The Joint Effort for Data assimilation Integration (JEDI)

Yannick Trémolet
Joint Center for Satellite Data Assimilation (JCSDA)

17 May 2021

Joint ECMWF/OceanPredict workshop on Advances in Ocean Data Assimilation

JEDI Objectives

The Joint Effort for Data assimilation Integration (JEDI) is a collaborative development between JCSDA partners

Develop a unified data assimilation system:
- From toy models to Earth system coupled models
- Unified observation (forward) operators (UFO)
- For research and operations (including O2R2O)
- Share as much as possible without imposing one approach

Design driven by scientific vision for data assimilation
- Not trying to salvage old codes at all cost

JEDI is a Joint Effort

JEDI core-team: A. Shlyaeva, B. Ménétrier, C. Gibert, C. Gas, D. Holdaway, E. Lingerfelt, M. Miesch, M. Olah, M. Abdi-Oskouei, R. Grubin, R. Honeyager, S. Herbener, S. Vahl, Y. Trémolet

JEDI contributors (2020): A. Weinbren, B. Johnson, BJ Jung, C. Pelley, C. Harrop, C. Martin, C. Thomas, D. Davies, D. Simonin, E. Liu, F. Diniz, F. Vandenberghe, G. Vernières, G. Thompson, H. Zhang, H. Ebrahimi, H. Kershaw, H. Shao, I. Genkova, J. McCreight, J. Rosinski, J. Guerrette, J. Ban, M. Wlasak, M. Pagowski, M. Cooke, N. Bowler, O. Lomax, R. Mahajan, R. Todling, S. King, S. Frolov, S. Sandbach, T. Sluka, V. Buchard, W. Śmigaj, X. Zhang, and more…

JEDI collaborators: C. Snyder, D. Kleist, D. Dee, N. Baker, R. Gelaro, T. Auligné, …

Representing: JCSDA, NOAA/EMC, NOAA/ESRL, NASA/GMAO, NRL, USAF, NCAR, UKMO

And more than 200 padawans who attended five JEDI Academies

October 2020: First JEDI Release

Releases

JEDI-FV3 RELEASE

CRTM RELEASE

JEDI-FV3 Release

The JEDI-FV3 release contains everything needed to emulate a simplified version of the JCSDA near-real-time (NRT) applications websites with JEDI-GFS and JEDI-GEOS. This includes interfaces for processing a wide variety of observational data and a library of observation operators for conventional, satellite radiance and GNSS radio-occultation measurements that can be run from both FV3-based GFS and GEOS models.

Version	Download	Quick Start - Tutorials	Support	Date
1.0.0	Code	Run JEDI-FV3 in a Container	Documentation	2020-10-28
Release Notes		Simulate Observations Like JCSDA NRT Application	Forums	
		Build and Test JEDI-FV3		

Next release coming soon!

Available at `https://www.jcsda.org/jedi-fv3-release`

JEDI Working Practices

- **Project methodology inspired by Agile/SCRUM**
  - Adapted to distributed teams and part time members
  - Work in small manageable increments with constant feedback
- **Collaborative environment**
  - Easy access to up-to-date source code (github)
  - Easy exchange of information (zenhub)
  - Weekly meetings by video (40 to 60 developers from at least 7 organizations)
  - Code sprints (8-10 developers working together on a specific topic, virtual in 2020/21)
- **Enforce software quality**
  - Automated testing: correctness, coding norms, efficiency
  - Code reviews (all developers actively involved)
- **Portability is part of the development**
  - Automatic tests with several compilers
  - JEDI available in containers (docker, singularity, charliecloud)
- **Modern programming technologies** (generic, object oriented)

JEDI is active:
> 200 PR merged / month on average

Y. Trémolet, JCSDA

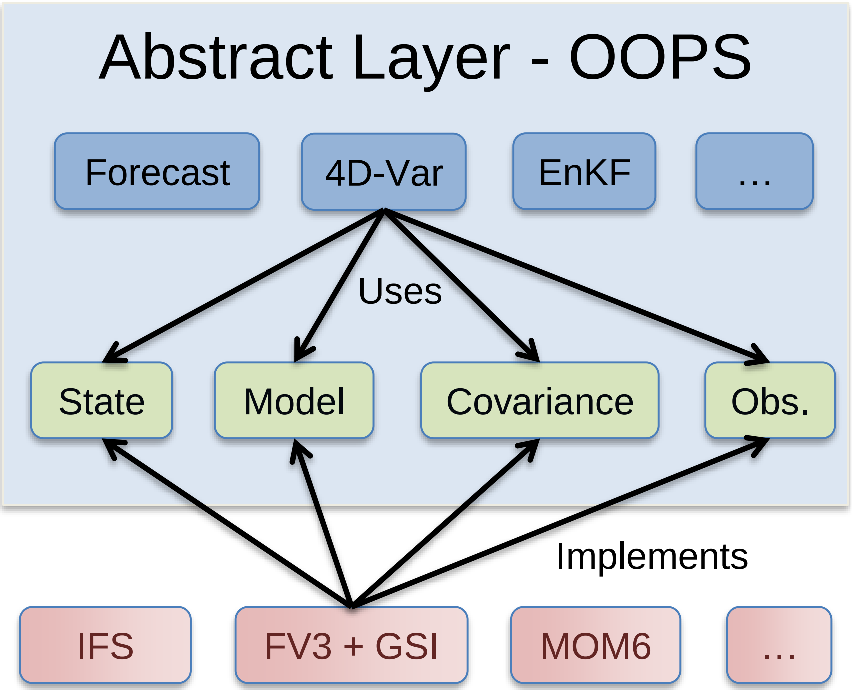
Object Oriented Prediction System (OOPS)

