Data Reorganization for Optimal Time Series Data Access, Analysis, and Visualization

William Teng[1,2], Hualan Rui[1,2], Richard Strub[1,2], and Bruce Vollmer[1]

[1]NASA GSFC, GES DISC

[2]ADNET Systems, Inc.

Work supported by NASA ROSES NNH11ZDA001N-ACCESS

Outline

- "Digital Divide" problem
- Solution for bridging the Divide
- Applications

More details: IN23B-1500, today 1:40 pm

Overall ACCESS Project: IN31A-1494, Wed. 8 am

“Digital Divide” Problem

- Data archived in the form of all variables one time step per file
- Users often need long time series for single variables at single grid “points”
- Access is orthogonal to archive → Inefficient

(Acronymic) Background

- CUAHSI (Consortium of Universities for the Advancement of Hydrologic Sciences, Inc.)
- HIS (Hydrologic Information System) Univ. of Texas-Austin (Co-I ACCESS project)
- NLDAS (North American Land Data Assimilation System)
- HSL (Hydrological Sciences Laboratory, GSFC) (Co-I ACCESS project)

NLDAS/GLDAS Data Files and Total Volume

Data sets	Temporal		Spatial		Dim	Total grids	# Files per data set	Total vol
	Res	Cov	Res	Cov				
NLDAS	hourly	1979-present	0.125°	N. Amer	224x464	93542*	289080	~ 4.8 TB
GLDAS	3-hourly	1948-present	0.25°	Global	600x1440	259200*	96360	~ 1.6 TB

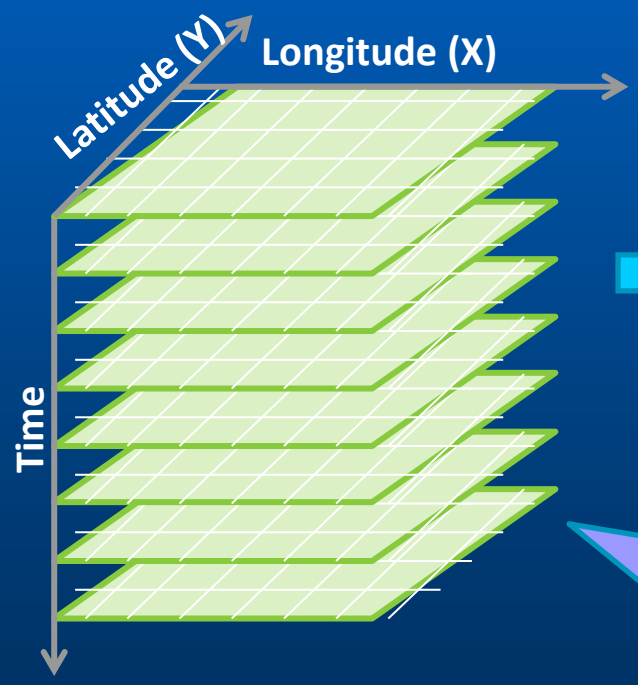
* Per parameter; reduced by NLDAS land ~ 0.9; GLDAS land ~ 0.3

BASINS Prototype

- EPA BASINS (Better Assessment Science Integrating point & Non-point Sources) – Multi-purpose environmental analysis system at watershed level.
- Data access requirement – Routine downloading of entire time period of NLDAS as time series.
- GDS (GrADS Data Server) can return NLDAS time series for ~20 days (~500 time steps).
- Solution: Partial reorganization (yearly latitudinal files) enables GDS to return entire NLDAS time series.

