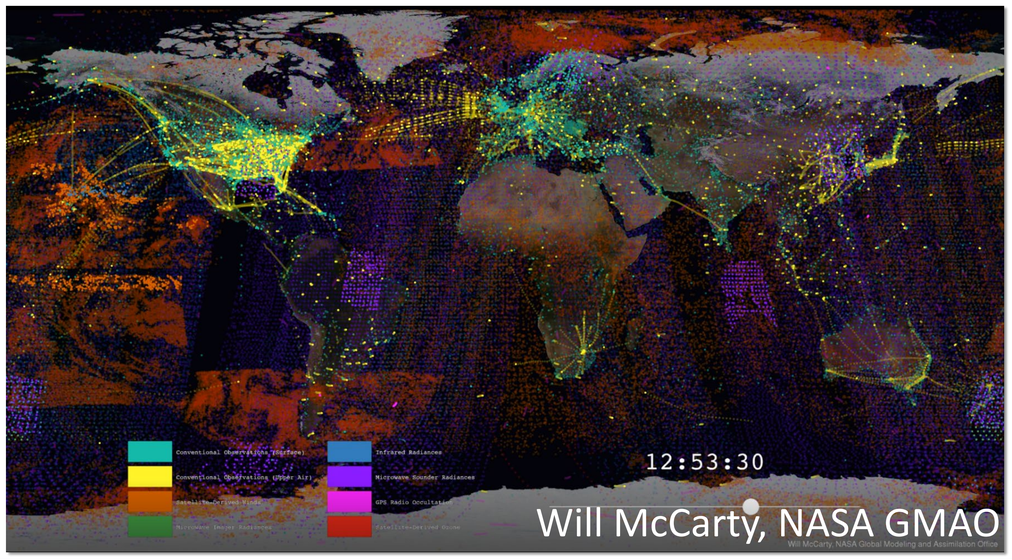
Unified Observation Processing

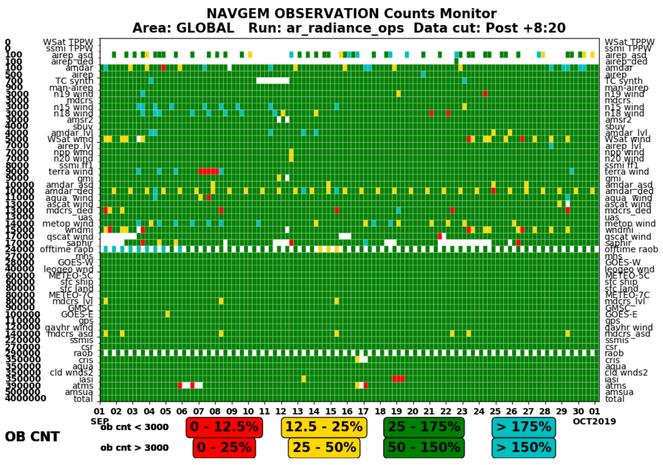
Benjamin Ruston[1] (ben.ruston@nrlmry.navy.mil), Nancy Baker[1], Patricia Pauley[1], Sarah King[1] and Eric Simon[2]
[1]Naval Research Laboratory, Monterey, CA [2]University Corporation for Atmospheric Research

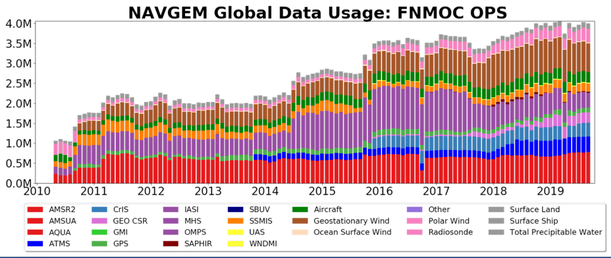
Earth Observing System

- Conventional (in-situ) systems:
  - Reporting not always standardized; irregular timing and coverage, variable meta-data, local quality control crucial
- Satellite systems:
  - Larger volumes, data reports more standardized, variable meta-data, processing from level-0 adds some quality control

The Naval Research Laboratory (NRL) develops and delivers operational environmental forecasting systems to Fleet Numerical Meteorology and Oceanography Center (FNMOC). These systems are driven by a large Earth Observing System (EOS). Initiatives undertaken across labs such as JCSDA JEDI Interface for Observation Data Access (IODA) and ECMWF ODB-C will be leveraged as we look to new techniques such as continuous data assimilation. Current efforts involve modifying the current processing chain, and refactoring when necessary to abstract functionality concepts allowing components to be used in common for the various data types used.





