Between a Map and a Data Rod

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.

University of Texas at Austin: David Maidment, Tim Whiteaker, and David Arctur

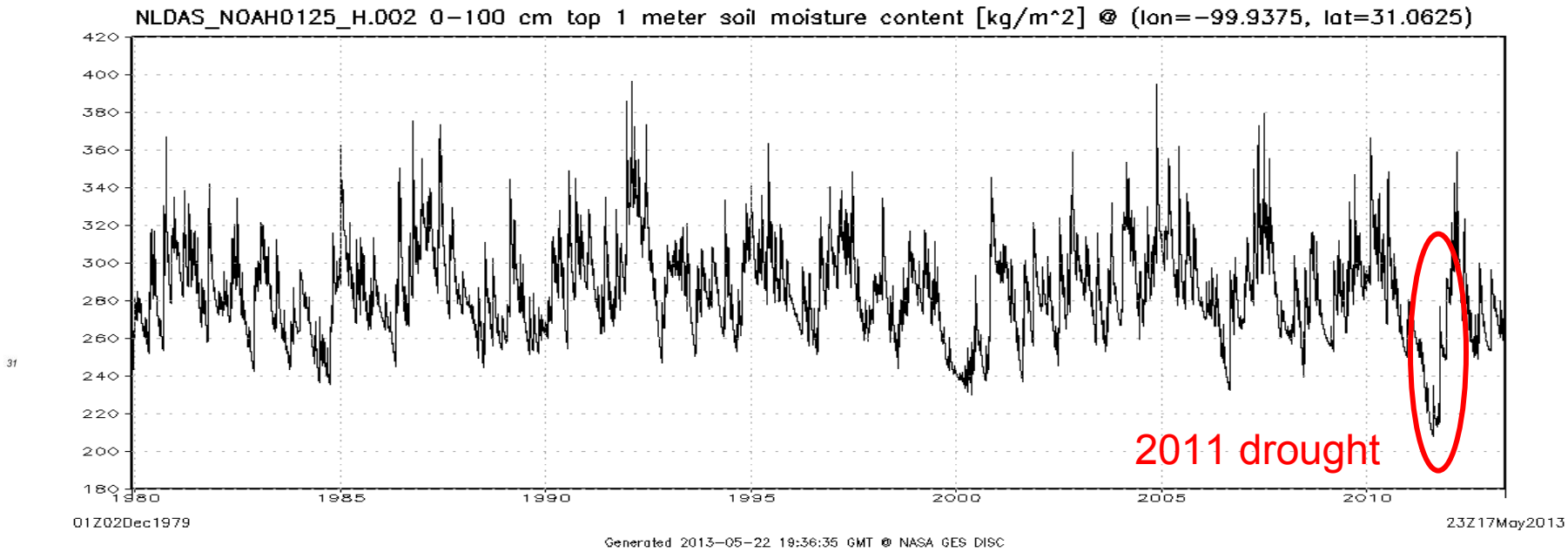
GSFC/Hydrological Sciences Lab: Christa Peters-Lidard, David Mocko, Dalia Kirschbaum, Matthew Rodell

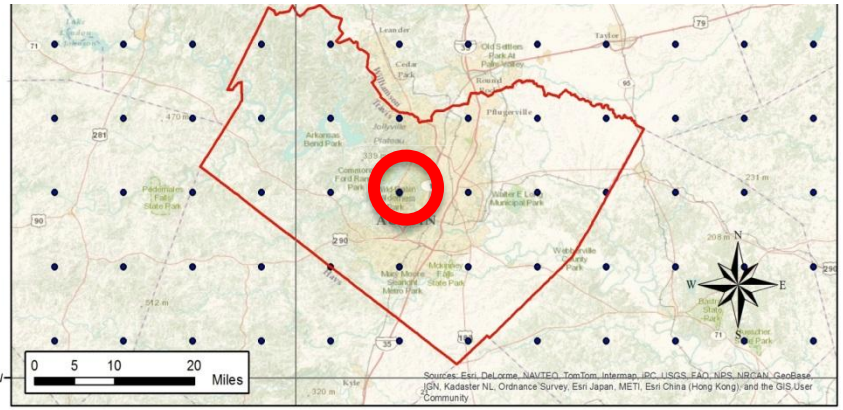
Brigham Young University: Daniel Ames

NASA ACCESS Program
NNH11ZDA001N-ACCESS
NNH13ZDA001N-ACCESS

Outline

- Motivation and background
- “Digital Divide” problem
- Solution: Pre-generated vs. on-the-fly
- Tiling, between a map and a data rod
- Summary and ongoing work