Generic, portable, model-agnostic DA system

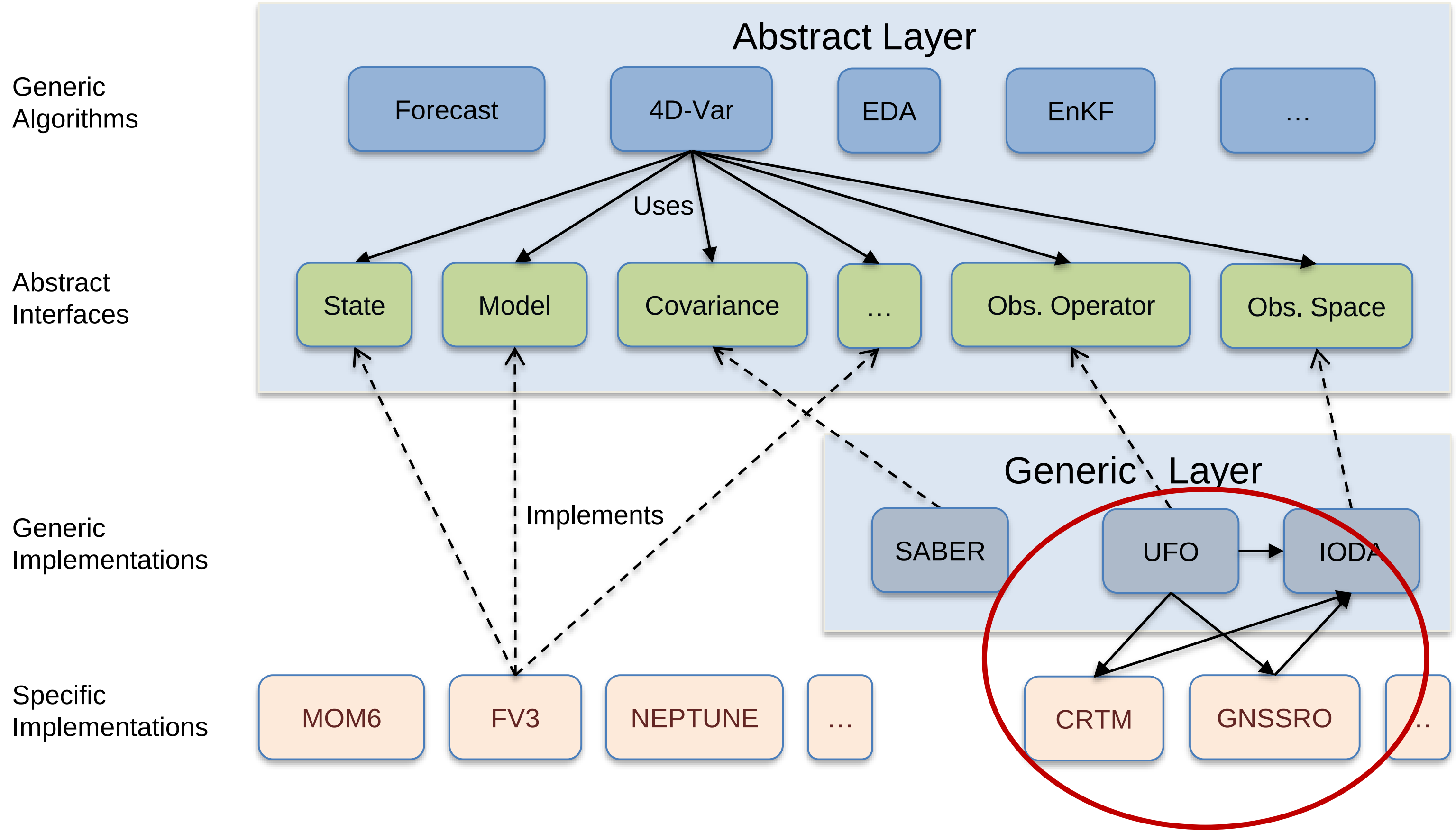
Use **object-oriented** and **generic** programing