ACCESS Project Solution for Bridging the Digital Divide

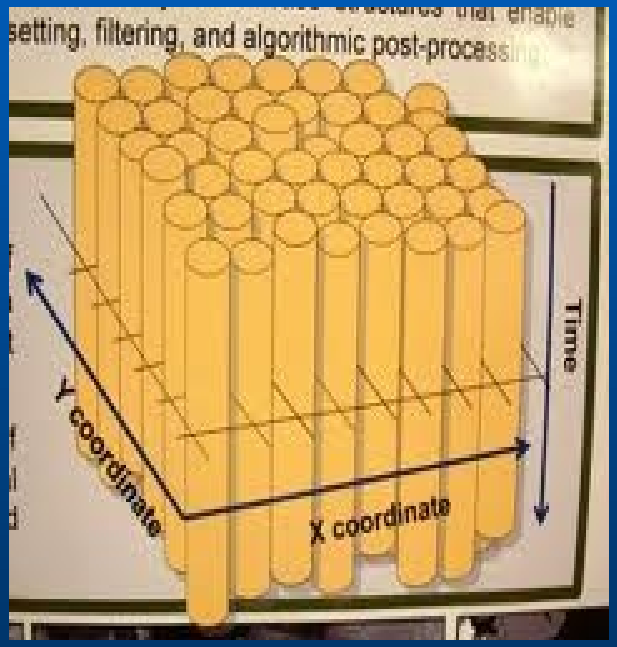
Original Data Archive



One variable
one grid point
all time steps
per file

All variables
all grid points
one time step
per file

Reorganized Data Archive



Time Series File Structure

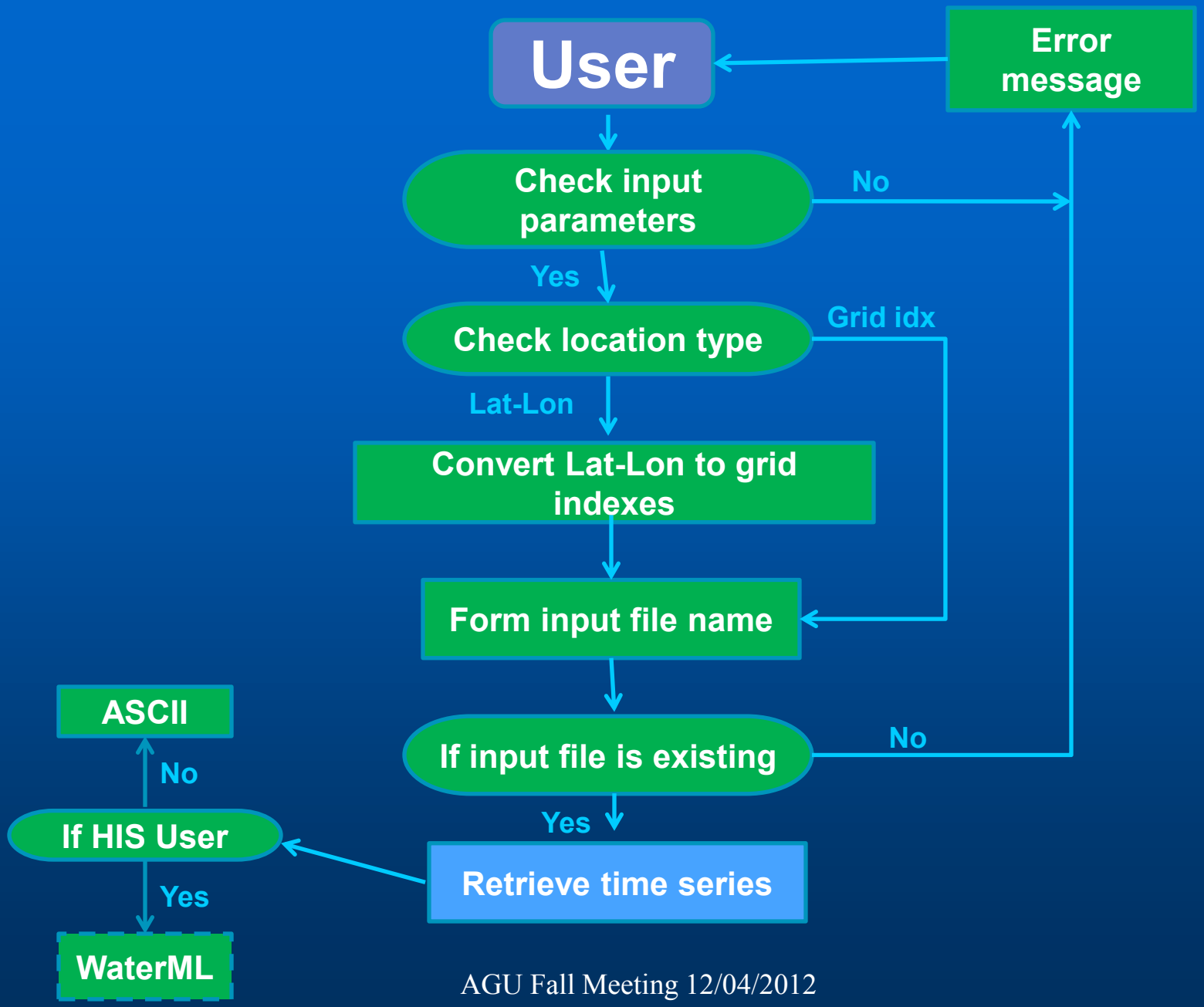
File header (300 bytes)

```
char prod_name[50];
char param_short_name[20];
char param_name[50];
char unit[20]
int grid_y, grid_x;
char begin_time[14];
char end_time[14];
unsigned int tot_record;
float min,max,mean,std;
int dt;
int ydim, xdim;
float start_lat, start_lon;
float dlat, dlon;
char last_updt[20];
char spare[60];
```

