

Environmental Science Workshop MSR Data Segment Collaboration

Rob Fatland

Microsoft Research

FAPESP – 11 November 2010

Overview

- Technology development in service to science motivation
- Completeness in technology
- MSR hopes to help build research projects with impact
 - ...on both methods and results
 - ...that cross over to public discourse

Example driver question:

How is our work at Microsoft Research providing

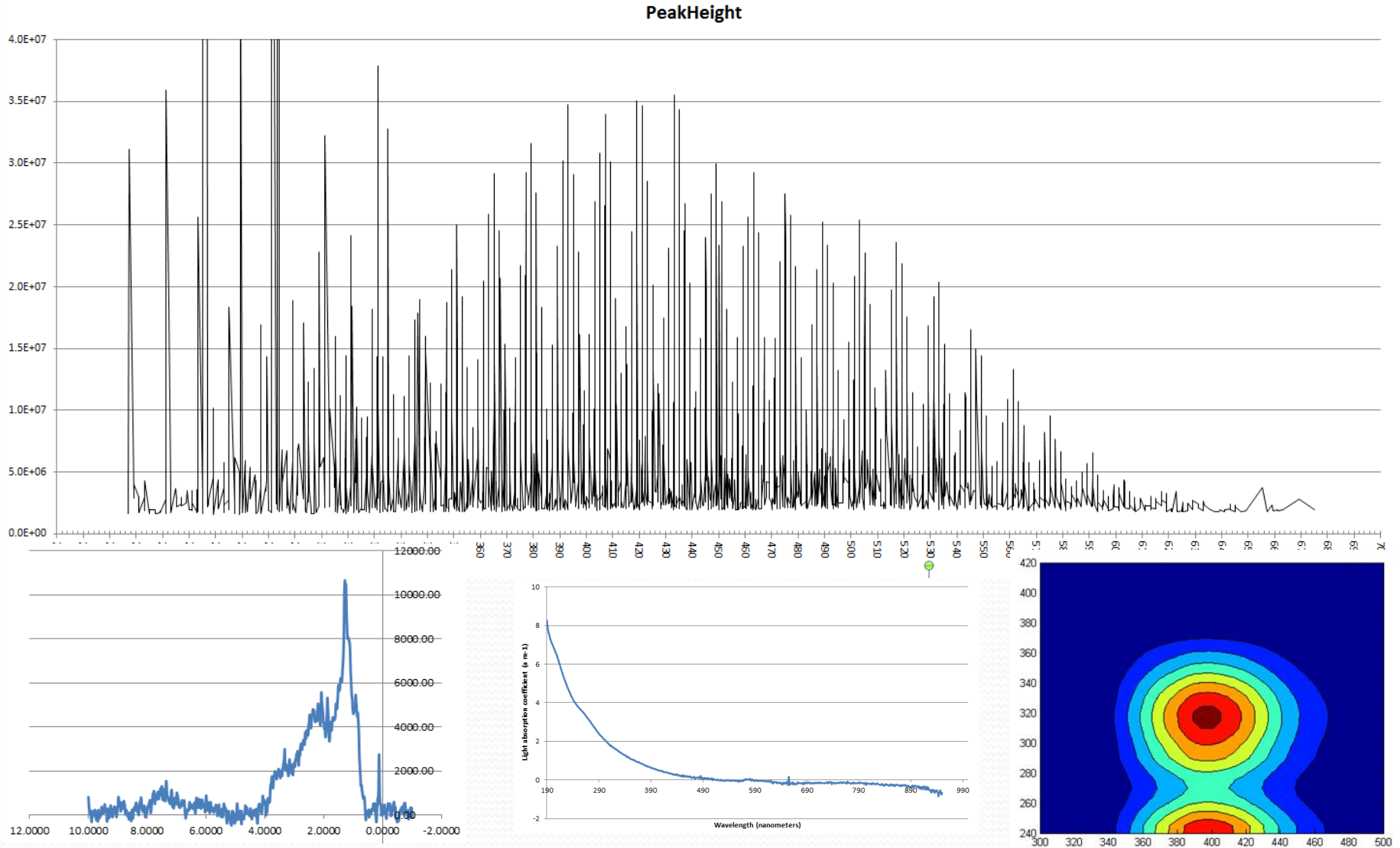
[hydrologists / ecologists / citizen scientists / policy makers]

the opportunity to collect and work with

[real-time / scientific / useful] data at

[new / larger / watershed] scales?

Complexity of Single Data Sources



Complexity of multiple data sources

- Scaling up to N data sources
- Without a priori understanding: How big is the relational problem? (Self, pairwise, ..., N)

$$\binom{N}{0} + \binom{N}{1} + \binom{N}{2} + \dots + \binom{N}{N} = 2^N$$

The implied processing tasks are interesting

- Raw: SVD on 10,000 x 3,000,000 observations
- Practical: Not so easy

Let us look beyond “Another Portal”

Breaking data archive barriers

sciscope

GEOSPATIAL DATA DISCOVERY
HAS NEVER BEEN EASIER...
START YOUR SEARCH HERE >>

LAYERS

- Watershed
- Ecoregion
- Geology

SEARCH

DATA

- Geographic area
- Time Frame
- Keyword

SUBMIT RESET

Welcome to SciScope!

SciScope is a tool by Microsoft Research to help geoscientists discover data from numerous data repositories with ease through a single, intuitive interface. Users can display multiple map layers related to the **scope** of their study and interact with geographical features on the map including dams, rivers, water bodies, geology, aquifer systems, ecological regions and river basins.



The interface includes a 'LAYERS' section with 'Watershed' options, a 'SEARCH' section with filters for 'Geographic area', 'Time Frame', and 'Keyword', and a 'SITE INFORMATION' section for 'HOH RIVER AT US HIGHWAY 101 NEAR FORKS, WA'. The map shows the Hoh River area with several data points marked by red circles.

HOH RIVER AT US HIG... Processing
 Discharge, Daily av... Remove

HOH RIVER AT US HIG... Download
 Discharge, Daily av... Remove

Name	Type	Compressed size	Password ...	Size
7123156-1062010103807PM242	XML Document	24 KB	No	988 KB

