

Implementing an Observation Support System for the Data Assimilation Research Testbed

Jason Ngo¹, Jeffrey Anderson², Nancy Collins²

¹Haverford College, Haverford, PA, ²Data Assimilation Research Section, NCAR

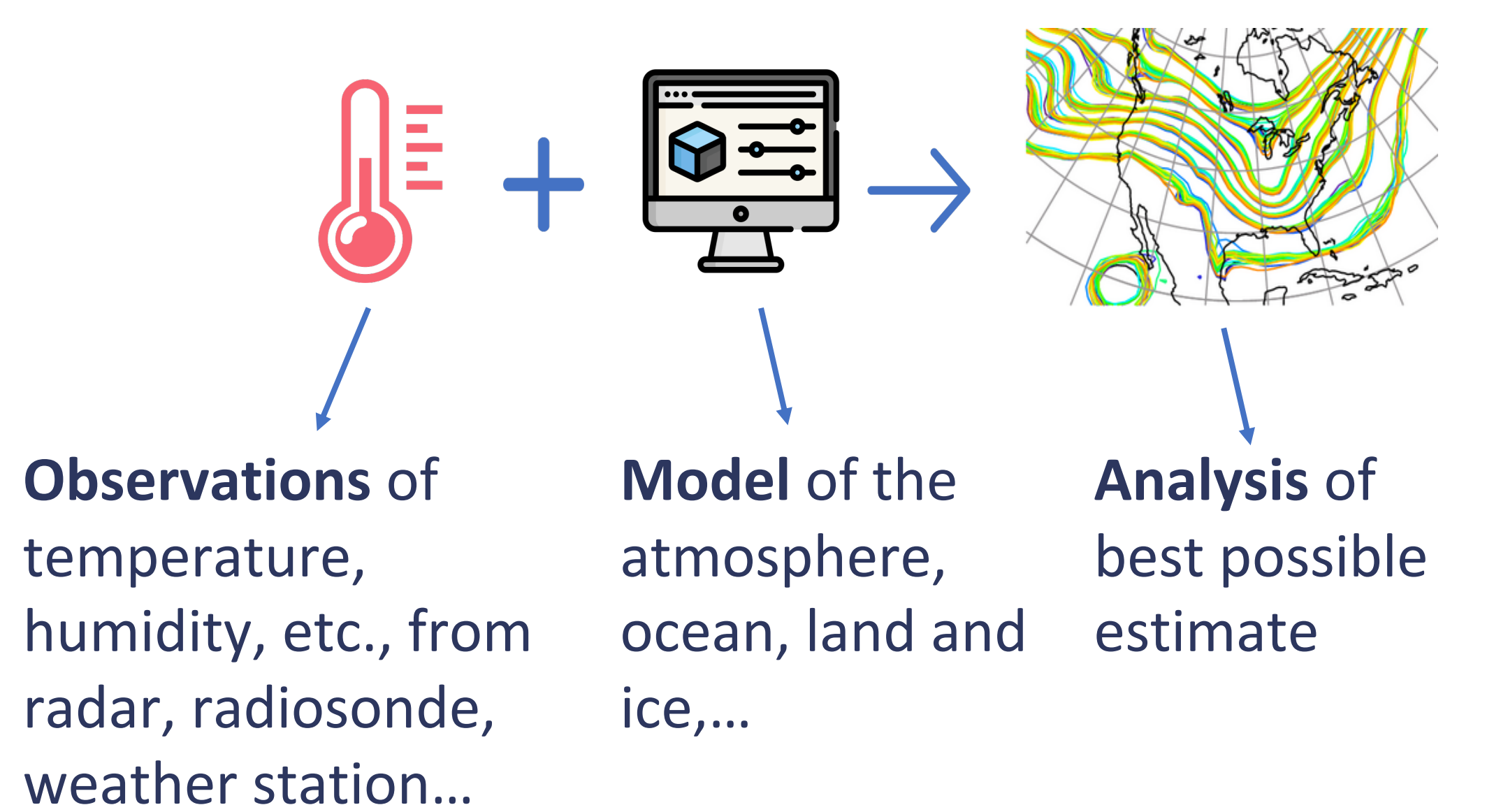
Correspondence: aqngo@haverford.edu



Introduction

Data Assimilation

- Observations combined with model forecast to produce an analysis
- Widely used in numerical weather prediction



Data Assimilation Research Testbed (DART)

- Data assimilation (DA) open-source software developed by the Data Assimilation Research Section (DARes) at NCAR
- DART provides modelers and observational scientists with powerful DA tools that are easy to implement and customize

Summary

Currently, DART stores information about weather observations in a proprietary format. Though containing exactly the information needed for assimilation, this format is non-standard and does not support database-type queries.

