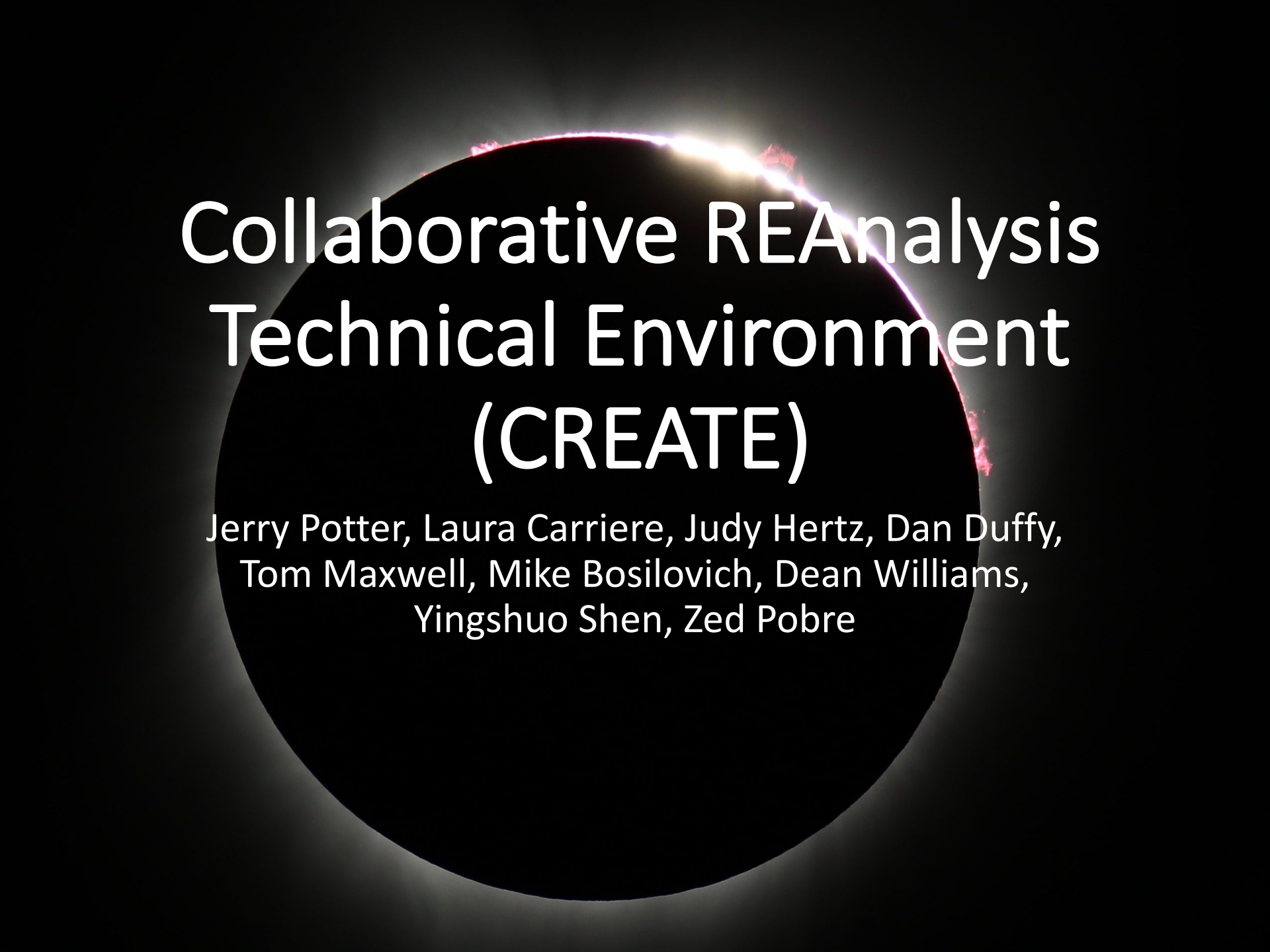


Collaborative REAnalysis Technical Environment (CREATE)

Jerry Potter, Laura Carriere, Judy Hertz, Dan Duffy, Tom Maxwell, Mike Bosilovich, Dean Williams, Yingshuo Shen, Zed Pobre

Outline

- Reanalysis
- Scope of data sets
- Workflow
- Tools
- Processing large data sets
- Examples
- Paper describing service
- Future

Repackaging Atmospheric Reanalysis Data Sets

- **Atmosphere Reanalysis**
- NASA MERRA
- NASA MERRA2
- ECMWF ERA-Interim
- NOAA/NCEP CFSR
- NOAA/ESRL 20CRv2c
- JMA JRA-25
- JMA JRA-55
- ERA20C and or CERA20C (under consideration)
- ERA5 (in progress)

The NCCS has published monthly and selected 6-hourly data from seven major atmosphere reanalysis projects from 1979 to 2017*. Up to 47 monthly variables were processed and published. Sixteen 6-hourly variables were processed and published.

