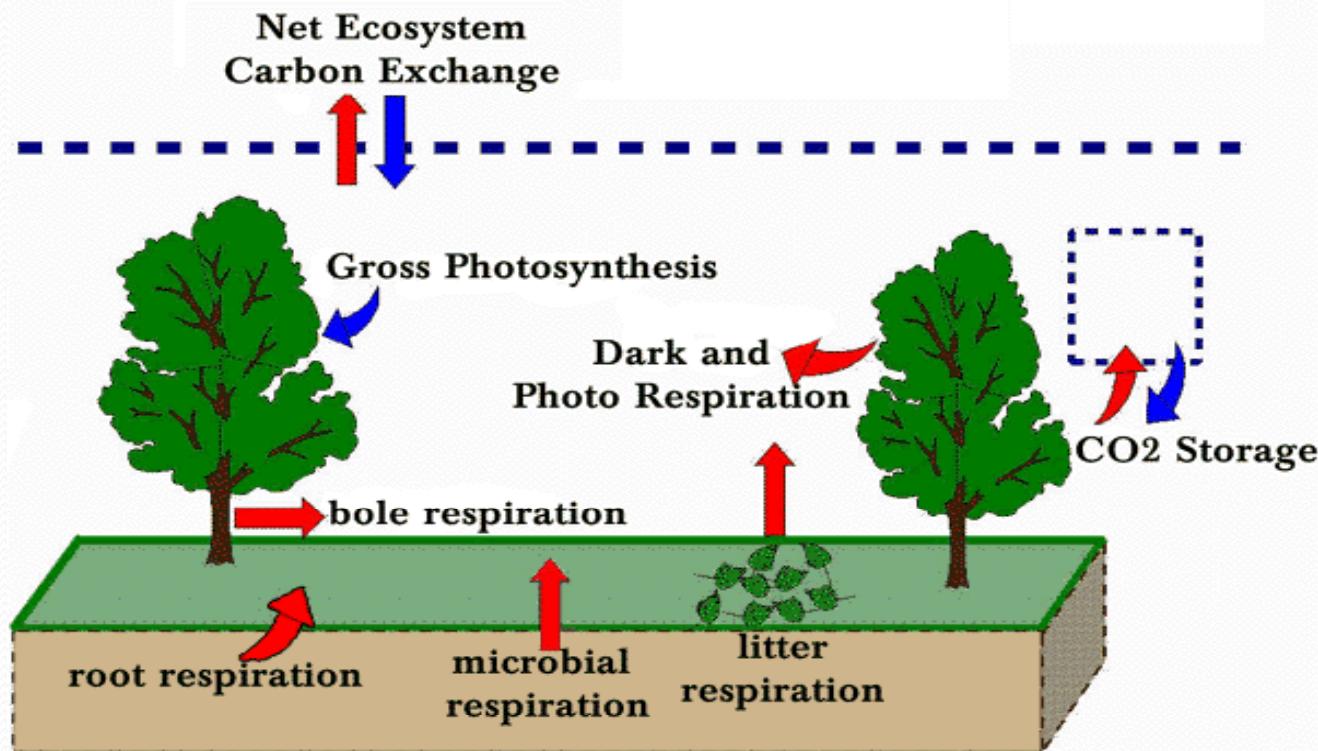


Enabling Carbon Flux Studies at the Regional and Ecosystem Scale

Deborah Agarwal
BWC technical team
16 July 2007

Canopy Carbon Balance

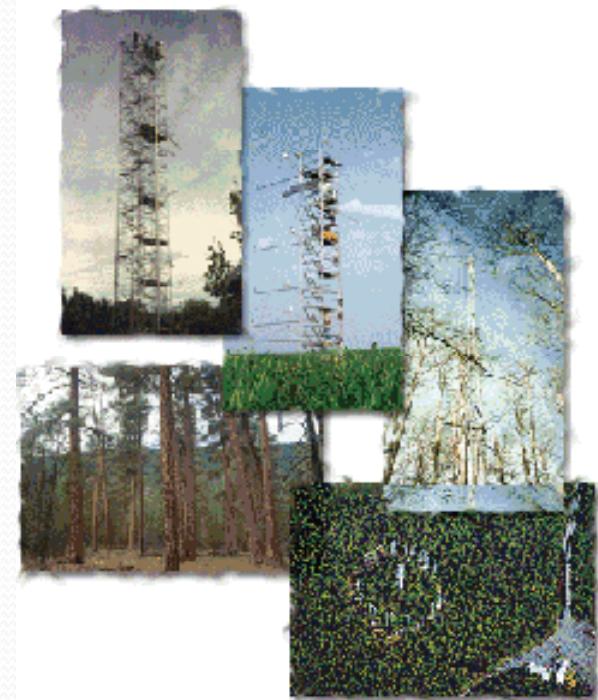


$$F_c + F_{\text{storage}} = -\text{NEE} = P_{\text{net}} + R_{\text{leaf}} + R_{\text{wood}} + R_{\text{roots}} + R_{\text{microbes}}$$

1. Applications of eddy covariance measurements, Part 1: Lecture on Analyzing and Interpreting CO₂ Flux Measurements, Dennis Baldocchi, CarboEurope Summer Course, 2006, Namur, Belgium (<http://nature.berkeley.edu/biometlab/lectures/>)

Carbon-Climate Analysis Goals

- Towers measure consistent carbon flux and micrometeorological parameters
- Tower researchers quality check data and then provide the data to regional archives.
- Regional and global carbon-climate analysis activities rely on data from regional archives
- Recent La Thuile workshop is gathering over 700 site-years of data available from over 200 sites around the world.