Time Series of top 1 meter soil moisture from NLDAS-2 Noah model, near the center of Texas (100W, 31N)

Courtesy of David R. Maidment
Center for Research in Water Resources
University of Texas at Austin

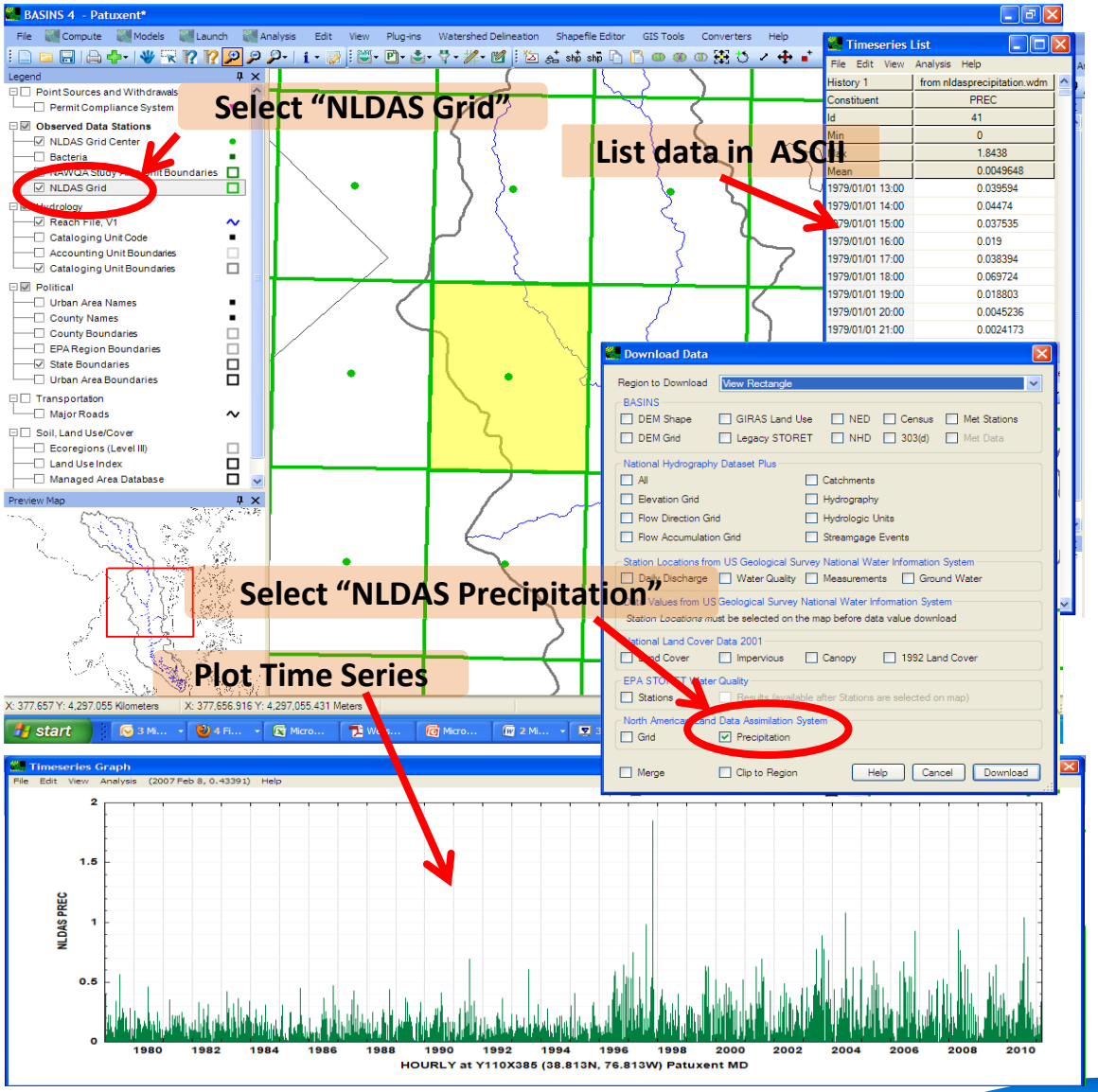
EPA BASINS[1] Prototype

Latitude x

DATA

Curtain

Time



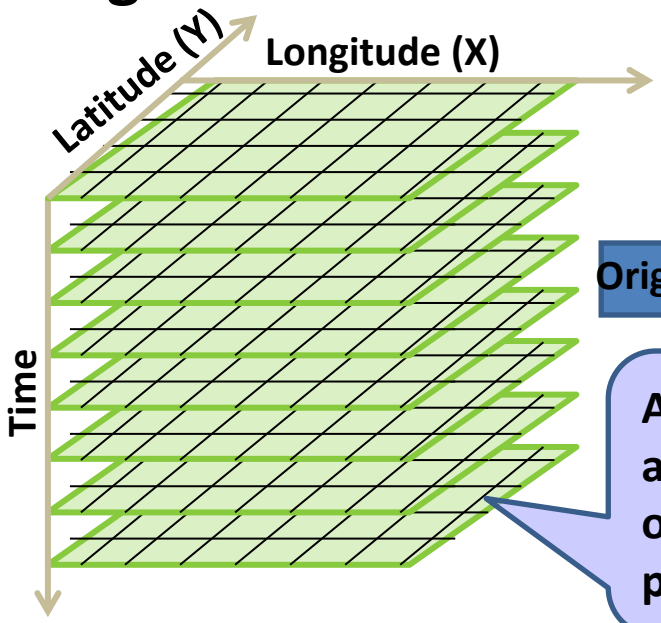
[1]Better Assessment Science Integrating Point and Nonpoint Sources

Digital Divide Problem … Orthogonal

	Original GRIB Files								Data Rods Binary Files				
