Data Assimilation Planning for the Unified Forecast System

Elements to discuss priorities for future DA strategy

Thomas Auligné. Director, Joint Center for Satellite Data Assimilation (JCSDA)

Arun Chawla (NCEP/EMC), Andrew Collard (NCEP/EMC), John Derber (NCEP/EMC), Ron Gelaro (NASA/GMAO), Dan Holdaway (NASA/GMAO), Daryl Kleist (NCEP/EMC), Rahul Mahajan (NCEP/EMC), Yannick Trémolet (JCSDA), Jeff Whitaker (ESRL/PSD)

Create best possible operational DA

Best operational forecast system **REQUIRES** best DA

Option 1: assemble the 'Dream Team', set S.M.A.R.T. goals, provide a *cocoon* for scientists, and get out of the way.

However, challenges for next-generation DA:

- DA systems becoming increasingly complex as science progresses
- Forecast improvement due to many obs. with **small individual impacts**
- Earth System Models are getting coupled to better account for interactions

Option 2: build collaborative infrastructure that scales for:

- More observations and higher model resolution
- More grids, sub-systems and applications
- More concurrent developments and distributed partners

JEDI: Motivations and Objectives

The Joint Effort for Data assimilation Integration (JEDI) is a collaborative development spearheaded by the JCSDA.

Develop a unified data assimilation system:

- From toy models to Earth system coupled models
- Unified observation (forward) operators (UFO)
- For research and operations (including R2O/O2R)
- Share as much as possible without imposing one approach

Grand Science Challenges for DA

…identified at NCAR/JCSDA ‘Blueprints’ Workshop (2016)

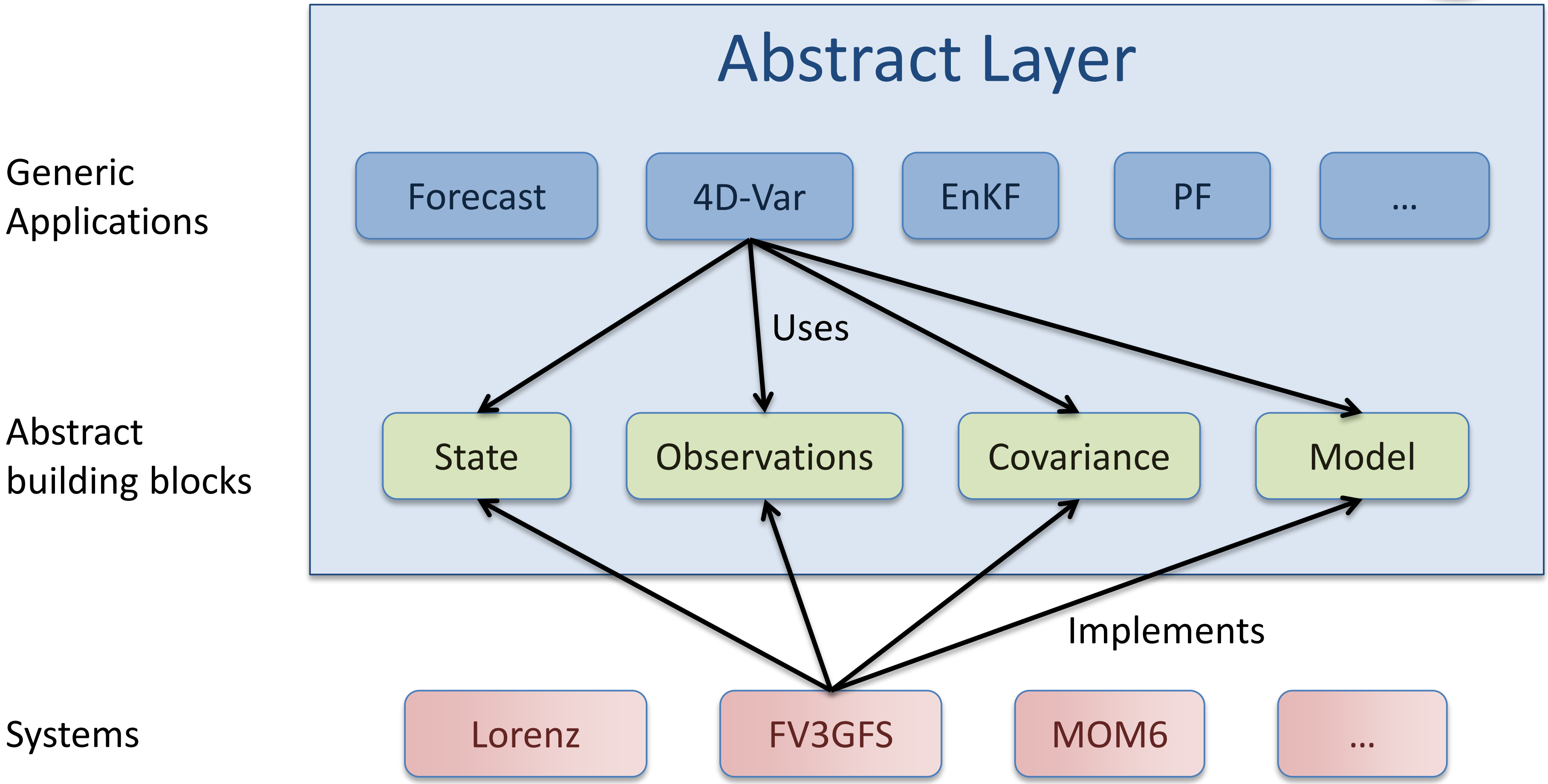
- Coupled DA across the Earth System
- Representation of model uncertainty
- Dealing with non-linearity and non-Gaussianity
- Multi-scale DA across temporal and spatial
- Dealing with massive increases in observations

“A data assimilation system is only as good as its weakest component” *John Derber*

Grand Science Challenges for DA

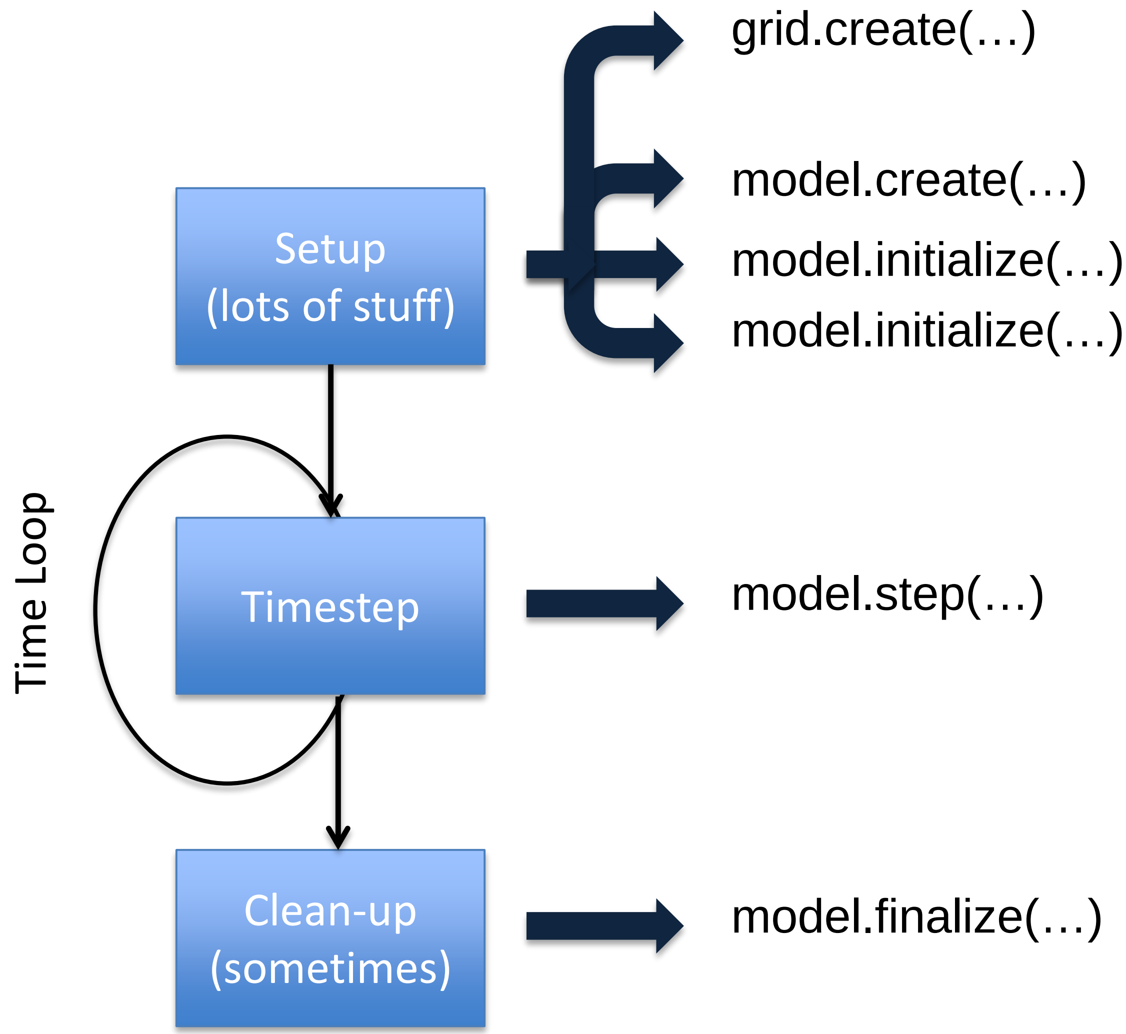
- Coupled DA across the Earth System
- Representation of model uncertainty
- Dealing with non-linearity and non-Gaussianity
- Multi-scale DA across temporal and spatial
- Dealing with massive increases in observations