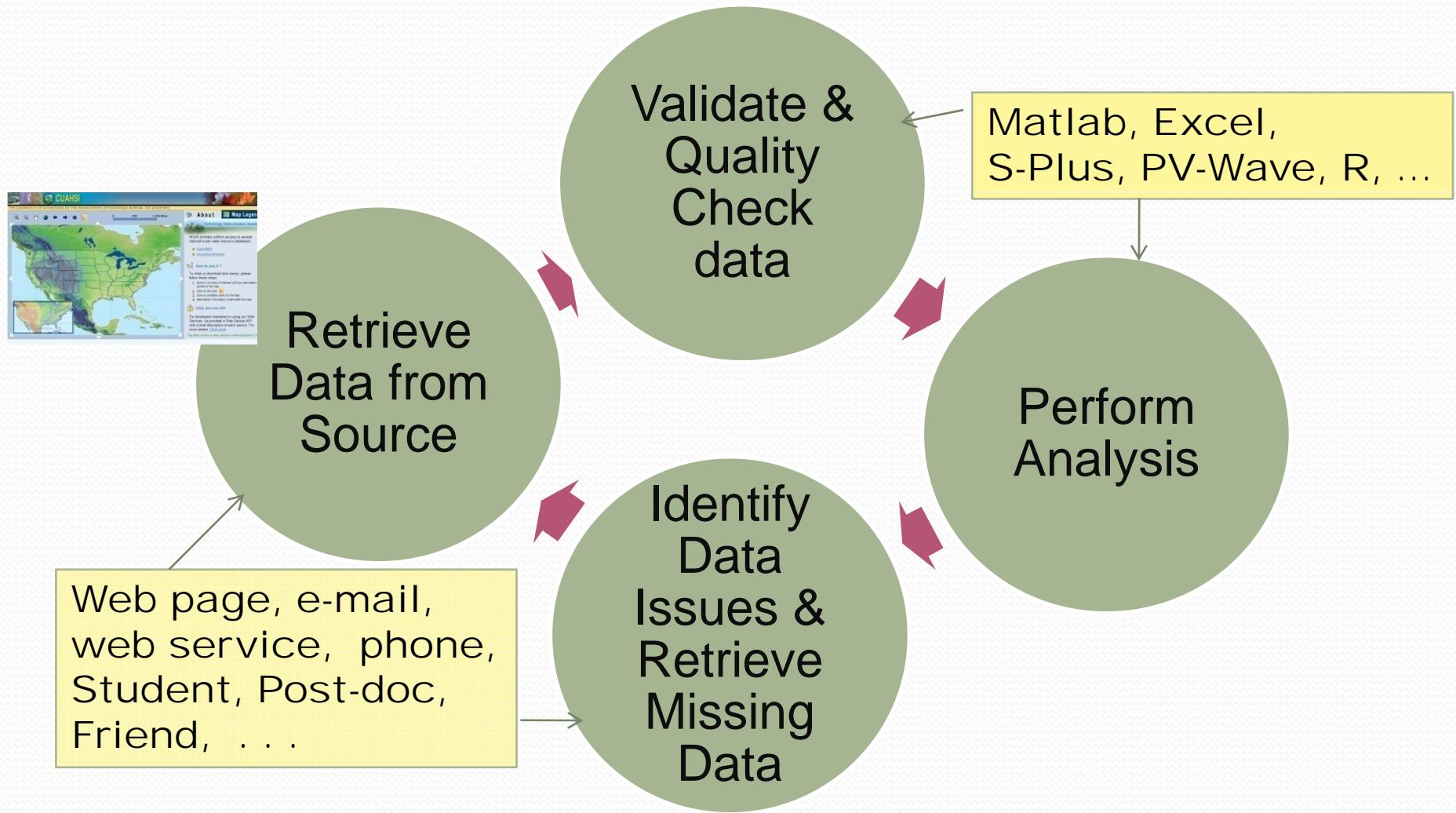


Measurements Are Often Not Simple or Complete

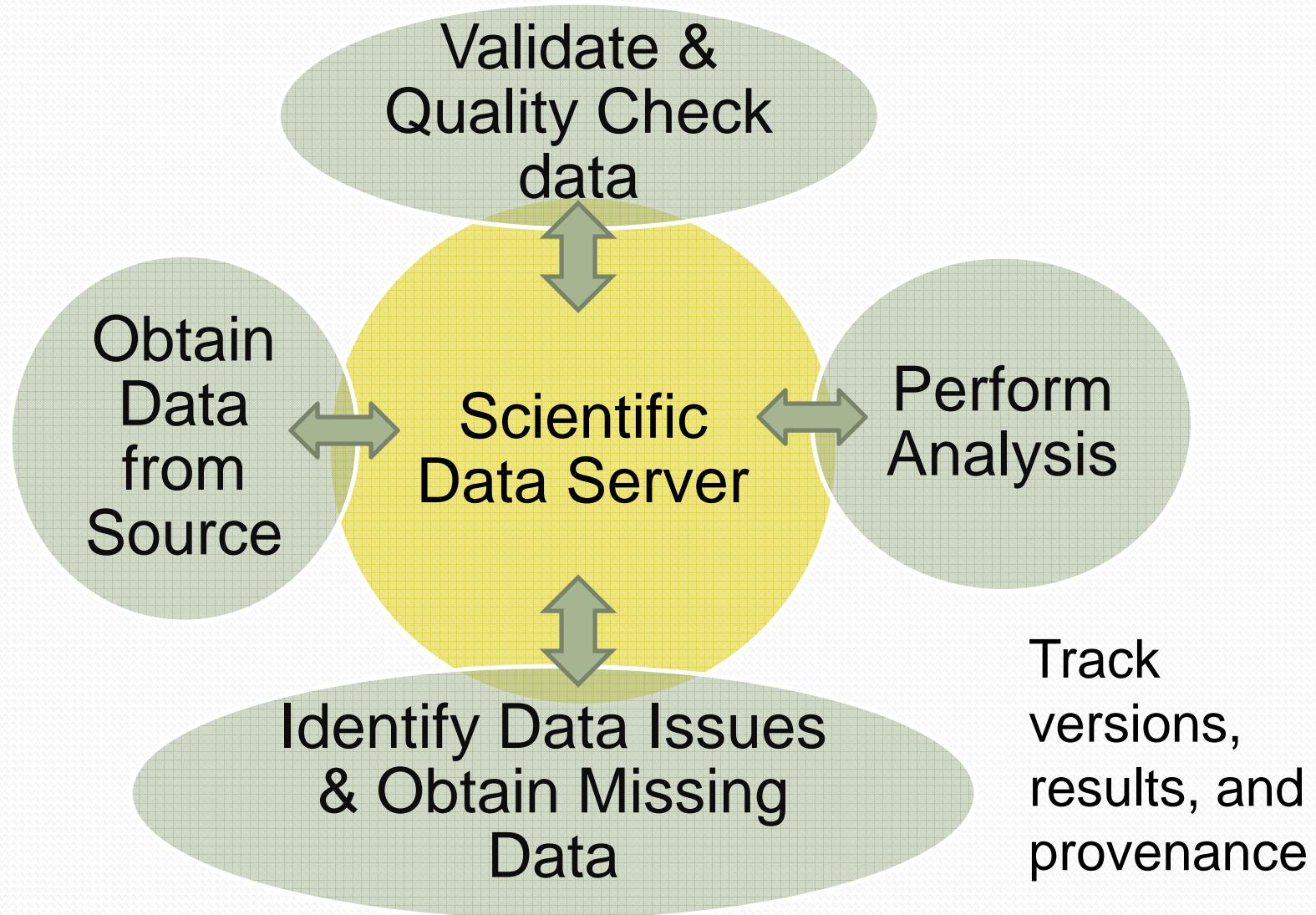
- Gaps in the data
 - Quiet nights
 - Bird poop
 - High winds
 -
- Difficult to make measurements
 - Leaf area index
 - Wood respiration
 - Soil respiration
 -
- Localized measurements – tower footprint
- Local investigator knowledge important
- PIs' science goals are not uniform across the towers