Noah LSM	Dimension lat x lon	Total # of Grids	# of /Files /day	Total # of years	Total # of Files	File Size (MB)	Total Vol (TB)	Land Fraction	# of Files /param	# of param	Total # of Files	File Size (MB)	Total Vol (TB)
NLDAS	224 x 464	103936	24	37	324120	6.8	2.2	0.7321	76088	21	1597848	1.295	2.07
GLDAS	600 x 1440	864000	8	16	46720	15.2	0.71	0.2813	243003	13	3159039	0.183	0.58
Total					370840		2.91				4756887		2.65

Data Rods: A Simple Solution for Bridging the Digital Divide

Original Data Archive

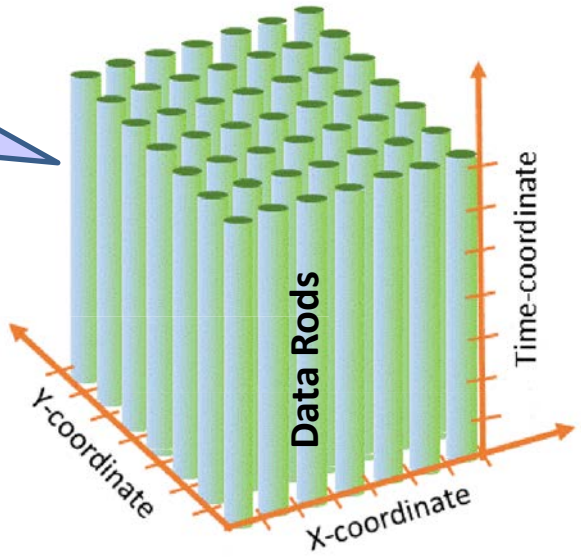


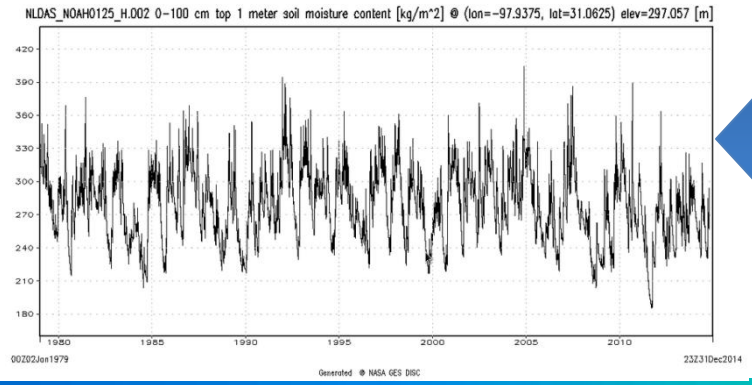
One variable
one grid point
all time steps
per file

