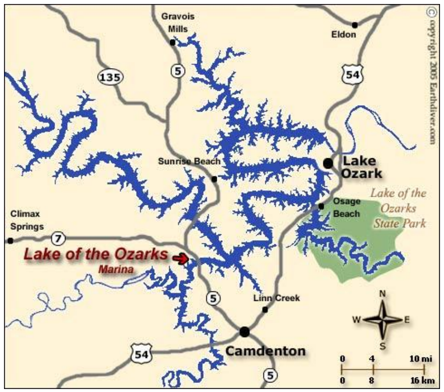
Scaling Science: Breathing Earth Science Simulator

Catharine van Ingen
Partner Architect
eScience Group, Microsoft Research
CSIRO 26 October 2010

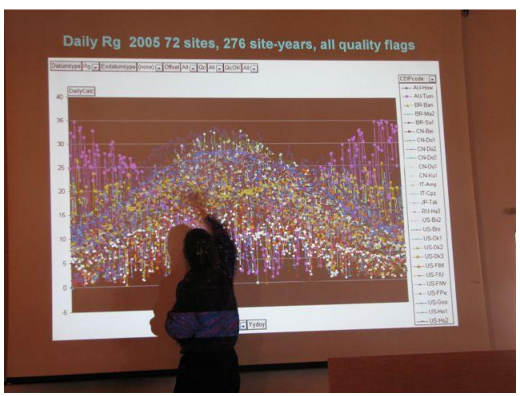
I've come full circle



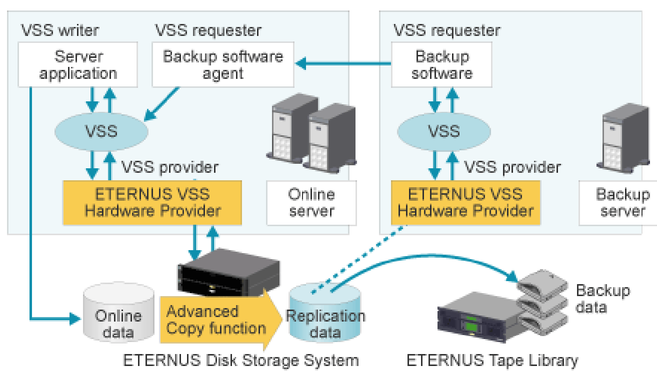
Modeler

Experimentalist

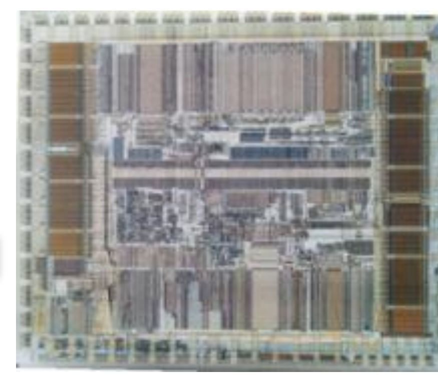
Detector integrator



Environmental eScientist



Storage architect



Computer architect

Introduction

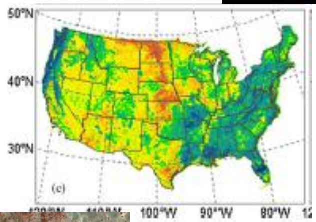
- The bulk of this talk is from two recent talks by Dennis Baldocchi and Youngryel Ryu.
  - I've been very fortunate in collaborating with and learning from them over the past 5 years.
  - The science insights are theirs, the mistakes are mine
- We're only just beginning to understand our recent results
  - We know we need help both in the validation as well as potential ideas for near term analyses

The Breathing of an Ecosystem is Defined by the Sum of an Array of Coupled, Non-Linear, Biophysical Processes that Operate across a Hierarchy/Spectrum of Time and Space Scales

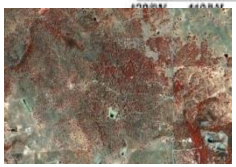
Spatial Scale of Inquiry

Spans 13–14 orders of Magnitude

Globe: 10,000 km (10^7 m)



