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

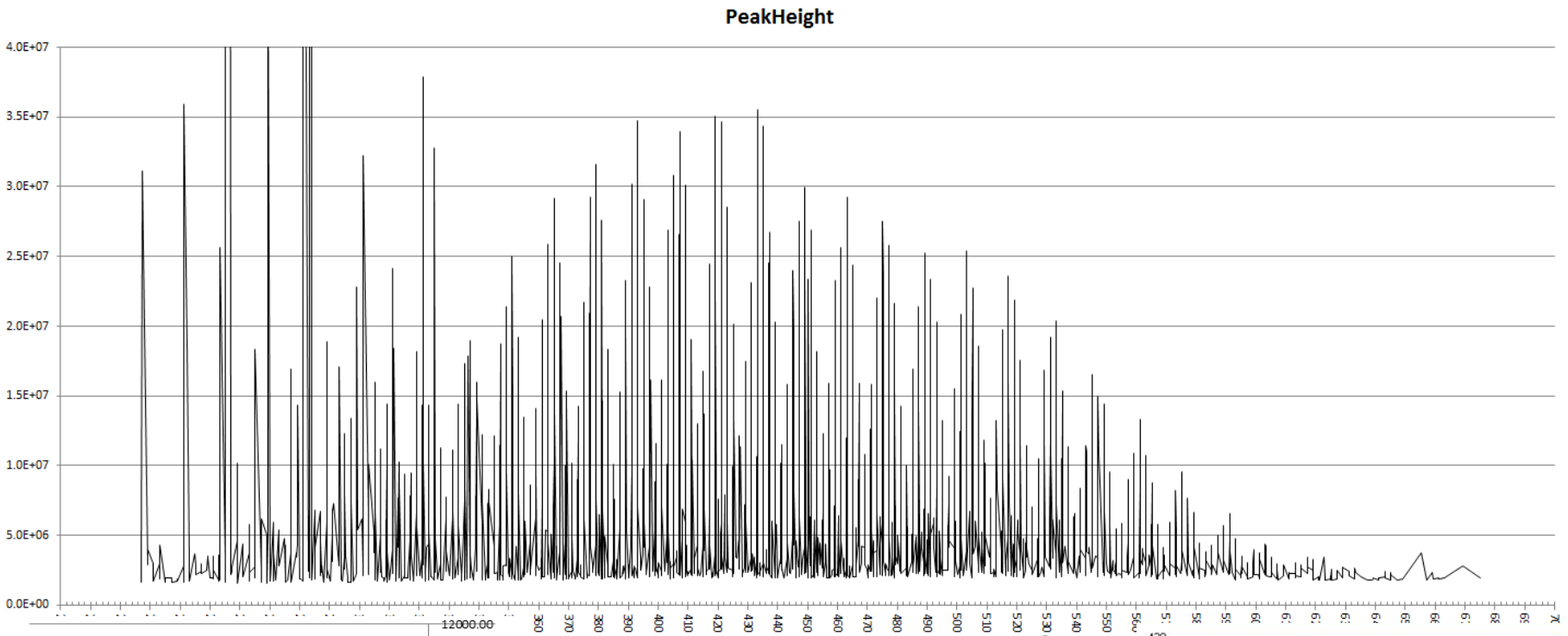
[hydrologists / ecologists / citizen scientists / policy makers]

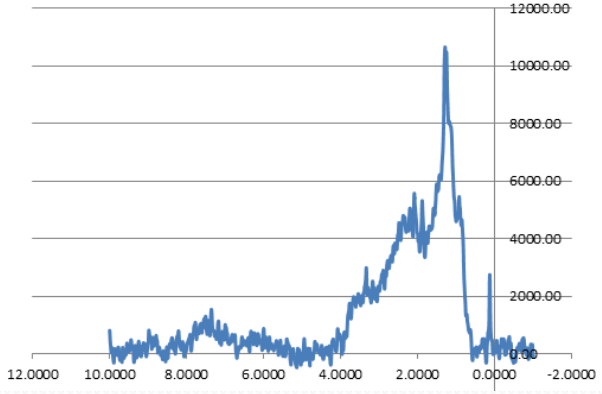
the opportunity to collect and work with

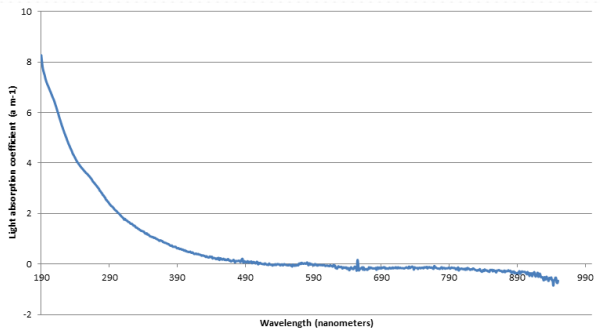
[real-time / scientific / useful] data at

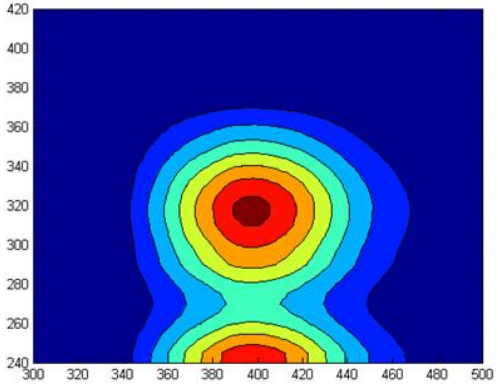
[new / larger / watershed] scales?

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)

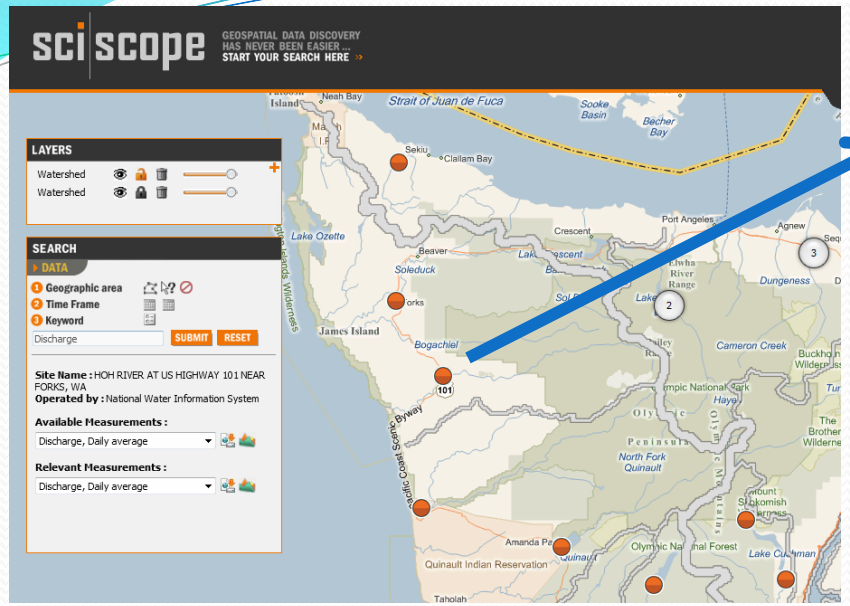
$$\binom{N}{0} + \binom{N}{1} + \binom{N}{2} + \ldots + \binom{N}{N} = 2^N$$

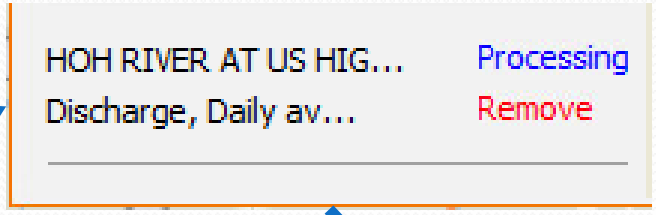
The implied processing tasks are interesting

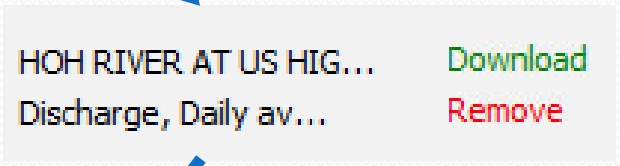
- Raw: SVD on 10,000 x 3,000,000 observations
- Practical: Not so easy

Let us look beyond “Another Portal”

Breaking data archive barriers







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

...