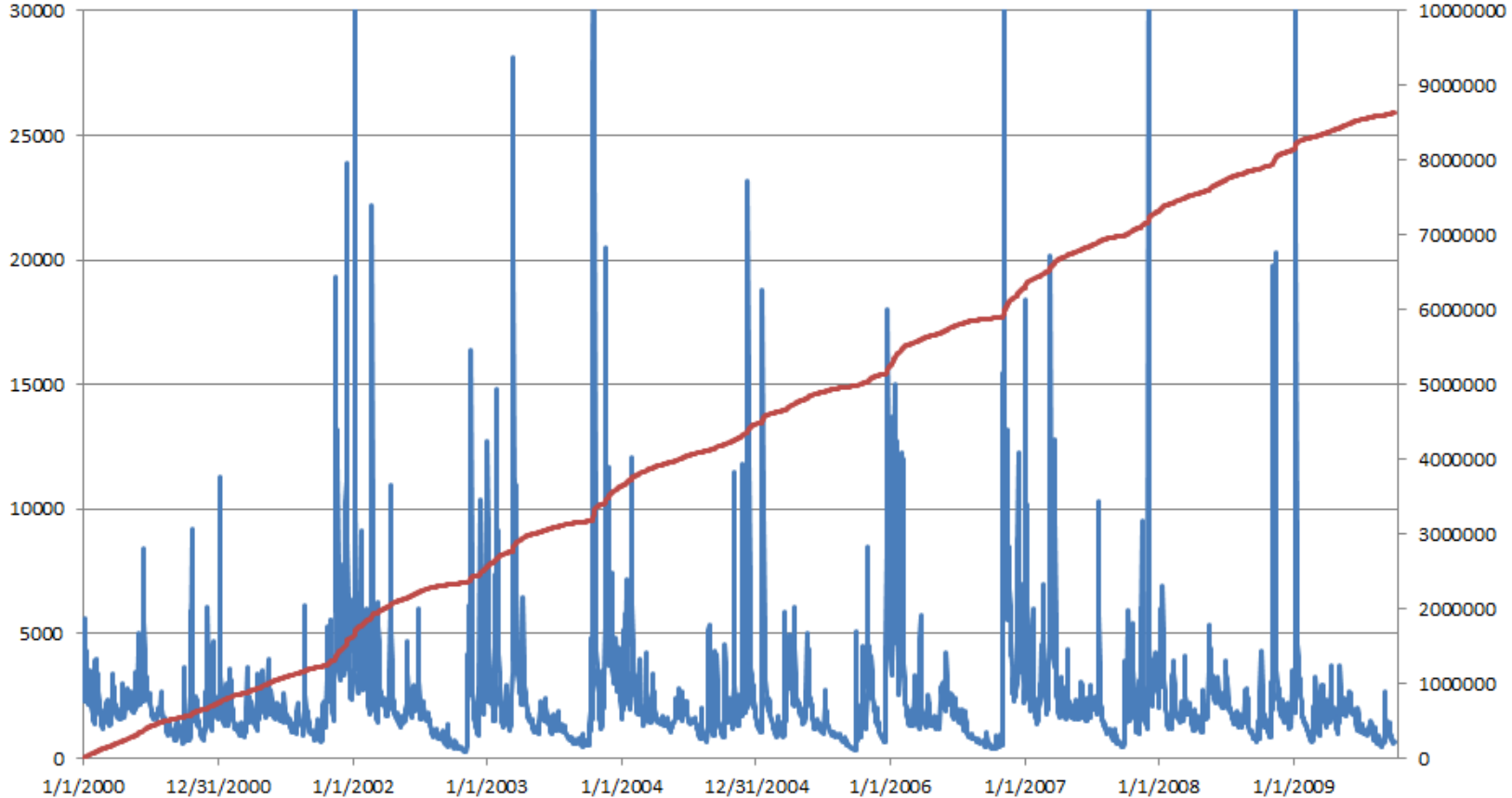
TimeSeries	
DateTime	Value
01/01/2000 12:00:00 PM	2440
01/02/2000 12:00:00 PM	2300
01/03/2000 12:00:00 PM	2500
01/04/2000 12:00:00 PM	5590
01/05/2000 12:00:00 PM	2570
...	
09/26/2009 12:00:00 PM	695
09/27/2009 12:00:00 PM	634
09/28/2009 12:00:00 PM	589
09/29/2009 12:00:00 PM	714
09/30/2009 12:00:00 PM	718
10/01/2009 12:00:00 PM	656

+

Location	
Name	HOH RIVER AT US HIGHWAY 101 NEAR FORKS, WA
Spatial Reference System	EPSG:4326
Position & Geometry	POINT (-124.251 47.8067)
Site Code	NWIS::12041200
Observed Variable	
Code	00060
Name	Discharge
Medium	Water
Units	cfs

=C2+B3

	A	B	C
1	TimeSeries		
2	DateTime	Value	Cell
3	01/01/2000 12:00:00 PM	2440	2440
4	01/02/2000 12:00:00 PM	2300	4740
5	01/03/2000 12:00:00 PM	2500	7240
6	01/04/2000 12:00:00 PM	5590	12830
7	01/05/2000 12:00:00 PM	3570	16400
8	01/06/2000 12:00:00 PM	30000	
9	01/07/2000 12:00:00 PM		



Mata Atlantica Micrometeorology

- Physical situation
- Sensor networks
- Data and results



Data manipulation using .NET

Simple Brazil Dataset Cleaner

Messages
19008: 2.35322 - 24.9 - 61.8
19009: 2.35356 - 25.0 - 61.0
at plotSourceX loop endRefTimeAsIndex is 14400
0 (6): Intrap 14398, 2Prior 0, PrnOnly 1, PostOnly 1, NO 0, nLines 26632

Main controls
Halt
Clear All
Set All
Go notes

iButton + met
Go iBtn 1
Go iBtn 2
Go Met 1
Go Met 2

mote subset config
Ref-0 2009 - 11 - 17 0 : 0 : 00.0
Year Month Day Hour Min Sec
Duration 5 52
decimal days max notes
Source Folder 2009 Data/datadrop 2010 04 20/
Data File Extension _External_4_19_2010.csv

iButton filename config
sub-folder 2009 Data/iButton pilot set 1 Nov 18 2009/
sources i sitemum/ i sitemum_idx_t (h)

Met stations + ancillary notes
Met 1 = (1 min) Air { T, RH, V, Dir }
Met 2 = (10 min avg) { Precip, Solar, Barom }
55 = Base (TB)
70 = GPS 1
71 = GPS 2
12 spare 11 @ Vexcel
53 spare 15 @ Vexcel
69 spare (was 10) 16 @ Vexcel
23 @ Vexcel
NaN @ Vexcel (was 67) 54 @ Vexcel