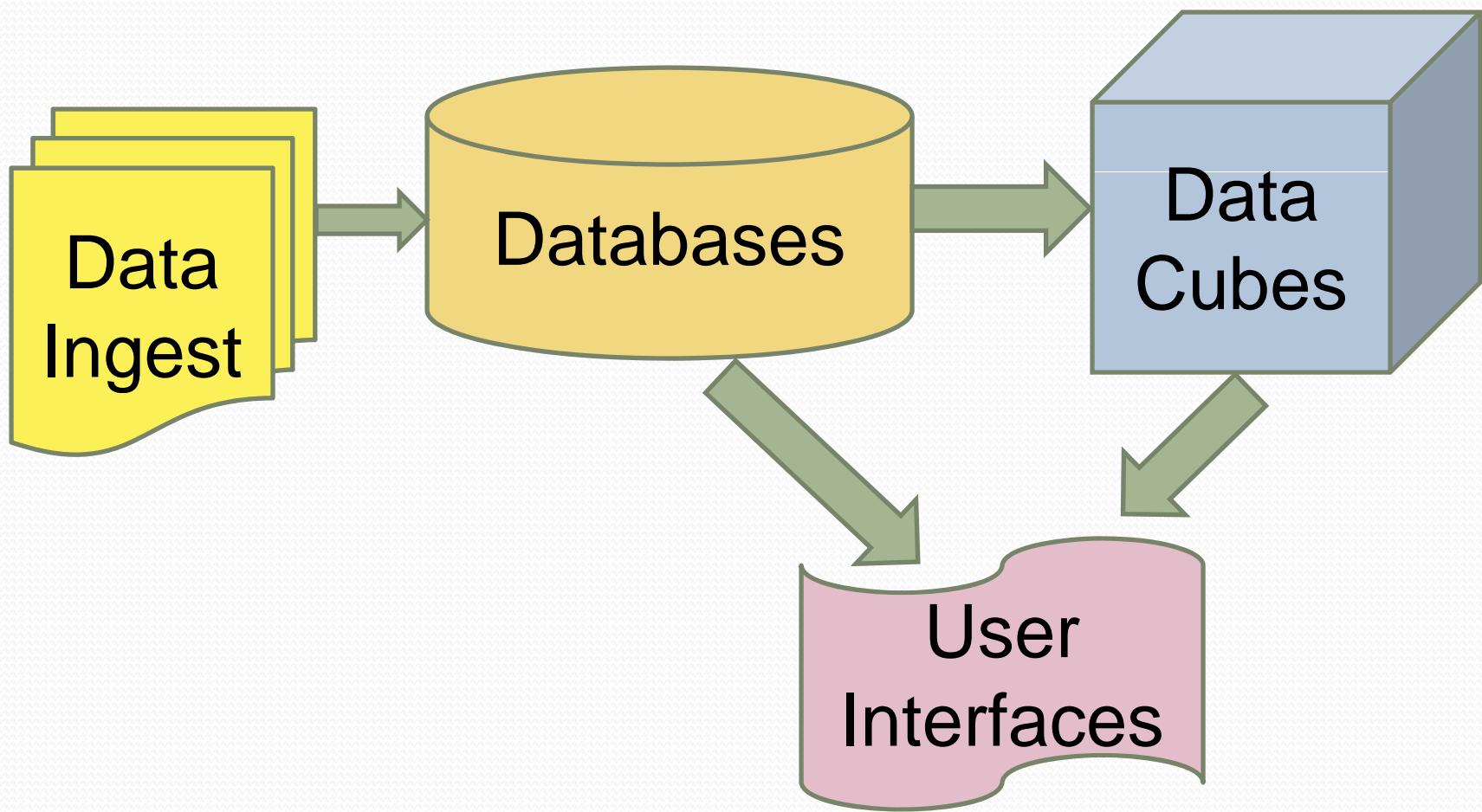
Typical Data Analysis Flow Today



Target Data Analysis Flow

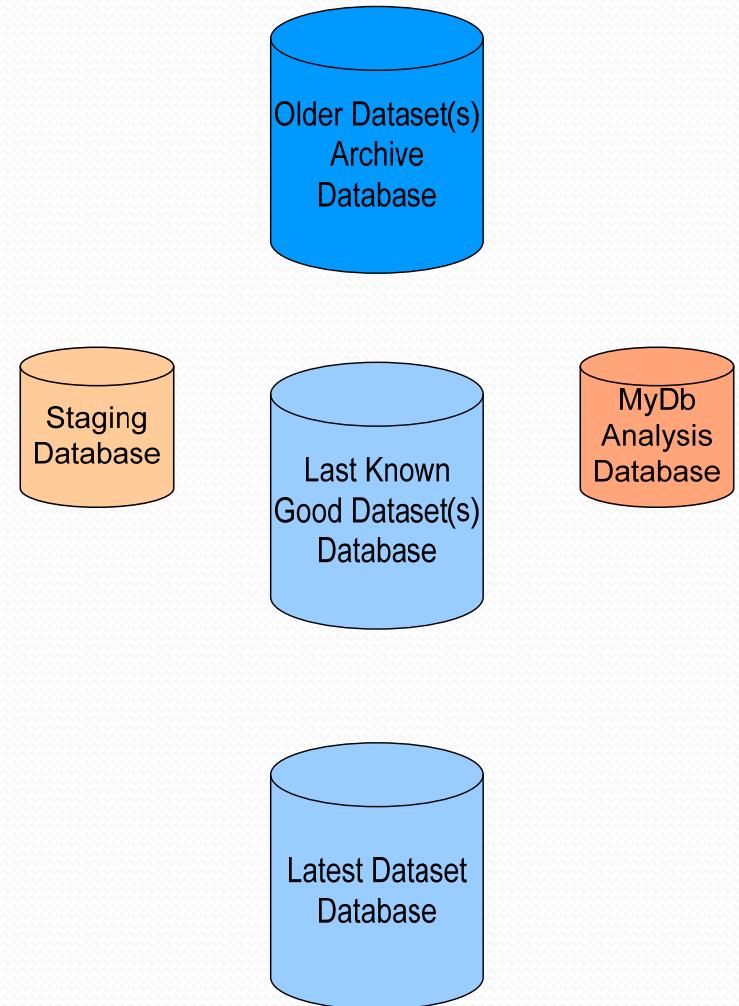


Scientific Data Server - Overview



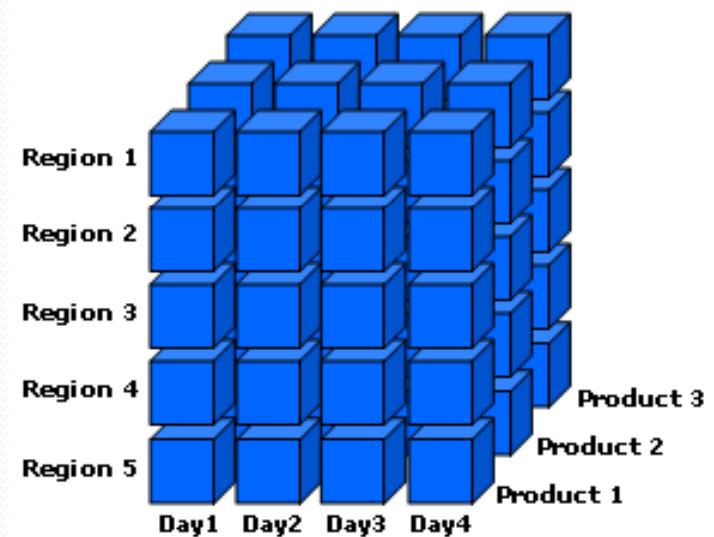
Database

- All descriptive metadata and data held in relational databases
 - Metadata is important too!
- While separate databases are shown, the datasets may actually reside in a single database
 - Mapping is transparent to the scientist
 - Separate databases used for performance
 - Unified databases used for simplicity
- New metadata and data are staged with a temporary database
 - Minimal quality checks applied
 - All name and unit conversions