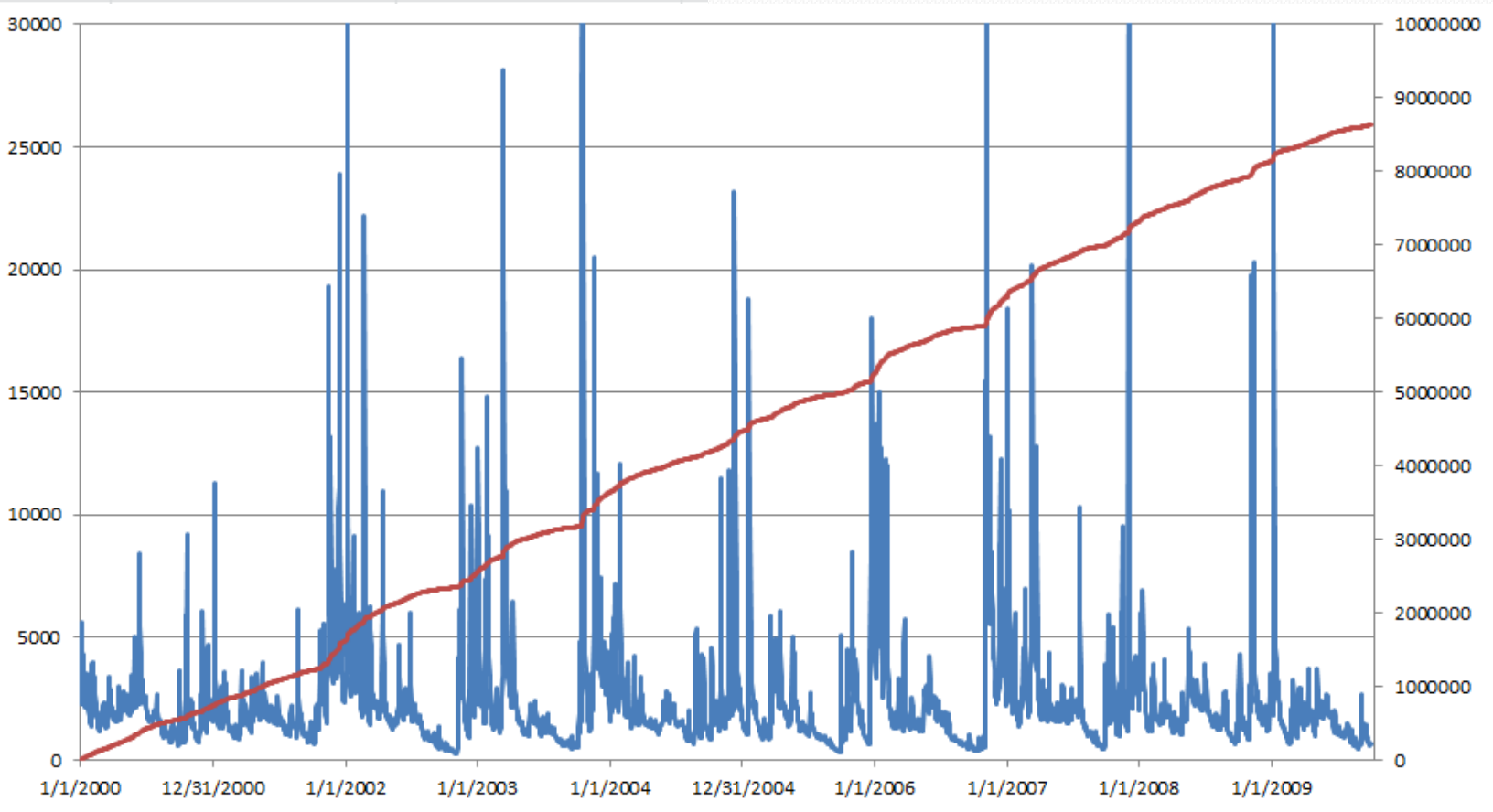
09/26/2009 12:00:00 PM	695
09/27/2009 12:00:00 PM	634
09/28/2009 12:00:00 PM	589
09/29/2009 12:00:00 PM	714
09/30/2009 12:00:00 PM	718
10/01/2009 12:00:00 PM	656

+

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

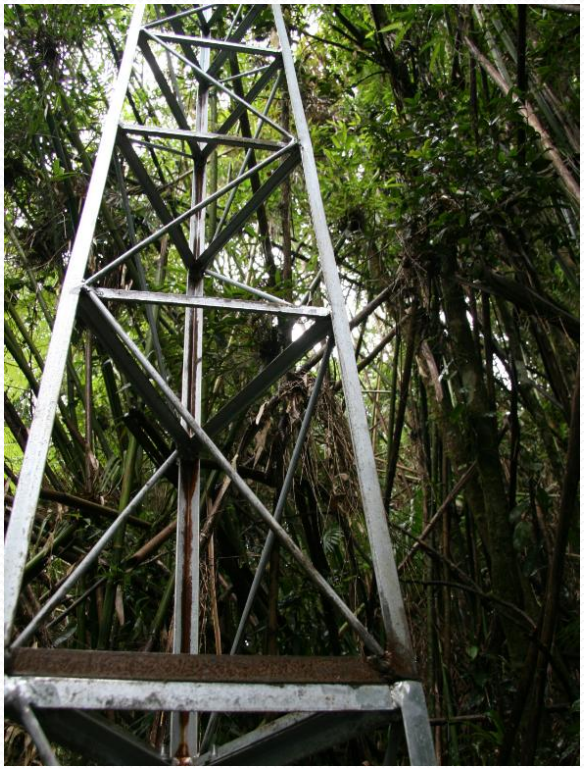
=C2+B3

	A	B	C	
1	TimeSeries			
2	DateTime	Value		Ce
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			
9	01/07/2000 12			