My project therefore implements a converter from the proprietary format to self-describing netCDF format and adds support for database management.

The project also develops a cross-platform GUI to help users visualize the contents of those netCDF files and understand the impact of observations on their model in a user-friendly fashion.

Motivations

Database Problem

DART currently stores observation information in obs_sequence files.

```
10 ADCP_U_CURRENT_COMPONENT
11 ADCP_V_CURRENT_COMPONENT
16 DRIFTER_U_CURRENT_COMPONENT
17 DRIFTER_V_CURRENT_COMPONENT
23 GLIDER_TEMPERATURE
29 SATELLITE_INFRARED_SST
30 SATELLITE_SSH
num_copies: 1 num_qc: 1
num_obs: 559502 max_num_obs: 5595
observation QC value
first: 1 last: 559502
OBS
-0.176900000000000000
1.000000000000000000
-1 2 -1
obdef loc3d
4.62424 0.32550 0.00000 -1
kind
30
3600 144270
2.5000000000000000E-003
OBS
-0.177800000000000000
1.000000000000000000
1 3 -1
obdef loc3d
4.61726 0.32724 0.00000 -1
kind
30
3600 144270
2.5000000000000000E-003
```

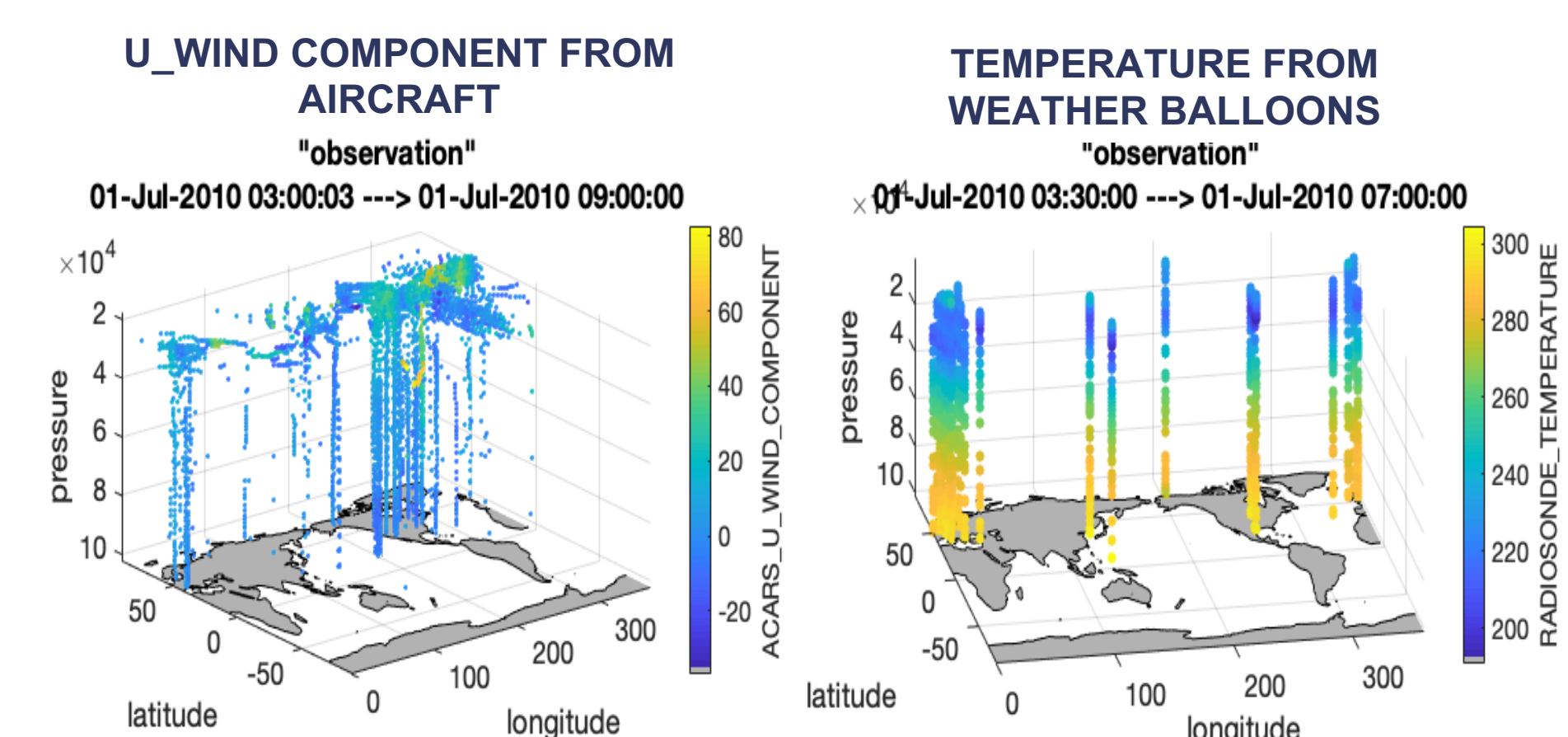
obs_sequence file structure

- Pros: contains exactly needed info, support extensible metadata
- Cons: difficult to query, non-standard

Current Visualization Infrastructure

The 3D point-data obs_sequence files are not supported by current visualization software.