Abstract Design: separation of concerns



Abstract interfaces are the most important aspect of the design

JEDI Model Interface [$x_t=M(x_0)$]



JEDI Thin Wrapper

Grid and parallel distribution built from stripped down version of main FV3 construct: same I/O routines and data structures (non-intrusive, lightweight, efficient code)

Works for any model, including coupled models

JEDI Build System

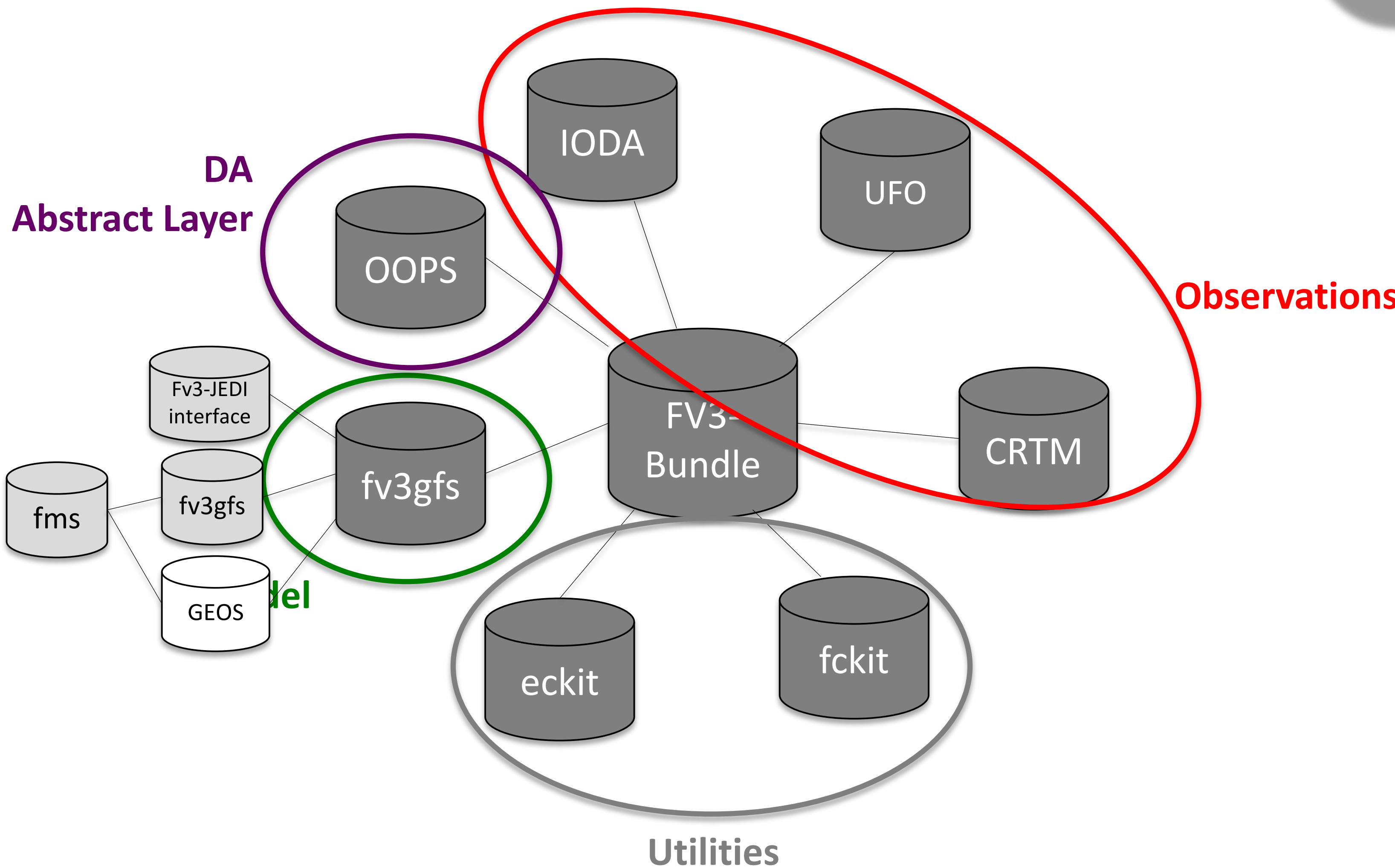
Open-development software-model: in addition to supported releases, community developers can obtain and collaborate on latest development branches

~20 repositories on Github.com incl. FV3GFS, GEOS, MPAS, LFriC, MOM5/CICE5, [MOM6/SIS2], [WRF]

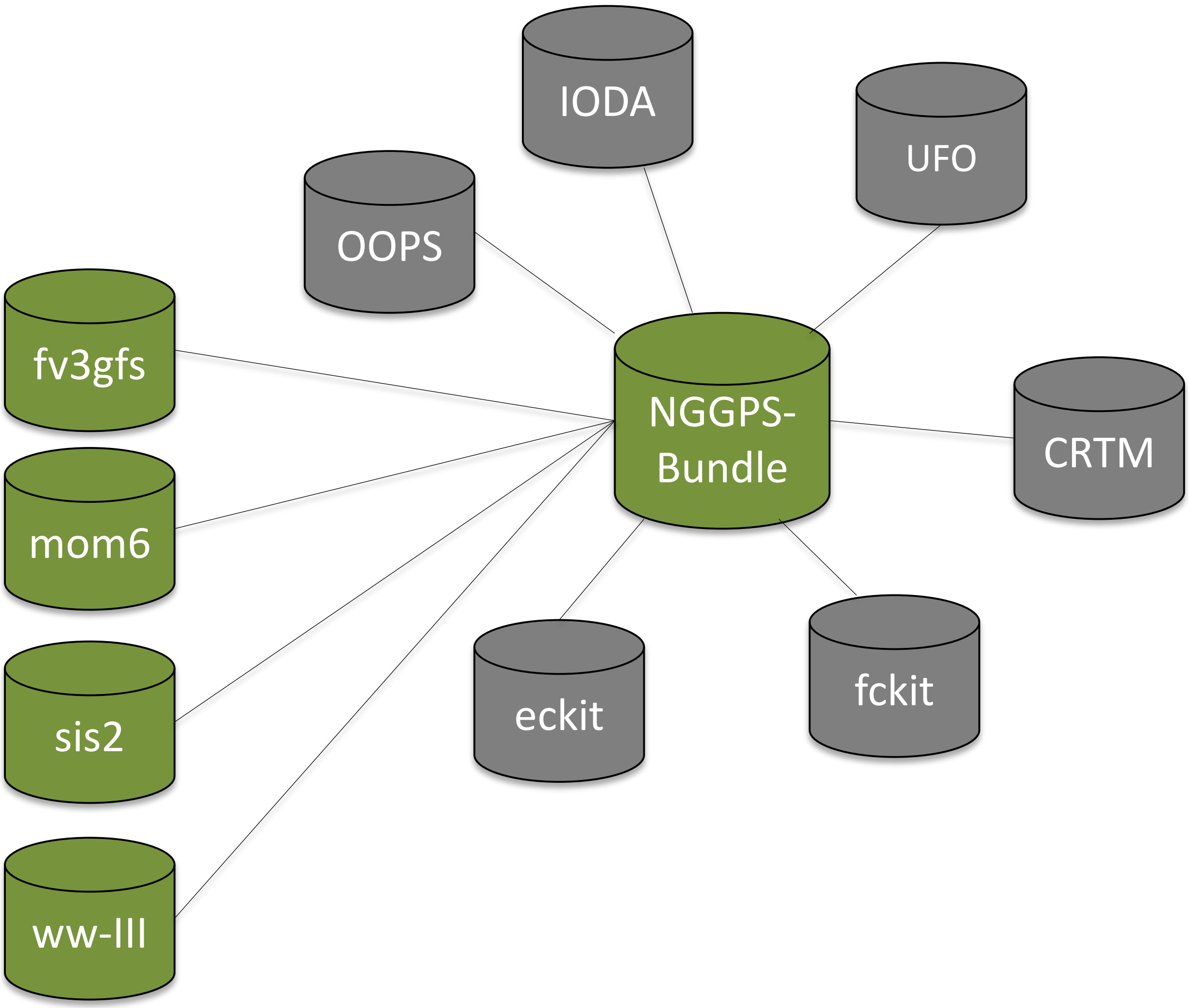
```
project( fv3gfs-bundle C CXX Fortran )
ecbuild_bundle_initialize()
ecbuild_requires_macro_version( 2.7 )
…
ecbuild_bundle( PROJECT eckit  GIT "https://github.com/UCAR/eckit.git"  TAG  0.18.0 )
ecbuild_bundle( PROJECT fckit  GIT "https://github.com/UCAR/fckit.git"  TAG  0.4.1 )
ecbuild_bundle( PROJECT crtm   GIT "https://github.com/UCAR/crtm.git"   BRANCH master    UPDATE )
ecbuild_bundle( PROJECT oops   GIT "https://github.com/UCAR/oops.git"   BRANCH feature/nicas UPDATE )
ecbuild_bundle( PROJECT ioda   GIT "https://github.com/UCAR/ioda.git"   BRANCH develop   UPDATE )
ecbuild_bundle( PROJECT ufo    GIT "https://github.com/UCAR/ufo.git"    BRANCH develop   UPDATE )
ecbuild_bundle( PROJECT fv3gfs GIT "https://github.com/UCAR/fv3gfs.git" BRANCH develop   UPDATE )
  ecbuild_bundle( PROJECT fv3gfs GIT "https://github.com/UCAR/mom6.git" BRANCH develop   UPDATE )
ecbuild_bundle_finalize()
```

```
> git clone https://github.com/UCAR/fv3gfs-bundle.git
> ecbuild fv3gfs-bundle
> make –j4
> ctest
```