* 20CRv2c (1851-2012) and CERA20C (1901-2010)

6-hourly Atmospheric Variables Processed and Published in ESGF

6-Hourly Variables

Total Cloud Fraction

Evaporation

Relative Humidity 3D

Specific Humidity 3D

Precipitation

Precipitable Water

Surface Pressure

Sea Level Pressure

Air Temperature 3D

Surface Air Temperature

Ozone Mole Fraction 3D

Eastward Wind 3D

Near Surface Eastward Wind

Northward Wind 3D

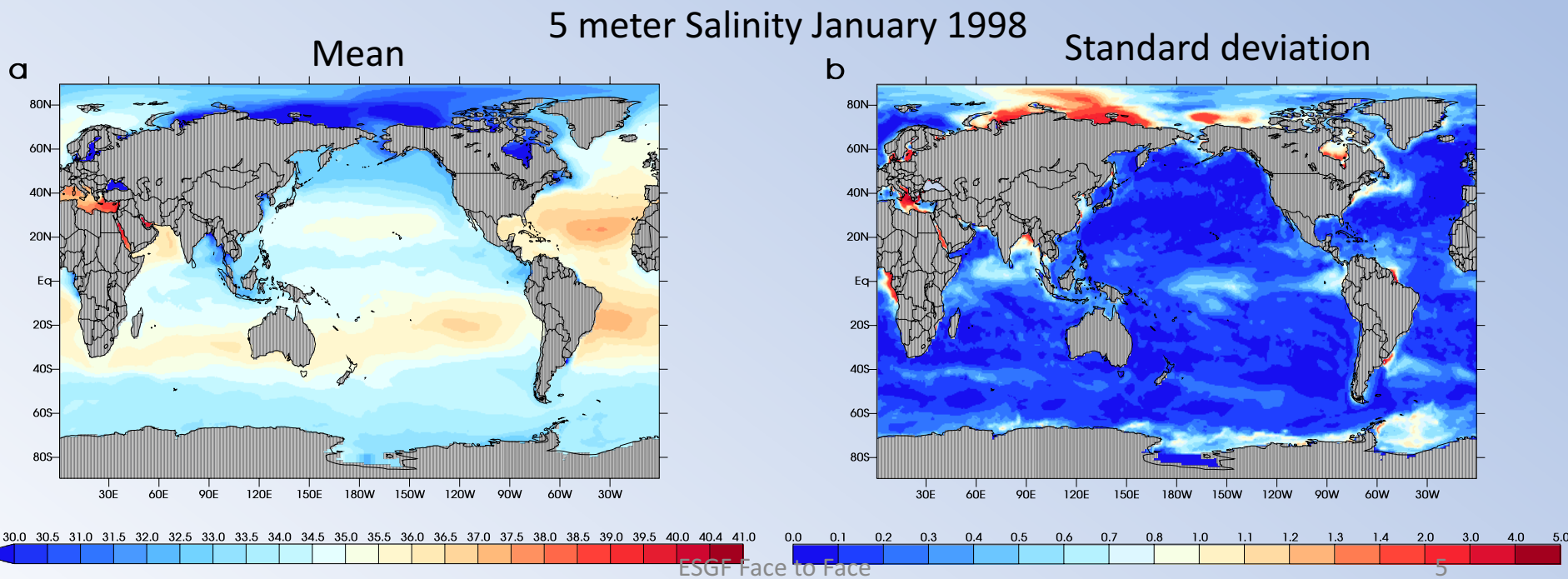
Near Surface Northward Wind

Geopotential Height 3D

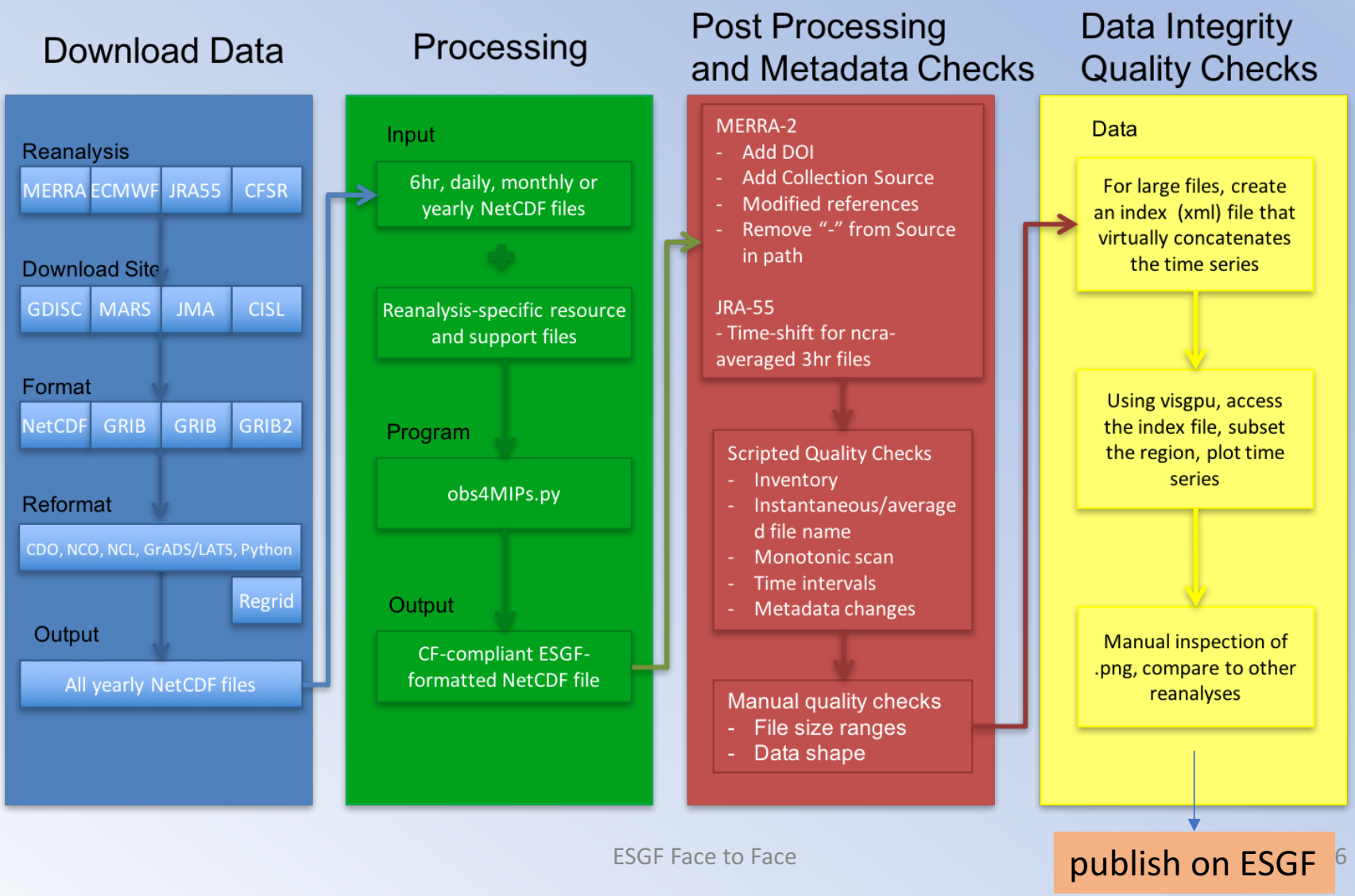
3 Dimensional 6 hour data poses a data handling challenge. 1TB/month.

8 Ocean Reanalyses Included in CREATE on ESGF

- Native grid
- Regridded to common 1 degree grid and common depths
- Generated ensemble average and standard deviation

