Contribution ID: **53** Type: **Oral presentation**

Design of a Generic Workflow Generator for the JEDI Data Assimilation System

Tuesday, 15 October 2019 11:00 (20 minutes)

The JEDI (Joint Effort in Data assimilation Integration) is a collaborative project that provides a generic interface to data assimilation algorithms and observation operators for atmospheric, marine and other Earth system models, allowing these components to be easily and dynamically composed into complete data-assimilation and forecast-cycling systems. In this work we present the design of a generic workflow generation system that allows users to easily configure the JEDI components to produce custom data analysis toolchains with full cycling capability. Like the JEDI system itself, the workflow component is designed as a dynamically composable system of generic applications. An important point is that the JEDI workflow system is a generic workflow generation system, designed to programmatically produce workflow descriptions for a range of production-quality workflow management software engines including ecFlow, Cylc, and Apache Airflow. Configuration of the JEDI executables, the Python applications that control them, and the connection of applications into larger workflow specifications, is entirely accomplished with YAML-syntax configuration files using the Jinja templating engine. The combination of YAML and Jinja is simultaneously powerful, simple, and easily editable, allowing the user to quickly reconfigure workflow descriptions. A user can change model parameters, DA algorithms, covariance models, observation operators, and observation QC filtering algorithms, as well as the entire workflow graph structure, all without writing any shell scripts, editing any code, or recompiling any packages. Another key focus of the JEDI workflow system is data provenance and experiment reproducibility. Execution reproducibility is accomplished through elimination of unportable shell scripting in favor of Python-3; reliance on version control systems; universal use of checksum verification of all input products; and archiving of all relevant configuration and state as human-readable and editable YAML files.

Primary authors: Dr OLAH, Mark J. (UCAR / JCSDA); Dr TRÉMOLET, Yannick (UCAR / JCSDA)

Presenter: Dr OLAH, Mark J. (UCAR / JCSDA)

Track Classification: Workshop: Building reproducible workflows for earth sciences