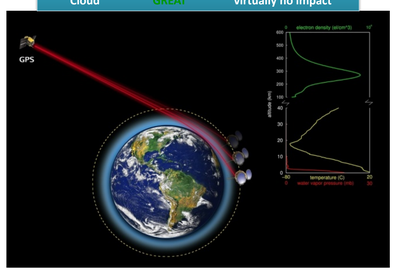


Data Usage

The amount of data assimilated by the environmental systems is growing rapidly quadrupling in the last decade. The strategies for using this data need to be re-examine for reusability and portability. Alignment with JCSDA JEDI IODA and the ECMWF ODB-C will be critical to quickly adopt and accelerate data usage across the centers. The data schema for the various data types must allow for flexibility, while maintaining family structures, to allow for APIs to access the data quickly and efficiently.

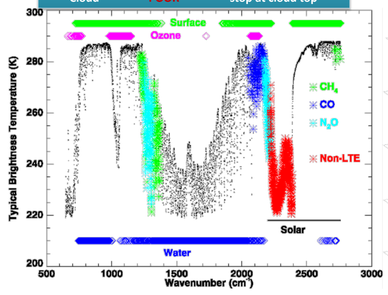
GPS-RO

Horizontal	POOR	200km
Vertical	GOOD	~8km-60km x200m
Cloud	GREAT	virtually no impact



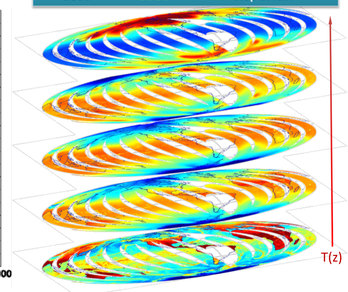
Infrared

Horizontal	GREAT	10km
Vertical	GOOD	surface-40km x2km
Cloud	POOR	stop at cloud top



Microwave

Horizontal	GOOD	25km
Vertical	GREAT	surface-90km x5km
Cloud	GOOD	stop at rain



Observation Synergy

Observations from *in-situ* and satellite platforms, and for various environmental systems like oceanic or ionospheric, all contain common traits. The handling of the ingest and the basic understanding of the measurement is required, but a focus on defining common attributes the data possess and defining attribute families is a focus going forward. An API which can perform actions on family attributes can transform stove-piped strategies into flexible ones for operations such as data thinning and error assignment, and allow decisions to be made by a particular application. Schema must prepare for heterogeneous metadata, and the API must maintain flexibility and parallelism.

Data format: blocks

- ODB API file consists of one or more blocks
- Each block consists of a header (metadata) and data section

ODB Governance
Report type:
16001 - Automatic Land SYNOP
varno:
110 - surface pressure,
58 - 2m relative humidity

andate :integer	antime :integer	reportype :integer	date :integer	time :integer	varno :integer	obsvalue :real	....	
20170108	120000	16001	20170108	150000	110	101220.00		Header
20170108	120000	16001	20170108	150000	58	0.992837		Data
20170108	120000	16001	20170108	150000	110	100110.00		
20170108	120000	16001	20170108	150000	110	100870.00		
20170108	120000	16001	20170108	150000	110	100980.00		
andate :integer	antime :integer	reportype :integer	date :integer	time :integer	varno :integer	obsvalue :real	....	
20170108	120000	16001	20170108	150000	110	101230.00		Header
20170108	120000	16001	20170108	150000	58	0.992852		Data
20170108	120000	16001	20170108	150000	110	100110.00		
20170108	120000	16001	20170108	150000	110	100860.00		
20170108	120000	16001	20170108	150000	110	100550.00		

Files with heterogeneous metadata

- Over time new columns are added to ODB. When retrieving from MARS data spanning multiple cycles it may happen that part of the retrieved data has more columns than the rest.

andate :integer	antime :integer	reportype :integer	date :integer	time :integer	varno :integer	obsvalue :real	
20170108	120000	16001	20170108	150000	110	101230.00	
20170108	120000	16001	20170108	150000	58	0.992852	
20170108	120000	16001	20170108	150000	110	100110.00	
20170108	120000	16001	20170108	150000	110	100860.00	
20170108	120000	16001	20170108	150000	110	100550.00	
andate :integer	antime :integer	reportype :integer	date :integer	time :integer	varno :integer	obsvalue :real	new_flag :bitfield
20170109	120000	16001	20170109	150000	110	101230.00	1
20170109	120000	16001	20170109	150000	58	0.992852	1
20170108	120000	16001	20170108	150000	110	100110.00	0
20170108	120000	16001	20170108	150000	110	100860.00	0
20170108	120000	16001	20170108	150000	110	100550.00	1