Data records

N = tot_record written in the file header

```
float data1
float data2
……
float dataN
```


User
Error message
Check input parameters
No
Yes
Check location type
Grid idx
Lat-Lon
Convert Lat-Lon to grid indexes
Form input file name
If input file is existing
No
Yes
Retrieve time series
If HIS User
No
ASCII
Yes
WaterML

Early time series service of NLDAS Noah 0-100 cm soil moisture, e.g.,

http://hydro1.sci.gsfc.nasa.gov/daac-bin/access/timeseries.cgi?*variable*=NLDAS:NLDAS_NOAH0125_H.002:SOILM0-100cm&*startDate*=1979-01-02T00&*endDate*=2012-09-30T23&*location*=NLDAS:X217-Y042&type=plot

```
Metadata of the Time Series file:

prod_name=NLDAS_FORA0125_H.002
param_short_name=APCPsfc
param_name=Precipitation hourly total
unit=kg/m^2
begin_time=1979/01/01/13
end_time=2012/11/19/12
time_interval(hour)=1
tot_record=297024
grid_y=113
grid_x=402
ydim=224
xdim=464
start_lat=  25.0625
start_lon=-124.9375
dlat=0.125000
dlon=0.125000
undef=  9.9990e+20
Last_update=Fri Nov 23 16:14:02 2012

Metadata for Requested Time Series:

prod_name=NLDAS_FORA0125_H.002
param_short_name=APCPsfc
param_name=Precipitation hourly total
unit=kg/m^2
begin_time=2012/10/28/12
end_time=2012/10/31/00
begin_time_index=296495
end_time_index=296555
lat=  39.1875
lon= -74.6875
grid_y=113
grid_x=402
tot_record=61
Request_time=Mon Nov 26 18:08:12 2012

Orig-rec # reqt-rec #          Date&Time        Data
    296495          0      2012-10-28 12Z    0.000000
    296496          1      2012-10-28 13Z    0.007600
    296497          2      2012-10-28 14Z    0.009100
    296498          3      2012-10-28 15Z    0.027600
    296499          4      2012-10-28 16Z    0.000000
    296500          5      2012-10-28 17Z    0.000000
    296501          6      2012-10-28 18Z    0.703200
    296502          7      2012-10-28 19Z    0.922000
    296503          8      2012-10-28 20Z    0.635100
    296504          9      2012-10-28 21Z    1.976800
    296505         10      2012-10-28 22Z    2.262700
    296506         11      2012-10-28 23Z    2.530000
    296507         12      2012-10-29 00Z    5.037700
```

Metadata for the original time series

Metadata for the requested time series

Time step referenced from the beginning of the original time series

Time step referenced from the beginning of the requested time series

Directory Structure and file name convention for reorganized time series files

One directory per parameter
One file per grid point (time series)

Directory Name Convention:
prod_name.vid.param_short_name

File Name Convention:
prod_name.vid.param_short_name.Y###-X###.bin

Example for NLDAS Precipitation (APCPsfc):