Continent: 1000 km (10^6 m)



Landscape: 1–100 km

Canopy: 100–1000 m

Plant: 1–10 m

Leaf: 0.01–0.1 m

Stomata: 10^{-5} m

Bacteria/Chloroplast: 10^{-6} m

Temporal Scale of Inquiry

Spans 12-15 orders of Magnitude

- Seconds/Hours/Day, 10-6 to 10-3 yr
  - Physiology
    - Photosynthesis, Respiration, Transpiration, Stomatal Conductance
- Seasonal & Annual, 10^0 yr
  - Net and Gross Primary Productivity
  - Autotrophic and Heterotrophic Respiration and Decompos
  - Plant Acclimation
  - Mineralization and Immobilization

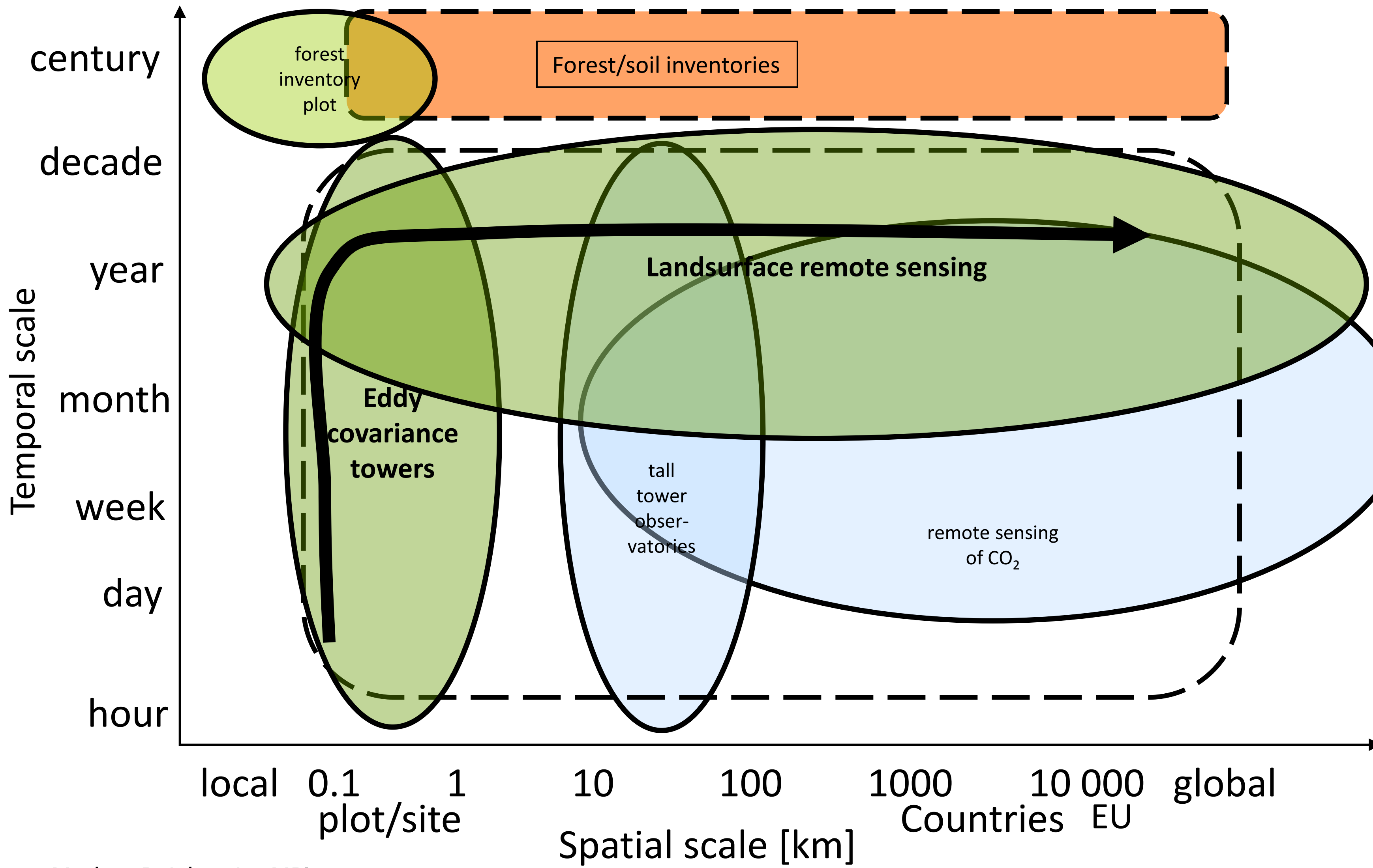
- Decadal, 10^1 yr
  - Competition, Gap-Replacement, Stand Dynamics
  - Changes in Soil Organic Matter

