Joint Effort for Data assimilation Integration

Unified Software Infrastructure for Data Assimilation

Yannick Trémolet
Tom Auligné

Joint Center for Satellite Data Assimilation (JCSDA)
JEDI: Motivations and 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 R2O/O2R)
- Share as much as possible without imposing one approach
Modern Software Engineering

We want a very flexible, reliable, efficient, generic, readable code.

This is not specific to NWP: the software industry has moved to generic and object-oriented programming 20 years ago.

The key idea is separation of concerns:
- All aspects exist but scientists focus on one aspect at a time.
- Different concepts are treated in different parts of the code.
- Nobody can know it all!

Good software engineering does not solve scientific problems: it is a tool to express and manage computations more efficiently.
Abstract interfaces are the most important aspect of the design.
JEDI Abstract Design

Abstract building blocks

Shared Components

Systems

Abstract Layer

State

Model

Atmosphere

Ocean

WRF

FV3

... 

MOM

... 

There is potential for sharing code at many levels (domain, technical...)
Model Design: $x_t = M(x_0)$

Works for any model, including coupled models.
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
- One objective is to build a community “operator-store”
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...)

Goal: one interface, possibly several implementations
Top-Down and Bottom-Up

Keep the computational parts of the existing codes and reuse them in a re-designed flexible structure.

This can be achieved by a top-down and bottom-up approach:
– From the top: Develop a new, modern, flexible structure
– From the bottom: Progressively create self-contained units of code by refactoring legacy code
– Put the two together

Interfaces are the key.

From a Fortran point of view, this implies:
– Control via interfaces (derived types passed by arguments),
– No global variables.
Collaborating: Repositories

Permission to fork repository are very easy to obtain
Contributing code is very controlled:
- Governance body to approve what features should be shared
- Pull request triggers code review before merging to higher level branch
Project Status

Infrastructure:
- Testing framework
- IODA

Models:
- FV3GFS/GOES-5
- WRF
- MPAS
- MOM, CICE
- ...

Observations from:
- GSI
- NAVDAS
- ...

Partners:
- NOAA/EMC
- NOAA/ESRL
- NASA/GMAO
- NRL
- NCAR
- ...

Distributed partners:
- Modern collaborative tools
- Regular code sprints

Ambitious timeline:
GSI-equivalent capability with FV3GFS in Spring 2019
Final comments

The JEDI project will provide a software infrastructure for data assimilation that
- Is model agnostic
- Does not impose one specific DA methodology or algorithm
- Encourages implementation of model independent observation operators (community “operator store”)
- Provides a unified Interface for Observation Data Access

The keys to success are separation of concerns and interfaces

JEDI is for scientific exploration and operational forecasting
... and exchanges between them.
Thanks to all involved in JEDI (so far)

Amal El Akkraoui, Anna Shlyaeva, Ben Ruston, Benjamin Ménétrier, Bob Oehmke, Bryan Flynt, Bryan Karpowicz, Clara Draper, Dan Holdaway, François Vandenberghe, Guillaume Vernières, Jeff Whitaker, Jing Guo, John Michalakes, Julie Schramm, Mariusz Pagowski, Mark Potts, Ming Hu, Rahul Mahajan, Ricardo Todling, Scott Gregory, Soyoung Ha, Tom Auligné, Will McCarty, Xin Zhang

The JEDI abstract layer is based on OOPS (ECMWF)

Join the force! There are currently 9 open positions at JCSDA in Maryland and Colorado. Additionally, open enrollment for JCSDA Visiting Scientist Program. See Tom Auligné or Yannick Trémolet.