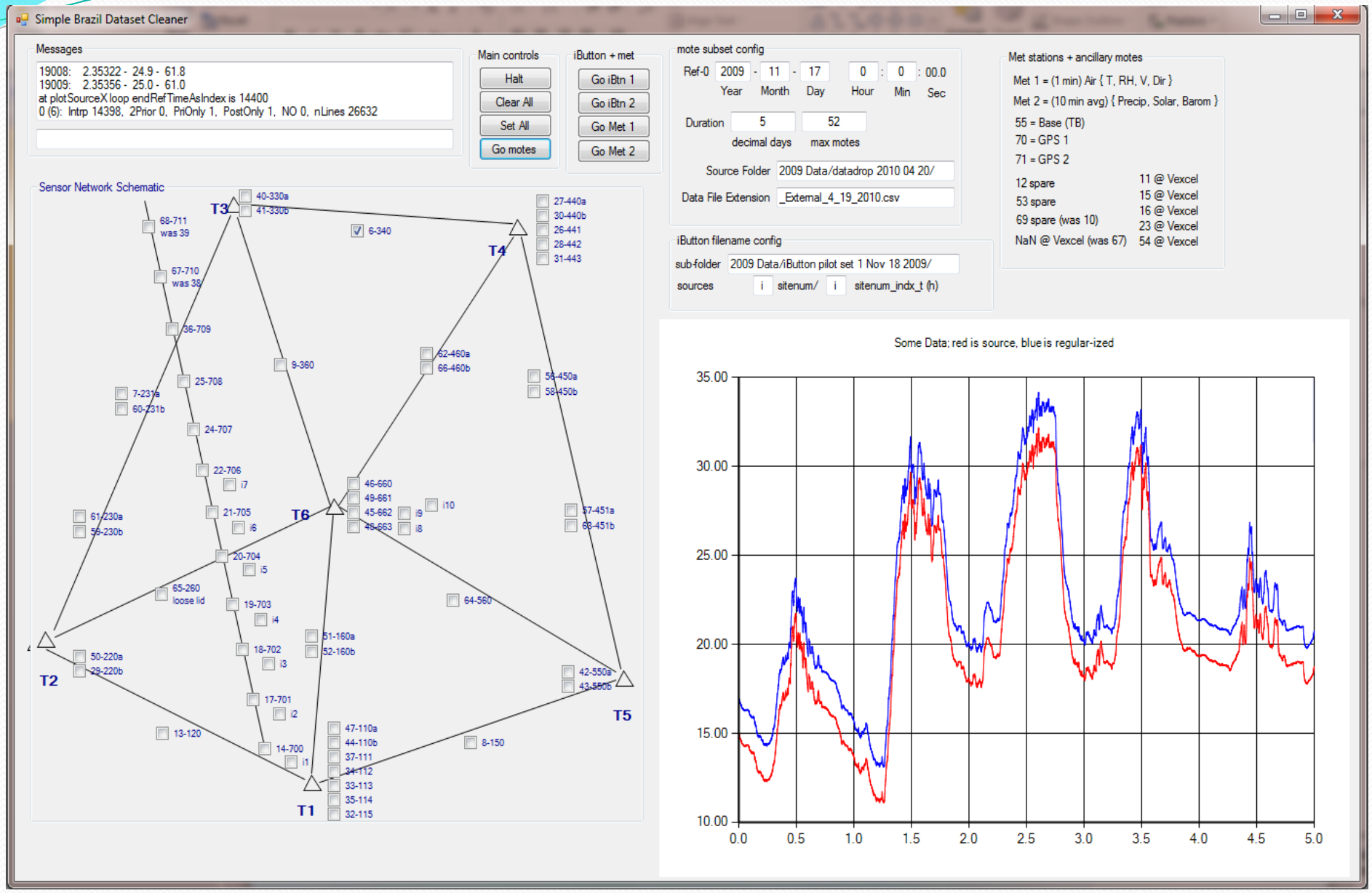


Mata Atlantica Micrometeorology

- Physical situation
- Sensor networks
- Data and results



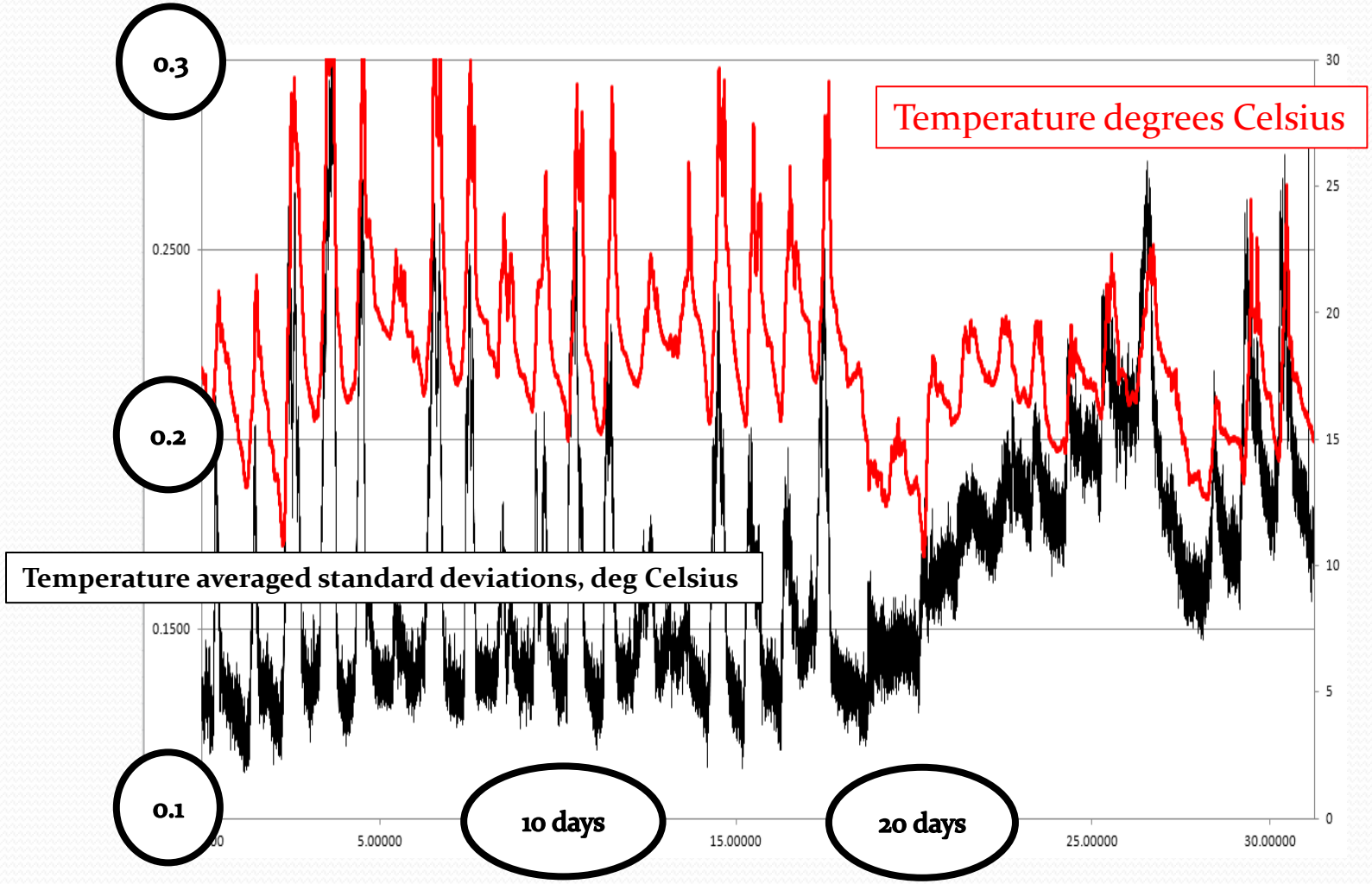
Data manipulation using .NET



Scientific validity of data