Each model implements pre-defined abstract interfaces

Separation of concerns



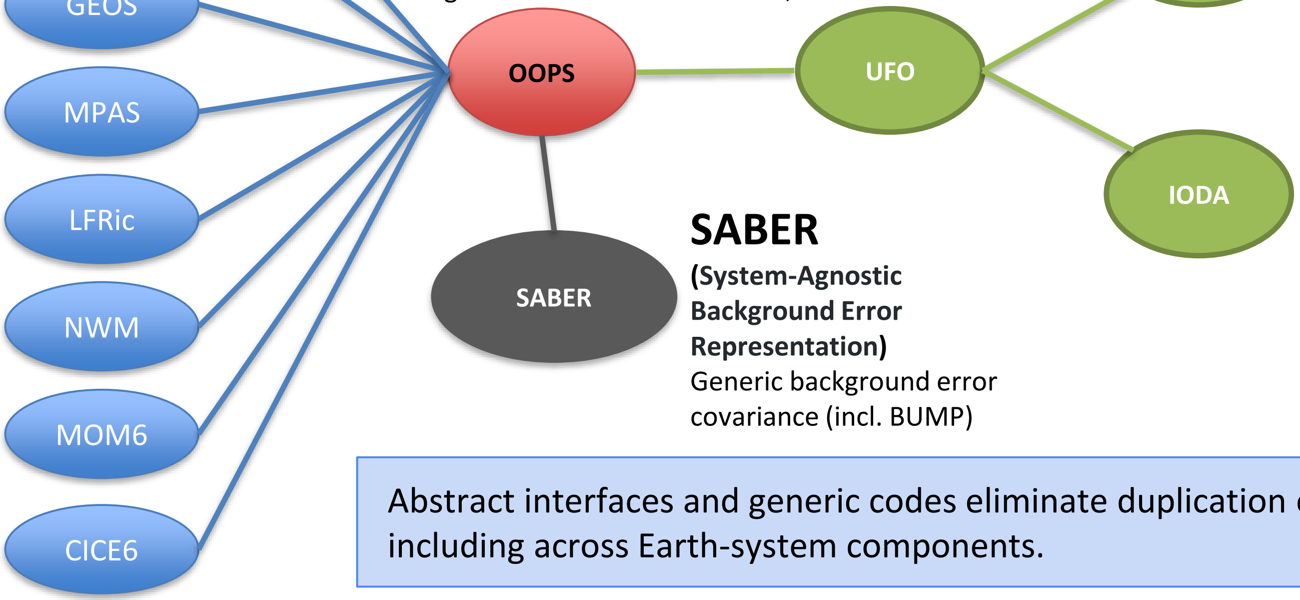
JEDI: Abstraction and Genericity



Abstract, model-agnostic DA system

OOPS is complemented by generic (shared) components.

JEDI Structure



OOPS

(Object Oriented Prediction System)
State-of-the-art data assimilation generic algorithms

UFO

(Unified Forward Operator)
The 'app-store' of model-agnostic observation operators, QC, bias correction

CRTM

(Community Radiative Transfer Model)
Accurately and efficiently simulate satellite radiances

IODA

(Interface for Observation Data Access)
Performs all the I/O of the observations

SABER

(System-Agnostic Background Error Representation)
Generic background error covariance (incl. BUMP)

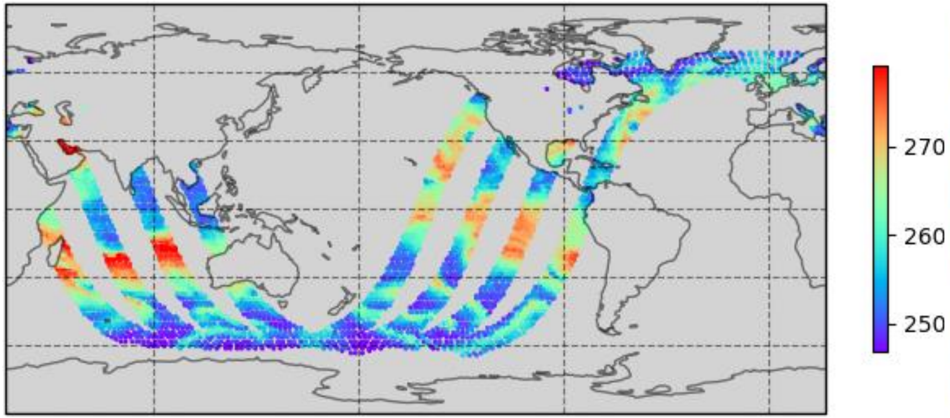
Abstract interfaces and generic codes eliminate duplication of code and effort, including across Earth-system components.