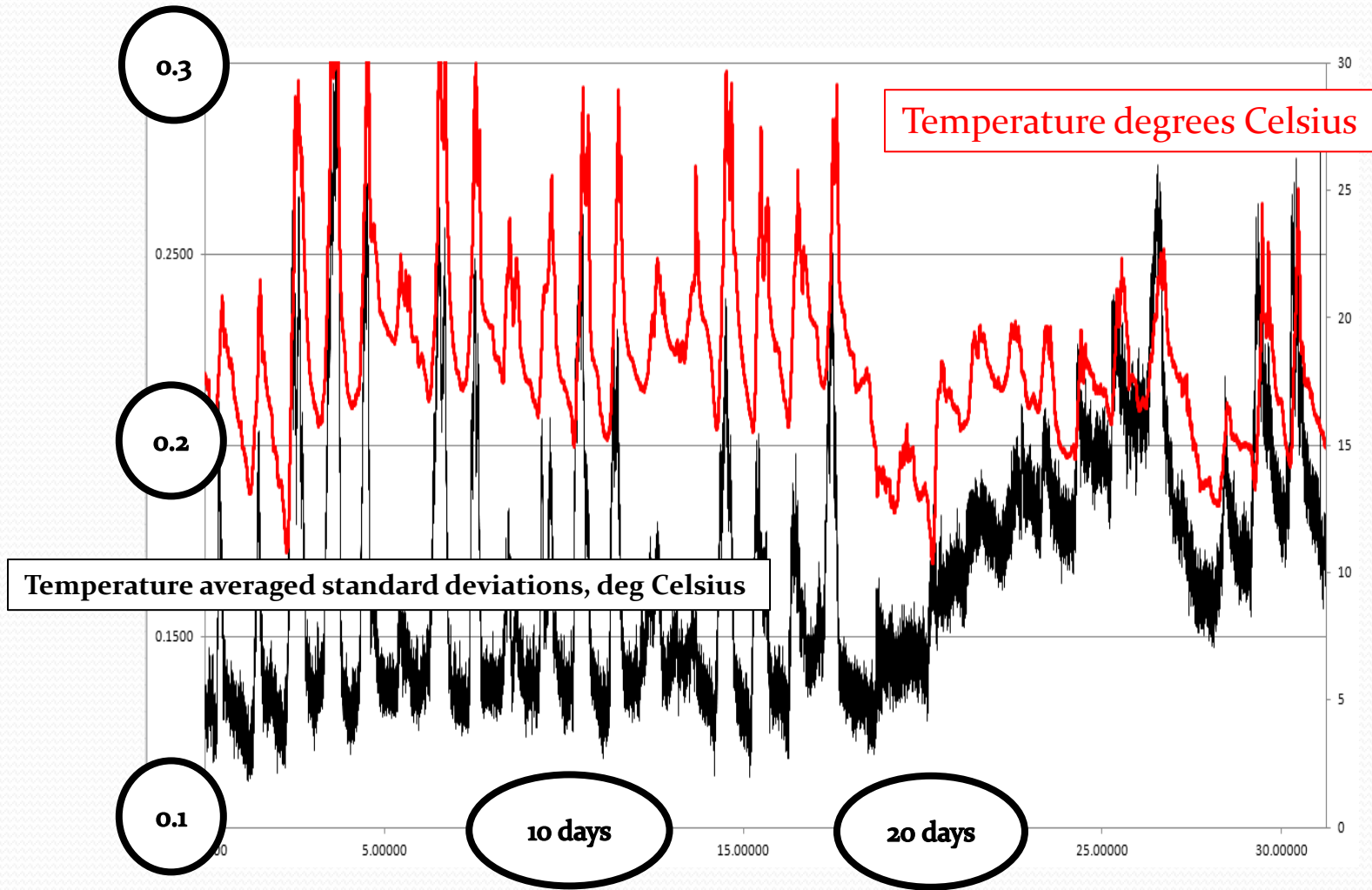
Sensor Network Schematic

Some Data: red is source, blue is regular-ized

Scientific validity of data

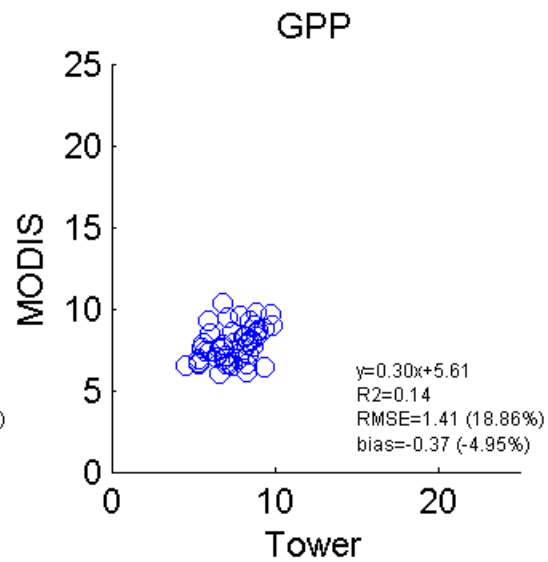
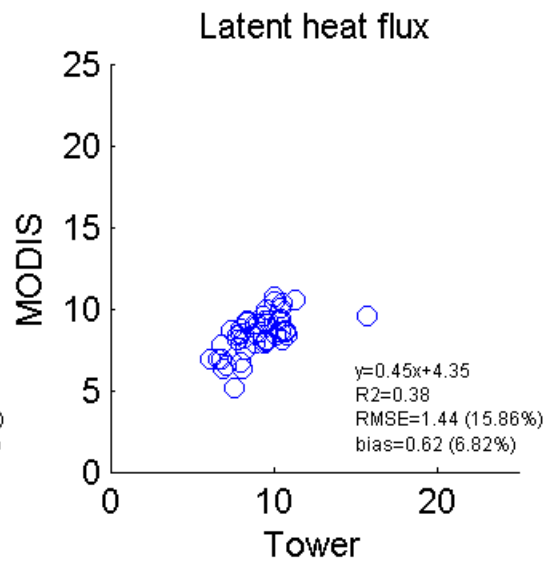
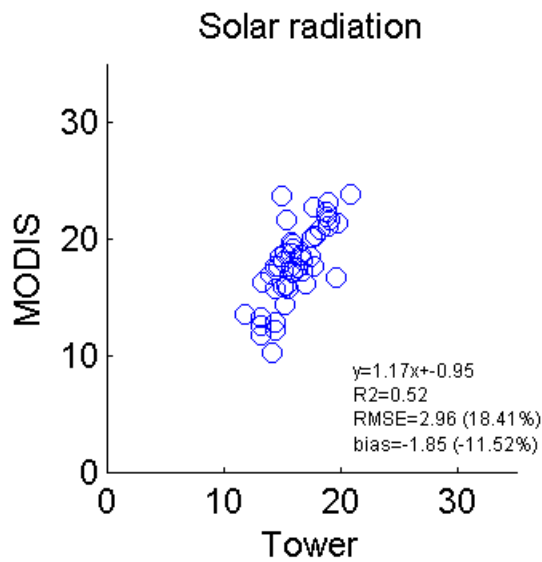
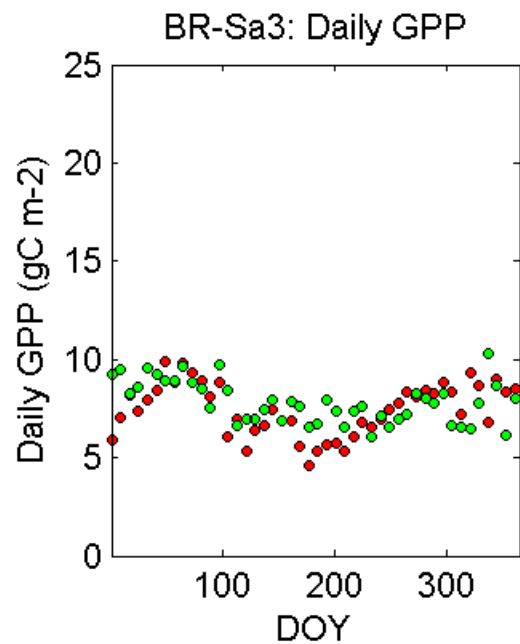
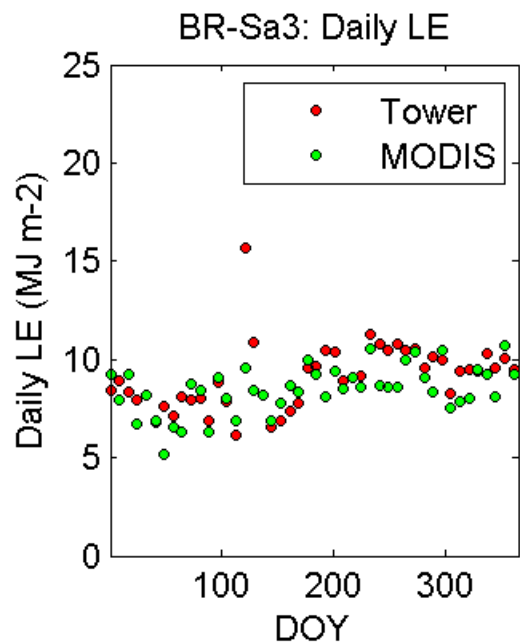
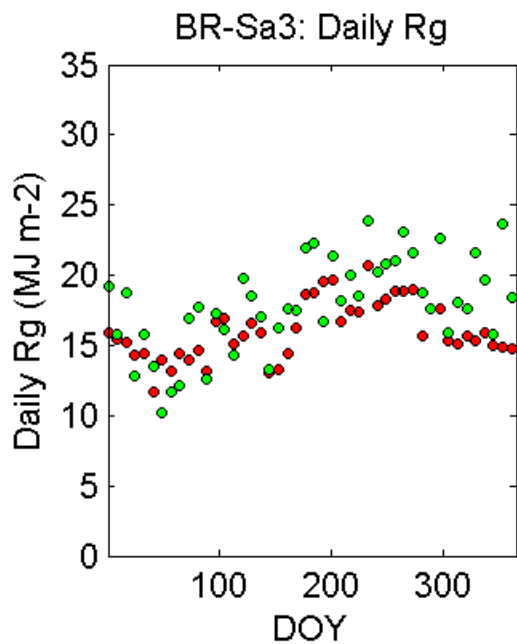
Black: Over 32 days how do 3 adjacent sensor vary compared to 0.1 deg Celsius?