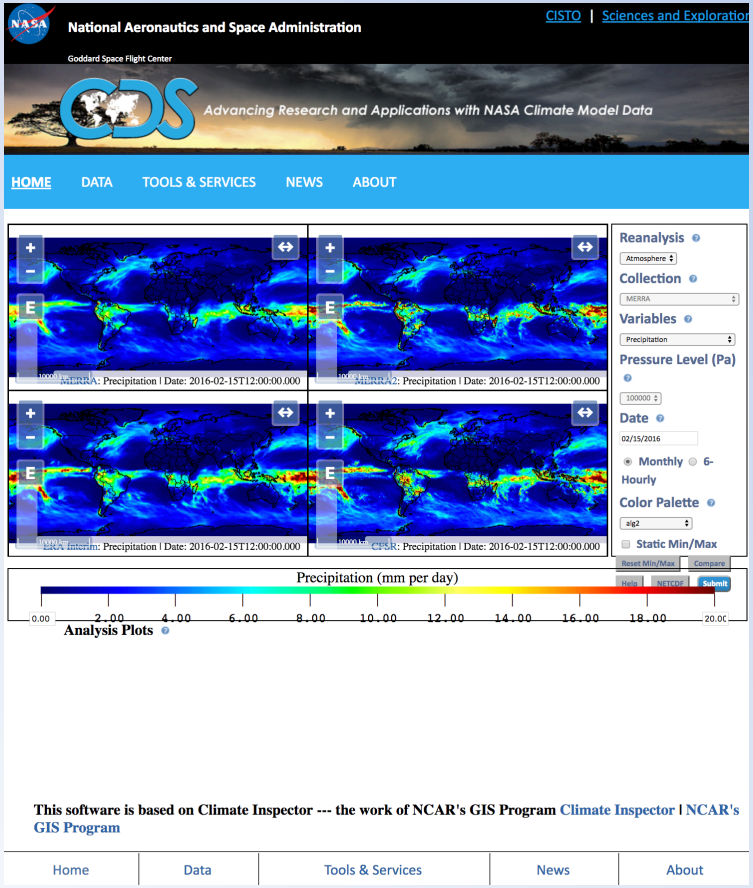


CREATE – Workflow for Atmospheric Reanalyses

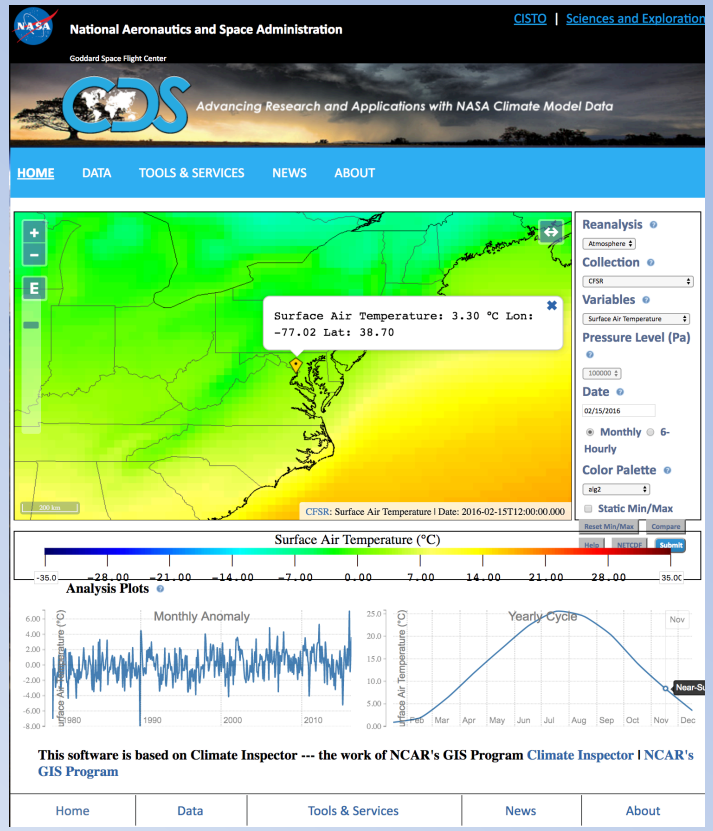


Tools for Visualization

- CREATE-V quick look view

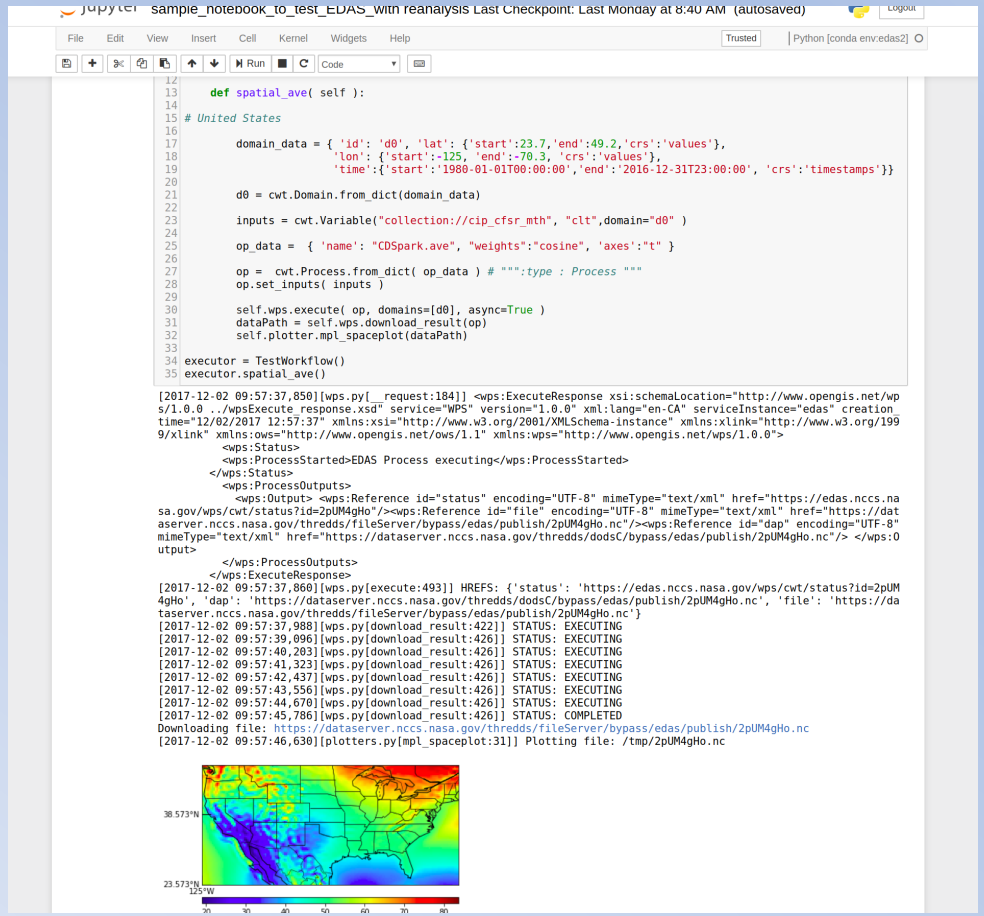


- Multiple reanalyses
- 6 hour data at selected time points
- Anomaly at selected grid points