GEOS in-core Data Assimilation

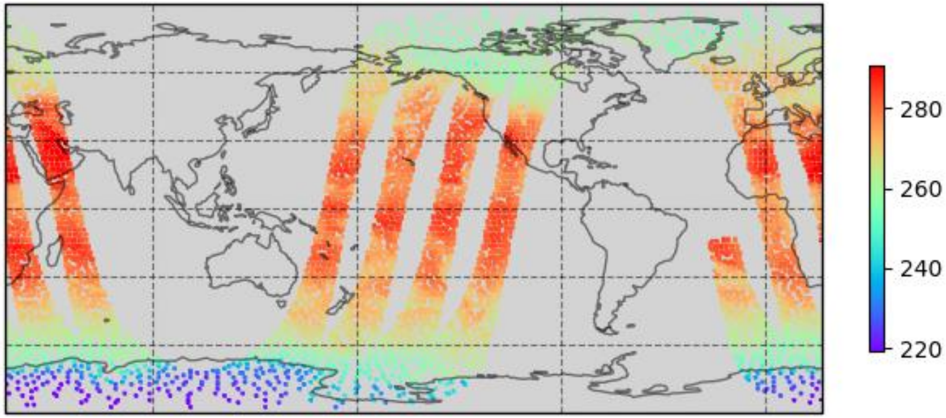
- Model runs in-core with FV3-JEDI.
- States can be exchanged from the model and be passed back to the model.
- Model can be rewound and used in outer loops.

Figures below show the in-core four-dimensional observation simulations for GMAO's production system. This is part of a continuously running near real time workflow.

GPM GMI Channel 6 h(x)
2020-10-17T09:00:00 - 2020-10-17T15:00:00



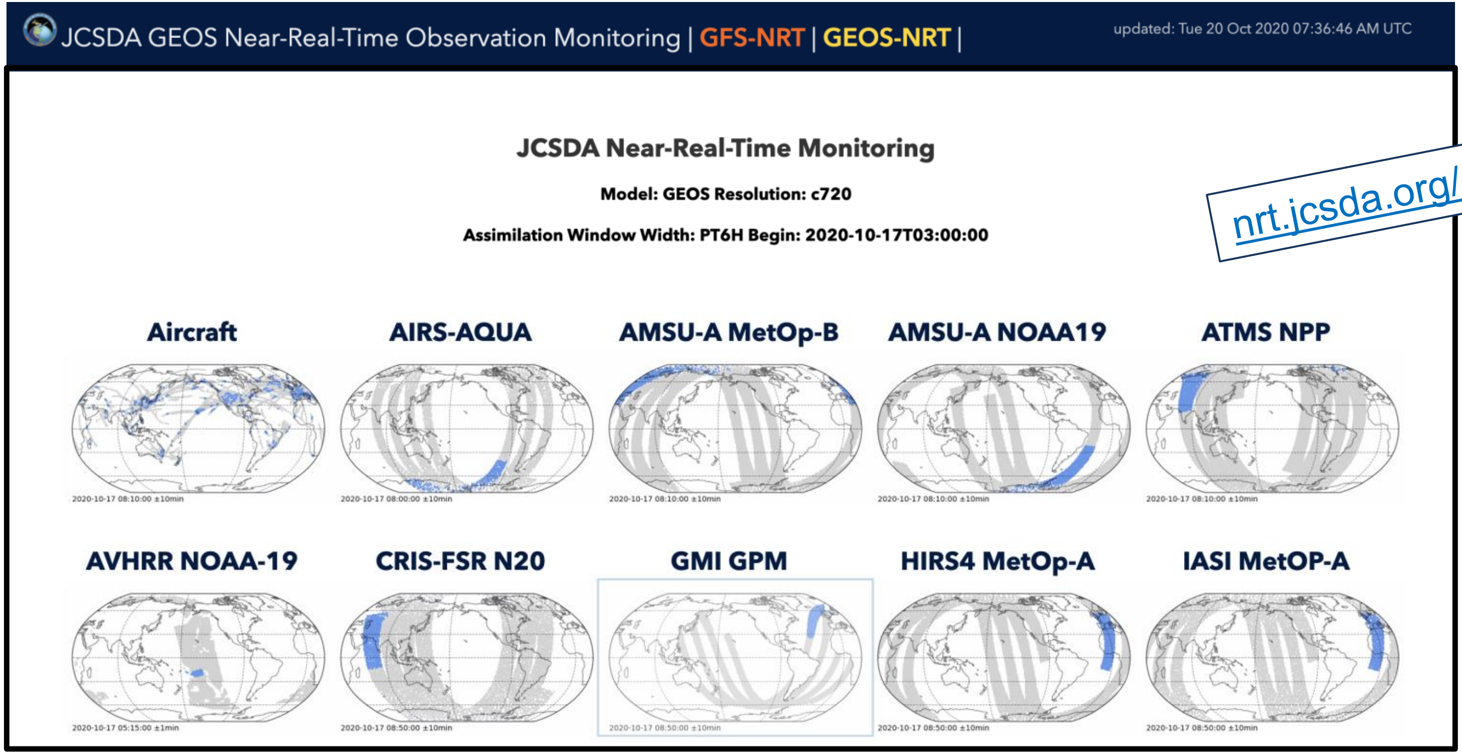
AQUA-AIRS Channel 1371 h(x)
2020-10-17T09:00:00 - 2020-10-17T15:00:00