Red (right vertical axis): What is the temperature in the canopy?

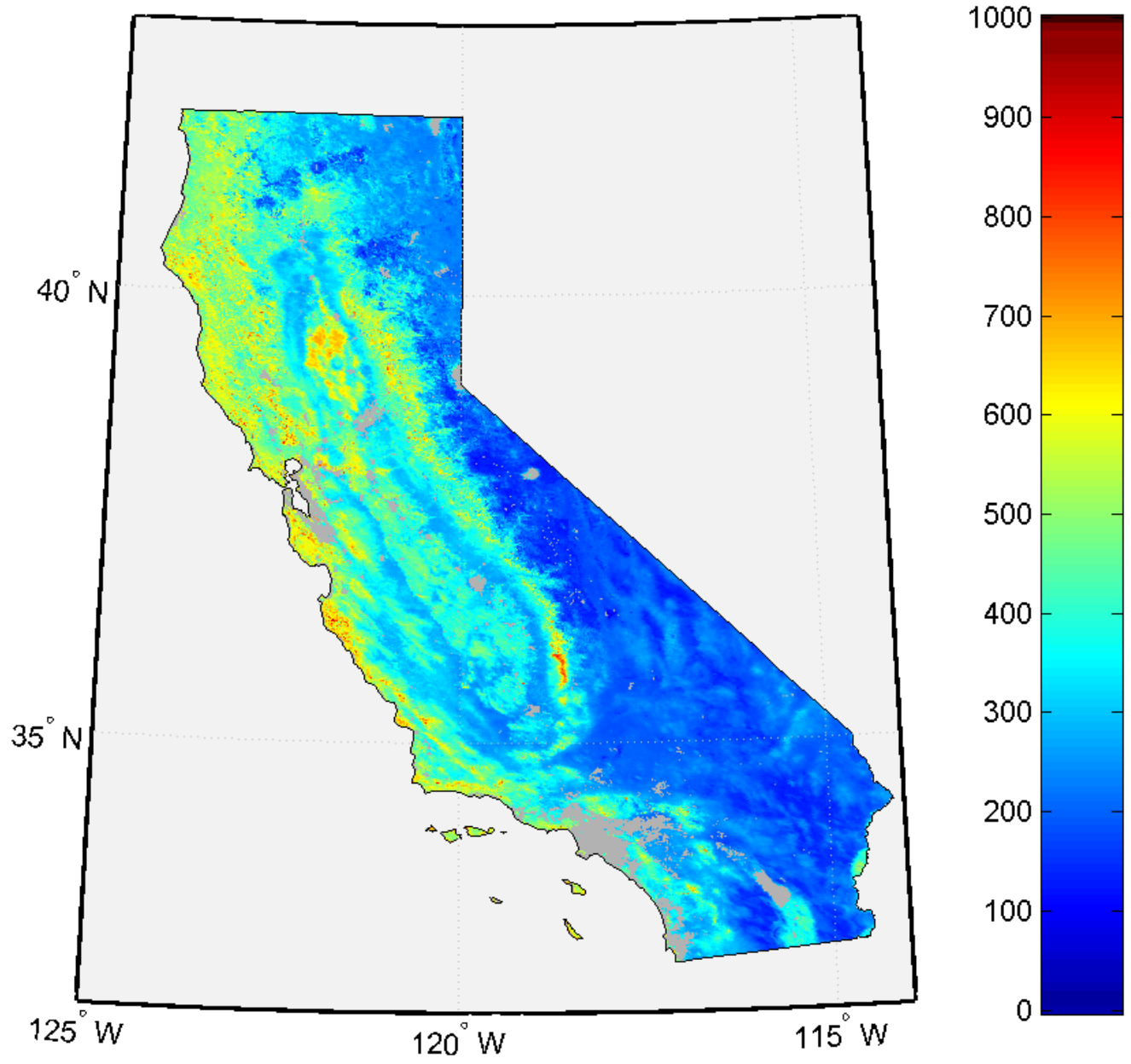


Remote sensing example

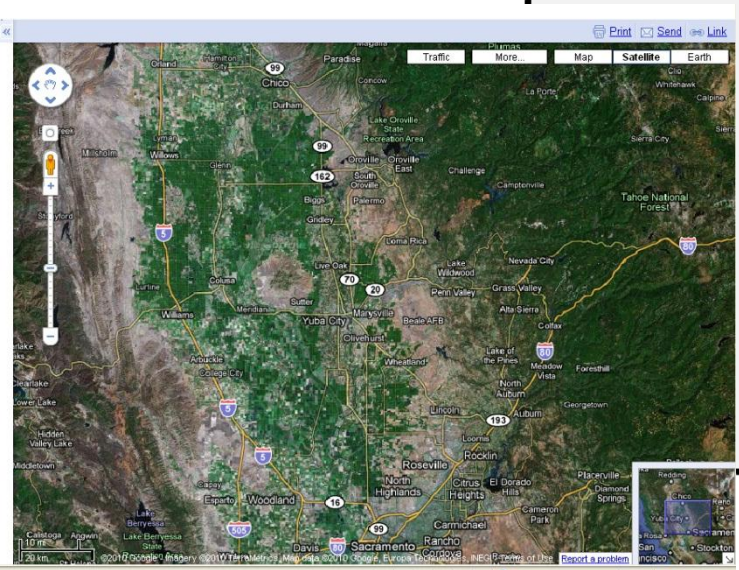
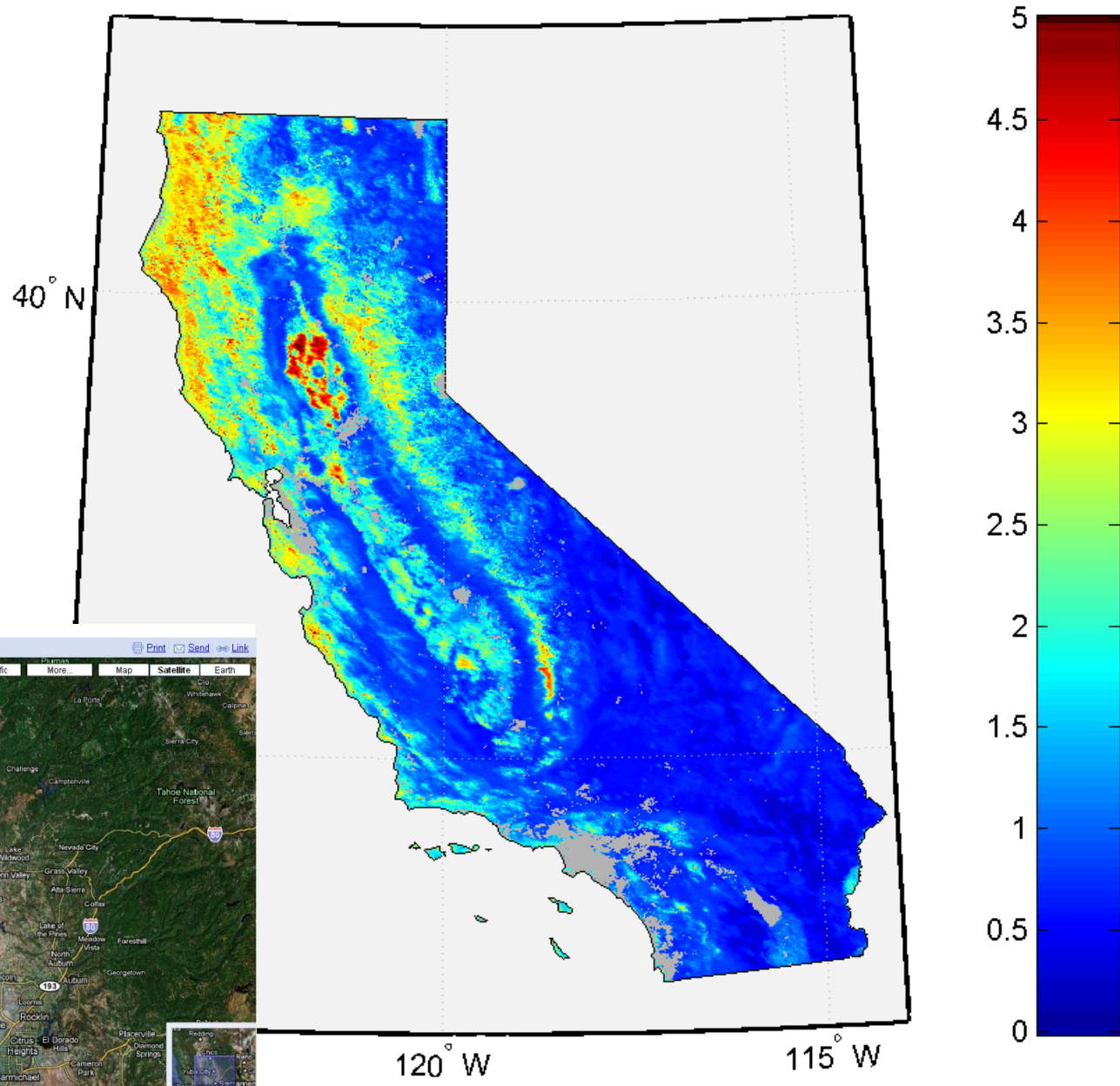
- MODIS has 36 spectral bands
 - MOD04 (aerosol)
 - MOD05 (precipitable water)
 - MOD06 (cloud)
 - MOD07 (atmospheric profile)
 - MOD11 (land surface temperature)
 - MCD12 (land cover)
 - MOD13 (vegetation index)
 - MCD15 (LAI)
 - MCD43 (albedo)