JEDI Build System: FV3GFS-Bundle



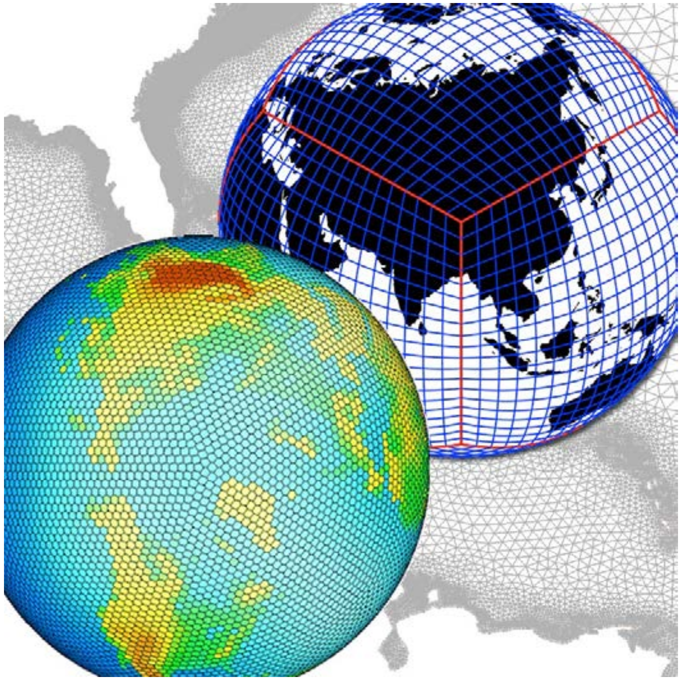
JEDI Build System: Coupled DA



Grand Science Challenges for DA

- Coupled DA across the Earth System
- Representation of model uncertainty
- Dealing with non-linearity and non-Gaussianity
- Multi-scale DA across temporal and spatial
- Dealing with massive increases in observations

Representation of Model Uncertainty



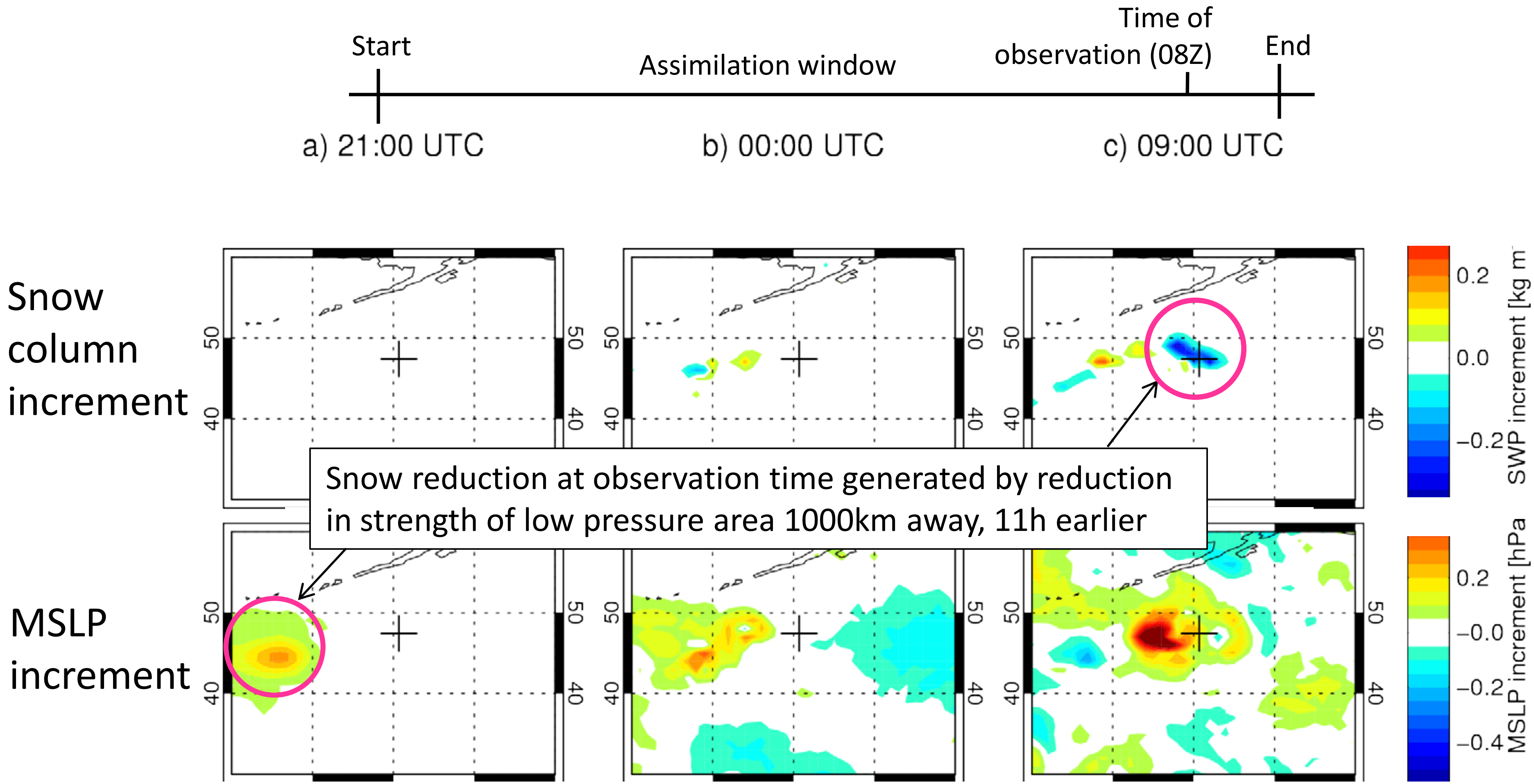
- **Use of ensemble information for B**
  - Train (full rank) B operators from ensemble of the day: 3D-4DVar
  - Localized Ensemble: 3D/4DEnVar, EnKF, LETKF
  - Hybrid: Hybrid-diag, NICAS (native)
- **DA algorithms to update ensemble**
  - Localized algorithms: EnKF, LETKF, MLEF
  - Global algorithms: Ensemble of DAs (EDA), EVIL, Randomization
- **Systematic model error**
  - Stochastic physics: SPPT/SHUM, SKEB
  - Weak constraint 4DVar: Q covariance matrix

Grand Science Challenges for DA

- Coupled DA across the Earth System
- Representation of model uncertainty
- Dealing with non-linearity and non-Gaussianity
- Multi-scale DA across temporal and spatial
- Dealing with massive increases in observations

Non-linearity and non-Gaussianity

Algorithms: Particle filters, 4DVar with outer-loops (weak NL)

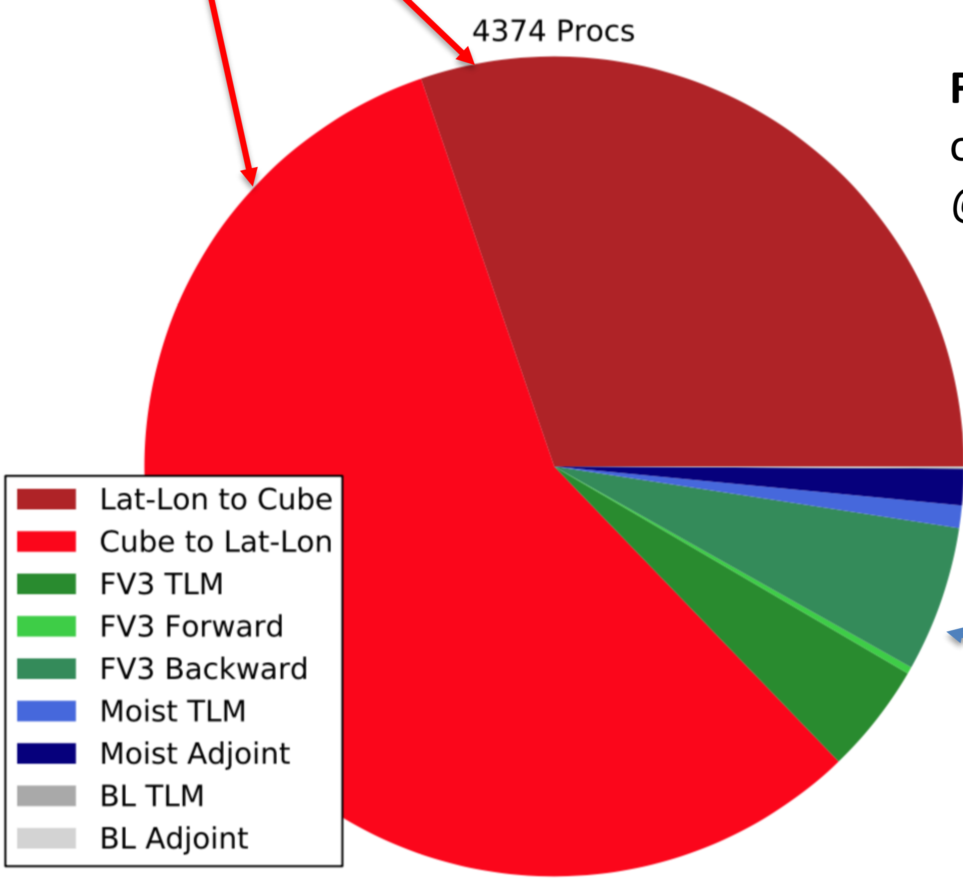


Frontal cloud and precipitation – 190 GHz

Source: Geer (2015)