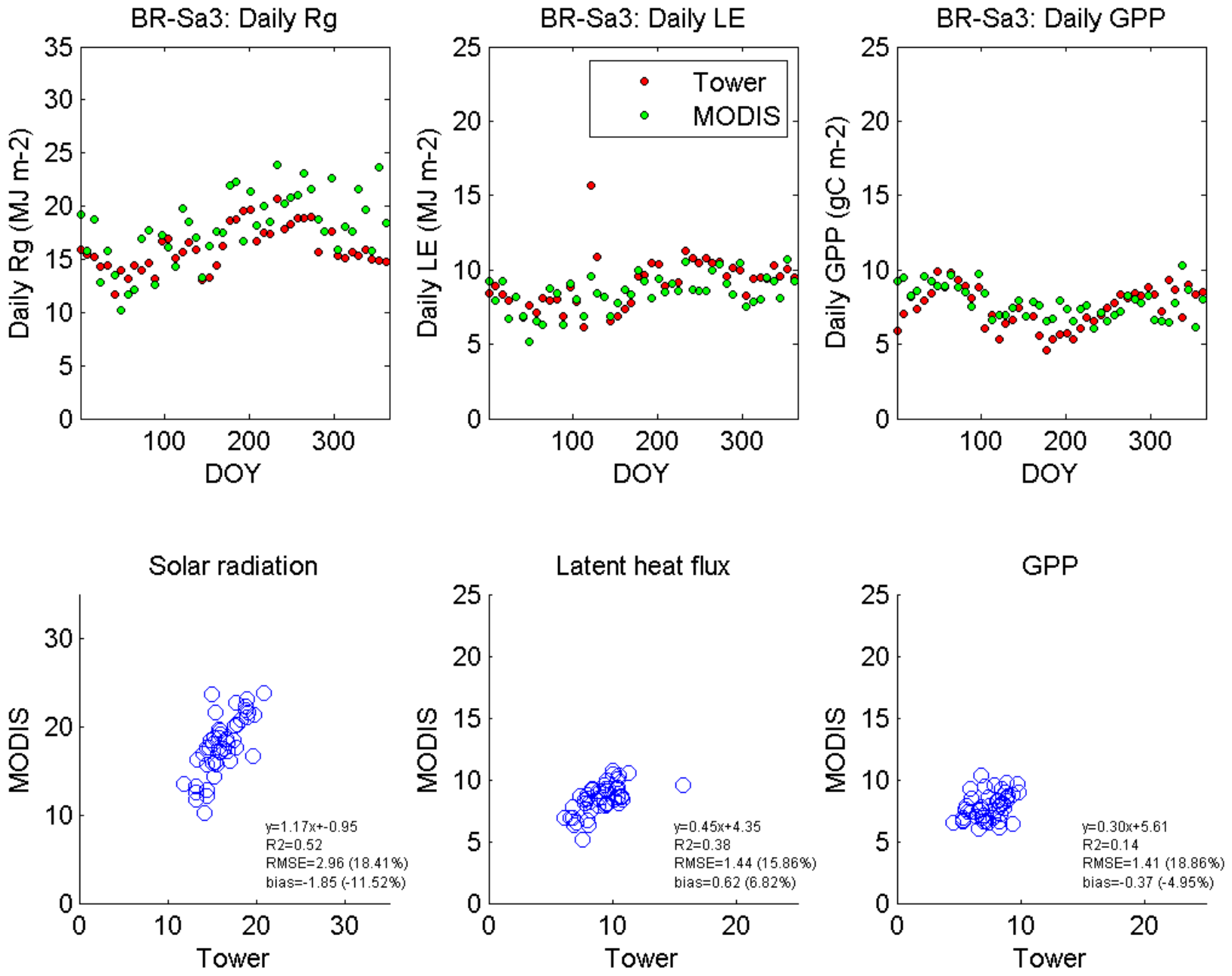
Black: Over 32 days how do 3 adjacent sensor vary compared to 0.1 deg Celsius?

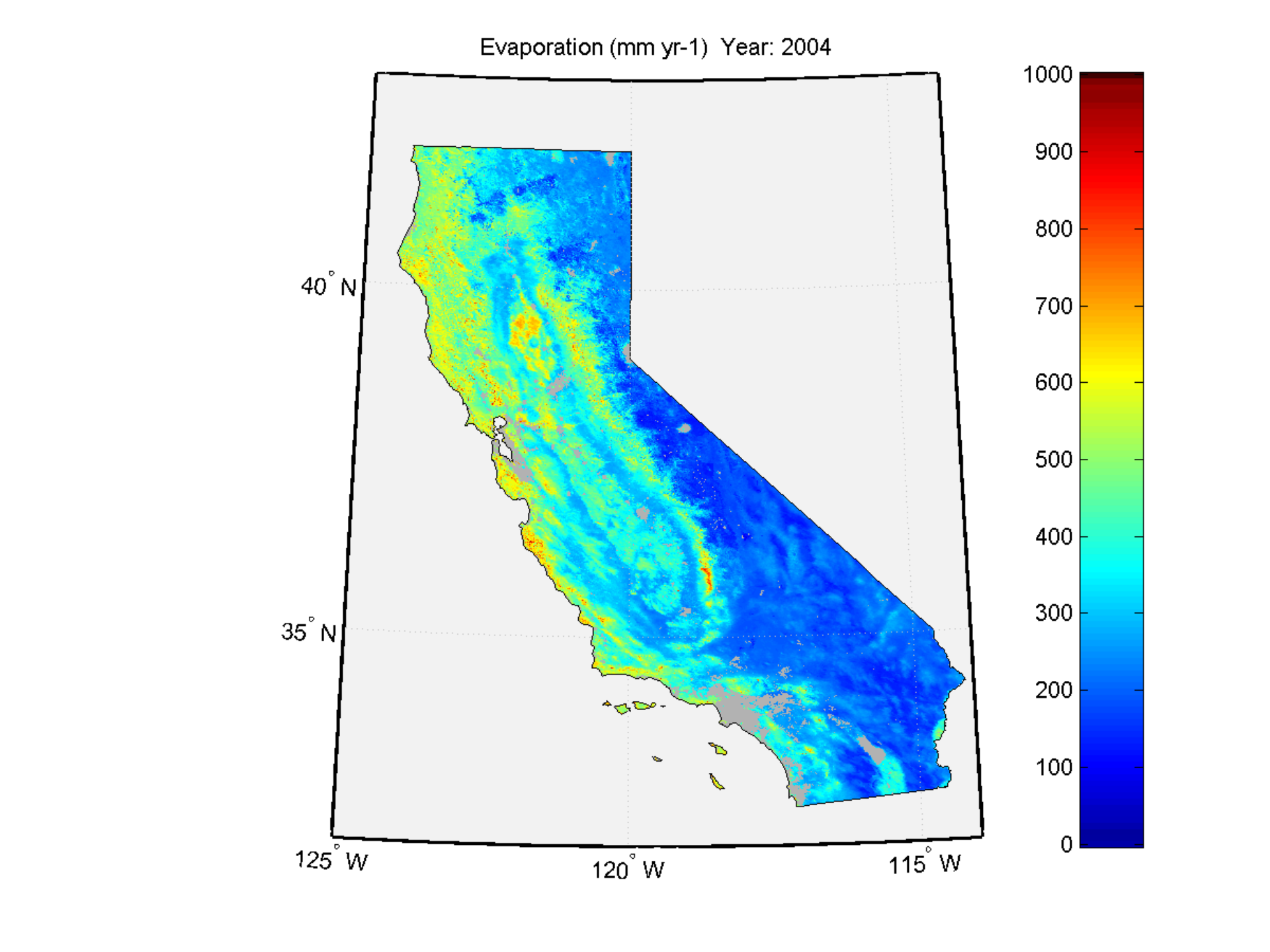
Red (right vertical axis): What is the temperature in the canopy?