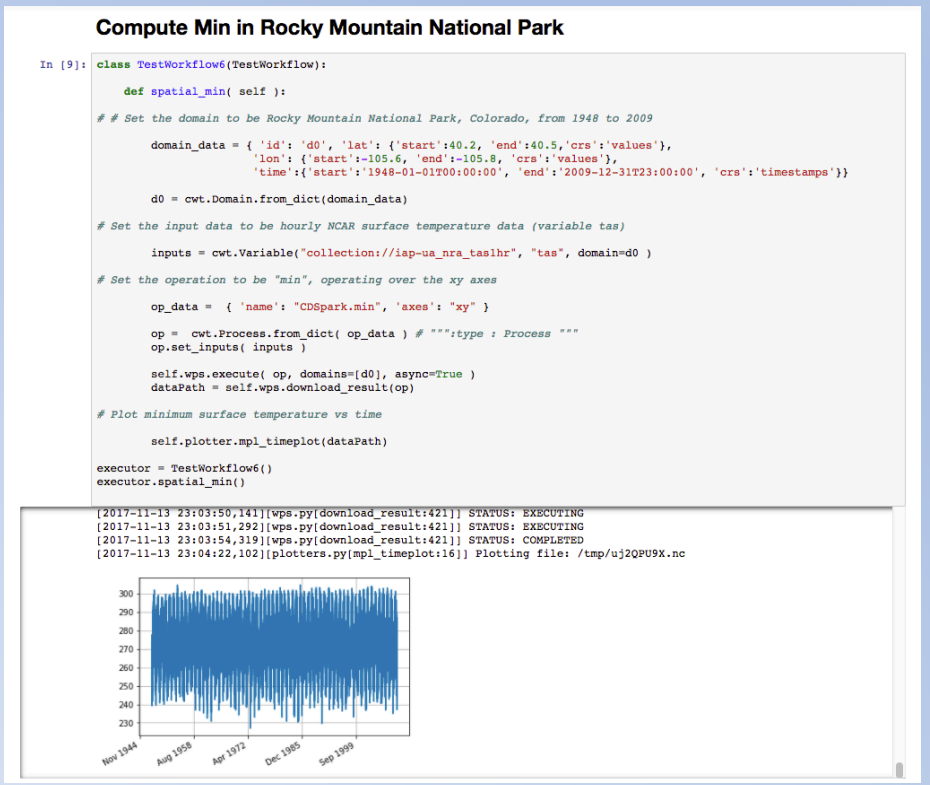
Tools for Analysis

- Earth Data Analytics Service (EDAS)
  - Server side analytics – Tom Maxwell presentation
  - Speeds operations by a factor of ~15x to 50x faster than standard python calls
- Calls to operations can be made via WPS from Jupyter Notebook on a local computer or any Python script



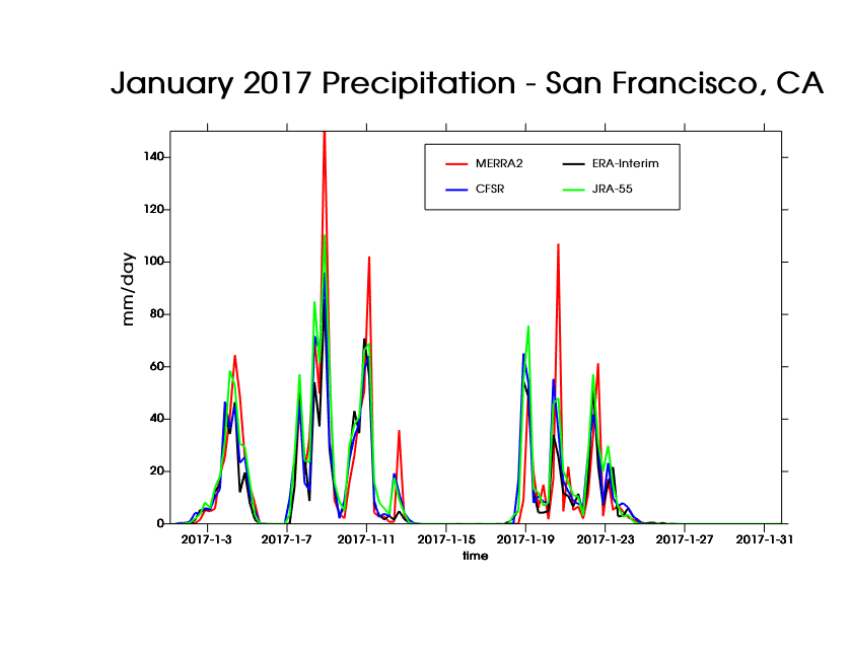
Hourly Surface Temperature Data

- Four temperature datasets combined with CRUTS data
  - NASA MERRA (1979-2009)
  - ERA-Interim (1979-2009)
  - ERA-40 (1958-2001)
  - NCEP-NCAR (1948-2009)
- QC wasn't possible on our original system, had to wait for the new analytics tool EDAS.
- Can evaluate 60 years of hourly data in 10 minutes.