Evaporation (mm yr-1) Year: 2004



Evaporation (mm day⁻¹) Mon: 7 Year: 2004



Internet 100%

9:37 AM Friday 8/13/2010

KO

- 4 Microsoft...
- EndNote X...
- MODIS Sou...
- 4 Internet...
- UltraEdit-3...
- 11 Adobe...

MATLAB 7...

Inbox - Thu...

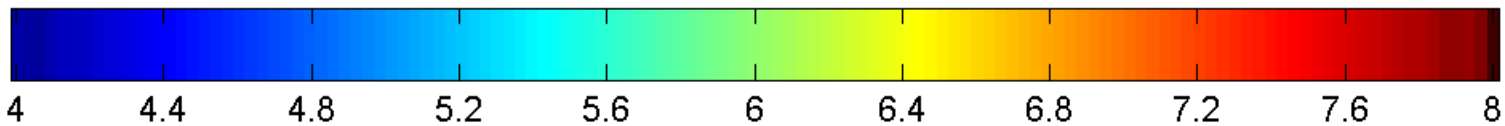
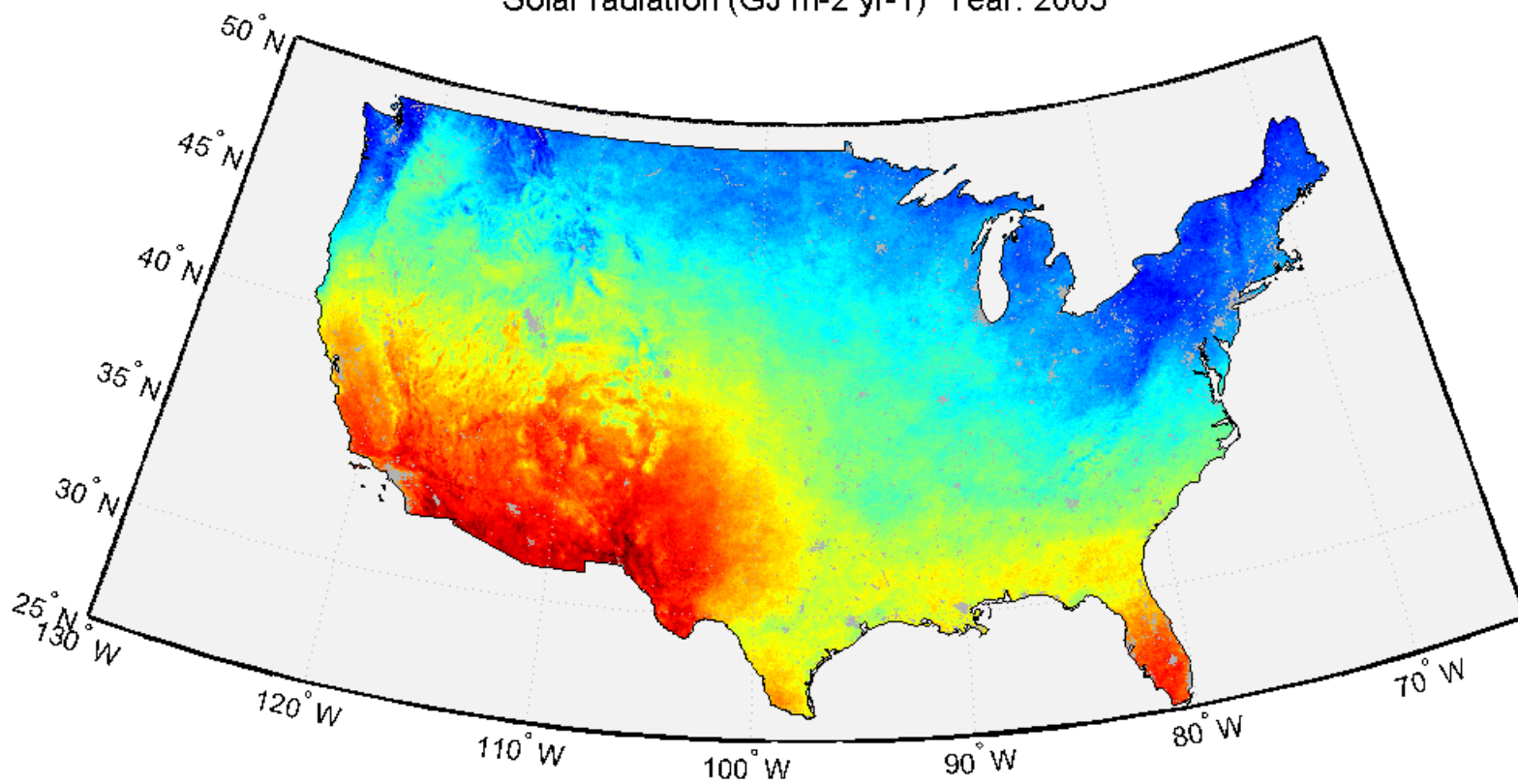
2 Notepad...