- Data may be exported to flat file, copied to a private MyDb database, directly accessed programmatically, or ?

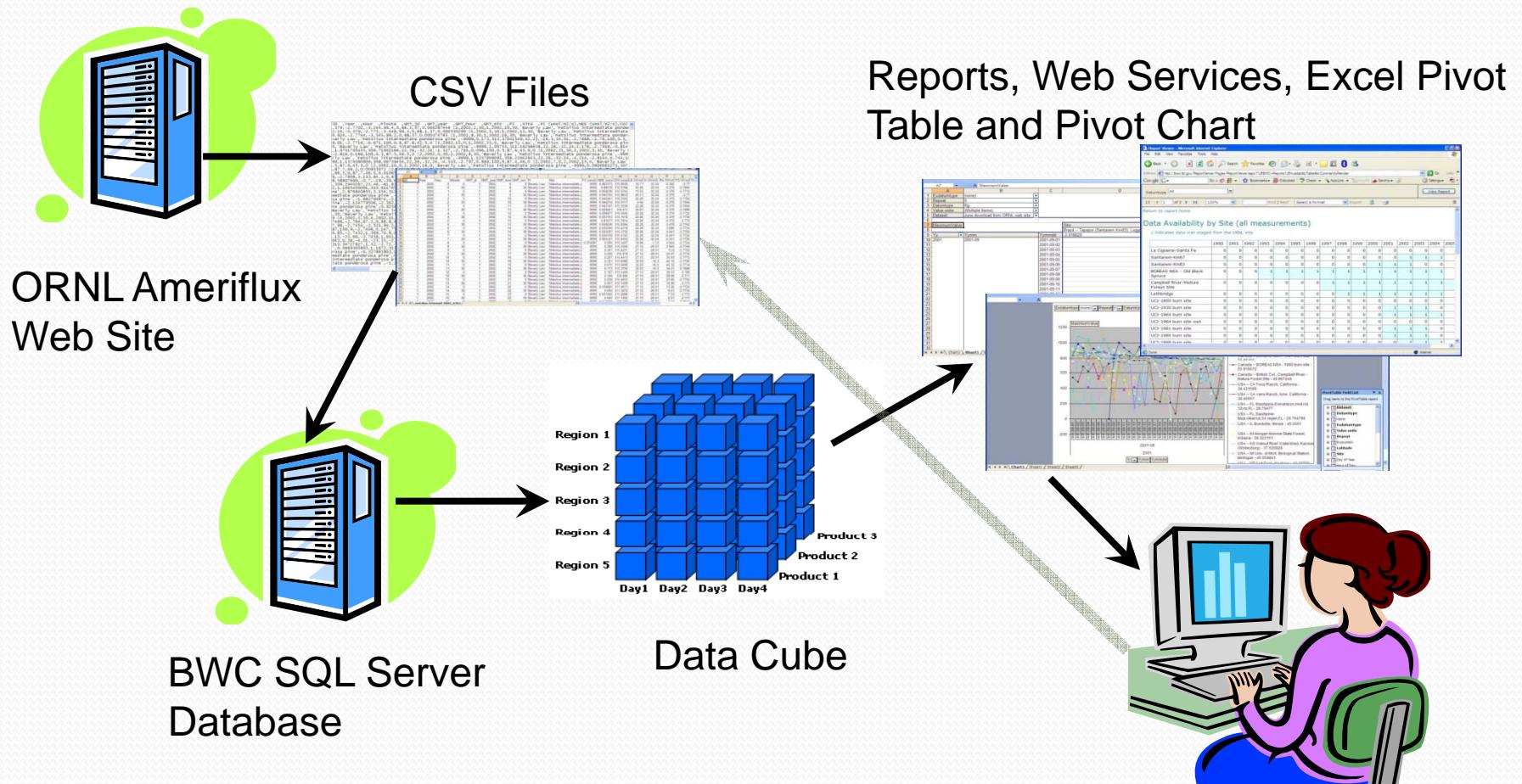


Data Cubes

- A data cube is a database specifically for data mining (OLAP)
 - Simple *aggregations* (sum, min, or max) can be pre-computed for speed
 - Additional calculations can be computed dynamically or pre-computed
 - Both operate along *dimensions* such as time, site, or datumtype
 - Constructed from a relational database
 - A specialized query language (MDX) is used
- Client tool integration is evolving
 - Excel PivotTables allow simple data viewing
 - More powerful analysis and plotting using Matlab and statistics software