Original data reorganized as …

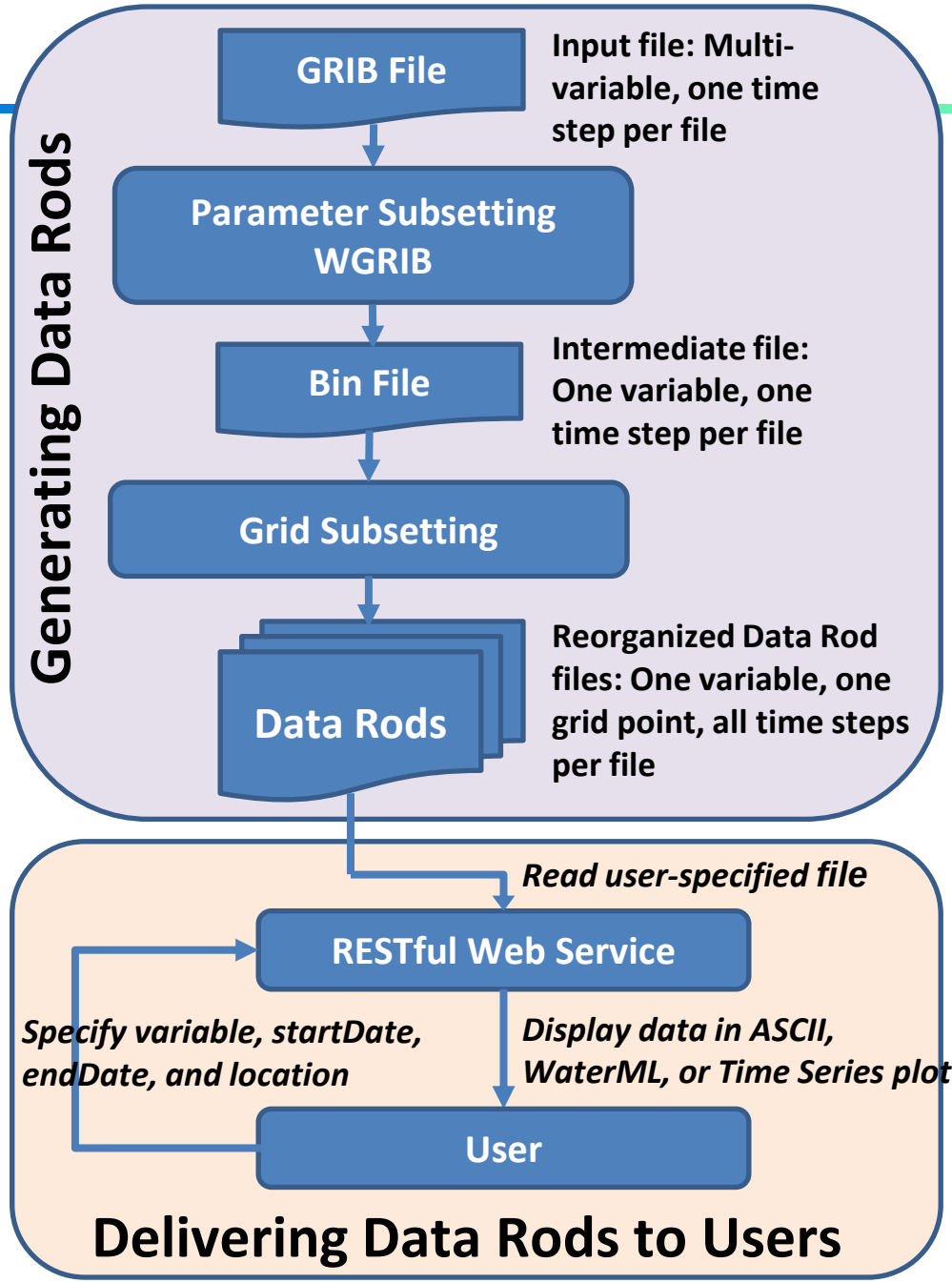
All variables
all grid points
one time step
per file