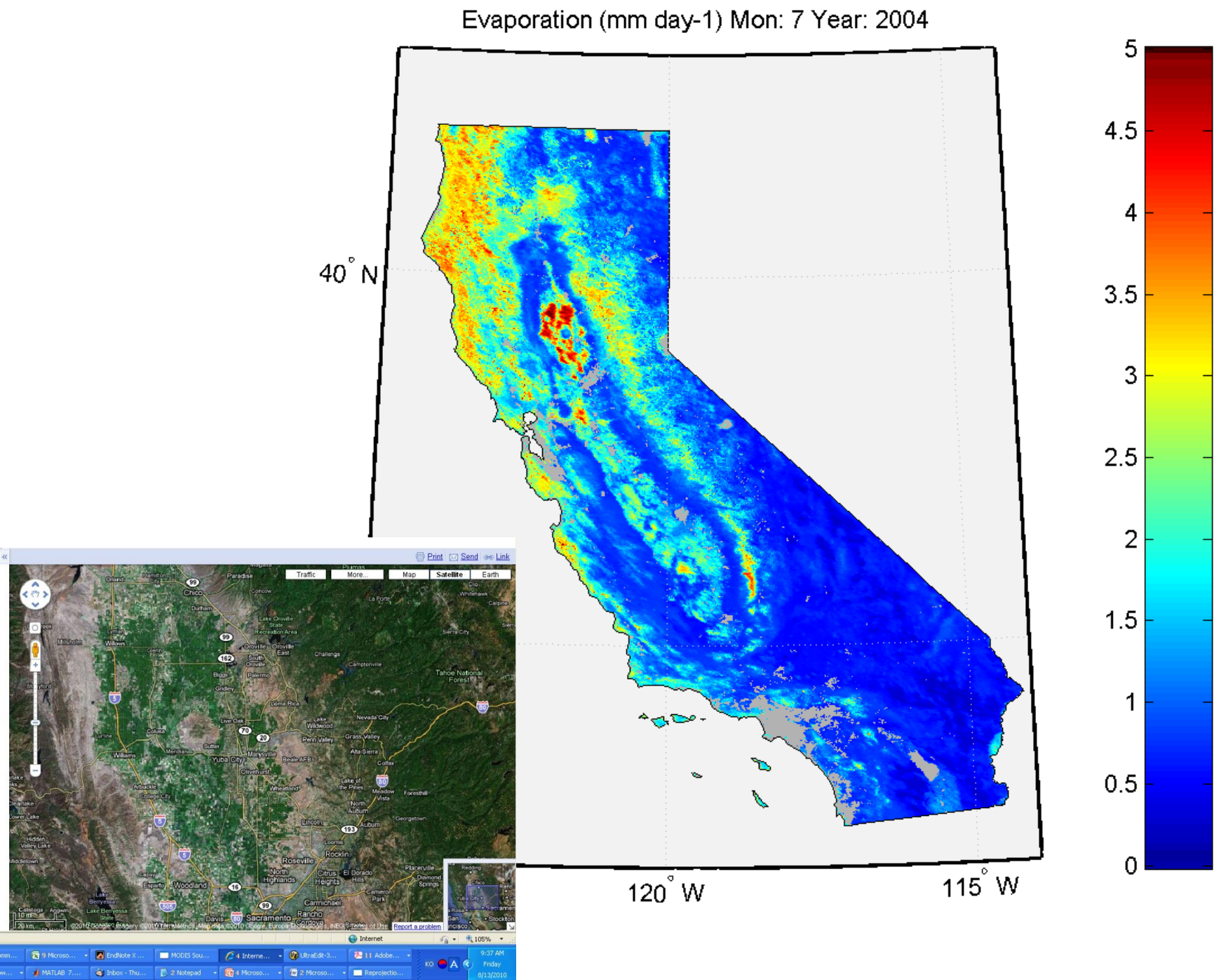


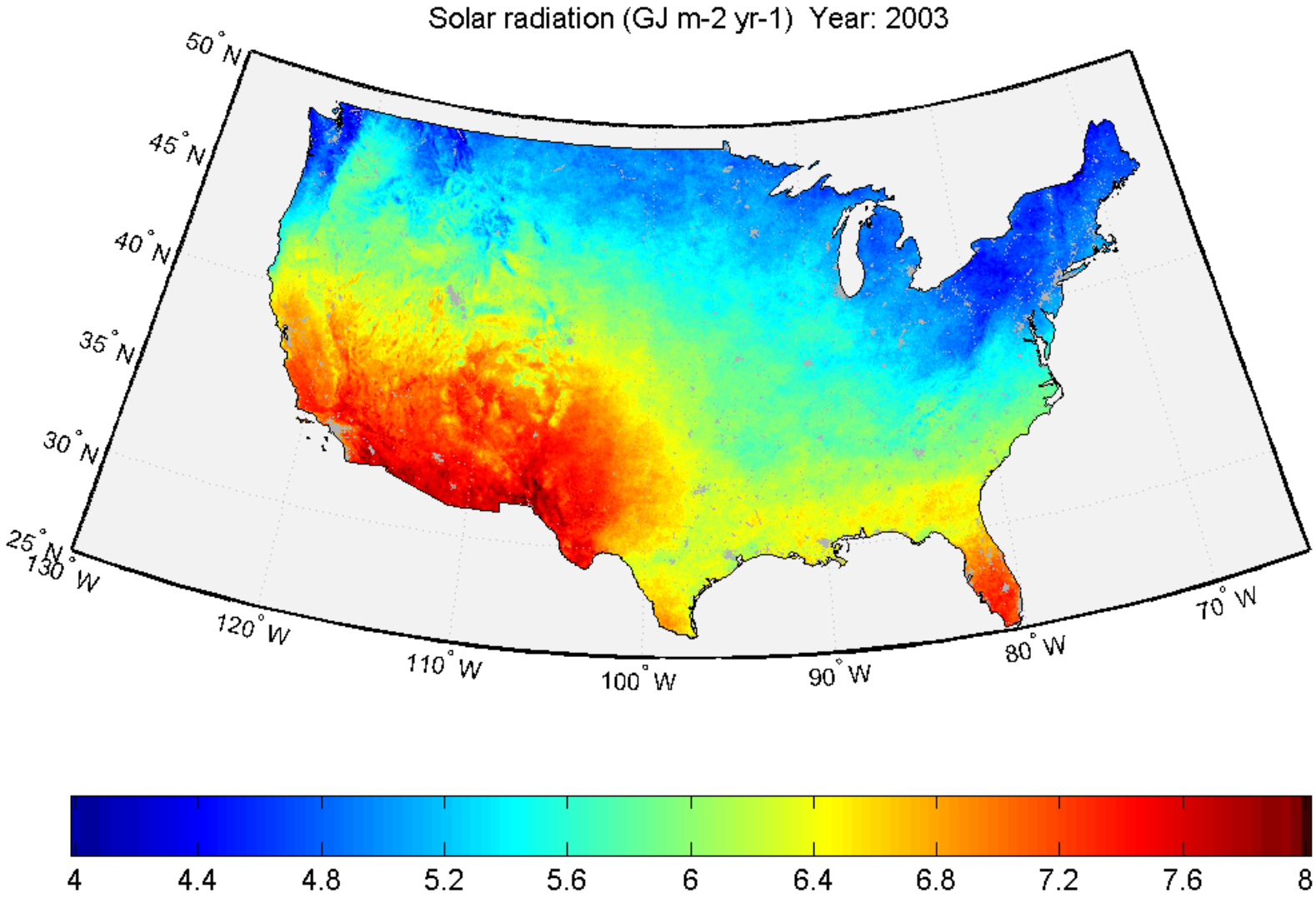
Remote sensing example

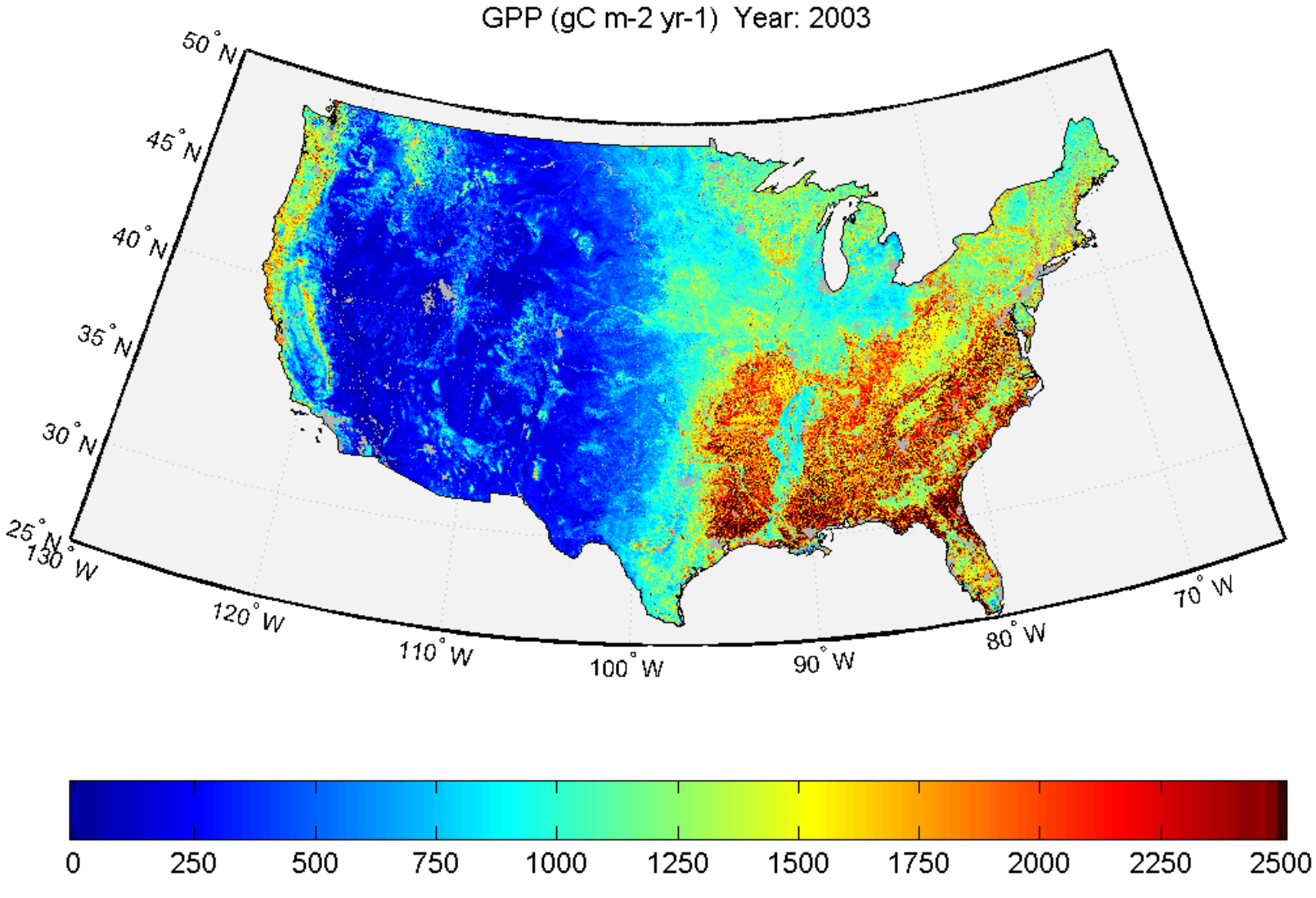
- MODIS has 36 spectral bands
  - MOD04 (aerosol)
  - MOD05 (precipitable water)
  - MOD06 (cloud)
  - MOD07 (atmospheric profile)
  - MOD11 (land surface temperature)
  - MCD12 (land cover)
  - MOD13 (vegetation index)
  - MCD15 (LAI)
  - MCD43 (albedo)