DARes therefore converts obs_sequence to netCDF and implements MATLAB to plot:



Plotted with MATLAB

- Pros: customizable, pretty
- Cons: expensive MATLAB license, difficult to maintain, does not support extensible metadata

Solutions

NetCDF Converter

Convert to standard netCDF with Fortran & Python scripts

```
types:
    extensible Metadata;
dimensions:
    obs;
variables:
    time(obs);
    lat(obs);
    lon(obs);
    salinity(obs);
    Metadata(obs);
// global attributes
groups: dummy {
    obs_id = 0,2,3...
    // attributes
}
```

netCDF file structure

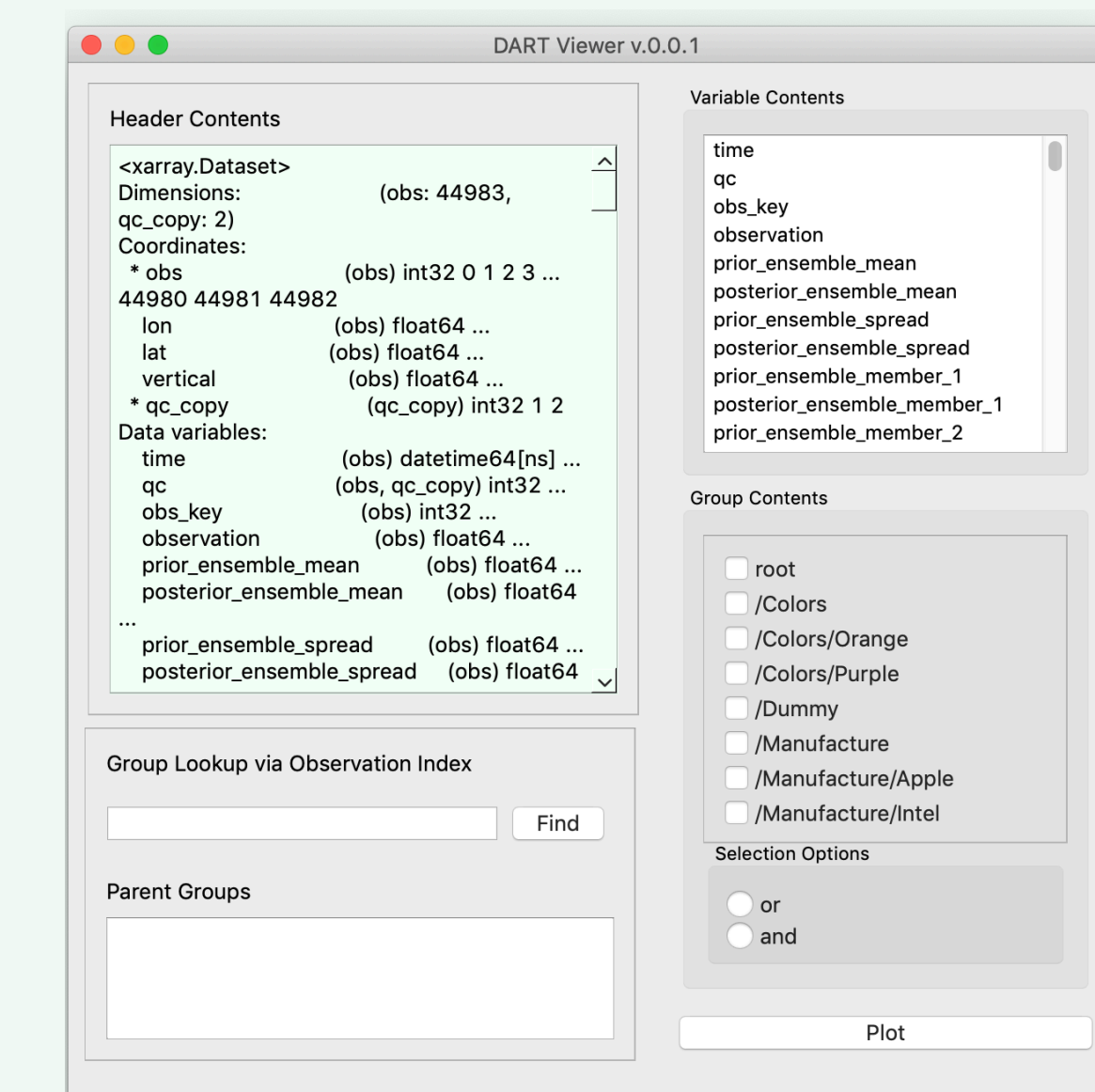
- Preserve point-data structure
- Support extensible metadata with varying-length variables
- Support subsets
- Follow data standards (CF Conventions)