Y. Trémolet, JCSDA

GEOS NRT observation simulation

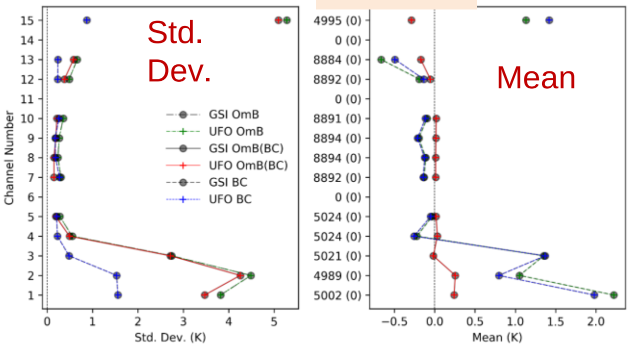
Continuously running in-core 4D H(x) using restarts and a subset of observations from GMAO's production system. Results are posted in near real time to the JCSDA website:



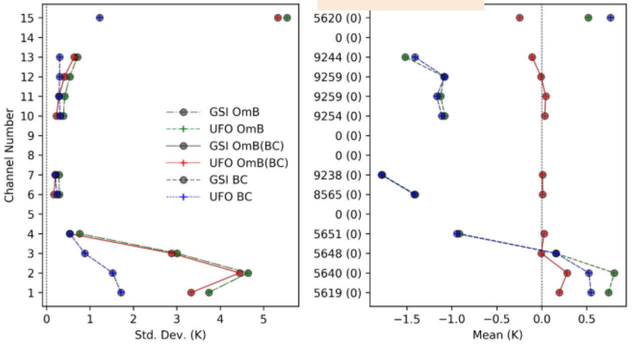
nrt.jcsda.org/geos

UFO Validation against GSI (ASMU-A)

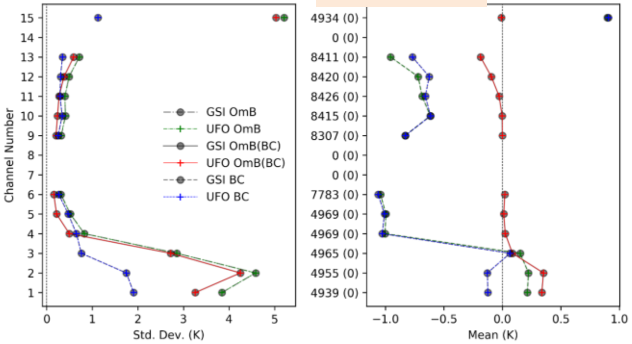
NOAA-15



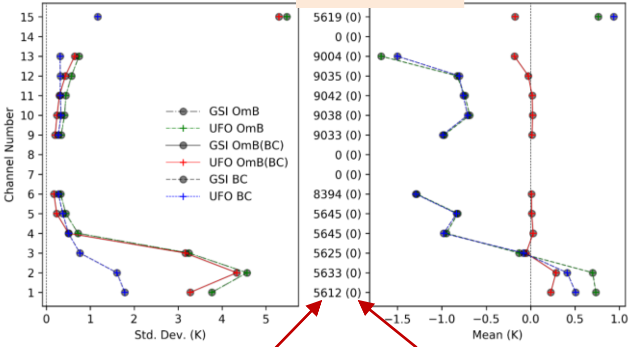
NOAA-18



NOAA-19



METOP-A



GDAS operational platforms/sensors (~GFSv15.3) were tested by JCSDA for one cycle using EMC provided obs and geovals

Code/configurations are in github repo (`develop`)

Figures from JCSDA OBS team (H. Shao *et al.*)

GSI and UFO values on top of each other

Obs # passed GSI QC

difference between JEDI and GSI Qced observations

JEDI-GODAS: Observations

Generic Quality Control: No Coding!

```
obs filters:
- filter: Domain Check
  where:
  - variable: {name: sea_area_fraction@GeoVaLs}
    minvalue: 0.9
- filter: Domain Check
  where:
  - variable: { name: sea_surface_temperature@GeoVaLs}
    minvalue: 5.0
- filter: Background Check
  absolute threshold: 0.2
- filter: Bounds Check
  minvalue: -2.0
  maxvalue: 2.0
  action:
    name: assign error
    error function:
      name: LinearCombination@ObsFunction
      options:
        variables: [mesoscale_representation_error@GeoVaLs,
                    obs_absolute_dynamic_topography@ObsError]
        coefs: [0.1,
                0.01]
- filter: BlackList
  where:
  - variable:
      name: latitude@MetaData
    minvalue: -65
    maxvalue: -30
  - variable:
      name: longitude@MetaData
    minvalue: -125
    maxvalue: -90
- filter: BlackList
  where:
```