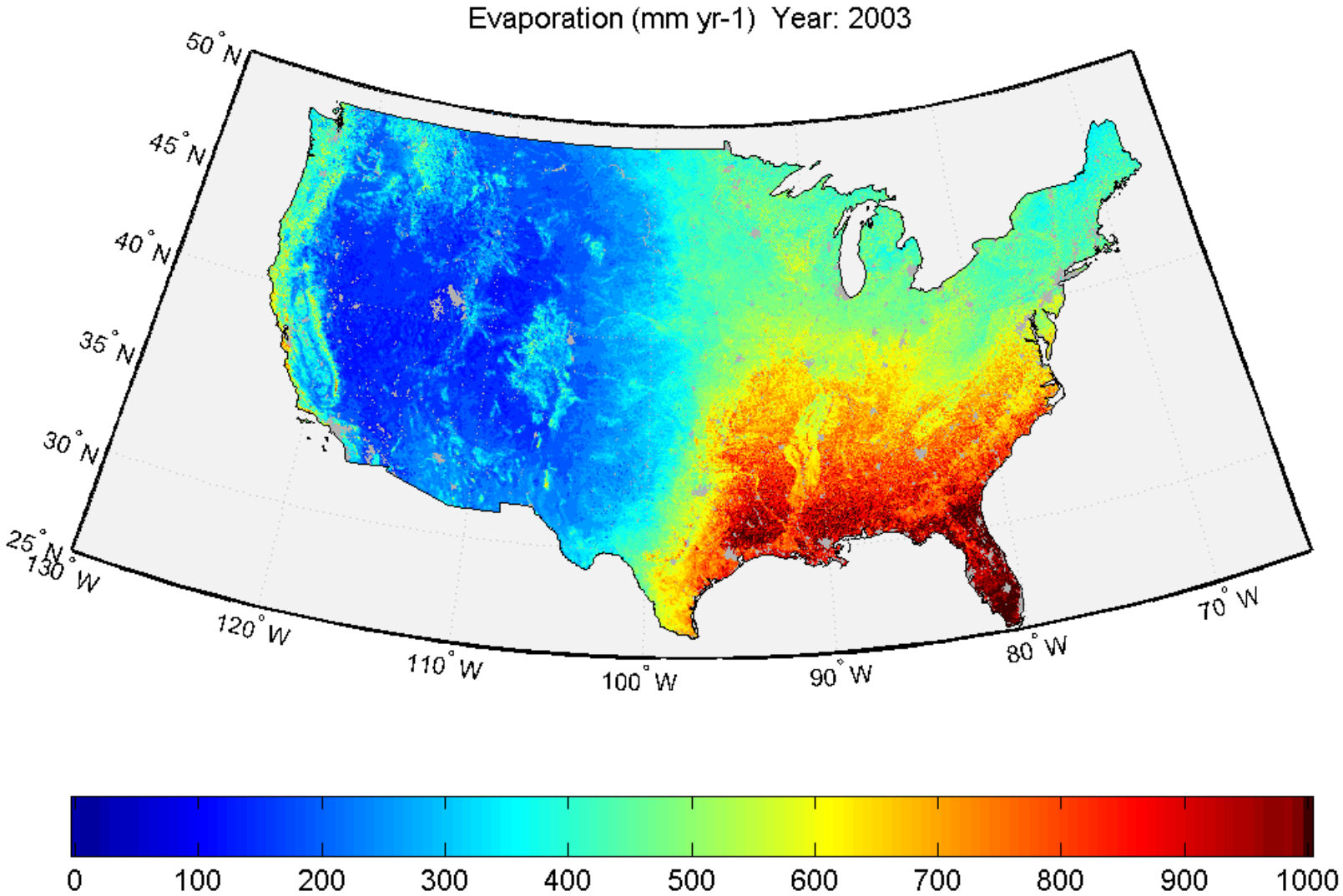

BR-Sa3: Daily Rg
Daily Rg (MJ m-2)
DOY
BR-Sa3: Daily LE
Daily LE (MJ m-2)
Tower
MODIS
DOY
BR-Sa3: Daily GPP
Daily GPP (gC m-2)
DOY
Solar radiation
MODIS
Tower
y=1.17x+-0.95
R2=0.52
RMSE=2.96 (18.41%)
bias=-1.85 (-11.52%)
Latent heat flux
MODIS
Tower
y=0.45x+4.35
R2=0.38
RMSE=1.44 (15.86%)
bias=0.62 (6.82%)
GPP
MODIS
Tower
y=0.30x+5.61
R2=0.14
RMSE=1.41 (18.86%)
bias=-0.37 (-4.95%)

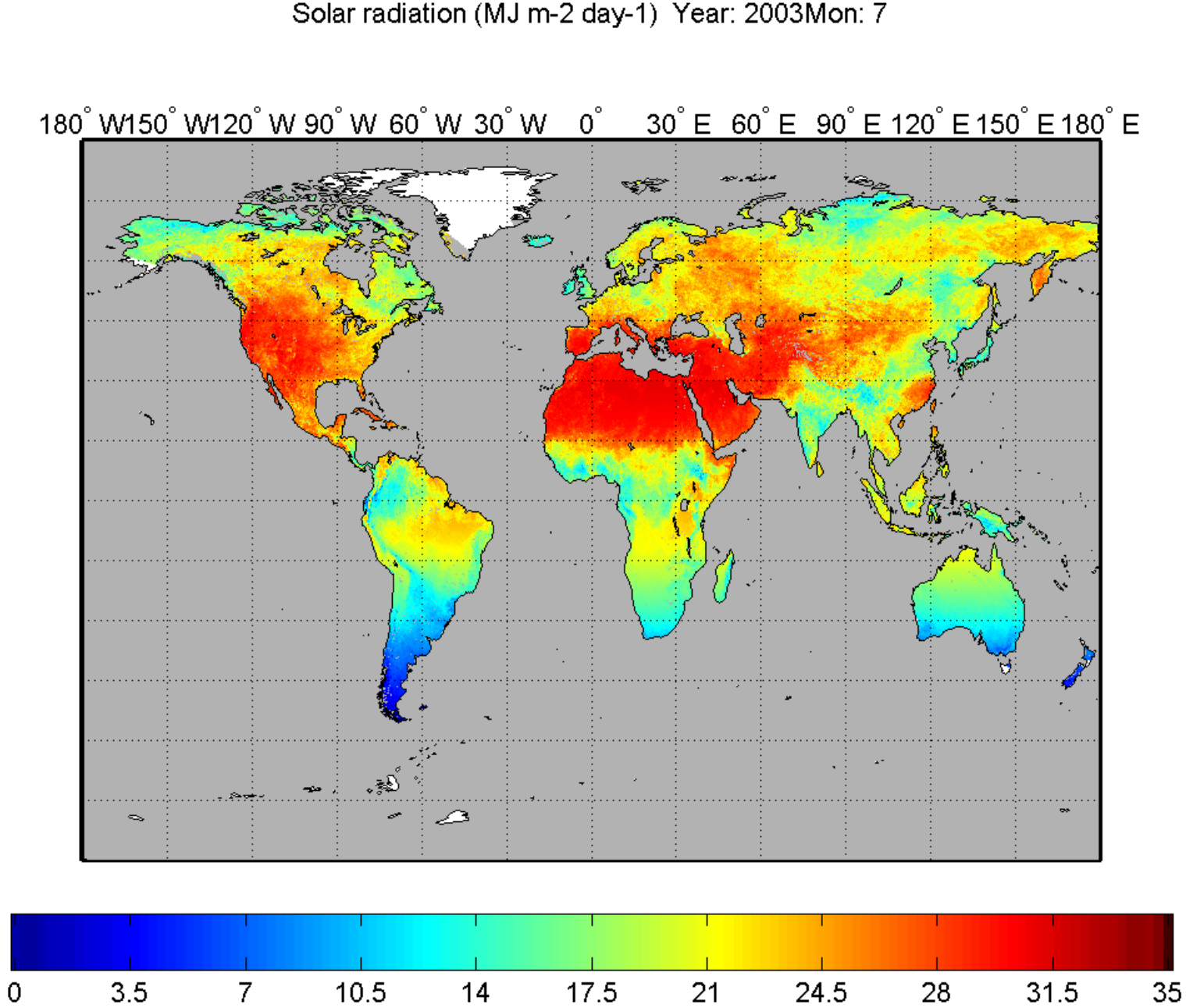

Evaporation (mm yr-1) Year: 2004
40° N
35° N
125° W
120° W
115° W
1000
900
800
700
600
500
400
300
200
100
0