4 Microso...

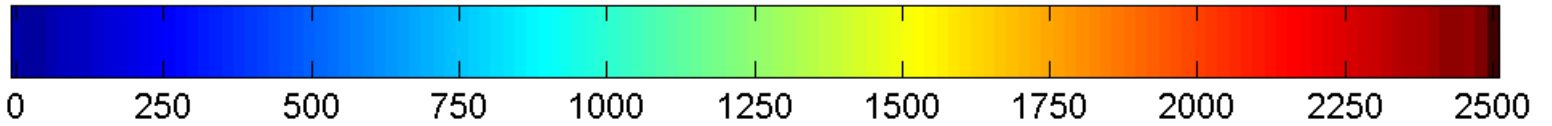
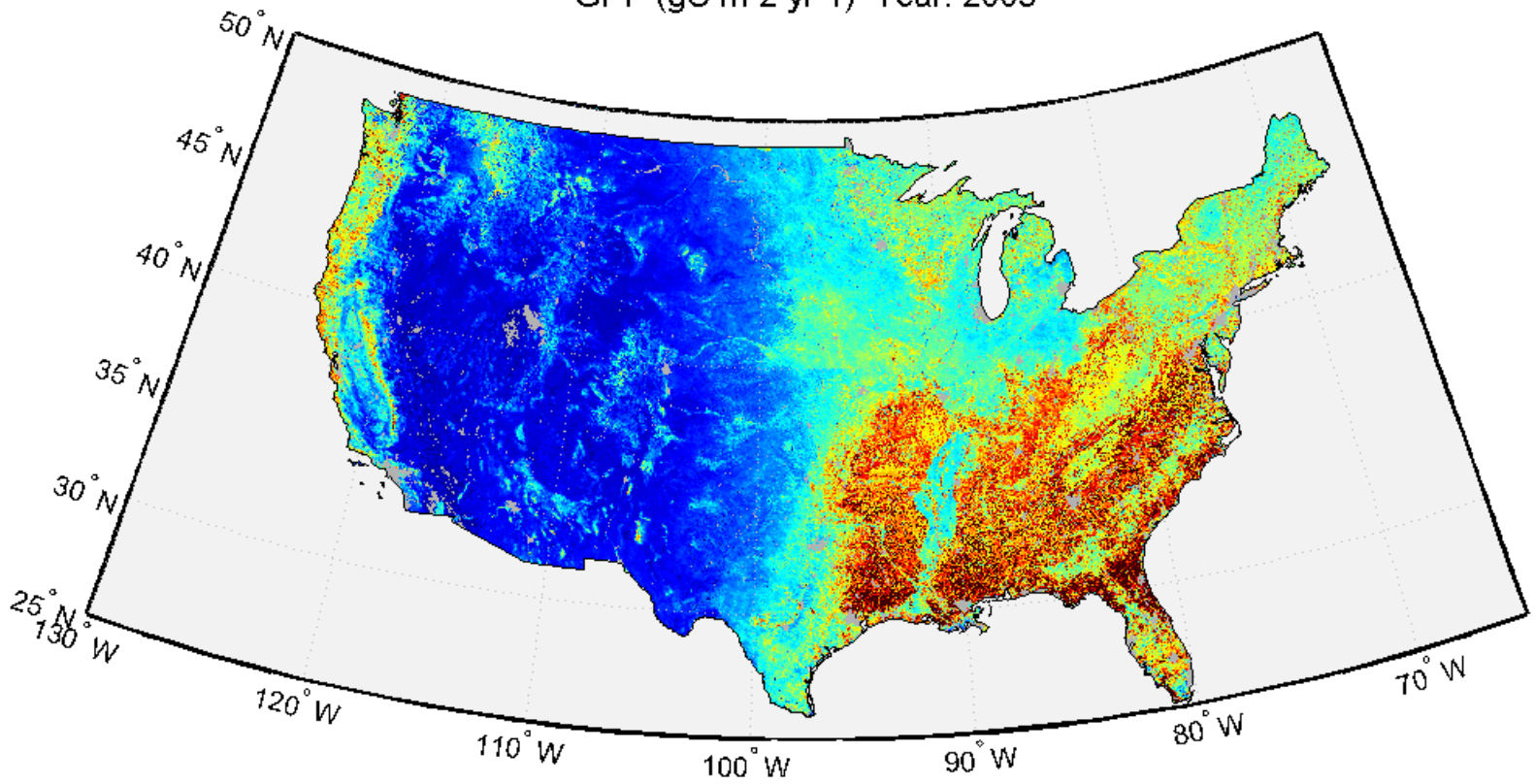
2 Microso...

Reprojecto...

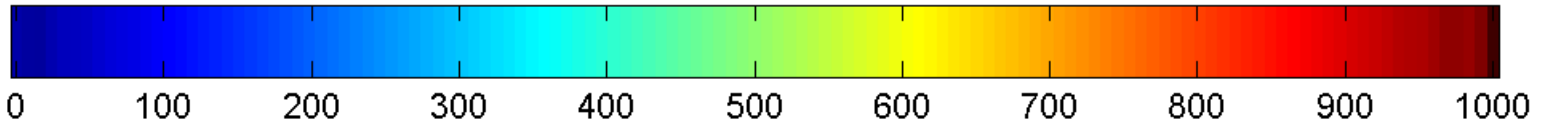
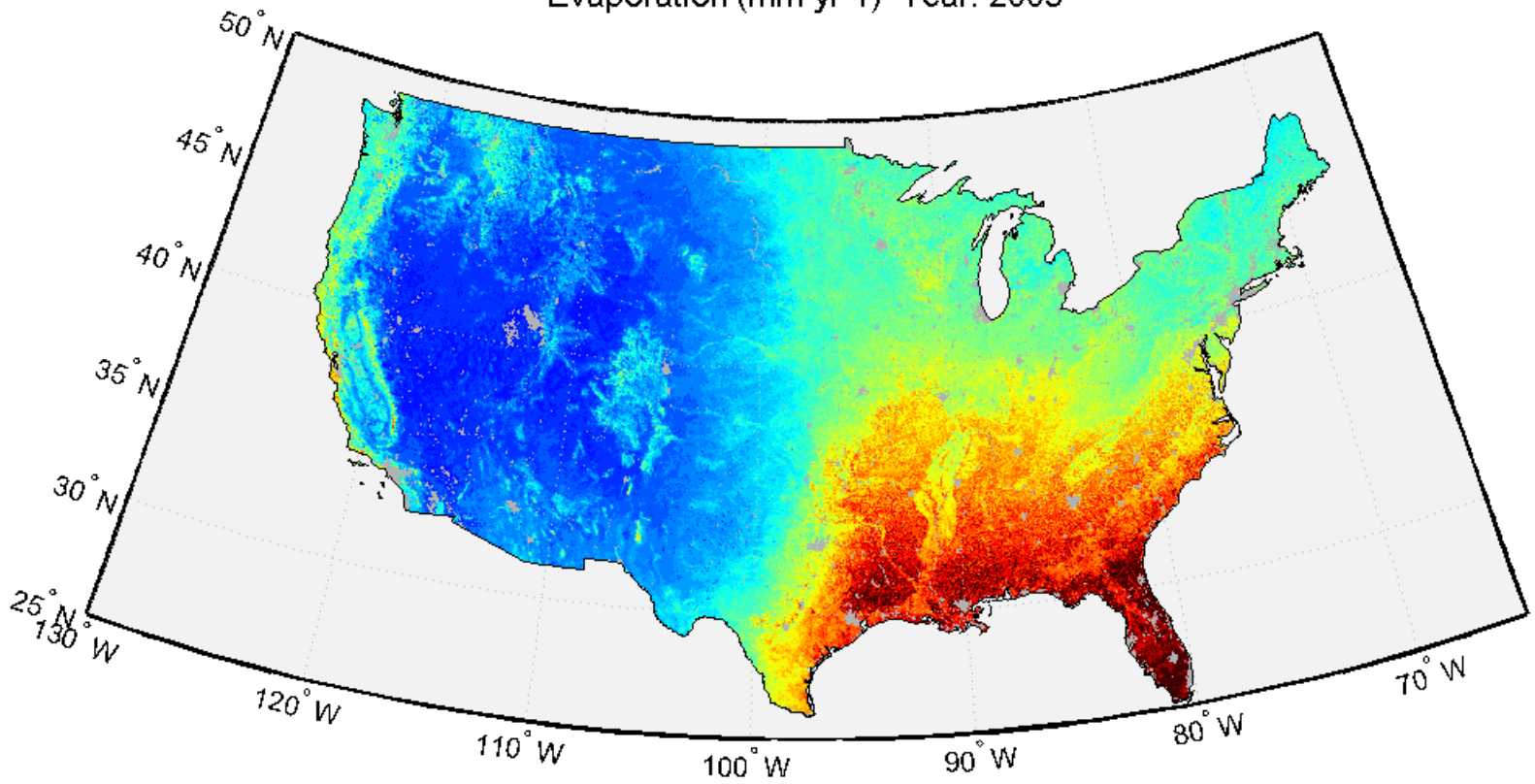
Solar radiation (GJ m⁻² yr⁻¹) Year: 2003



