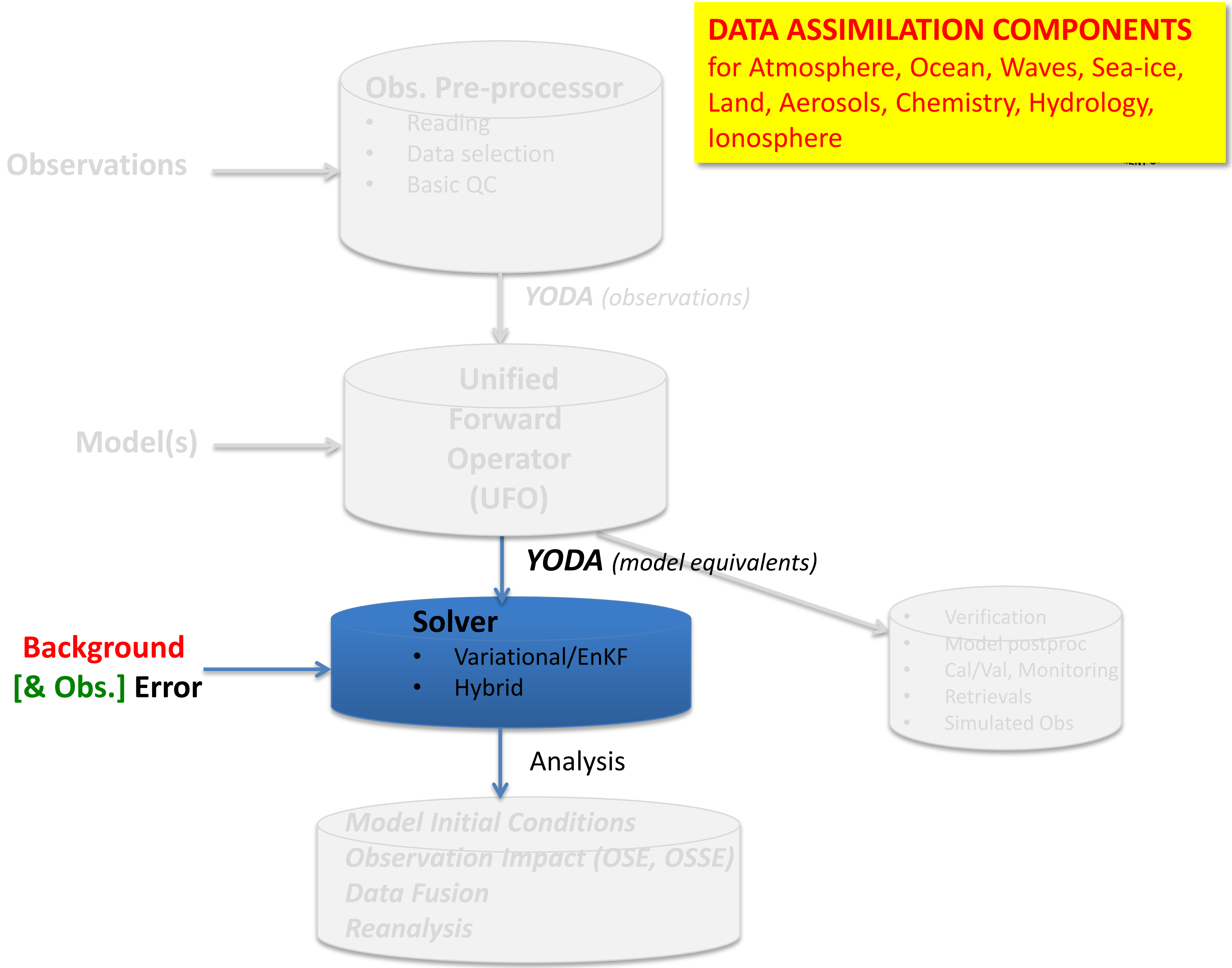

DATA ASSIMILATION COMPONENTS
for Atmosphere, Ocean, Waves, Sea-ice, Land, Aerosols, Chemistry, Hydrology, Ionosphere
Observations
Obs. Pre-processor
Reading
Data selection
Basic QC
YODA (observations)
Model(s)
Unified Forward Operator (UFO)
YODA (model equivalents)
Background [& Obs.] Error
Solver
Variational/EnKF
Hybrid
Verification
Model postproc
Cal/Val, Monitoring
Retrievals
Simulated Obs
Analysis
Model Initial Conditions
Observation Impact (OSE, OSSE)
Data Fusion
Reanalysis

Observation Pre-Processor



- Flexible architecture = series of filters
- Can be done as soon as data is available




DATA ASSIMILATION COMPONENTS
for Atmosphere, Ocean, Waves, Sea-ice, Land, Aerosols, Chemistry, Hydrology, Ionosphere
Obs. Pre-processor
Reading
Data selection
Basic QC
Observations
YODA (observations)
Unified Forward Operator (UFO)
Model(s)
YODA (model equivalents)
Solver
Variational/EnKF
Hybrid
Background
[& Obs.] Error
Verification
Model postproc
Cal/Val, Monitoring
Retrievals
Simulated Obs
Analysis
Model Initial Conditions
Observation Impact (OSE, OSSE)
Data Fusion
Reanalysis

JEDI: Milestones

- **Phase 1: Q2 FY18** Technical specs + Prototyping
  - **UFO (Unified Forward Operator)**: Prototype development (limited obs. and models)
  - **YODA (Your Observation database API)**: requirements & tech. specs.

    New features: - observations simulated from FV3 native grid (outside & inside model),

    - standardized access to observations

- **Phase 2: Q4 FY19** Mature system demonstration
  - **UFO**: complete set of observations and models
  - **YODA**: Development of next-generation database
  - **Solver**: Externalized B and localization operators
  - **Pre-processor**: generic filters
  - **Software infrastructure**: unified code repository, regression tests, documentation

    New features: - strongly coupled earth prediction system,

    - compare solvers,

    - simplified code management

- **Phase 3: Q2 FY21** Transition to operations

Foreseen Challenges (5-10 years)

Model

- NWP evolving into Earth System Prediction
  (Ocean, Waves, Cryosphere, Land, Hydrology, Aerosols, Atmospheric composition, Whole Atmosphere, etc.)

- (Strongly) coupled system
- "Seamless" analysis across scales
  - Situational awareness -> Nowcasting -> NWP -> Seasonal -> Climate
  - Convection permitting -> Synoptic -> Global

Observations

- Big Data = more volume, variety, velocity

- Accumulation of small improvements from each instrument
- Need lots of "smarts" in data mining

DA Algorithms

- No obvious winner in sight
- Uncertainty w/r HPC scalability

- Need flexibility to keep options open
- Optimum may be application/machine dependent

NGGPS Atmospheric Data Assimilation - Summary

- **Major Accomplishment in FY16:**
  - Operational implementation of 4d-hybrid EnVar system and use of cloudy radiances
- **Priority Foci for FY17**
  - Operational implementations. Modifications for changes in model resolution
  - Inclusion of GOES-R, JPSS-1 and other new data sources
  - Improvements to 4d hybrid, observation error specification, bias correction, Enhanced QC, all-sky radiances (including model balance issues), computational optimization
  - Preparation for Nation unified next-generation Data Assimilation, incl. NGGPS dynamic core.
- **Key Issue**
  - Resources (computational and human) are not sufficient to meet expectations.