Hybrid 4DVar is ambitious, yet achievable

Necessary for GSI, but will disappear in JEDI (native grid)



FV3 GEOS TLM/ADM called from DA on 4374 cores for 70 iterations @50km resolution

Wallclock time: 17.78 min
(10min for ECMWF oper resolution)

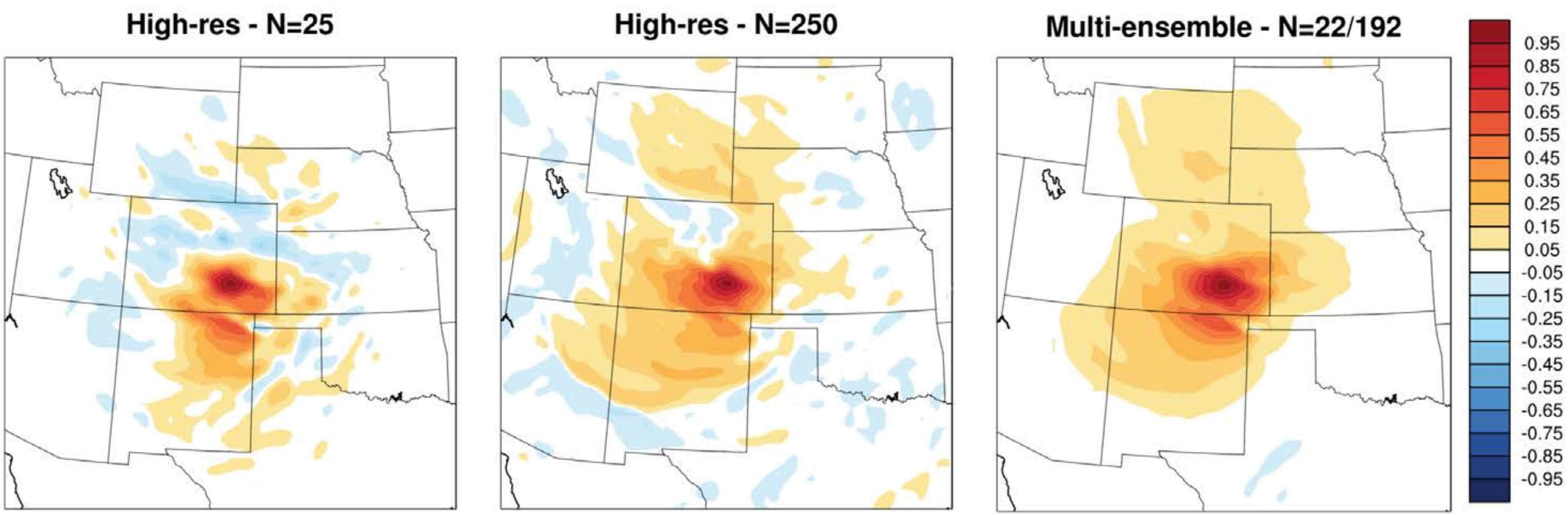
Source: Dan Holdaway (GMAO)

Grand Science Challenges for DA

- Coupled DA across the Earth System
- Representation of model uncertainty
- Dealing with non-linearity and non-Gaussianity
- Multi-scale DA across temporal and spatial
- Dealing with massive increases in observations

Multiscale DA across temporal and spatial
from global to convective-allowing resolution

- Balance particularly challenging at convective scale
- Displacement DA can reduce non Gaussianity
- Hydrometeors at smaller spatiotemporal scales
- RUC with overlapping windows. Continuous integration
- Across scales? Wave-band localization, multigrid DA, multi-res. ensemble



Source: Ménétrier and Auligné

Grand Science Challenges for DA

- Coupled DA across the Earth System
- Representation of model uncertainty
- Dealing with non-linearity and non-Gaussianity
- Multi-scale DA across temporal and spatial
- Dealing with massive increases in observations

Dealing …

…parti

ng

ime

selection, blacklisting

or superobbing

and detection

nal) Quality Control

(Most) curr

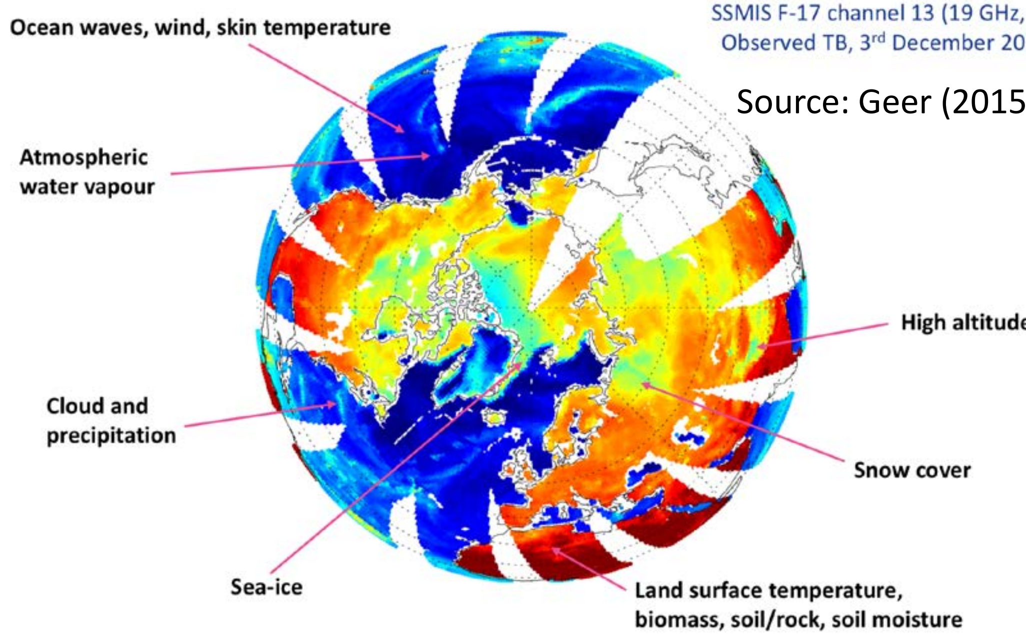
- few obs.
- slow evo

May 2002):
d ~0.05%

- Today: (satelli
- Tomorr

Obs. Processing in the Big Data era

- Improved Radiative Transfer Model (CRTM)
  - Code efficiency, scalability
  - Science: all-sky, all-surface, multiple sub-systems



SSMIS F-17 channel 13 (19 GHz, v) Observed TB, 3rd December 2014

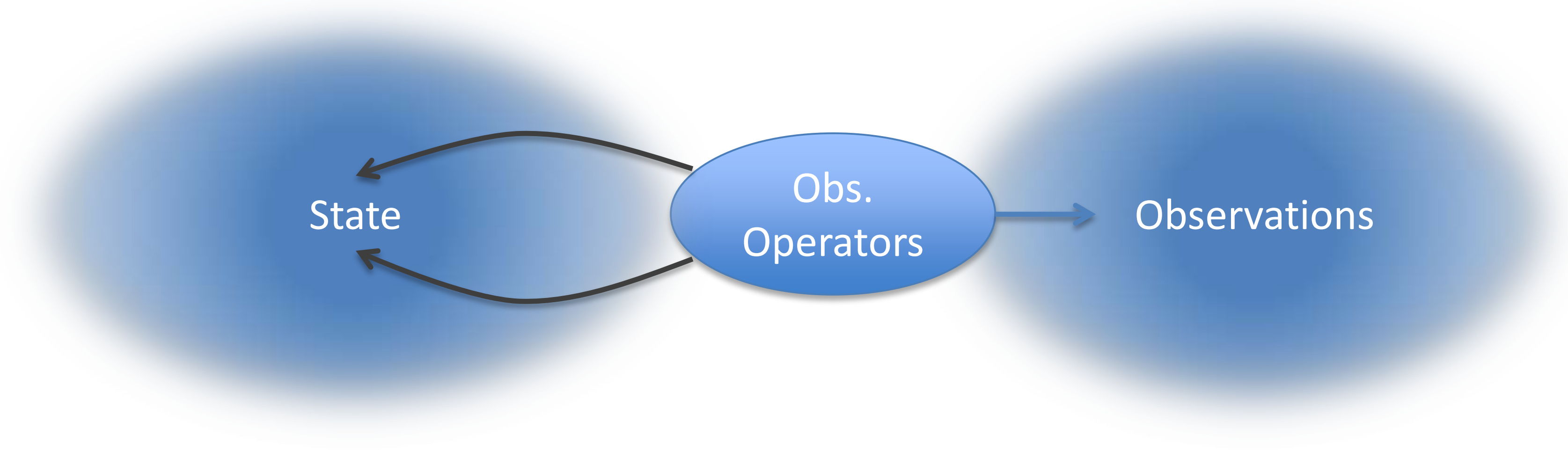
Source: Geer (2015)