Reorganized Data Archive





More than 324,000 time steps
(37 years) plotted in ~ 1 second


Generating Data Rods
GRIB File
Input file: Multi-variable, one time step per file
Parameter Subsetting WGRIB
Bin File
Intermediate file: One variable, one time step per file
Grid Subsetting
Data Rods
Reorganized Data Rod files: One variable, one grid point, all time steps per file
Read user-specified file
RESTful Web Service
Specify variable, startDate, endDate, and location
Display data in ASCII, WaterML, or Time Series plot
User
Delivering Data Rods to Users

Global Level 3 (Gridded) Single Variable NASA Earth Science Data To Time Series ("Data Rods") Using NCO, NetCDF, and Giovanni

F. Files for previous years: Stored across several file systems (for parallel I/O) as separate, month-long, global lat-lon data cubes, with 3rd dimension along time

Data cubes

e.q., last 4 years of data at lat=42.5, lon=-124.9

G. *NCO subsets time series of single grid point along time dimension from each cube*

E. *Requests time series for years x-1, x-2, x-3, …*

A. *Client (e.g., HydroDesktop) sends data request with spatial-temporal constraints*

B. *Requests time series for current year x*

D. *NCO subsets time series of single grid point along time dimension from all files in current year*

H. *NCO concatenates all the responses. Custom software delivers resulting single time series as ASCII, WaterML1, WaterML2, or plot*

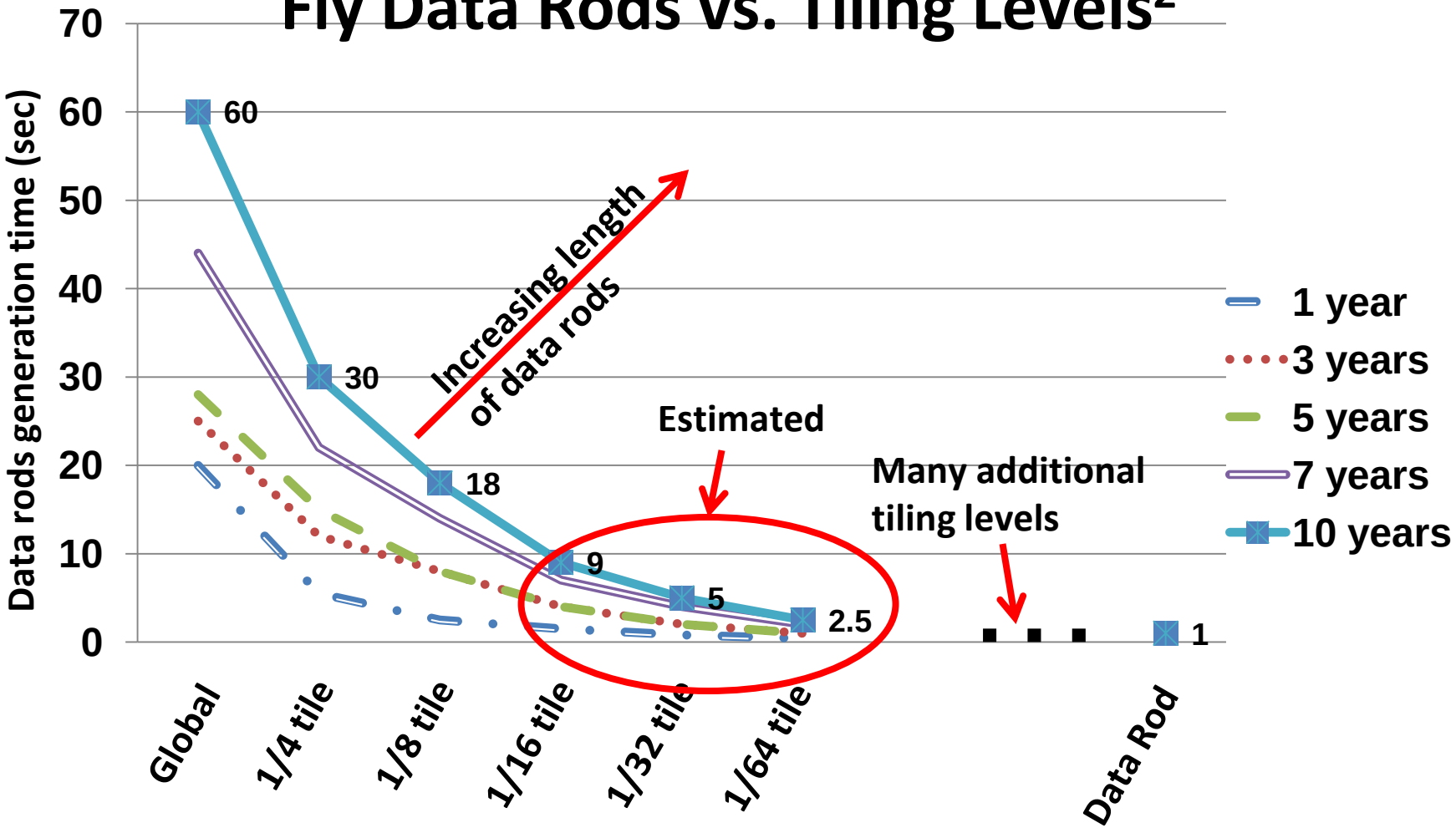
J. Files for previous years in Giovanni Cache, concatenated into data cubes, and stored in other file systems