Land mask

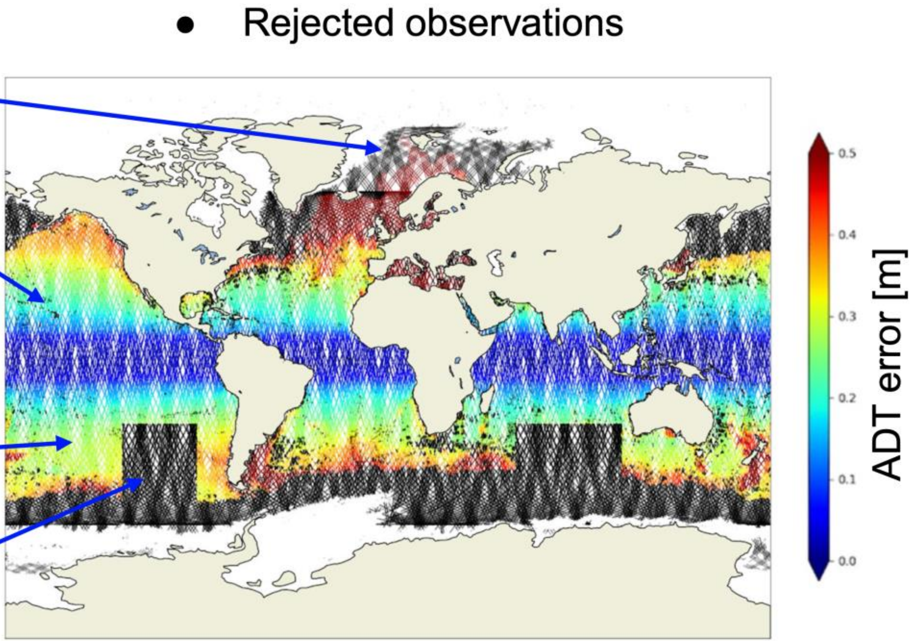
Reject ADT obs if SST<5°C

Reject ADT obs if |Obs-Bkg|<0.2 m

Reject ADT obs outside of [-2.0m, 2.0m]

Assign ADT obs error [m]

Reject ADT obs in specific region

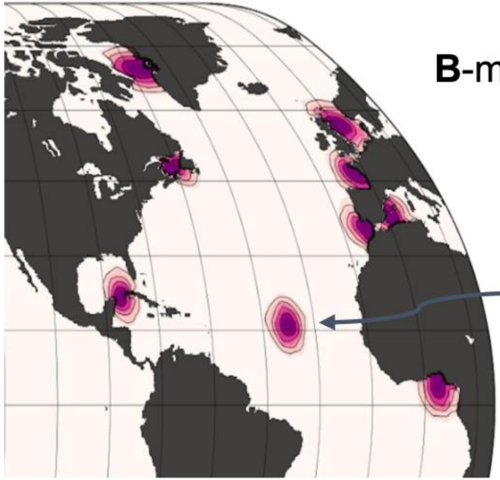


Figures from SOCA team

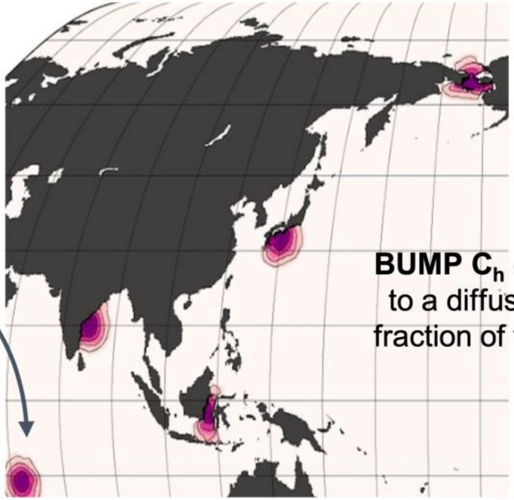
JEDI-GODAS: Covariance Modeling

$$B = K F_h^{\frac{1}{2}} D_p D_f C_v^{\frac{1}{2}} C_h C_v^{\frac{1}{2}T} D_f D_p F_h^{\frac{1}{2}T} K^T$$

B-matrix on **U**nstructured **M**esh **P**ackage (BUMP, Benjamin Menetrier)



Impulse response of $\mathbf{C_h}$ to dirac delta functions

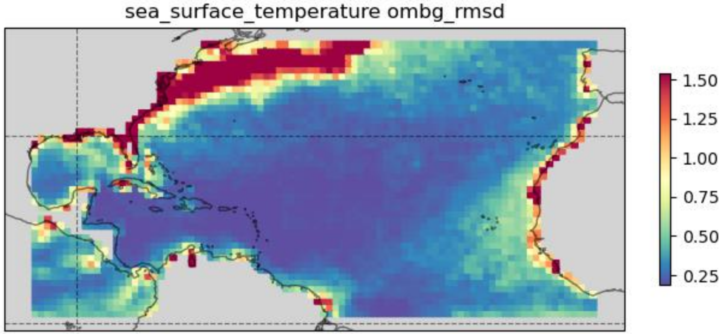


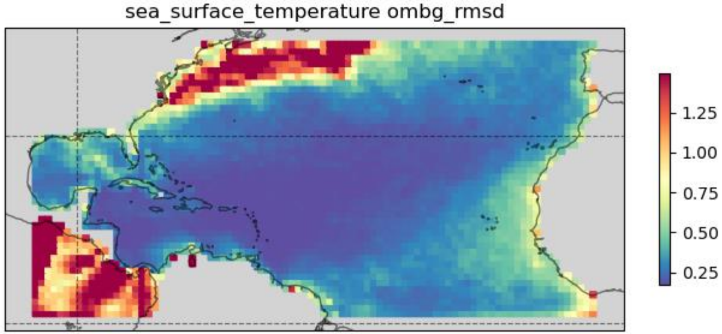
BUMP $\mathbf{C_h}$ operator is similar to a diffusion operator at a fraction of the computational cost