- Different obs. selections in ensemble of DAs, and overlapping windows

- Fast dev./oper cycle + expert adaptive systems for obs. processing (VarBC, VarQC)

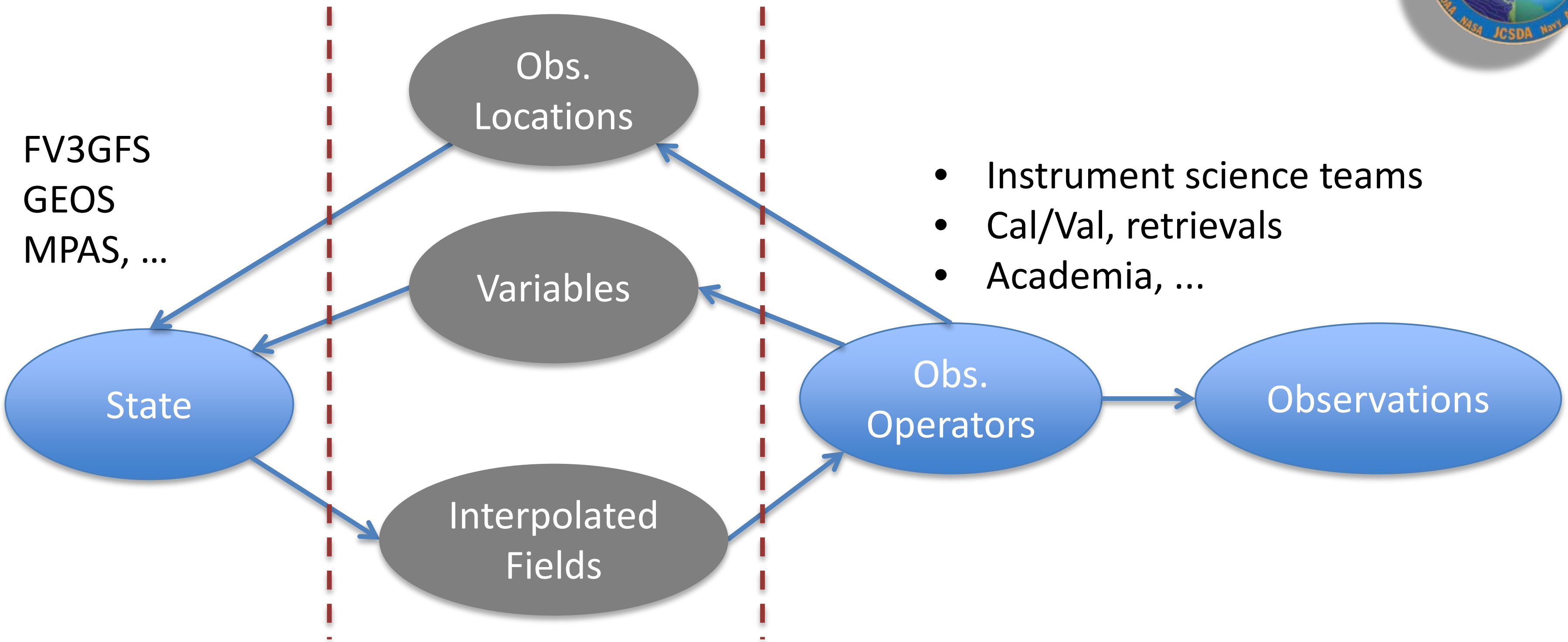
- JEDI-UFO, the ‘App Store’ of Observation Operators

Observation Operators



- In most existing systems, observation operators directly access state/model data
- Observation operators, and as a result DA systems, are very model specific

UFO: the interface advantage



- JEDI/UFO introduces standard interfaces between the model and observation worlds
- Observation operators are independent of the model and can easily be shared, exchanged, compared

STATUS: Interpolation from model native grid to observation locations

Prototype for radiosonde T, radiances (AMSUA), sea-ice fraction, [aircraft]

Interface for Observation Data Access (IODA)

Interface to isolate science code from data storage

Three levels:
- Long term storage (historic database)
- Files on disk (one DA cycle)
- In memory handling of observations (hardware specific?)

Two environments:
- Plotting, analyzing, verifying on workstation
- DA and other HPC applications (MPI, threads, GPUs…)

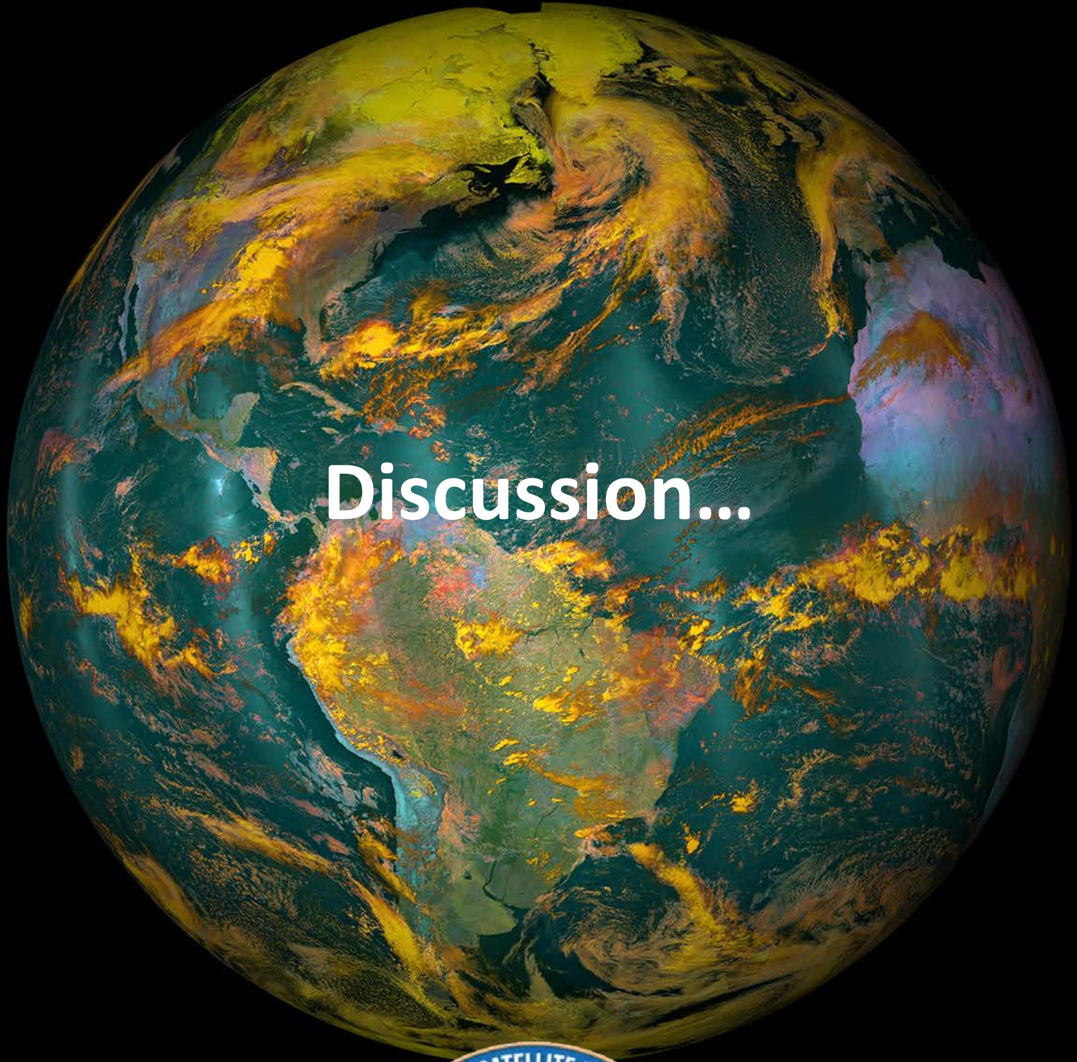
Can one interface do it all? Can one implementation do it all?