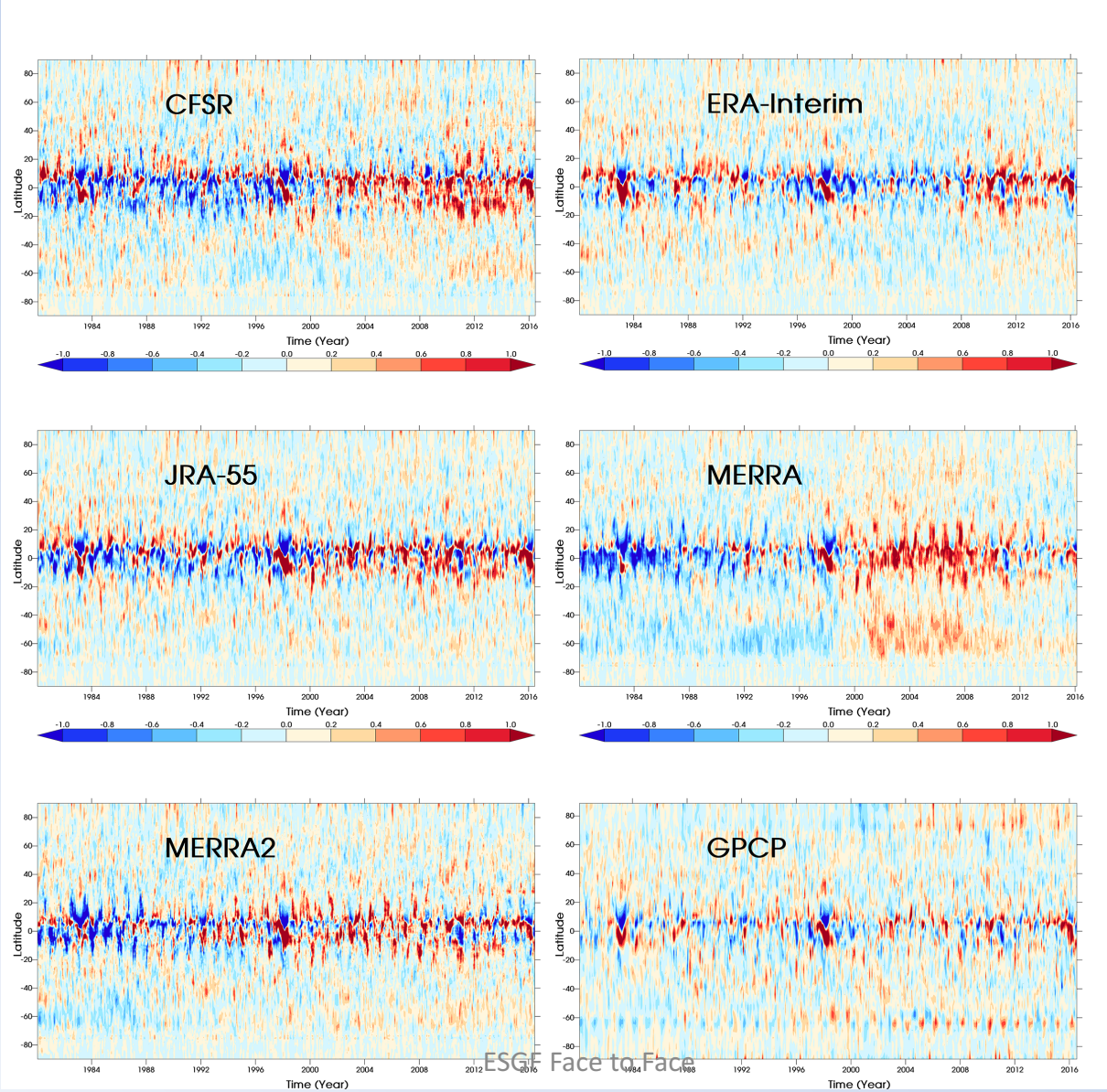


Issues About Using Reanalysis for Model Validation

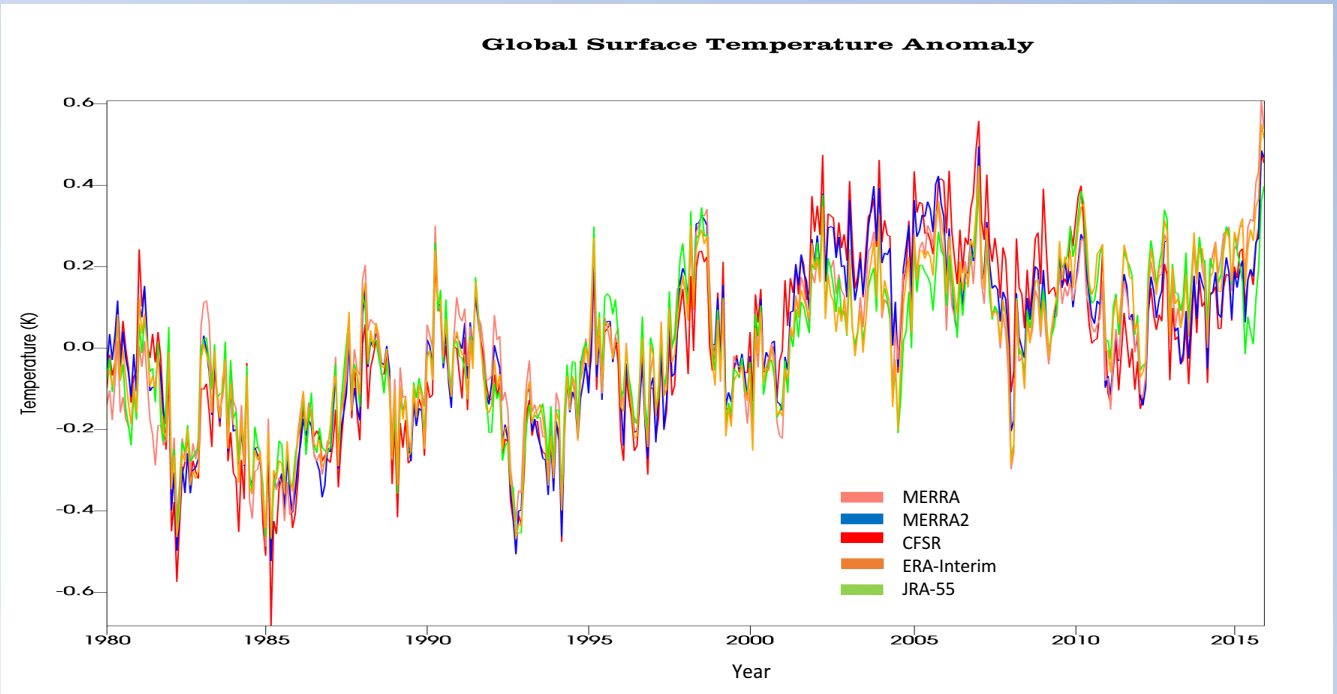
- Traditionally reanalysis limited to a few variables
  - winds – and associated fields
  - temperature
  - humidity
  - surface pressure
  - geopotential height
  - are there more?
- Not viewed as useful - model influenced
  - radiative fluxes
  - precipitation