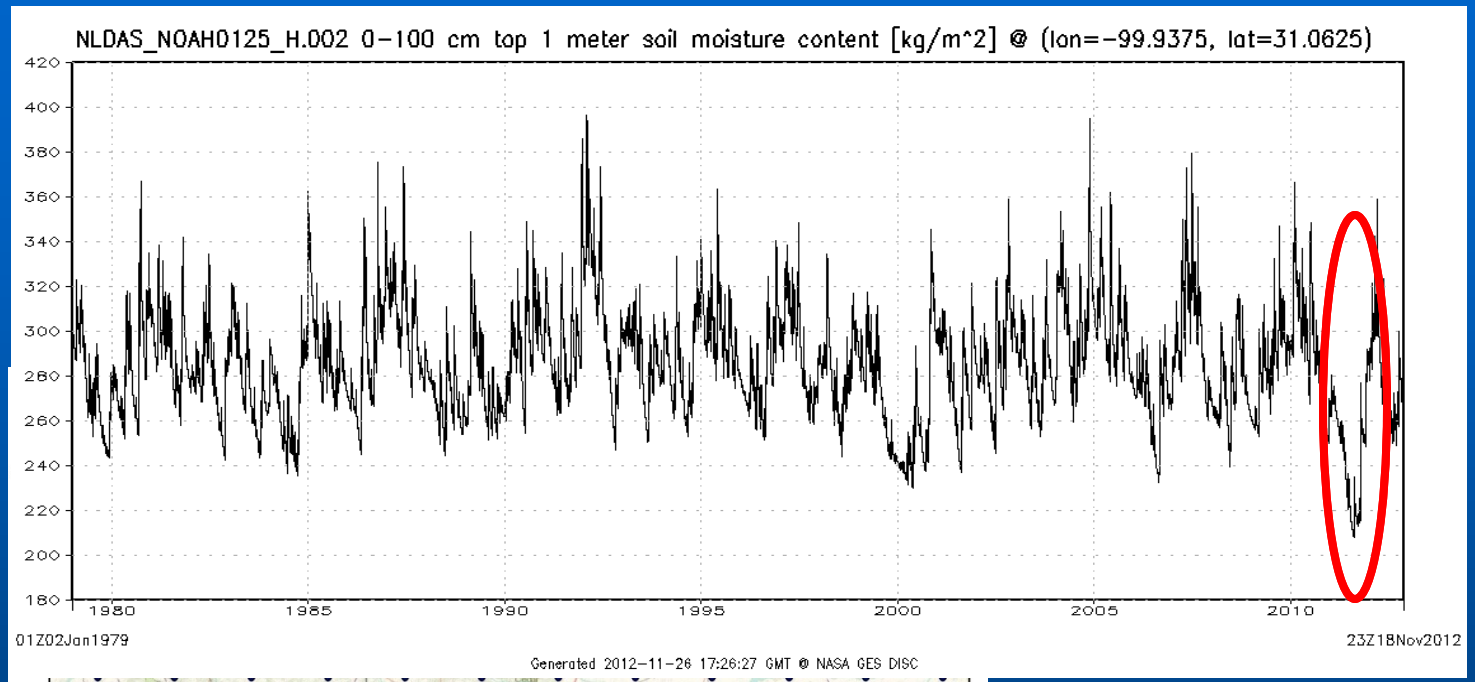
Directory Name:

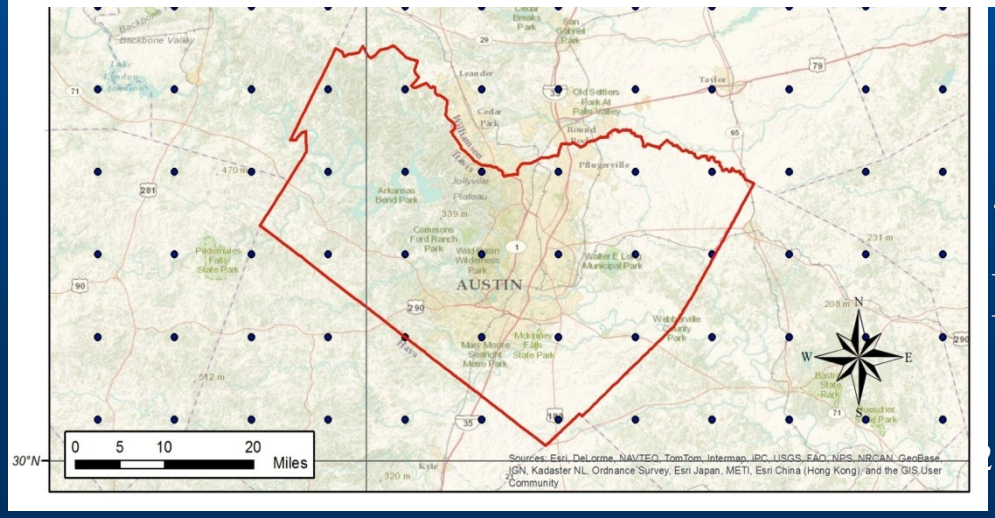
NLDAS.FORA012_H.002.APCPsfc

File Names:

NLDAS.FORA012_H.002.APCPsfc.Y000-X001.bin
NLDAS.FORA012_H.002.APCPsfc.Y000-X002.bin
…….
NLDAS.FORA012_H.002.APCPsfc.Y080-X300.bin
…….
NLDAS.FORA012_H.002.APCPsfc.Y223-X463.bin

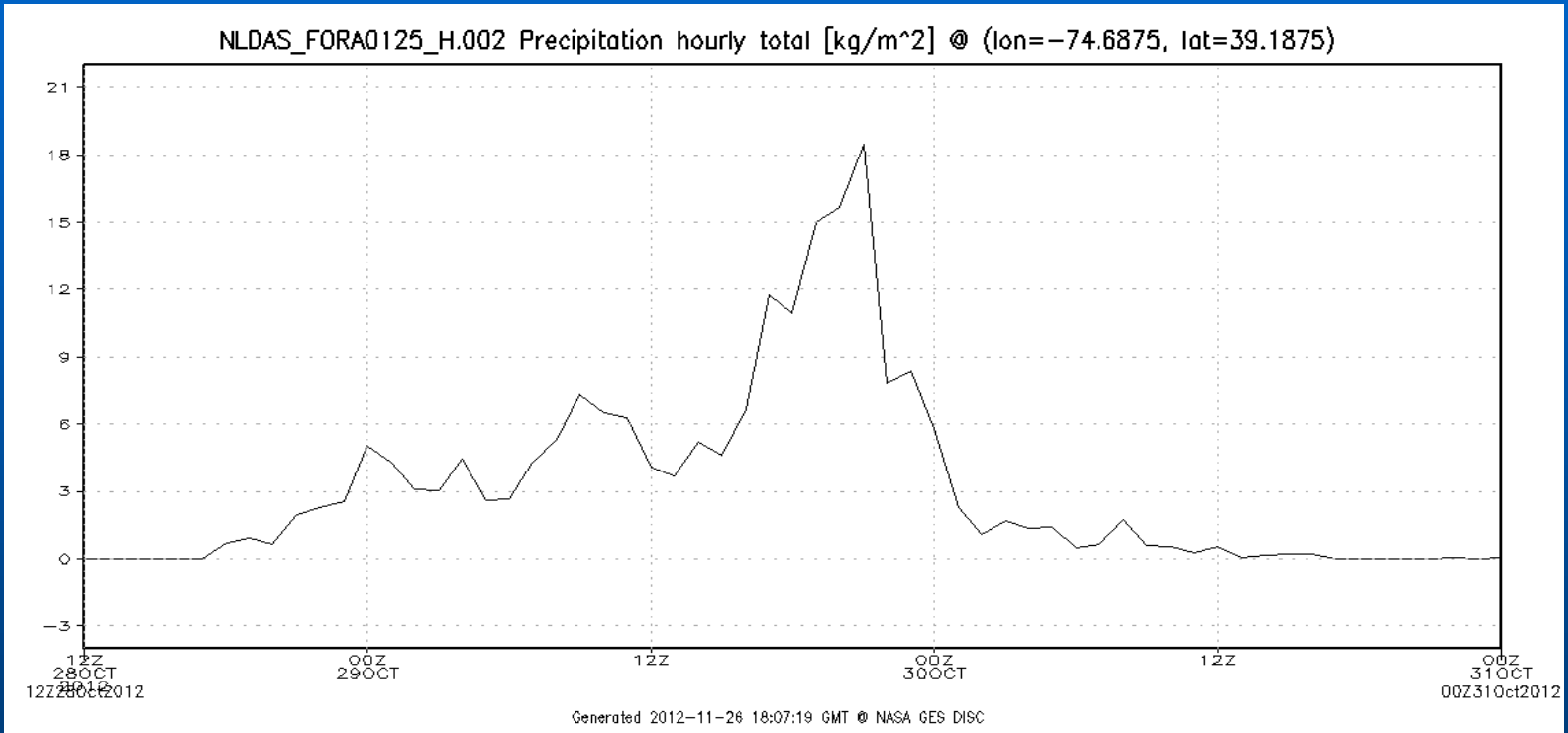
TNRIS* Use Case



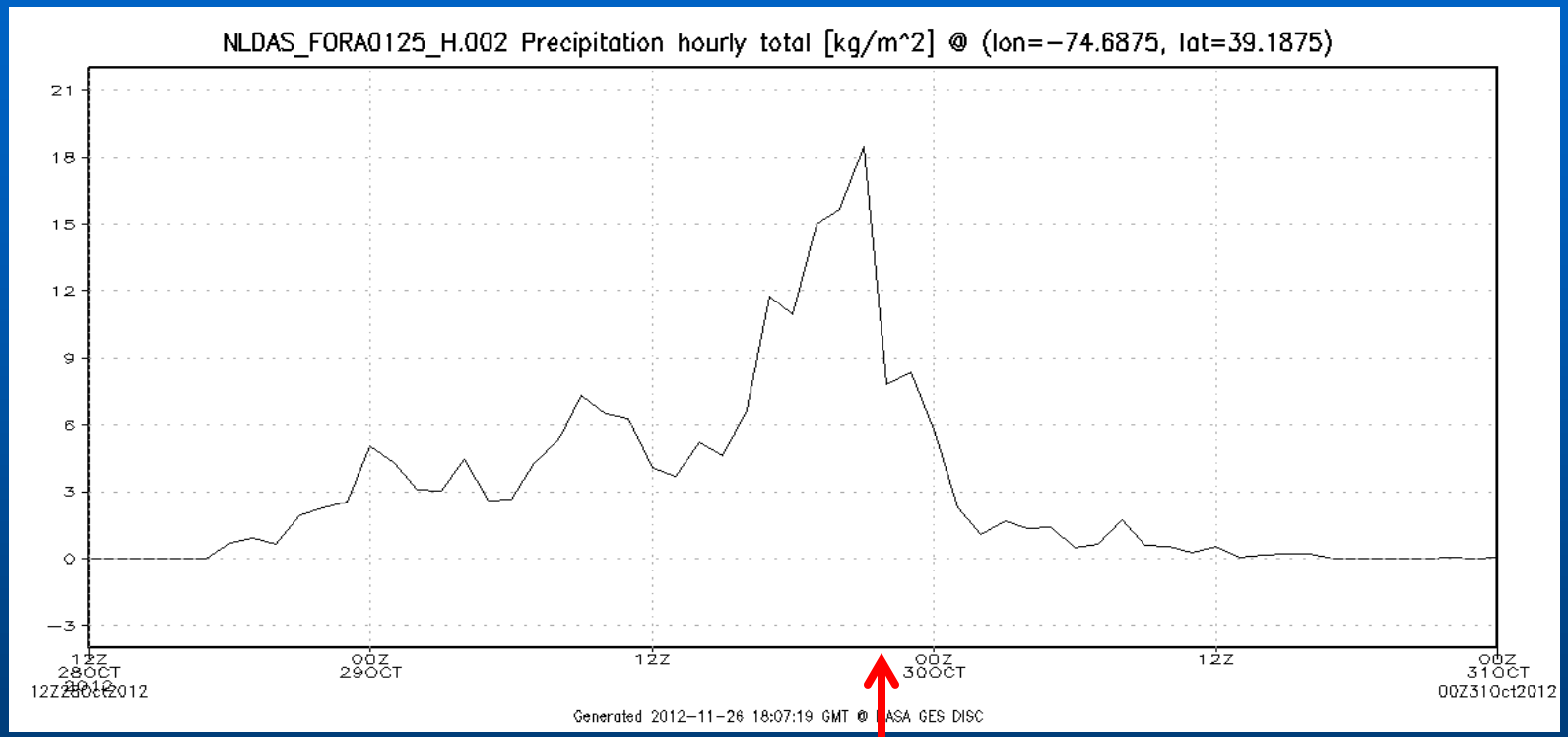


*Texas Natural Resources Information System

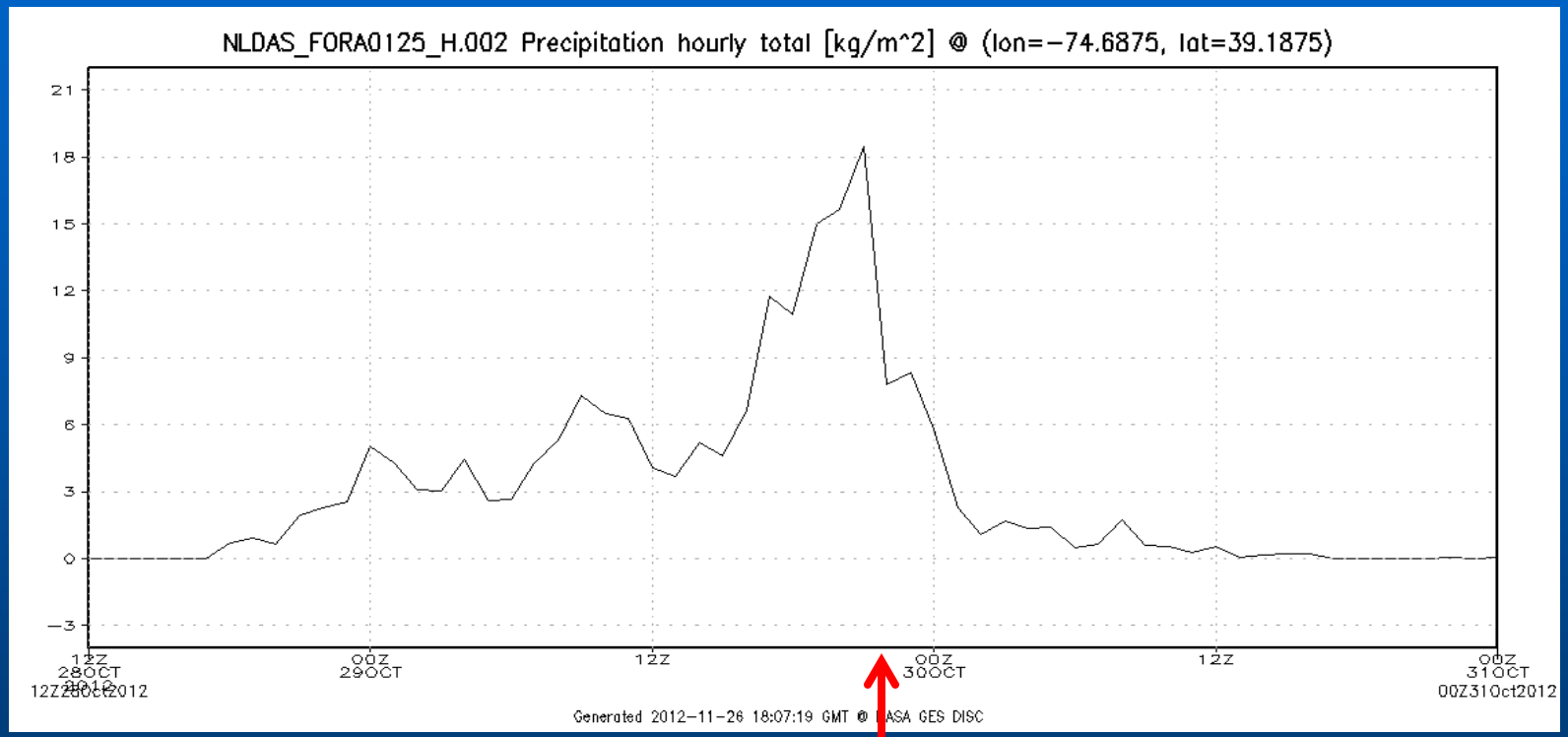
Hurricane Sandy



Hurricane Sandy



Hurricane Sandy



Sump pump failed

NLDAS parameters with time series access available (0.125° hourly, 1979-present)

Forcing

- Precipitation hourly total
- 2-m above ground temperature
- 10-m above ground zonal wind speed
- 10-m above ground meridional wind speed
- Potential evaporation

Noah

- 0-100 cm top 1 meter soil moisture content
- 0-10 cm soil temperature
- Surface runoff
- Total evapotranspiration

Summary and a Look Ahead

- Data reorganization as time series is a solution for the "Digital Divide" problem.
- Solves a longstanding problem of CUAHSI user community in accessing NASA data.
- Scaling issue w/ this solution.
- Looking beyond the ACCESS project, need a solution that does not involve data reorganization.
- Parallel processing (e.g., Hadoop)

Questions?

More details: IN23B-1500, today 1:40 pm

Overall Project: IN31A-1494, Wednesday 8:00 am