Elements for discussion on DA Strategy

1. DA = critical component of the UFS (incl. verif., postproc., model dev.)
2. Hybrid 4DVar ambitious but achievable within 3-5 years
3. Major opportunity in massive number of untapped observations
4. Unified DA to explore science challenges of Coupled DA
5. JEDI = game changer to speed up disruptive developments

For info., ECMWF Strategy for Data Assimilation (Bonavita et al., 2017)

1. The OOPS project operational implementation
2. Ensemble of DAs (EDA): lower cost but equally effective
3. 4DVar: weak constraint, quasi-continuous 4D, saddle point, TLM/ADM
4. Coupled DA: atmos., ocean, SST, sea-ice, [waves, land surface]
5. More observations: different subsets for EDA & overlapping windows



Discussion…

Thanks to all involved in JEDI (so far)

Credits: Amal El Akkraoui, Anna Shlyaeva, Ben Johnson, Ben Ruston, Benjamin Ménétrier, BJ Jung, Bob Oehmke, Bryan Flynt, Bryan Karpowicz, Clara Draper, Dan Holdaway, Dom Heinzeller, François Vandenberghe, Gael Descombes, Guillaume Vernières, Jeff Whitaker, Jing Guo, John Michalakes, Julie Schramm, Mariusz Pagowski, Mark Miesch, Mark Potts, Ming Hu, Rahul Mahajan, Ricardo Todling, Scott Gregory, Soyoung Ha, Steven Herbener, Steve Sandbach, Tom Auligné, Will McCarty, Xin Zhang, Yannick Trémolet. The JEDI abstract layer is based on OOPS (ECMWF).

Model Interface

Initial implementation done for FV3GFS, GEOS5, MPAS, LFriC, CICE5, [MOM6], [WRF]

Status for FV3:

- Geometry: JEDI grid and parallel distribution built from stripped down version of main FV3 construct: same I/O routines and data structures (non-intrusive, lightweight, efficient code)
- Fields: read and write model restarts using FMS I/O (will be extend to use NEMS/MAPL/ESMF)
- Linear Algebra: all basic operations
- Interpolation: from FV3 native grid to observation locations (and B unstructured mesh)
- Model Interface: Nonlinear, tangent-linear, and adjoint interfaces installed. Trajectory implemented. Passed internal tests.

Plans:

- Interfaces: complete work for existing models. Extend to NEPTUNE, NAVGEM, WW-III
- Trajectory : read files for FSOI and 4DVar without complete model interface
- Testing: evaluate GEOS5 TLM/ADM from FV3GFS trajectory. Extend test suite.

Observations

Status:

- IODA: read and write observation values, locations, metadata for UFO, O-B, O-A from extended version of GMAO NetCDF diag files
- UFO: model independent operators for Radiosonde temperature, Aircraft (under testing), and satellite radiances with interface to CRTM (tested with AMSUA only).
- Scalability: testing multiple strategies for observation parallel distribution

Plans:

- IODA: develop API to access observation data
- UFO: clean current prototype. Extend observation locations and GeoVals to complex operators (e.g. GPSRO). Add new obs types: more radiances, conventional, GPSRO, AMVs, AOD, [radar].
- Quality Control: implement set of generic QC filters
- Observation Error: initial diagonal R, then correlated errors
- Bias Correction: develop VarBC.

Background Error Covariances

Status:

- NICAS: generic horizontal and vertical auto-correlation operator, applicable to ANY model grid. Dirac test with [FV3GFS, GEOS5], MPAS, WRF, CICE5, MOM5, LFrIC.
- Hybrid-diag: calculation of optimal horiz.+vert. localization length-scales and hybridization coefficients
- EnVar: prototype full-ensemble B matrix operator

Plans:

- NICAS: clean and complete current prototype. Test results and code performance on HPC
- Static: develop climatological B with flexible CVT
- Hybrid: implement optimal hybrid coefficients.
- State Augmentation: include bias parameters with Hessian preconditioning
- Displacement: develop generic field alignment transformation

Solver

Status:

- OOPS: currently 12 interchangeable options of variational solvers (primal space, dual space, saddle-point) with various preconditioning options.

Plans:

- 3DEnVar: testing and evaluation of prototype
- 4DEnVar: investigate time-evolving localization
- Hybrid-4DVar: build single executable with FV3GFS NL model and GOES5 TLM/ADM
- Ensemble: implement EnKF, LETKF, etc.
- FSOI: develop ensemble- and adjoint-based Forecast Sensitivity – Observation Impact

Infrastructure

Status:

- Governance: finalizing document detailing the project, partners, governance, decision-making, roles and responsibilities,
- Code Management: defined working practices using modern collaborative development tools. Started series of successful code sprints
- Build System: Cmake-based build from multiple code repositories on Github.com
- Access and Portability: use of containers (Docker -> Singularity)
- Testing: implementing suite of tests in Continuous Integration environment

Plans:

- Testing: toolbox to create tests, and environment to run tests (mutualized with MPAS/WRF)
- Coding Standards: automatic testing of C++, Python, [and Fortran]
- Code Management: implement Agile procedures for code review and issue tracking
- Community Support: complete documentation, on-line tutorial, Helpdesk, JEDI Academy

Final comments

JEDI is for **scientific exploration** and **operational forecasting** … and O2R2O

- Model agnostic: ground up design to assimilate on native grid of any model, avoiding costly interpolations
- Scalability with increasing number of processors
- Does not impose one specific DA methodology or algorithm. Range of solvers including with time parallelization.
- Highly collaborative and easy to work together, due to good abstraction, separation of concerns, and interfaces
- Encourages implementation of model independent observation operators (community “operator store”)
- Will adapt well to strongly coupled DA

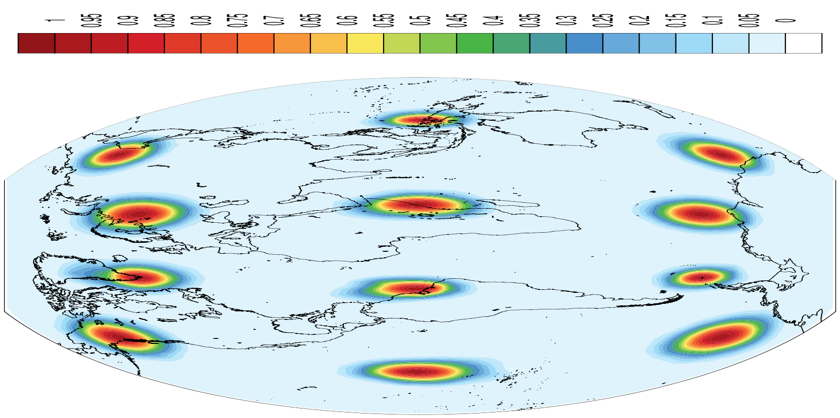
Code Prototyping

'B Matrix' Bootcamp – 01-21 Aug 2017 – Boulder, CO

Scope: Design, develop, and test a generic prototype software for modeling background error covariances in research and operations.

Participation: JCSDA, NCAR, GMAO, OAR, EMC, Météo-France, Met Office

Models:
- Lorenz, QG
- FV3-based GFS, GEOS5
- WRF
- MPAS
- MOM+CICE



JEDI Code Repositories on Github.com

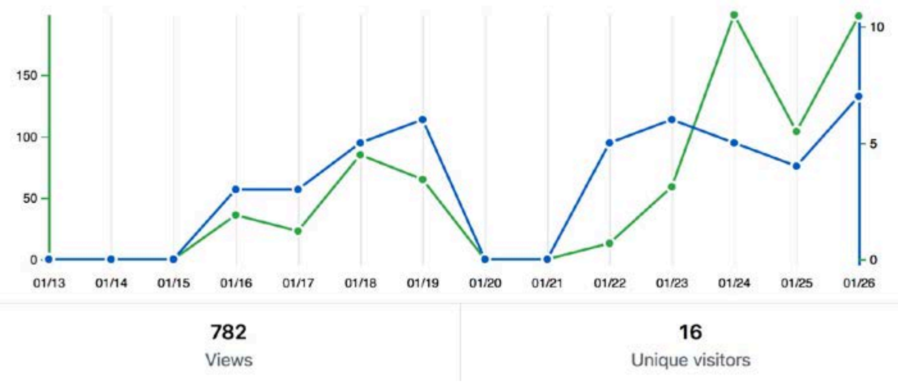
- **Oops** Object Oriented Prediction System
- **Ioda** Interface for Observation Data Access
- **Ufo** Unified Forward Operator
- **Crtm** Community Radiative Transfer Model

- **Fv3gfs** FV3GFS interfaces for OOPS/JEDI
- **Fv3gmao** fv3 code for GEOS5 TLM/ADM
- **Fms**

- **Soca** MOM6 and CICE5 interfaces
- **Padawav** WWIII interfaces (TBD)
- **Mpas** MPAS interfaces for OOPS/JEDI
- **Lfric** LFRic interfaces for OOPS/JEDI