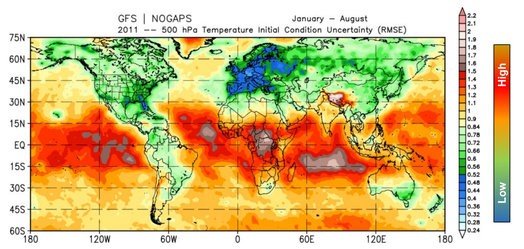
Piotr Kuchta ECMWF 22Feb2017
https://confluence.ecmwf.int/download/attachments/73008494/ODB%20API%20Intro.pdf?version=1&modificationDate=1488572989201&api=v2

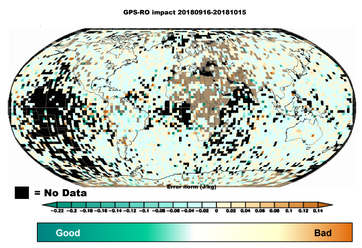
Commercial Data and Small Satellite Era

Environmental observation, particularly in the U.S. are increasingly reliant on commercial providers and the evolving small satellite era. These data will be large in volume, with poorly defined sources and fluctuating quality control approaches. Further we may expect format changes as companies (providers) may fail or be acquired. An adaptive strategy, and potentially artificial intelligence (AI) approaches to identify and if possibly remedy issues should be fully explored, and systems should readily adapt to new data types and be equipped with bias and error mitigation.

Continuous Data Processing (Ingest)

A more streamlined way to ingest data, and to preserve more data is being designed. As data arrives a continuous processing system stores key meta-data from the observation to a catalogue. At present, decisions are made with a goal to maintain the highest quality reports, but often other alternate reports are lost. Methodical cataloging will allow preservation of and be used to maintain currently discarded duplicate reports. Decisions regarding which report to use can be more dynamic, allowing secondary reports to be used for validation, or for alternate initialization. The ability to monitor observation impact on the system can further be used to feed the quality control decisions. Areas of dense observation show low errors and smaller differences between centers, analysis error monitoring with feedback to quality control may prove critical.

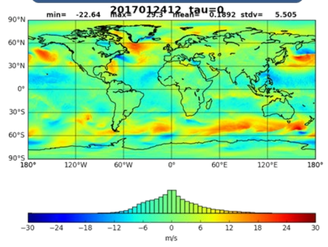




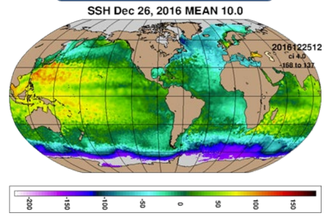
Multi-System Portability

Similar to design strategies for data assimilation systems where the solver can have a modular ability to use 4D-Var, Hybrid 4D-Var, or EnKF depending on the problem at hand. The handling of data streams should be thought of in a common way for a multitude of environmental system. The Navy Earth System Prediction Capability (ESPC) is an excellent testbed to explore a common strategy as the effort works towards running a coupled system combining top of the atmosphere to bottom of the ocean observation, data assimilation and modeling.

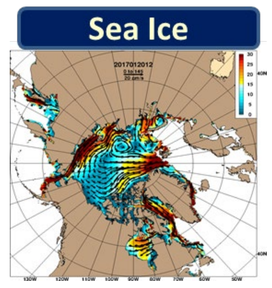
Atmosphere



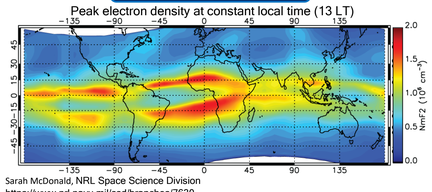
Ocean



Sea Ice



Ionosphere



Sarah McDonald, NRL Space Science Division
https://www.nrl.navy.mil/ssd/branches/7630

Conclusions & Future Directions

- **The Earth Observing System is evolving, with both commercial data and small satellites expanding volumes rapidly**
- **Data schema with families of common attributes can organize data while maintaining flexibility and parallelism**
- **Continuous data processing should be re-examined to reduce loss of data; coupling with analysis error estimates can increase and refine data usage**
- **Large coupled modelling systems are natural testbed for common data handling and rapid prototyping for signals-of-opportunity and new sensor technologies**