- Century, 10^2 yr
  - Succession, Mortality
- Millennia, 10^3 yr
  - Species migration
  - Soil Formation
- Geological Periods, 10^6 yr
  - Evolution, Speciation, Extinction, Climate Regimes
- Eons, 10^9 yr
  - Evolution of Life and the Formation of our Atmosphere

Big Picture Question Regarding Predicting and Quantifying the 'Breathing of the Biosphere':

How can We Be 'Everywhere All the Time?'

Components of an Integrated Earth System EXIST, but are Multi-Faceted



From: Markus Reichstein, MPI

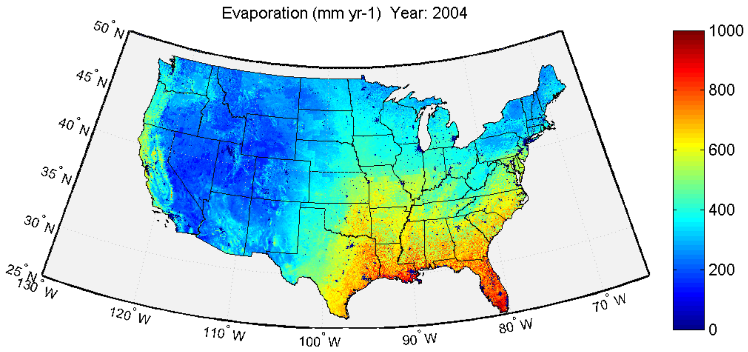
Challenges and Points of Collaboration between the Computational and Scientific Communities