GPP (gC m⁻² yr⁻¹) Year: 2003

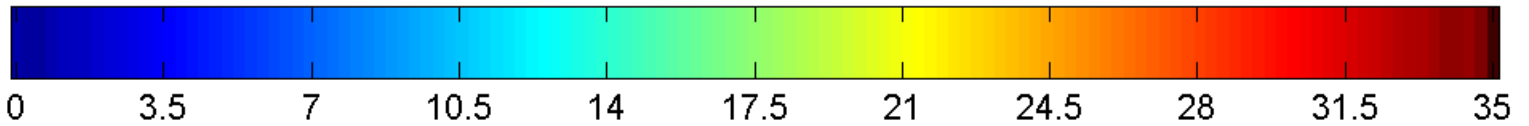
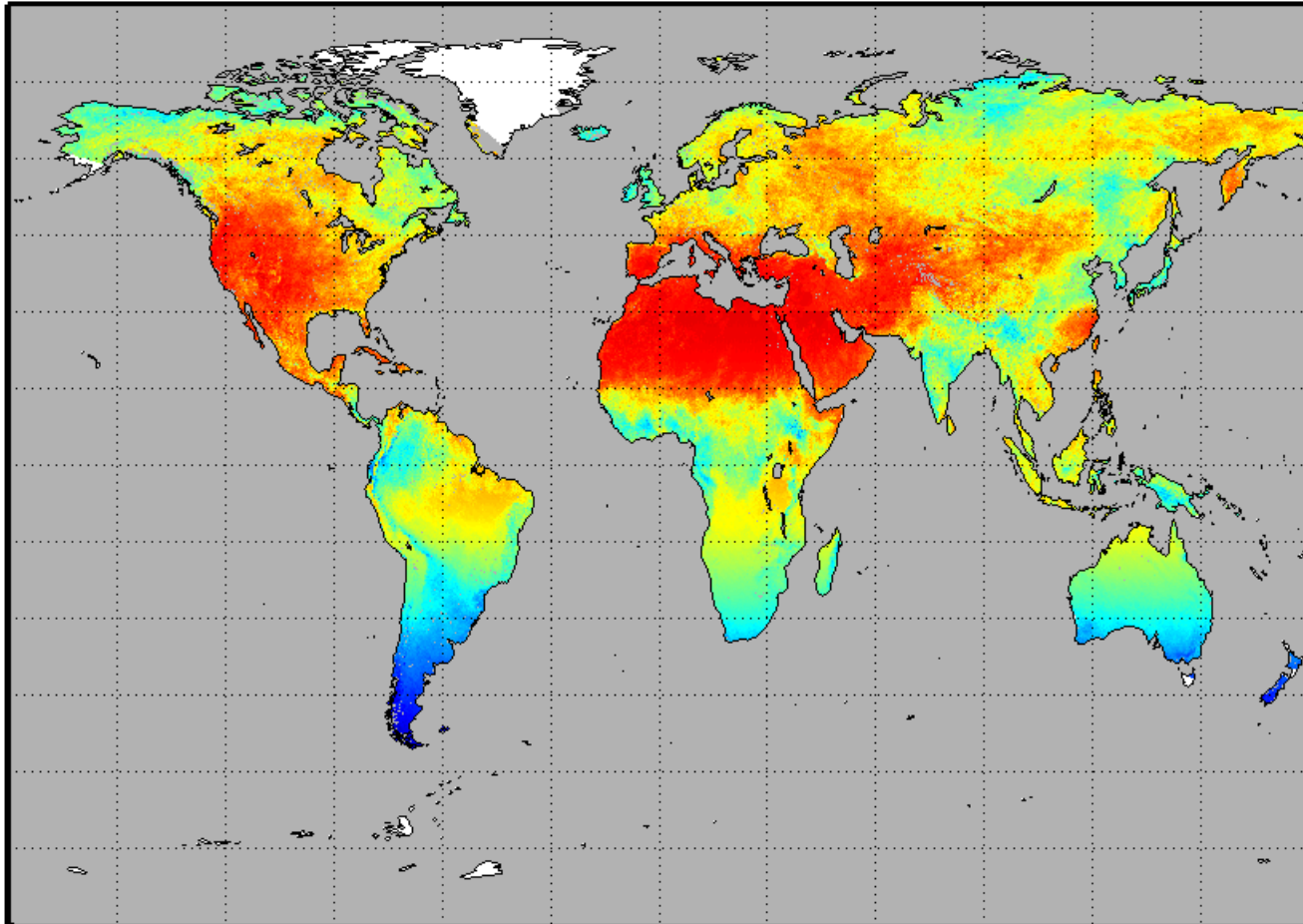


Evaporation (mm yr-1) Year: 2003



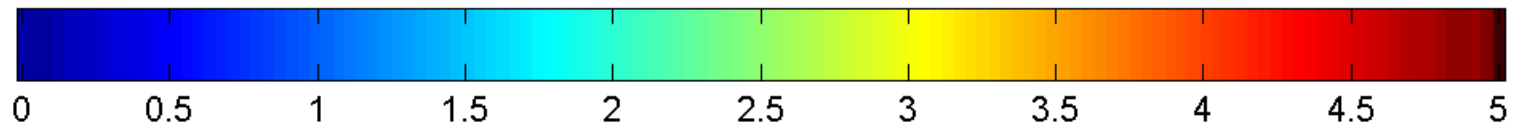
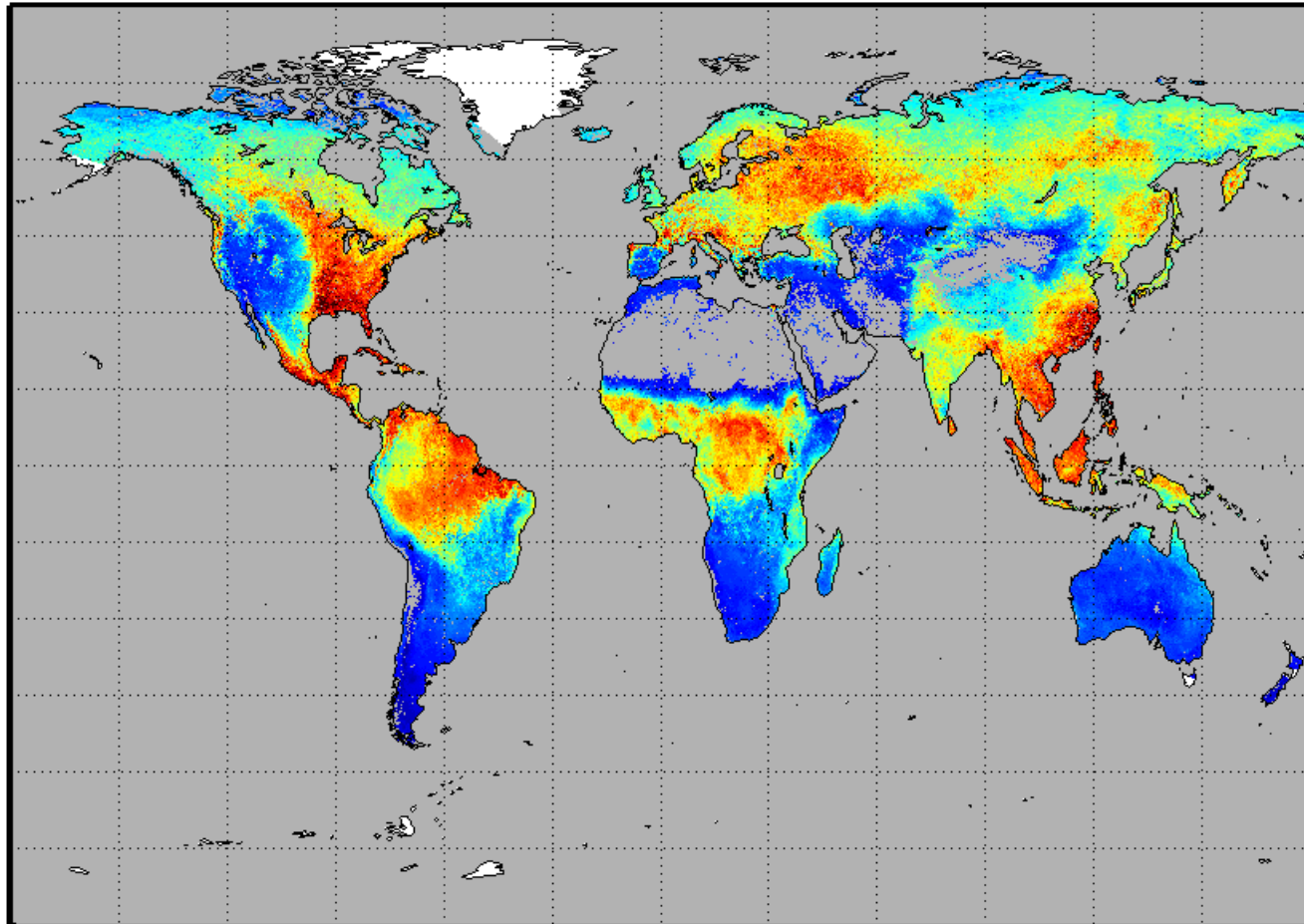
Solar radiation (MJ m⁻² day⁻¹) Year: 2003 Mon: 7