More in talk later in this session

Ocean Global and Regional DA

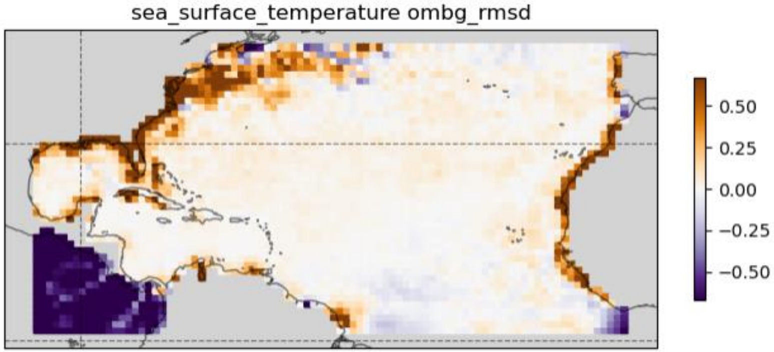
SST HAT10 vs Global UFS (1 degree)





2015-01-01 00Z to 2015-12-31 23Z

JCSDA ioda-plots



2015-01-01 00Z to 2015-12-31 23Z

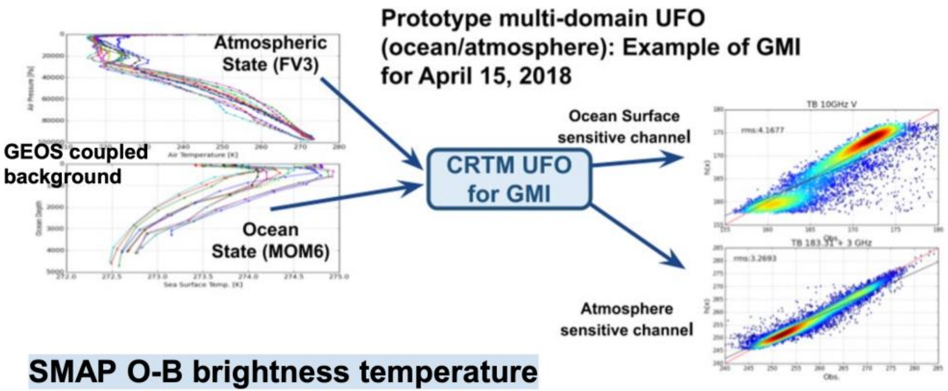
JCSDA ioda-plots

Figures from SOCA team

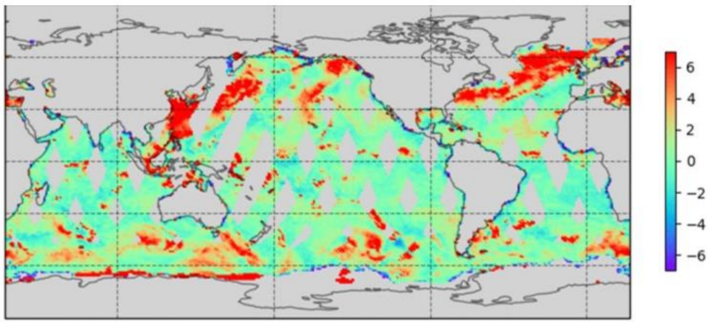
SOCA: Towards Coupled DA

Hamideh Ebrahimi (JCSDA)

- direct assimilation of MW brightness temperatures
  - SST (**GMI**)
  - SSS (**SMAP**)
- using CRTM with 2 domains (atmosphere & ocean)



SMAP O-B brightness temperature



2015-01-01 01Z to 2015-01-01 23Z

JCSDA ioda-plots

JCSDA Workflow Requirements

Models

- FV3-GFS / FV3-GEOS
- MOM6 / SOCA
- UFS
- MPAS
- Neptune
- LFRic
- …

Systems

- HPC (NOAA, NASA, Navy, USAF, NCAR, Met Office…)
- Cloud (AWS, …)
- Workstations, laptops (Linux, OSX)