Upscaling Gas Fluxes from Leaves to Landscapes

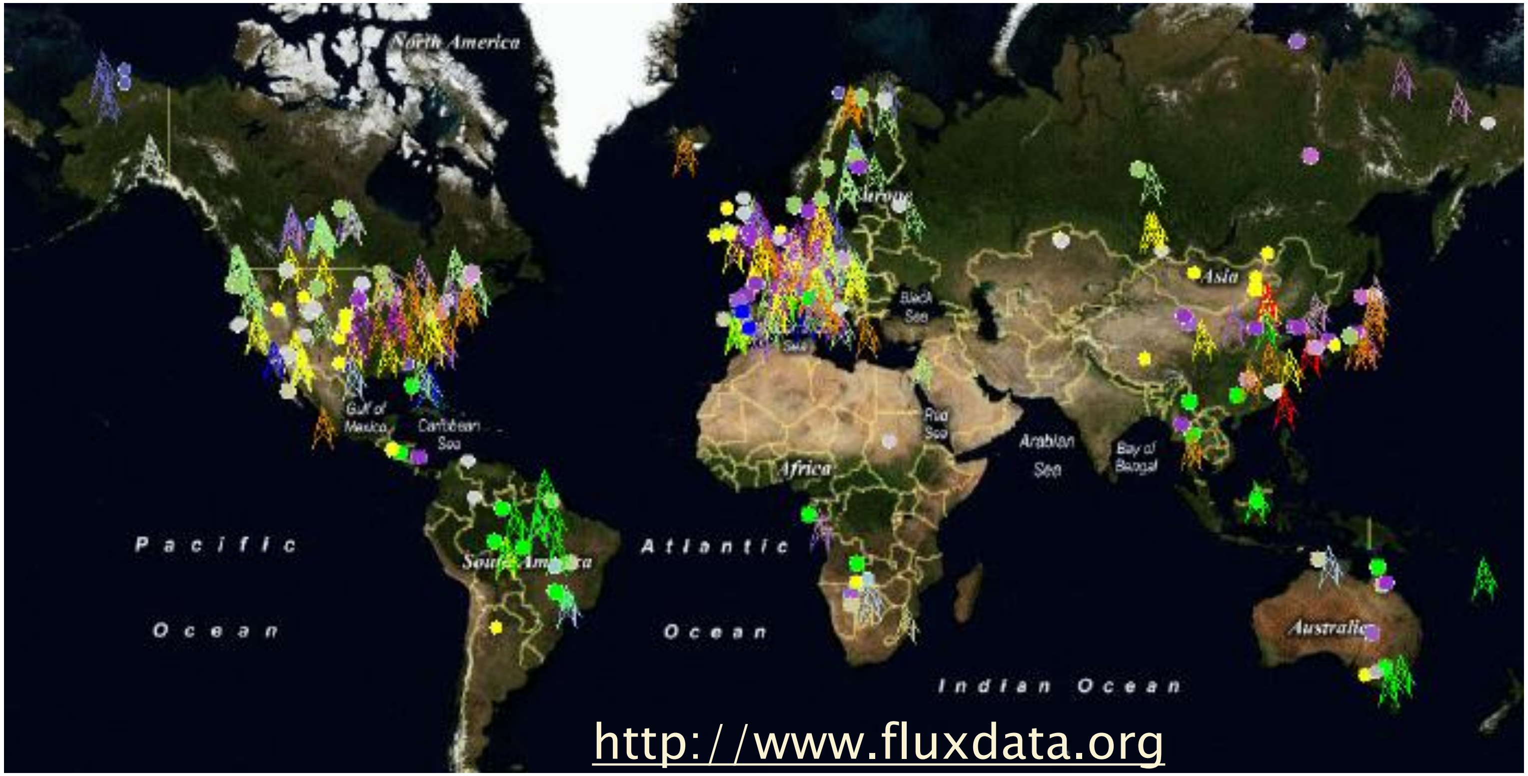
➢Requires Complex, Monte-Carlo Light-Rendering, like that used in 'A Bug's Life'

Upscaling Fluxes from Landscapes to the Globe

➢ Acquire, Process, Distill and Disseminate Vast Quantities of Flux Data for Model Ground-Truthing, Validation and Parameterization



Fluxnet : A Network of Networks



http://www.fluxdata.org

- 467 towers world wide
- 967 site-years of sensor data from 253 towers (~800K data points)
- ~20 sensor measurements per tower; 20 derived science variables
- 145 ancillary variables

- Original data set assembled and processed in 2007
- 20x larger than previous synthesis dataset
- Currently 85 paper teams with over 400 scientists