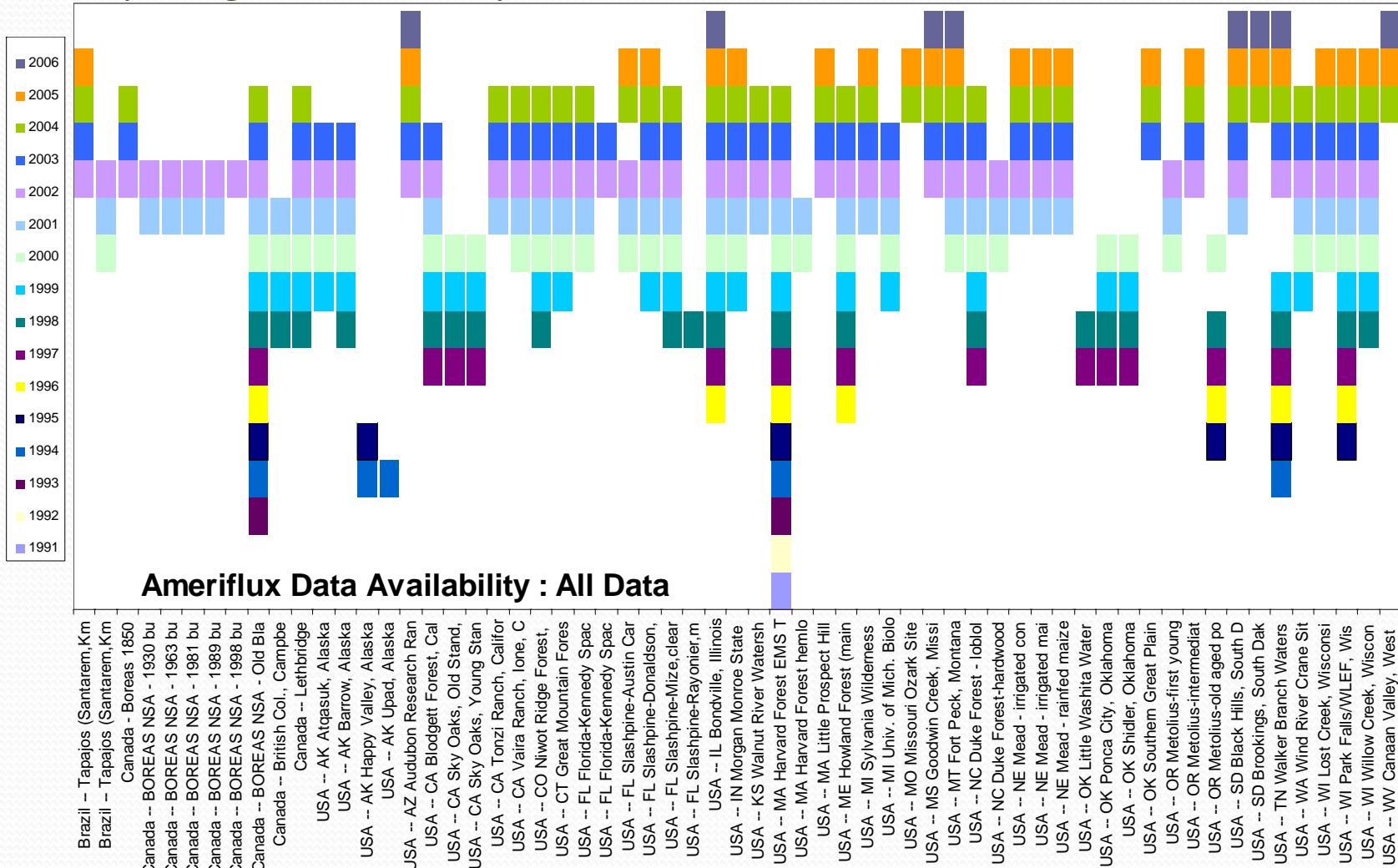


Scientific Data Server – User Interface



Browsing For Data Availability

Sites Reporting Data Colored by Year



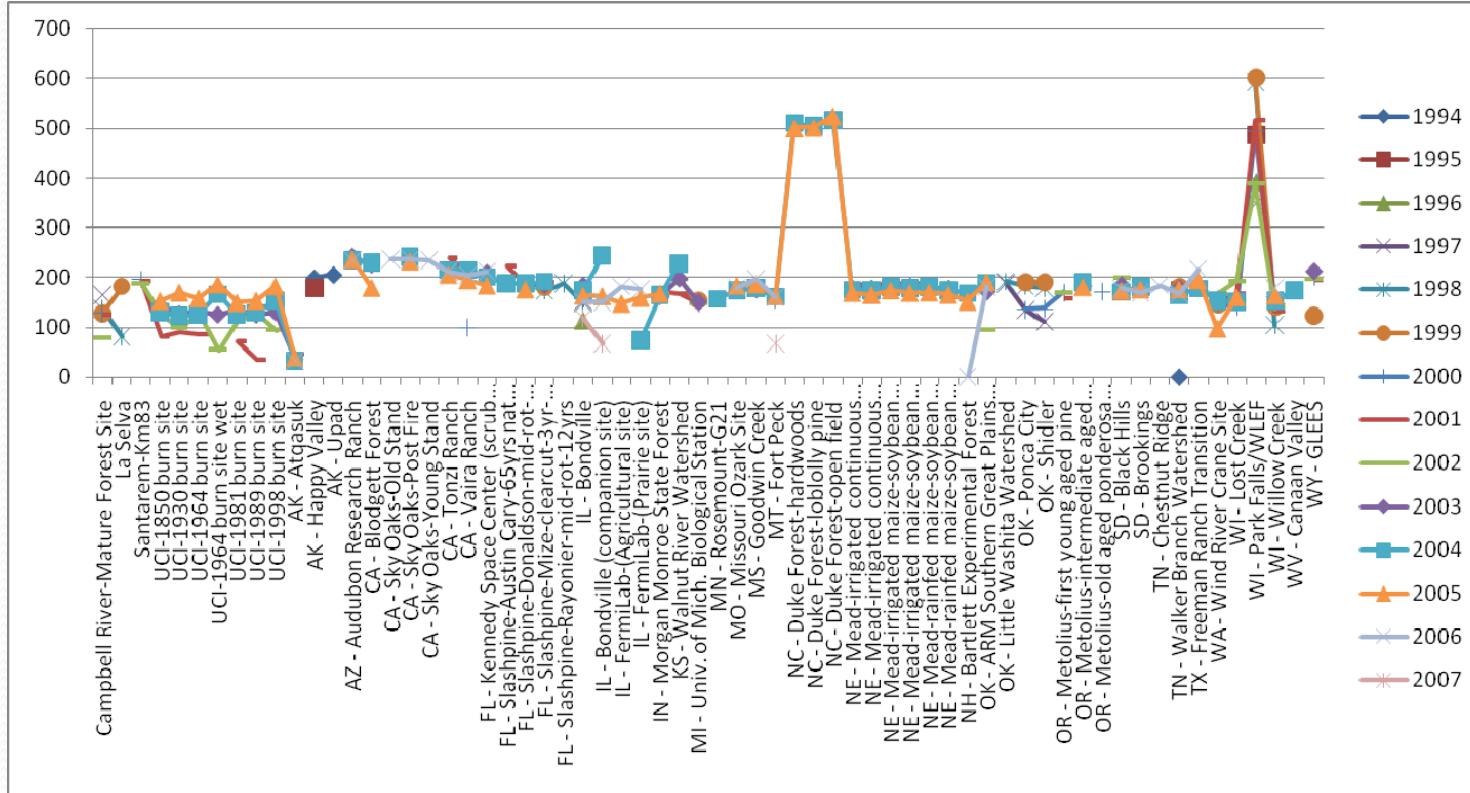
Required variable reporting by site by year

- Each row corresponds to one site-year
- Each cell corresponds to one site year of (FC, CO2 or SCO2, UST, PAR or Rg, TA, and Rh or H2O).
- Color indicates:
 - Red – likely not enough for processing - % < .3 reported (roughly less than 5K of 17.5K)
 - Green – likely not enough for processing .3<%<.999
 - Yellow – may not be good for processing due to gap-filling - % > .999
- Red CO2 (second column) can be ignored for cropland/grassland sites
- Sites shown are just a sample