180° W 150° W 120° W 90° W 60° W 30° W 0° 30° E 60° E 90° E 120° E 150° E 180° E



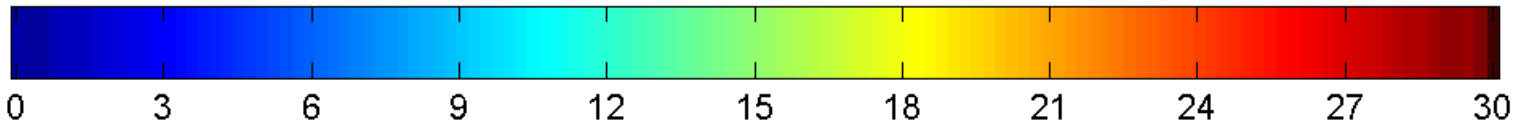
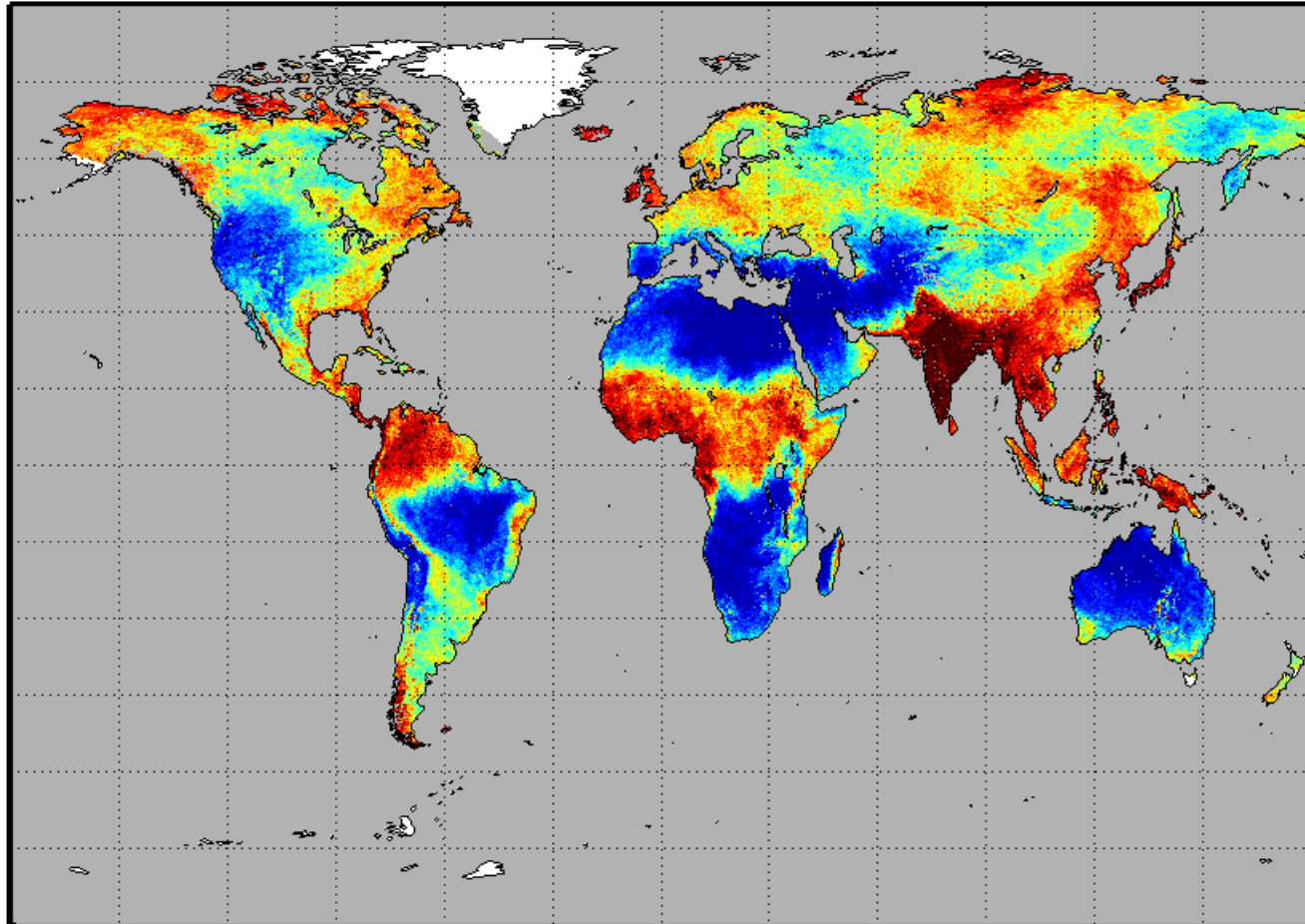
Evaporation (mm day⁻¹) Year: 2003 Mon: 7

180° W 150° W 120° W 90° W 60° W 30° W 0° 30° E 60° E 90° E 120° E 150° E 180° E



Cloudy days (days mon-1) Year: 2003 Mon: 7

180° W 150° W 120° W 90° W 60° W 30° W 0° 30° E 60° E 90° E 120° E 150° E 180° E



Conclusions

Technology / Tools / Data: Exciting, interesting, many gaps

Problem solving in the environmental space drives problem solving in the technology space.

The danger is in paying short-and-sufficient attention.

A wealth of examples and experience now exist to provide inspiration, motivation and guidance towards deep data segment solutions.

Ultimate importance: What do the scientists need from the technology?

rob.fatland@microsoft.com



Thanks

Richard Alley's AGU address on paleoclimate:

http://www.agu.org/meetings/fm09/lectures/lecture_videos/A23A.shtml

SciScope data catalog:

<http://www.sciscope.org>

Worldwide Telescope:

<http://www.worldwidetelescope.org>

Using Excel:

Robert de Levie. **Advanced Excel for scientific data analysis**. Second edition, Oxford, ISBN 978-0-19-537022-5, 2008.

Microsoft Research: <http://research.microsoft.com>

My contact info: rob.fatland@microsoft.com