Study Sites

Crops: soybeans, alfalfa, wheat, corn, rice

Grassland and peatland pastures

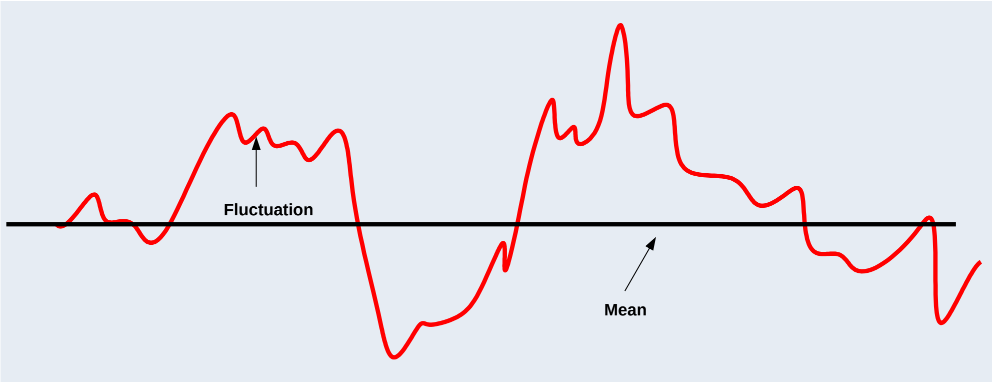
Boreal Conifer Forest

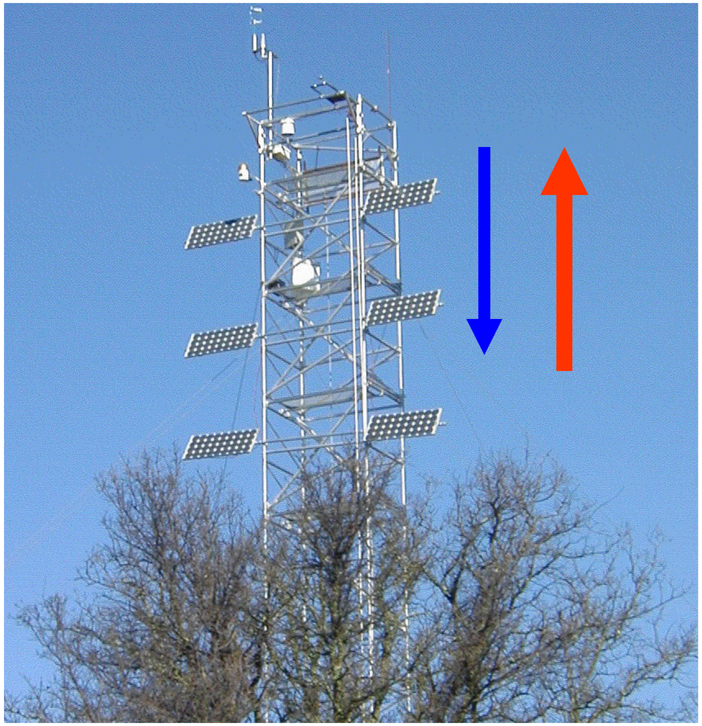
Deciduous Forest

Savanna Woodland

Eddy Covariance Fluxes

$$F = \overline{w'c'}$$





Net Flux is the sum of the mass flux moving across a plane (w C) in the up and down-drafts of Air and Wind Sampled 10 times per second

Motivation for BESS

- Current Global-Scale Remote Sensing Products tend to rely on
  - Highly-Tuned Light Use Efficiency Approach
    - GPP=PAR*fPAR*LUE (since Monteith 1960's)
  - Empirical approach (machine learning technique)
  - Some Forcings come Satellite Remote Sensing at fine scale
  - Others come from coarse reanalysis data (several tens of km resolution)
- Hypothesis: We can do Better by:
  - Applying the Principles which Reflect Intellectual Advances in these Fields over the past Decade
  - Merging Vast Environmental Databases Utilizing Cloud Computational Resources

Lessons Learned from CanOak

25+ years of Developing and Testing a Hierarchy of Scaling Models with Flux Measurements at Contrasting Oak Woodland Sites in Tennessee and California