Data Availability	Low Threshold			High Threshold			FC	CO2/SCO2	UST	PAR/RG	TA	RH/H2O
	0.3000	0.3000	0.3000	0.9990								
CA - Blodgett Forest												
1997	0.0000	0.2241	0.4149	0.4937	0.2443	0.4150	OR - Metolius-intermediate aged ponderosa pine					
1998	0.0000	0.2913	0.5228	0.6013	0.4150	0.6123	2003	0.4741	0.6901	0.8336	0.9493	1.0000
1999	0.5188	0.5124	0.8202	0.9888	0.6123	0.9207	2004	0.7551	0.8921	0.8678	0.9934	0.9493
2000	0.8081	0.8792	0.8504	0.9639	0.9207	0.9639	2005	0.7470	0.9664	0.0000	0.0000	0.9934
2001	0.8397	0.9112	0.863	0.9845	0.9639	0.9846	SD - Black Hills					
2002	0.8038	0.3247	0.8523	0.5623	0.9845	0.9259	2001	0.2195	0.0000	0.3359	0.3184	0.2321
2003	0.8389	0.9116	0.8327	0.9605	0.9246	0.7778	2002	0.0000	0.0000	0.4695	0.5733	0.3179
2004	0.7437	0.8284	0.9412	0.9697	0.7778	0.8087	2003	0.3329	0.0000	0.8141	0.8487	0.5733
2005	0.6702	0.6151	0.0000	0.0000	0.8307	0.0000	2004	0.6300	0.0000	0.8929	0.9305	0.8487
CA - Tonzi Ranch												
2001	0.5132	0.5616	0.8064	0.9999	0.7060	0.9172	2005	0.8682	0.0000	0.9522	0.9683	0.9305
2002	0.7325	0.8443	0.8754	1.0000	0.9983	1.0000	SD - Brookings	0.9233	0.0000	0.0000	0.9683	0.0000
2003	0.7857	0.8962	1.0000	1.0000	1.0000	1.0000	2004	0.4628	0.0000	0.8267	0.9918	0.6897
2004	1.0000	1.0000	0.8534	0.9911	1.0000	0.9999	2005	0.7296	0.0000	0.9072	0.9970	0.9918
2005	0.7490	0.8541	0.9786	0.9957	0.9999	0.9984	2006	0.8034	0.0000	0.0000	0.9973	0.0000
2006	0.7712	0.9526	0.0000	0.0000	0.9984	0.0000	TN - Walker Branch Watershed					
CO - Niwot Ridge Forest												
1998	0.1671	0.1671	1.0000	0.9818	0.0744	0.8447	1994	0.0001	0.0000	0.7726	1.0000	0.0001
1999	1.0000	1.0000	0.9995	0.9989	0.8725	0.9831	1995	0.5992	0.9698	0.8237	0.9999	0.9972
2000	1.0000	1.0000	1.0000	0.955	0.9926	0.9538	1996	0.6143	0.9423	0.7946	0.9999	1.0000
2001	1.0000	1.0000	0.9954	0.9890	0.9866	0.6631	1997	0.5758	0.9598	0.8030	0.9890	0.9999
2002	1.0000	0.9999	0.9822	1.0000	0.9974	0.7888	1998	0.5614	0.7850	0.9624	1.0000	0.9999
2003	1.0000	1.0000	1.0000	1.0000	0.9977	1.0000	1999	0.9052	0.9408	0.9038	0.9973	1.0000
2004	1.0000	1.0000	0.0000	0.0000	1.0000	0.0000	2000	0.8883	0.0000	0.8591	1.0000	0.9973
2005	0.8421	0.0000	0.0665	0.0745	1.0000	0.0000	2001	0.8421	0.0000	0.0000	1.0000	0.0745
IL - Bondville												
1996	0.2122	0.0000	0.8388	1.0000	0.3510	1.0000	2003	0.6162	0.0000	0.8028	0.8870	0.8326
1997	0.7856	0.0000	0.9127	0.9999	1.0000	1.0000	2004	0.7149	0.0000	0.6937	0.5785	0.8870
1998	0.7289	0.0000	0.9307	1.0000	1.0000	1.0000	2005	0.4349	0.0000	0.6907	0.8038	0.5791
1999	0.8182	0.0000	0.7981	1.0000	1.0000	1.0000	2006	0.6642	0.0000	0.0000	0.8038	0.0000
2000	0.7158	0.0000	0.8720	1.0000	1.0000	1.0000	WA - Wind River Crane Site					
2001	0.6979	0.0000	0.8740	0.9973	1.0000	0.9973	1998	0.4696	0.4665	0.8691	0.8253	0.7369
2002	0.7216	0.0000	0.8973	0.9839	0.9973	0.9839	1999	0.8105	0.8529	0.7735	0.9448	0.8902
2003	0.7765	0.0000	0.7486	0.9356	0.9839	0.9357	2000	0.6081	0.6532	0.8588	0.9400	0.9814
2004	0.5894	0.0000	0.8522	0.9608	0.9349	0.9608	2001	0.8578	0.8276	0.9346	0.9163	1.0000
2005	0.7518	0.0000	0.7910	0.8997	0.9608	0.8997	2002	0.9332	0.9376	0.9518	0.9995	1.0000
2006	0.7748	0.0000	0.2183	0.2301	0.8997	0.2301	2003	0.8640	0.8848	0.9091	1.0000	1.0000
2007	0.2179	0.0000	0.0000	0.0000	0.2301	0.0000	2004	0.8796	0.8838	0.2522	0.2736	1.0000
NC - Duke Forest-hardwoods												
2001	1.0000	0.8523	1.0000	1.0000	1.0000	1.0000	WI - Park Falls/WLEF					
2002	1.0000	0.9496	1.0000	1.0000	1.0000	1.0000	1995	0.0000	0.3983	0.1389	0.3899	0.2476
2003	1.0000	0.9475	1.0000	0.9402	1.0000	1.0000	1996	0.0000	0.4060	0.3042	0.3994	0.2160
2004	1.0000	0.9910	1.0000	0.9997	1.0000	1.0000	1997	0.0000	0.4603	0.2838	0.4253	0.2479
2005	1.0000	0.9792	0.0000	0.0000	1.0000	0.0000	1998	0.0000	0.4619	0.3565	0.4643	0.4349
OK - Shidler												
1997	0.2959	0.2685	0.9999	1.0000	0.2959	0.9999	1999	0.0000	0.4017	0.1054	0.4912	0.3647
1998	1.0000	0.8865	1.0000	1.0000	1.0000	1.0000	2000	0.0000	0.4055	0.1388	0.4122	0.3218
1999	1.0000	0.9660	0.2595	0.2595	1.0000	0.2563	2001	0.0000	0.4317	0.0000	0.3329	0.3531
2000	0.2595	0.2553	0.0000	0.0000	0.2595	0.0000	2002	0.0000	0.4182	0.2207	0.4813	0.2796
WI - Park Falls/WLEF												
2003	0.0000	0.3248	0.2359	0.4570	0.4355	0.4601	2004	0.0000	0.3248	0.2359	0.4570	0.4355

Of the 285 site years with good FC, 50 site years are missing one of (UST, PAR/Rg, and TA) and 79 sites have likely gap-filled data.