Comparing Reanalysis Precipitation Anomaly with Observations



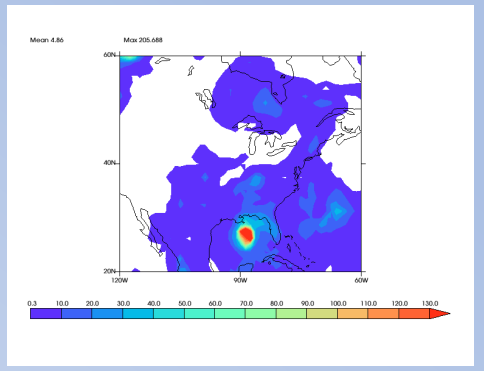
Surface Air Temperature Anomaly 1979-2015

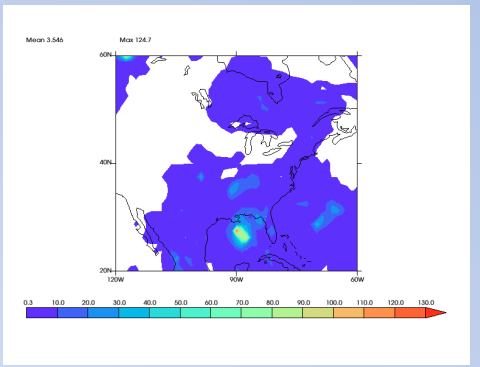


Monthly Global Surface Air temperature Anomaly (1980-2015)

Multiple Reanalysis Ensemble (MRE)

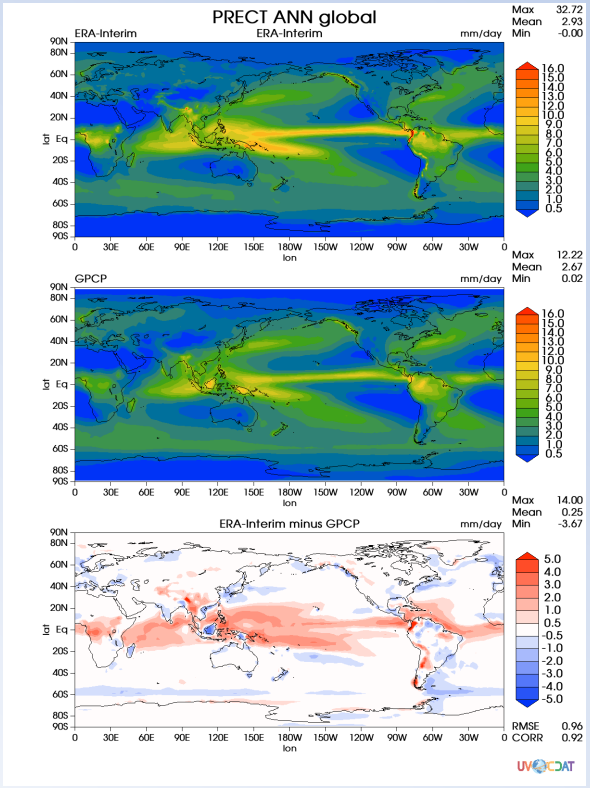
Mean (top) and standard deviation (bottom) of the MRE precipitation during Hurricane Katrina - August 30, 2005