Graphical User Interface (GUI)

Frontend: built with Qt, a cross-platform software development environment

GUI pros:

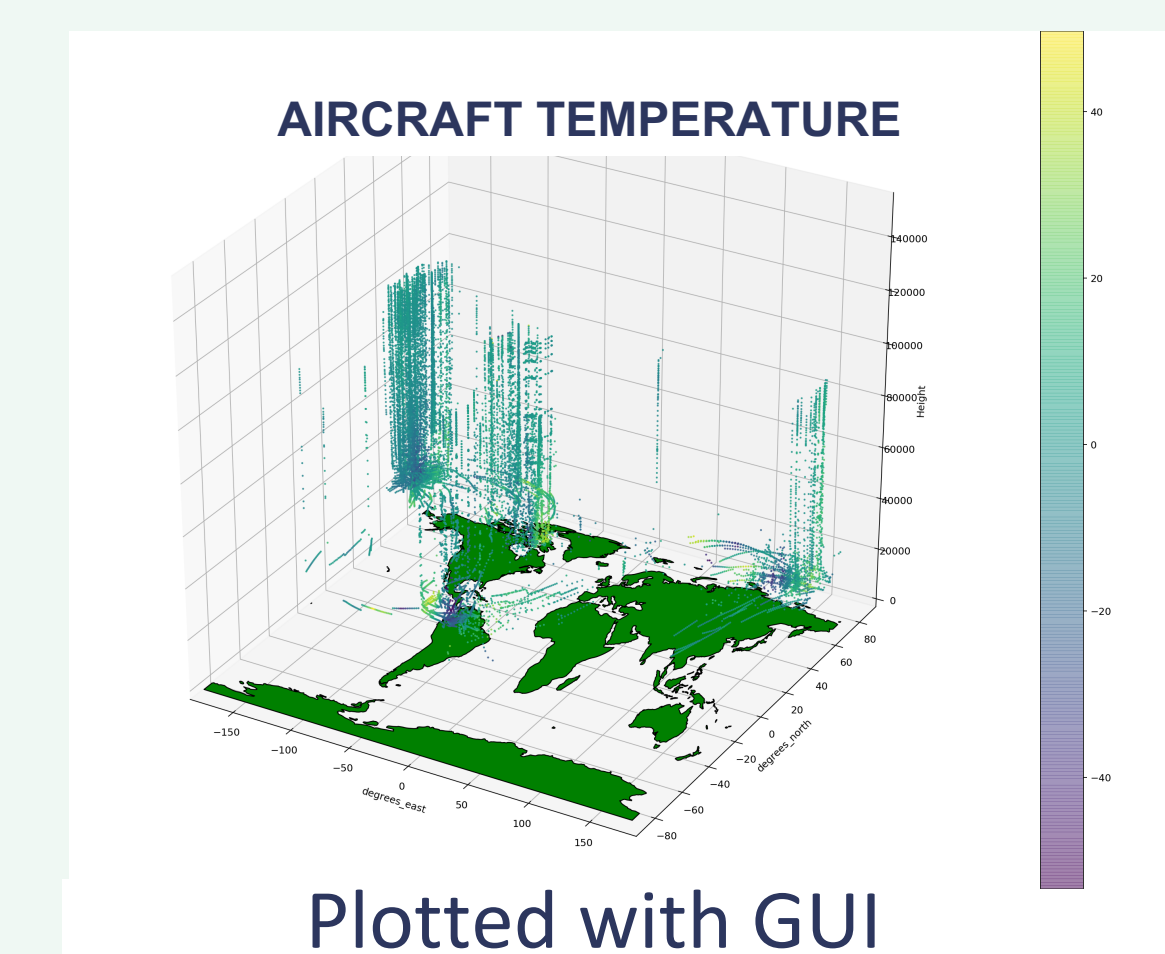
- Requires no programming languages
- Free
- Mac/Linux compatible



Python Backend:

Database: xarray, netcdf4

Visualization: matplotlib, cartopy



Future Work:

- Automated testing capabilities
- Customizable plots
- More intuitive interface