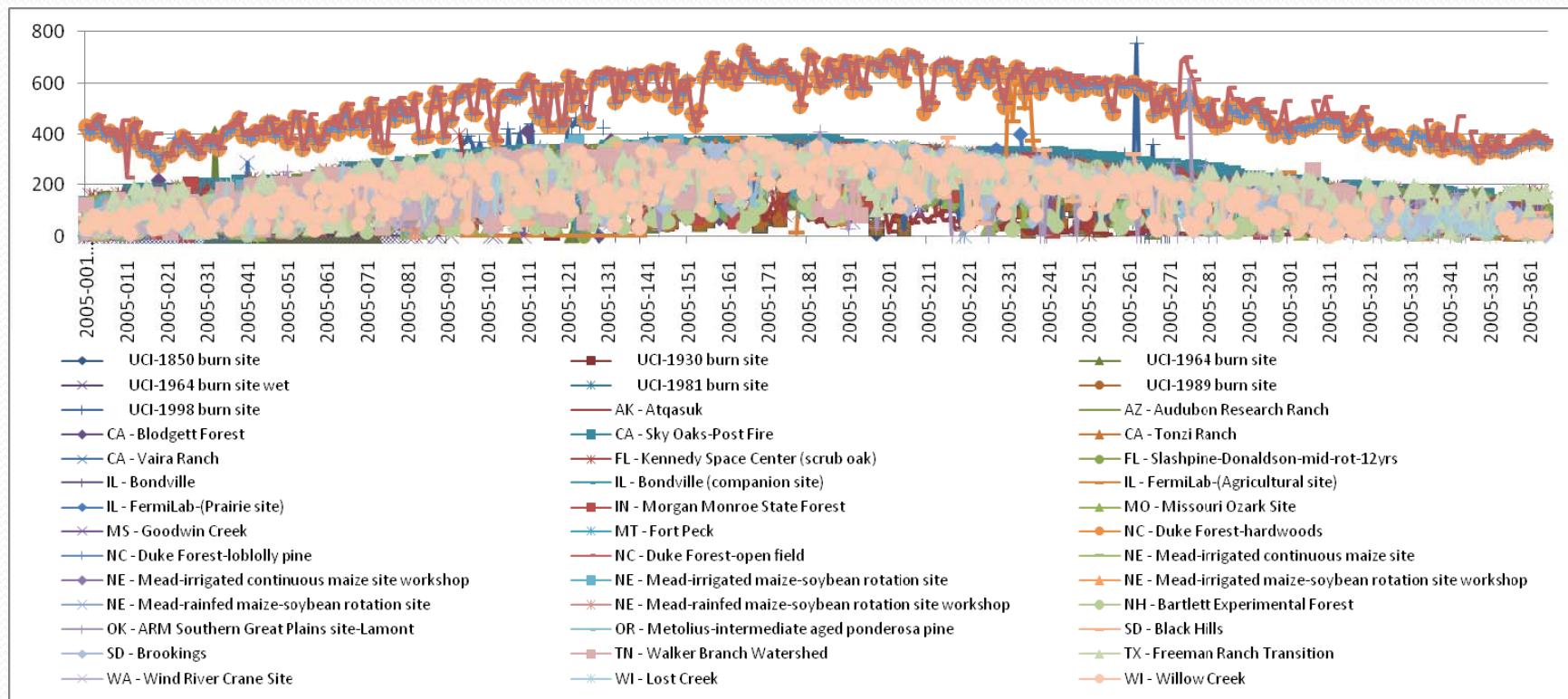
Obviously bad annual averages



Data cube used to browse average yearly Rg values across all site-years
16 additional likely problematic site-years at 5 sites