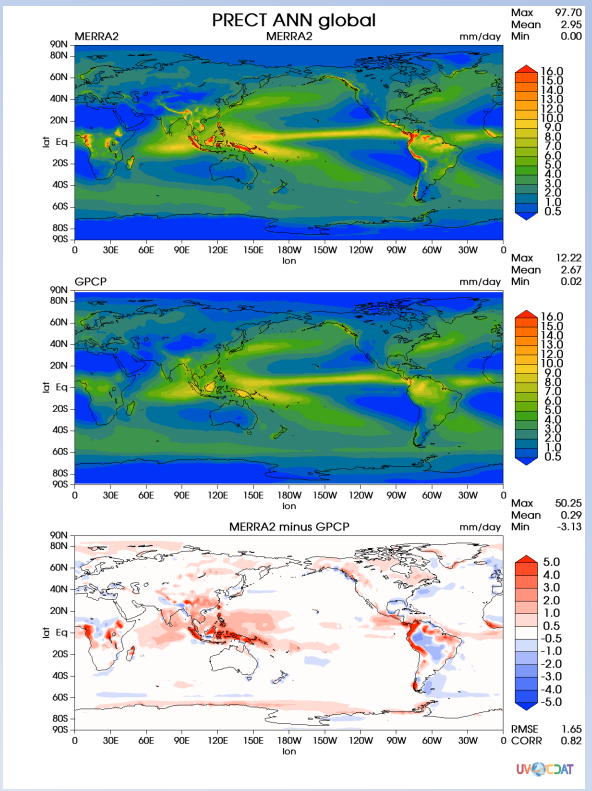


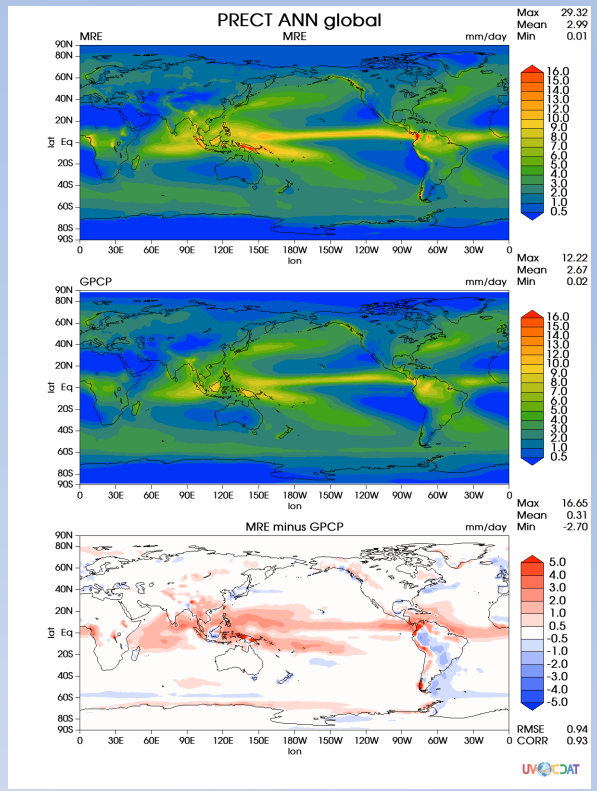


- Monthly data regridded – standard horizontal and vertical
  - Mean and standard deviation (unbiased)
- Monthly ensemble mean and standard deviation are manageable.
- Standard computing for 6-hour 0.5° grid 3D time series for each variable computing on the order of days.
  - Necessary to study events and the degree to which the reanalyses agree or disagree

MRE Annual Climatology Compared to ERA-Interim and MERRA2







Thanks to Zeshawn Shaheen, Chengzhu (Jill) Zhang, and Chris Golaz

Paper Submitted and Accepted by BAMS Describing CREATE

Short summary of CREATE to appear in “In Box” section of a future issue.

Enabling Reanalysis Research Using the Collaborative REAnalysis Technical Environment

(CREATE)

Gerald L. Potter[1]

Laura Carriere[1]

Judy Hertz[1]

Michael Bosilovich[2]

Daniel Duffy[1]

Tsengdar Lee[3]

Dean N. Williams[4]

The Next Big Thing

- CERA20 – 10 member ensemble of 1901-2010 surface pressure driven reanalysis
- MERRA – possible plans for new high resolution reanalysis
- 20CRv2c – (1871-2012) update with high resolution model – more ensembles
- ERA5 .25 degree 1 hour output at 137 levels
  - 40 PB for full archive
  - Latency 7 days, currently 2 months

Summary

Challenges and accomplishments

- Scope of data sets
- Workflow
- Tools

Preparing data for ESGF has become easier – still using CMOR2. CREATE-V provides quick look.

- Processing large data sets
- Examples

As data sets become larger, even QC has become a challenge. Necessary to have more advanced server side computing. EDAS is very promising.

- Paper describing service

Usage will increase after publication – ESGF/COG is difficult to use.

The paper emphasizes the Use of ESGF to Access Reanalysis Data

- Reviews were mostly about ESGF – use and searching difficulty.
- “it should be noted the ESG/CoG site can be very difficult to use. The most common issue is SSL errors and there tend to be opaque or non-existent error messages when things fail. Once successfully logged into the site, users can search for datasets but oddly not variables. Users can look for datasets that have certain variables but it seems one cannot look for files containing a certain variable. The different reanalysis are found by looking at 'experiments'. Variables are sorted in alphabetical order but include the type so that all surface files are under "S". ”
- Includes a demo Jupyter-notebook example requires installation of Anaconda uvcdat etc. to access NASA data on THREDDS server at GSFC.

We Pointed out the “help” Button on the Home Page