C. Files for current year: Stored in Giovanni Cache as single variable, single time-step, global lat-lon layers, updated as new data become available

I. Archived data, subset by variable, converted to "fast file" format, and stored in Giovanni Cache

GIOVANNI

Generation Time for TRMM[1] On-the-Fly Data Rods vs. Tiling Levels[2]



[1]Tropical Rainfall Measuring Mission

[2]Tiling: Dividing the data set grid into subgrids (e.g., ¼ tiling for TRMM divides its global grid into 4 equal subgrids).

Teng, W., H. Rui, R. Strub, and B. Vollmer. Optimal reorganization of NASA earth science data for enhanced accessibility and usability for the hydrological community, J. Amer. Water Resources Assoc. Forthcoming, 2016.

Data Rods Metrics

2013-01-01 to 2015-11-30

Product	Protocol	# Users	# Files	Volume (GB)
NLDAS_FORA0125_RODS	FTP	8	17,733,371	20,585
NLDAS_NOAH0125_RODS	FTP	5	16,741,164	19,580
GLDAS_NOAH025_RODS	FTP	13	39,654,230	5,946
NLDAS_FORA0125_RODS	NLDAS_FORA	445	241,470	92
NLDAS_NOAH0125_RODS	NLDAS_NOAH	286	187,923	95
GLDAS_NOAH025_RODS	GLDAS_NOAH	487	62,680	36
NLDAS_FORA0125_RODS	WEB_LDAS	300	118,561	1,784
NLDAS_NOAH0125_RODS	WEB_LDAS	336	452,667	6,401
GLDAS_NOAH025_RODS	WEB_LDAS	392	79,844	94
Total			75,271,910	54,613

Users: Number of distinct users

FTP: Get data rods via FTP

WEB_LDAS: Access data rods in ASCII or as Time Series plot via GES DISC Web services

Other protocols: Access data rods via CUAHSI HIS (HydroDesktop)

Summary and Ongoing Work

- Developed operational way to reorganize data that is optimal for user communities that are point-time series oriented.
- Solved the motivating problem presented by CUAHSI HIS: create time series of hourly data, for single grid cells for entire period of coverage.
- Key to all solutions is to reorganize data that is optimal for desired method of data access.
- Ongoing investigation into tiling of data set grids has yielded results that are very encouraging for significantly reducing the generation time for data rods.

Extras

Data Rods

http://disc.sci.gsfc.nasa.gov/hydrology/data-rods-time-series-data