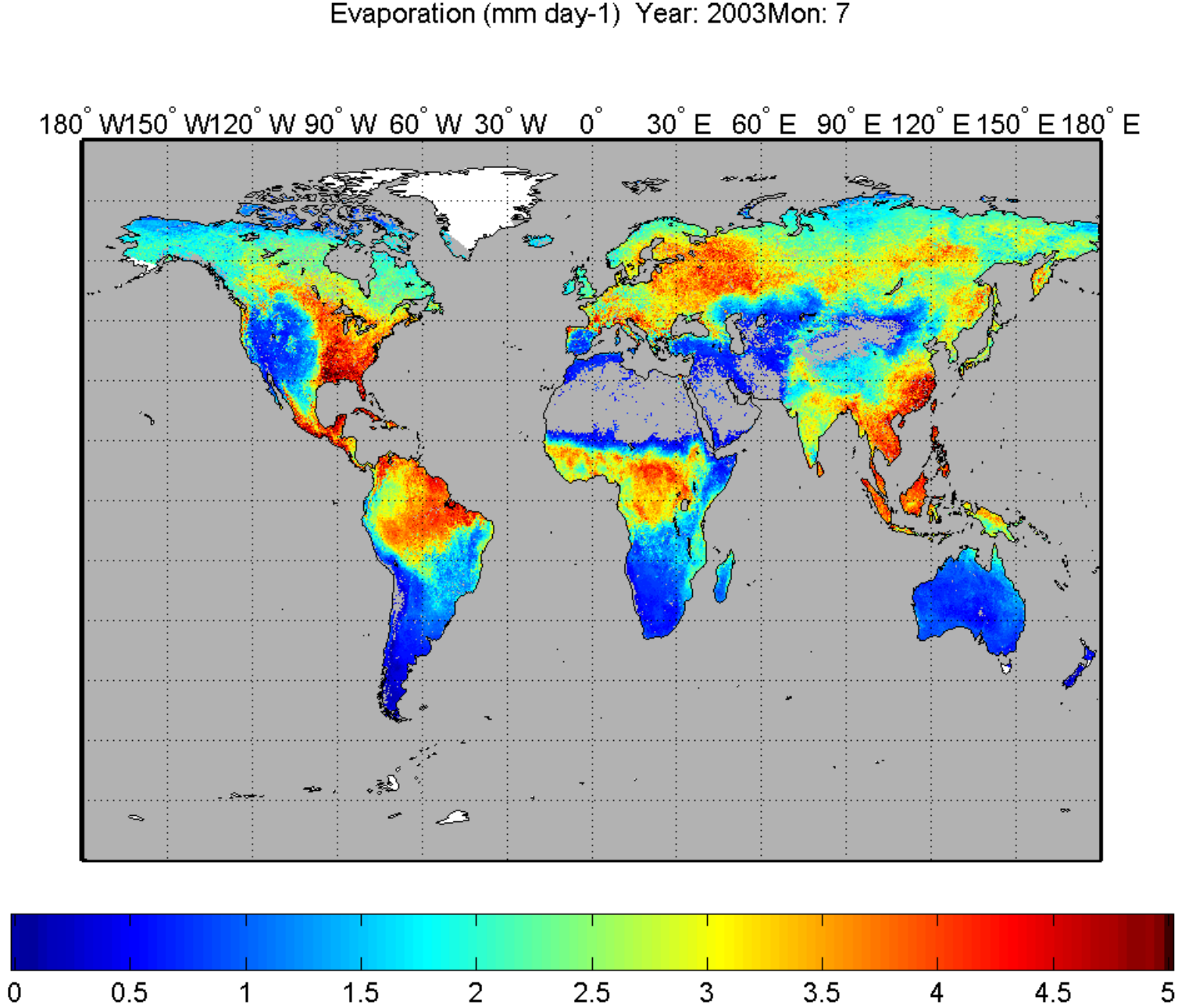

Evaporation (mm day-1) Mon: 7 Year: 2004
40° N
120° W
115° W
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
Print
Send
Link
Traffic
More...
Map
Satellite
Earth
Chico
Paradise
Oroville
Yuba City
Marysville
Roseville
Rocklin
Woodland
Davis
Sacramento
Rancho Cordova
Carmichael
Tahoe National Forest
Internet
105%
9:37 AM
Friday
8/13/2010

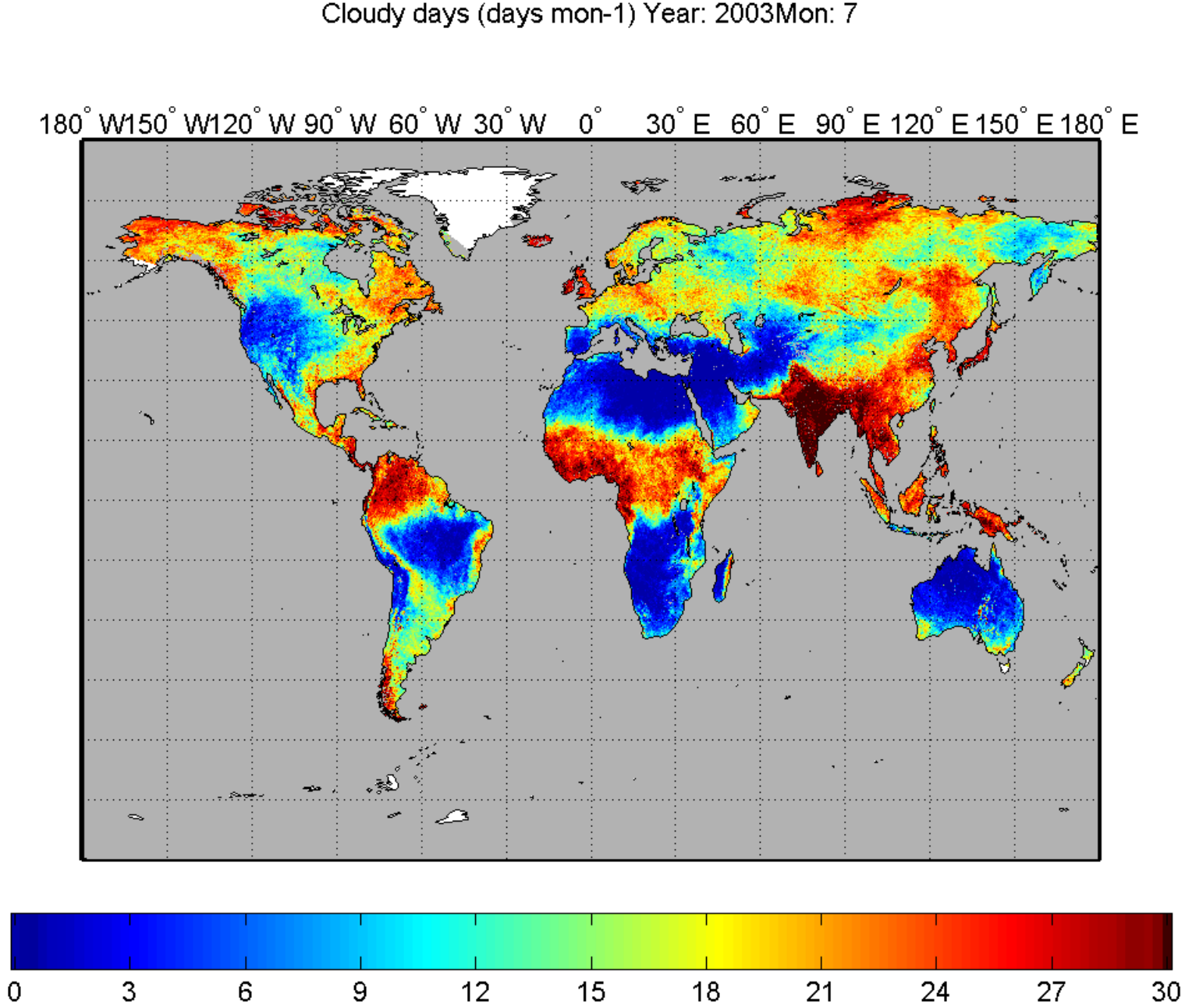

Solar radiation (GJ m-2 yr-1) Year: 2003
50˚N
45˚N
40˚N
35˚N
30˚N
25˚N
130˚W
120˚W
110˚W
100˚W
90˚W
80˚W
70˚W
4
4.4
4.8
5.2
5.6
6
6.4
6.8
7.2
7.6
8


GPP (gC m-2 yr-1) Year: 2003
50˚N
45˚N
40˚N
35˚N
30˚N
25˚N
130˚W
120˚W
110˚W
100˚W
90˚W
80˚W
70˚W
0
250
500
750
1000
1250
1500
1750
2000
2250
2500


Evaporation (mm yr-1) Year: 2003
50˚ N
45˚ N
40˚ N
35˚ N
30˚ N
25˚ N
130˚ W
120˚ W
110˚ W
100˚ W
90˚ W
80˚ W
70˚ W
0
100
200
300
400
500
600
700
800
900
1000


Solar radiation (MJ m-2 day-1) Year: 2003Mon: 7
180° W150° W120° W 90° W 60° W 30° W 0° 30° E 60° E 90° E 120° E 150° E 180° E
0 3.5 7 10.5 14 17.5 21 24.5 28 31.5 35


Evaporation (mm day-1) Year: 2003Mon: 7
180° W150° W120° W 90° W 60° W 30° W 0° 30° E 60° E 90° E 120° E 150° E 180° E
0 0.5 1 1.5 2 2.5 3 3.5 4 4.5 5


Cloudy days (days mon-1) Year: 2003Mon: 7
180° W150° W120° W 90° W 60° W 30° W 0° 30° E 60° E 90° E 120° E 150° E 180° E
0 3 6 9 12 15 18 21 24 27 30

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