Workflow engines

- ecFlow
- Cylc
- Airflow…

In addition to all *normal* operational and research requirements

EWOK + R2D2

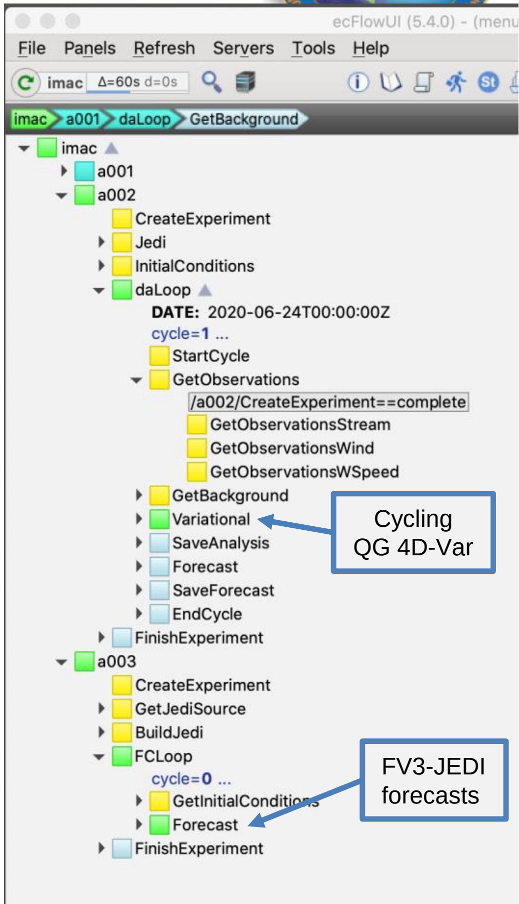
EWOK provides generic tools to describe suites = abstract algorithms (same concepts as OOPS/JEDI).

Tasks are generic, with model specific implementations only when required (same concepts as OOPS/JEDI).

Given specific parameters (experiment configuration) a suite is generated for a target model (UFS, GEOS), workflow engine (ecflow, cylc) and machine (Orion, Discover).

Data access is a major obstacle to portability and genericity,
R2D2 abstracts storage from the workflow:

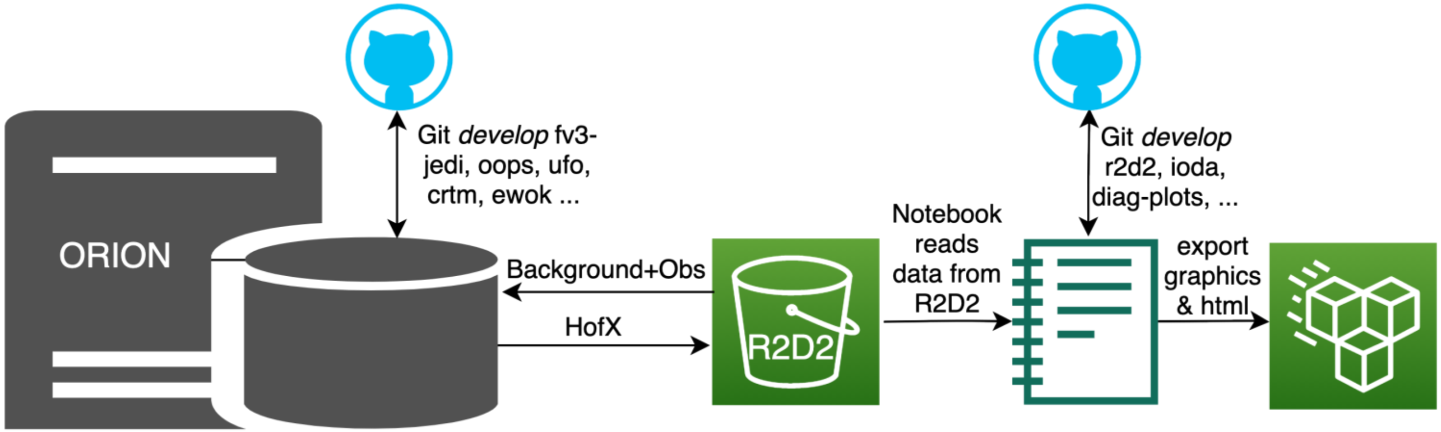
- Data is managed and accessed by key/values (not path or filename).
- Easy access for experiments (HPC, cloud) and diagnostics (desktop, cloud).
- It can store/fetch data in the cloud (S3) and locally, or a combination of both (local data store first, then remote).



Generic Diagnostics

Output of EWOK experiments are stored in R2D2 on AWS:

- Jupyter notebooks for interactive diagnostics
- Immerse scientists into the data
- Collective scientific diagnostics and inter-comparisons
- Easy interfacing with ML tools (SageMaker…)



Final Comments

JEDI is becoming more mature:

- "H(x)" computed in near real time for monitoring
- The system will evolve in non-cycling, then cycled data assimilation applications with more models

Next steps:

- Environment for running and managing experiments (HPC and cloud)
- Cloud-based interactive diagnostic tools
- Scientific validation (operational-grade cycling experiments)
- Coupled DA (Coupled H(x), Coupled B, Coupled solvers)
- Investigate integration of AI/ML

Y. Trémolet, JCSDA