- **Singularity** Singularity container for JEDI

Traffic on OOPS Repository

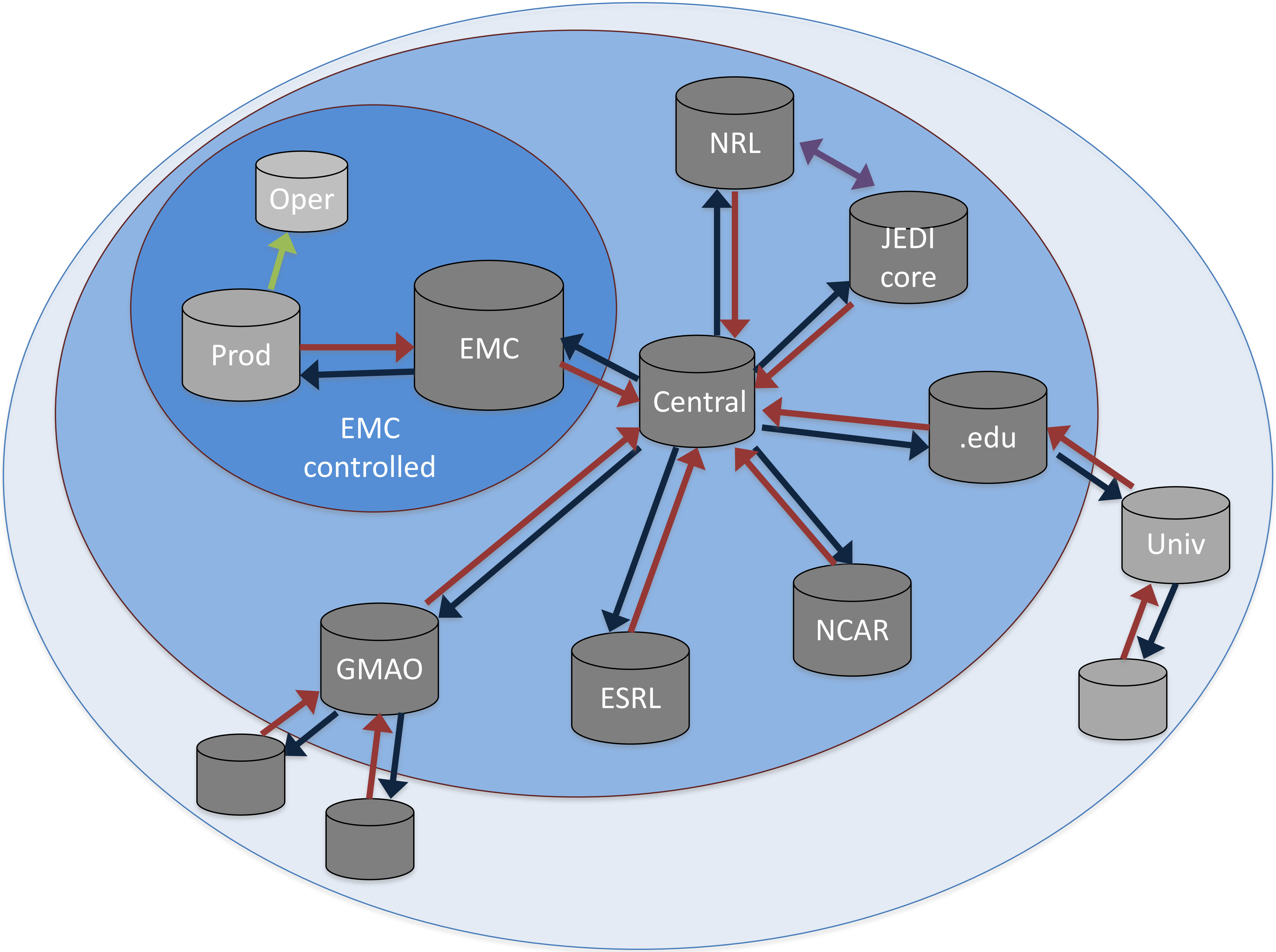


- **Ecbuild** Build system (cmake-based)
- **Eckit** C++ utilities for OOPS/JEDI
- **Fckit** Fortran utilities for OOPS/JEDI

- **fv3gfs-bundle**
- **mpas-bundle**
- **lfric-bundle**
- **jedi-bundle**

Objectives of this Presentation

- Review major science challenges for DA in next decade
- Associated software/techniques of interest
- Provide elements to discuss priorities for short-term (next 3y.) and longer term (5-10y.)



NRL
Oper
JEDI
core
Prod
EMC
Central
.edu
EMC
controlled
Univ
GMAO
ESRL
NCAR

JEDI Containers: Build and Experiments

GitHub version-controls the Dockerfile

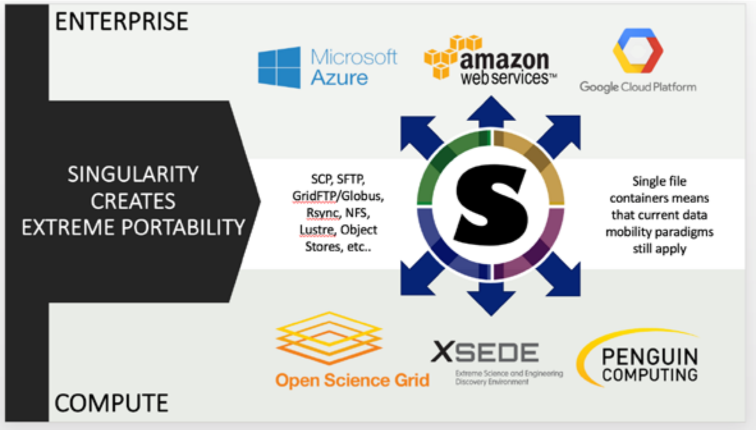
Download, build, and install the following libraries :

- open-mpi v2.1.0
- zlib v1.2.11
- szip v2.1.1
- jpeg v9b
- png v1.4.19
- jasper v1.900.2
- hdf5 v1.8.17
- freetype v2.5.5
- netcdf-c v4.4.11
- netcdf-fortran v4.4.4
- lapack v3.7.0
- parallel-netcdf v1.8.1
- xerces-c v3.1.4
- esmf v7.0.0
- udunites-2 v2.2.24
- nco v4.6.6
- grib_api v1.21.0
- cdo v1.8.2
- pio v1.7.1

The major NCEP libraries are also installed at :

- /nwprod/lib/bacio/v2.0.1/libbacio_v2.0.1_4.a
- /nwprod/lib/bacio/v2.0.1/libbacio_v2.0.1_8.a
- /nwprod/lib/ip/v2.0.0/libip_v2.0.0_4.a
- /nwprod/lib/ip/v2.0.0/libip_v2.0.0_8.a
- /nwprod/lib/ip/v2.0.0/libip_v2.0.0_d.a
- /nwprod/lib/sigio/v2.0.1/lib/libsigio_v2.0.1_4.a
- /nwprod/lib/sigio/v2.0.1/libsigio_v2.0.1_4.a
- /nwprod/lib/sp/v2.0.2/libsp_v2.0.2_4.a
- /nwprod/lib/sp/v2.0.2/libsp_v2.0.2_8.a
- /nwprod/lib/sp/v2.0.2/libsp_v2.0.2_d.a
- /nwprod/lib/w3emc/v2.2.0/libw3emc_v2.2.0_4.a
- /nwprod/lib/w3emc/v2.2.0/libw3emc_v2.2.0_8.a
- /nwprod/lib/w3emc/v2.2.0/libw3emc_v2.2.0_d.a
- /nwprod/lib/w3nco/v2.0.6/libw3nco_v2.0.6_4.a
- /nwprod/lib/w3nco/v2.0.6/libw3nco_v2.0.6_8.a
- /nwprod/lib/w3nco/v2.0.6/libw3nco_v2.0.6_d.a

Docker pulls the codes and scripts from GitHub



JEDI Project: Ambitious Timeline

Initial Design

- Jan 2017: Planning begins
- Feb 2017: Collect requirements
- Avr 2017: Unified DA Planning Meeting

Functional Prototype

YOU ARE HERE

- Aug 2017: 1st code sprint (B matrix)
- Nov 2017: 2nd code sprint (UFO)
- Feb 2018: Basic 3DEnVar prototype

Rapid Development

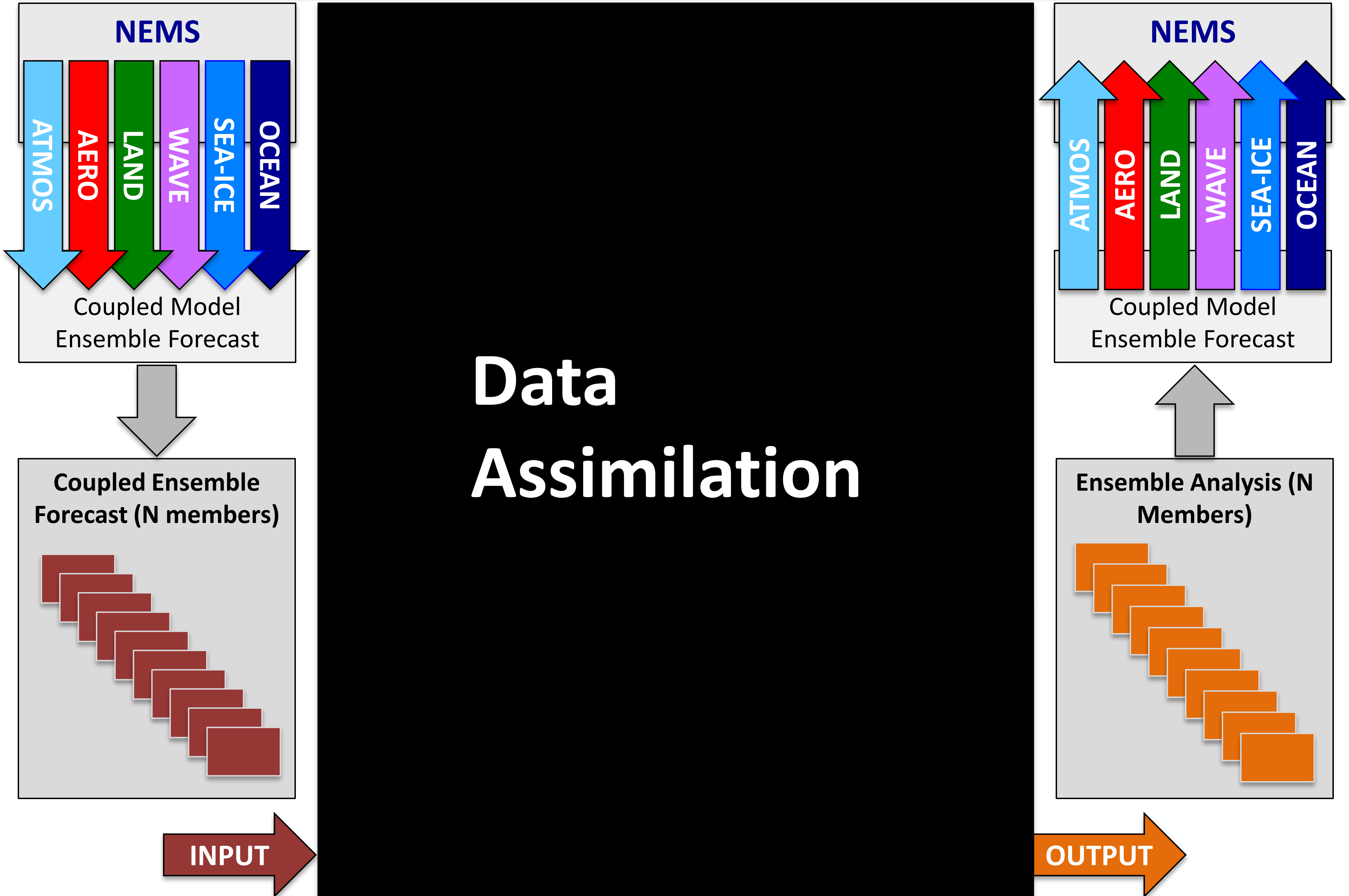
- Avr 2018: 4DEnVar prototype
- Jun 2018: Code training for Padawans
- Aug 2018: Hybrid 4DVar prototype
- Sep 2018: Prototype evaluation