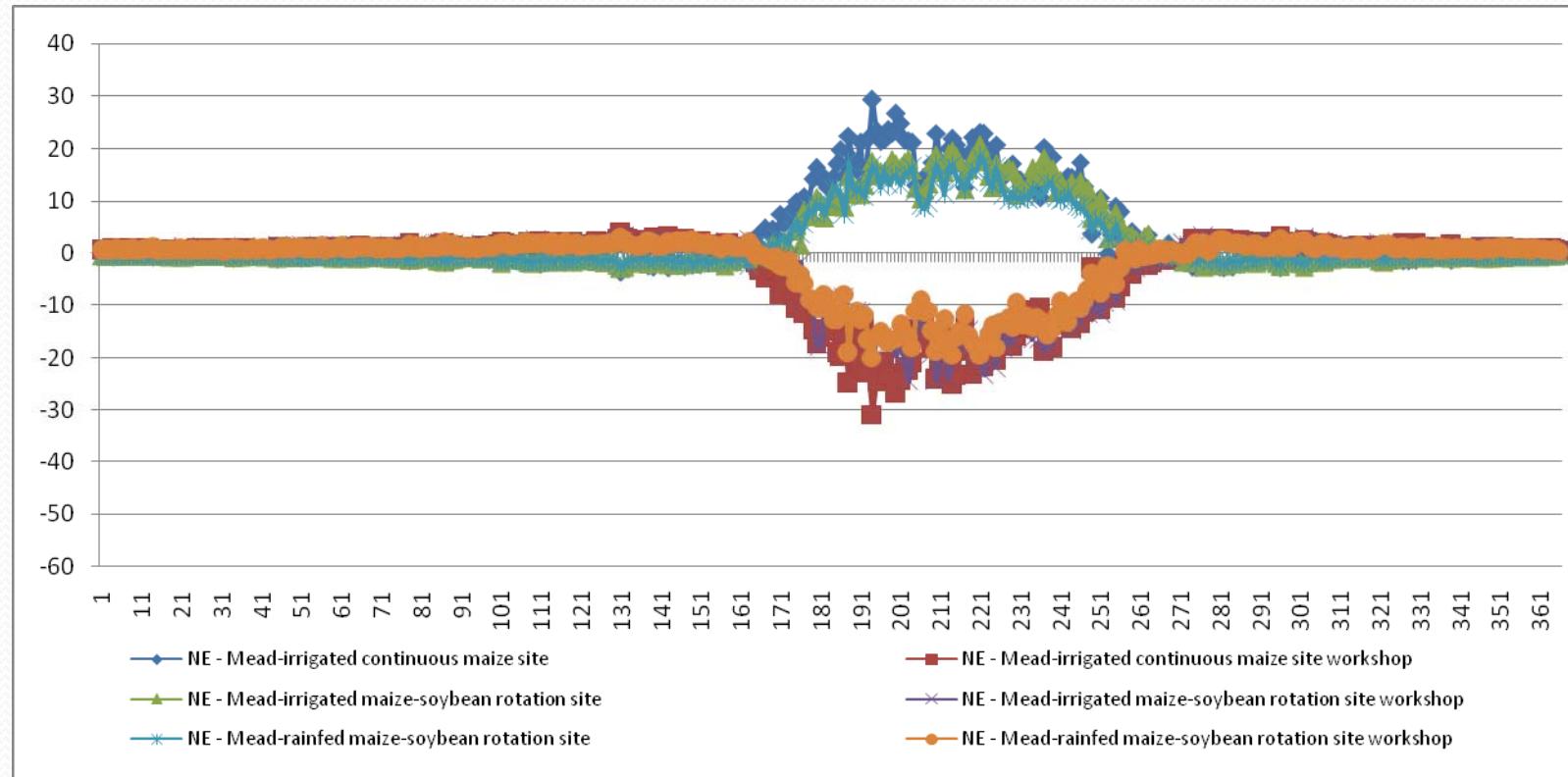


Drill down to consistent (bad) daily values



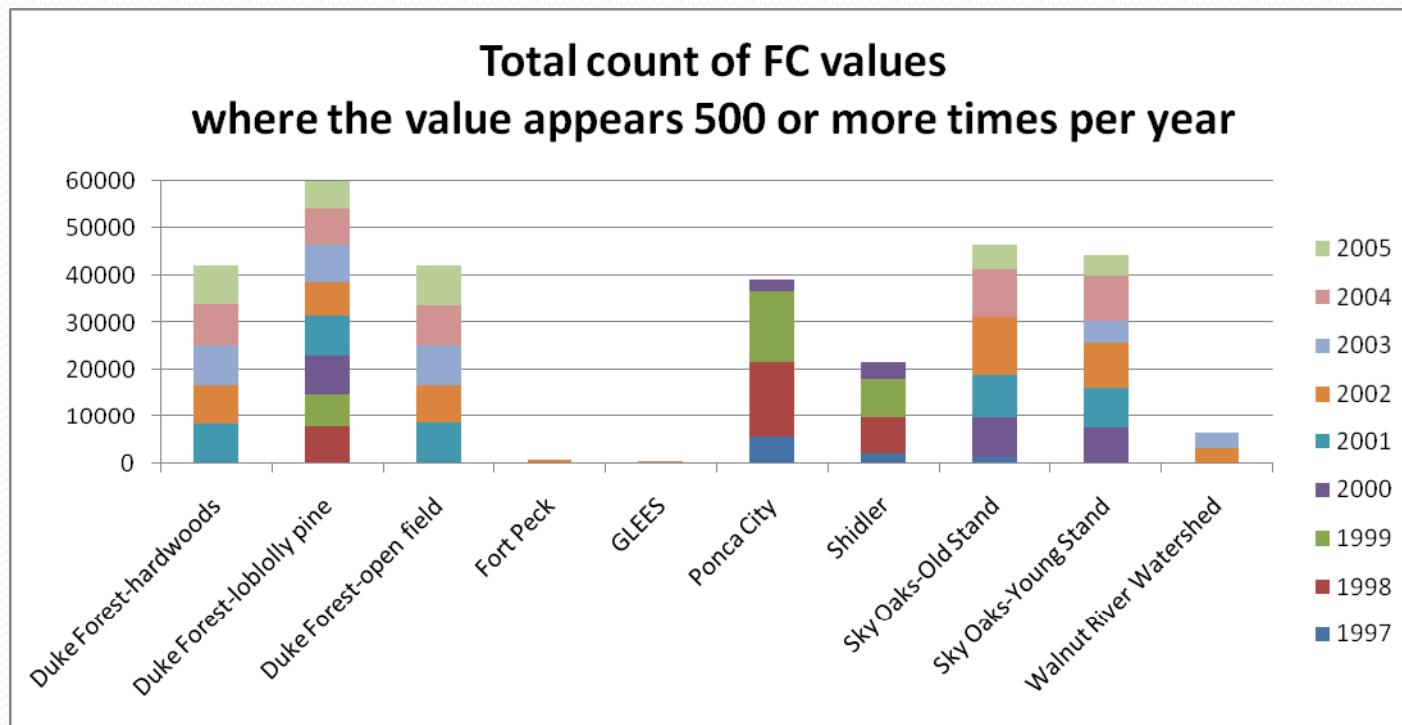
Data cube used to browse 2005 Rg values shows consistently high reporting (not just a few very large spikes) at Duke Forest sites

Drill down to Mead sites



Daily average FC at each site shows likely units and/or sign issues

Simple check: the same FC value reported too often



Database query returns count by (site, year, value) where count ≥ 500
Graph shows sum of (returned counts) by site and year

Sites with high sums likely report very few unique FC values

Project Status

Ameriflux data

- ~145 million daily values in cube
- Advanced calculations and statistics in cube
- Methods of providing $\frac{1}{2}$ hourly data access efficiently
- Data quality assessment
- Collection and incorporation of biological data

Fluxnet data

- ~210 million daily values in cube
- Waiting on gap filling and quality checking operations
- Data server, cubes, and reports will likely be the primary repository and access for researchers using the data

Russian River Data

- ~23 million values in cube
- Integration with Matlab and GIS
- Additional of data from other sources (fish, sediment size, ...)

On-going and Future Challenges

- Versioning of data and collection of data provenance
- Improved performance of database and cube
- Handling of biological data
- Advanced plotting capabilities
- Integration of collaborative tools
- Automation of data ingest, cube building, and report building
- MyDB and MyCube capabilities
- Integration of workflow capabilities
- Data server in a box
- Sociological change

Summary

- Ameriflux collaboration is adopting the Data Server architecture for the data repository
- BWC Data Server will be hosting the Fluxnet dataset which is expected to be the foundation for a broad range of research investigations
- Ecological measurements are often “messy”
- Applying the Data Server to watersheds introduces many additional challenges

Acknowledgements

Berkeley Water Center, University
of California, Berkeley, Lawrence
Berkeley Laboratory

Jim Hunt

Matt Rodriguez

Monte Good

Rebecca Leonardson (student)

Carolyn Remick

Susan Hubbard

Yoram Rubin

Microsoft

Catharine van Ingen

Jayant Gupchup (student)

Nolan Li (student)

Tony Hey

Dan Fay

Stuart Ozer

SQL product team

Jim Gray

Ameriflux Collaboration

Dennis Baldocchi

Beverly Law

Tara Stiefl (student)

Youngryel Ryu (postdoc)

Gretchen Miller (student)

Mattias Falk

Tom Boden

Bob Cook

CarboEurope Collaboration

Dario Papale

Markus Reichstein

*Project funded by Microsoft