Toward Operations

- Nov 2018: JEDI Academy
- Sep 2019: UFO ready
- Sep 2020: Solvers ready (Var + Ens.)

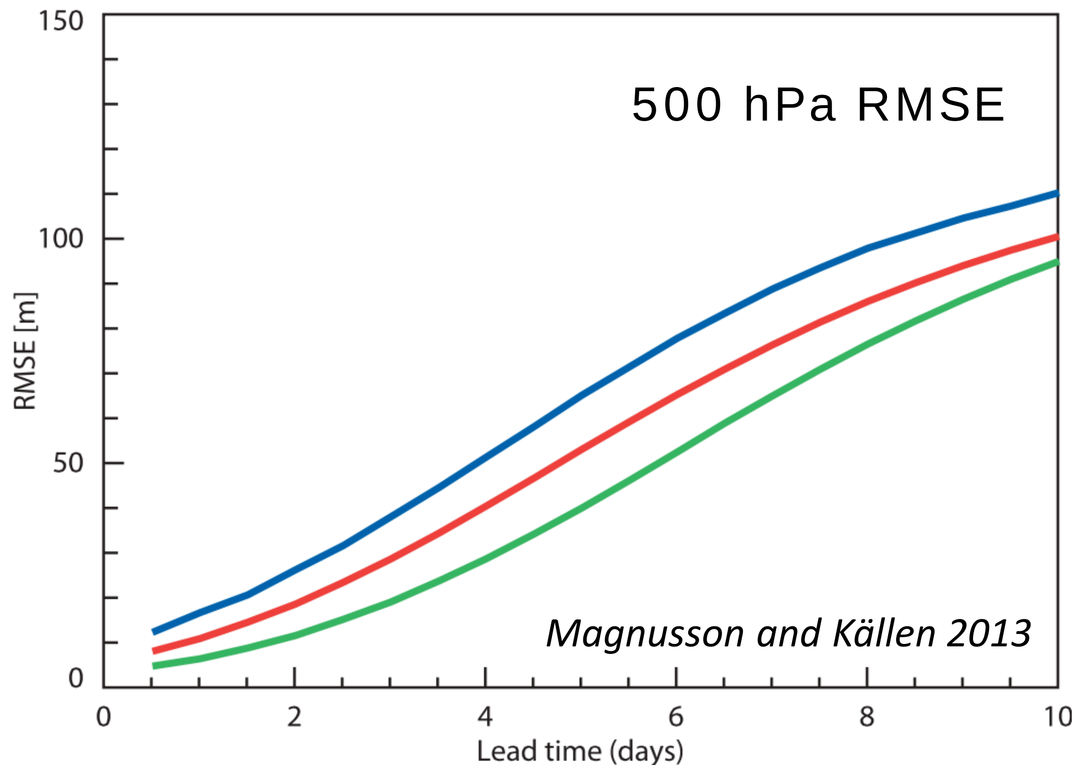
Project: 4 FTEs (Core) + 5 FTEs (In-Kind) - Lead: Yannick Trémolet (JCSDA)

NCEP Coupled Hybrid Data Assimilation and Forecast System



Adapted from Saha

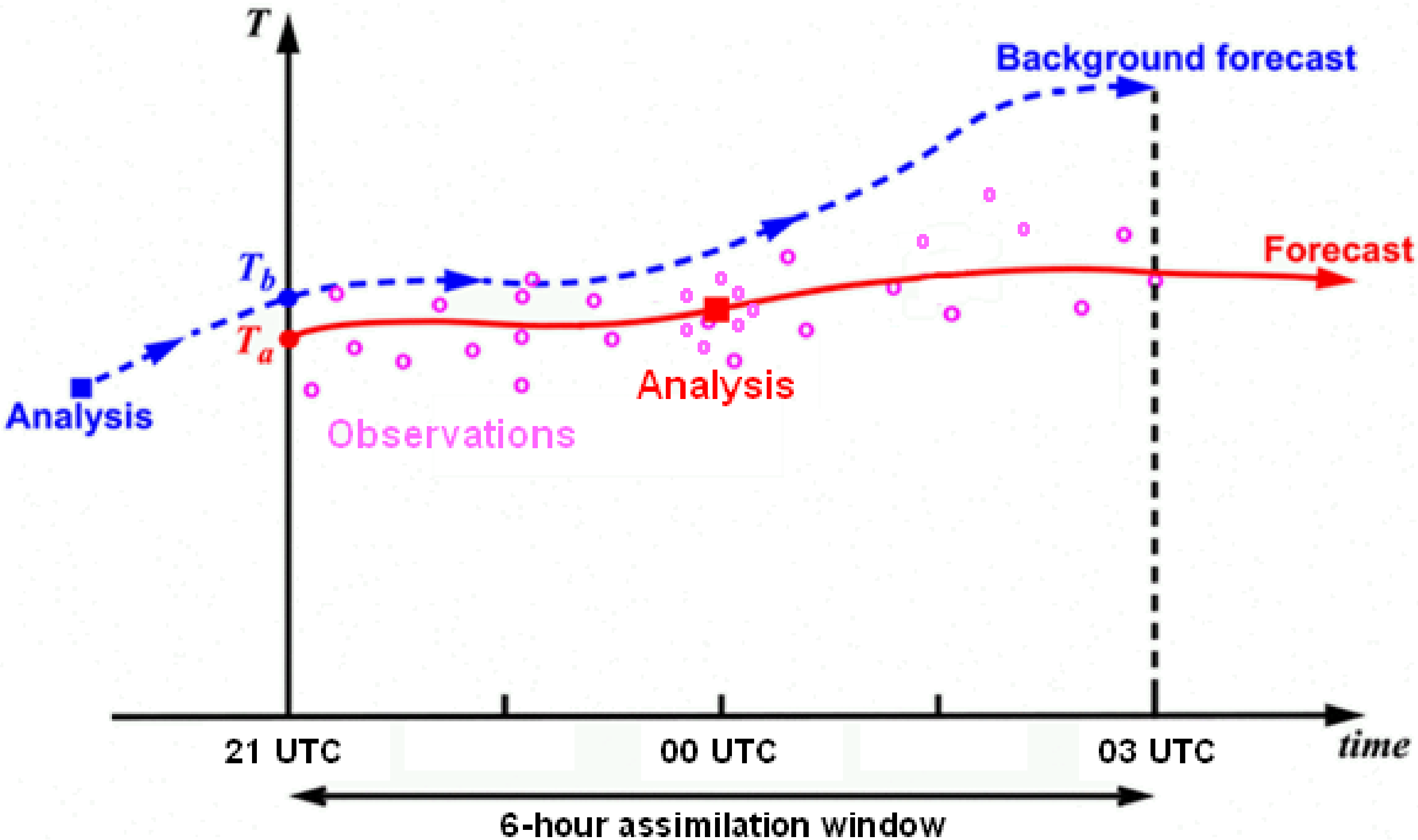
Intro: Data Assimilation matters… A LOT!



- Contributions to forecast error : 50% model error, 50% DA
- Requires balancing resources (FTEs, CPU, $$$)
- The best operational forecast system **REQUIRES** the best operational DA

Model and DA are fundamentally 4D

- Algorithms:
  - 4DVar with outer-loops for weak non-linearities
  - Particle filters



Source: ECMWF