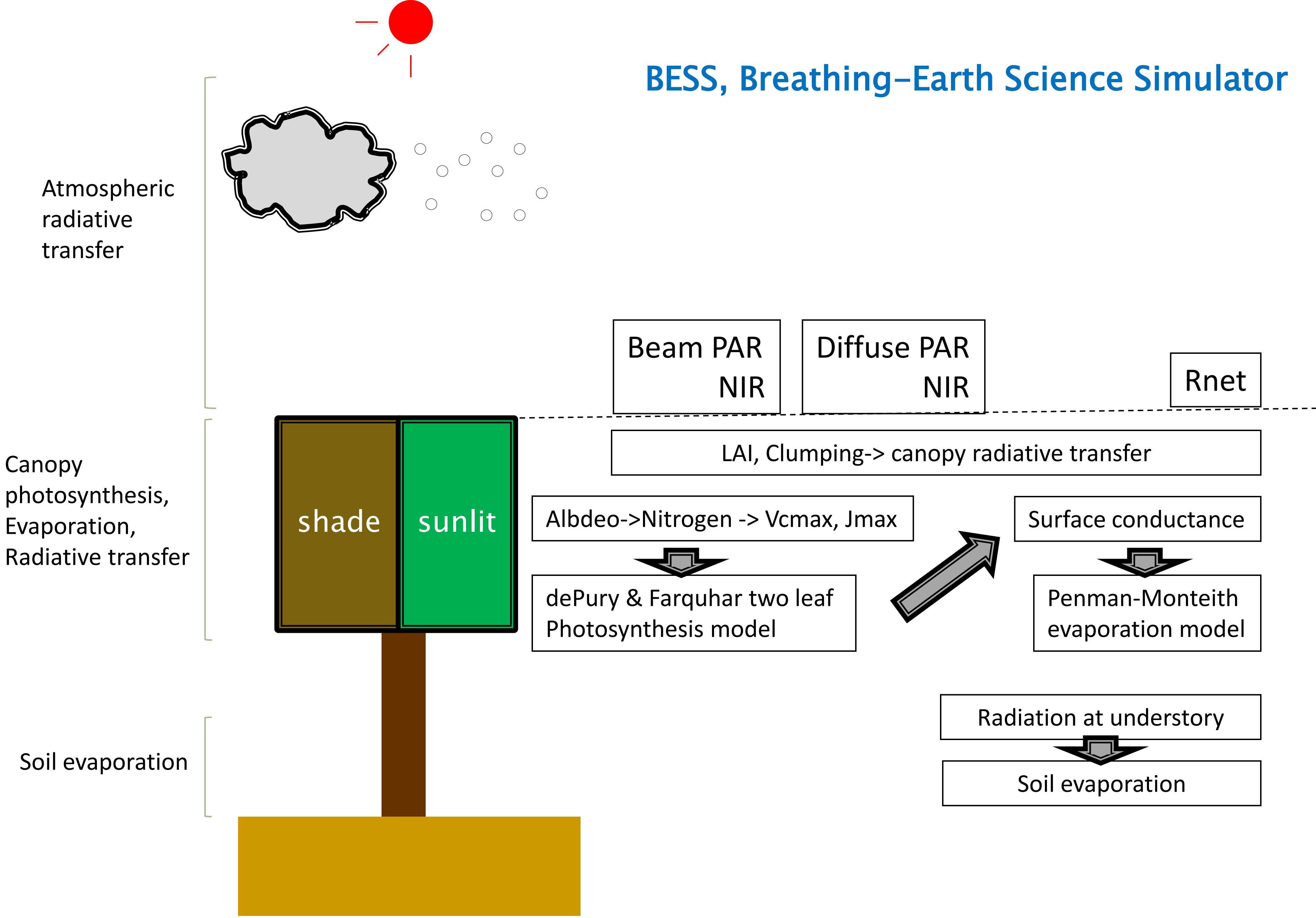
We Must:

- Couple Carbon and Water Fluxes
- Assess Non-Linear Functions with Leaf-Level Microclimate Conditions
- Consider Sun and Shade fractions separately
- Consider effects of Clumped Vegetation on Light Transfer
- Consider Seasonal Variations in Physiological Capacity of Leaves and Structure of the Canopy

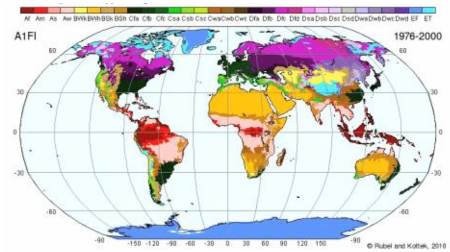
Necessary Attributes of the Next-Generation Global Biophysical Model, BESS

- Direct and Diffuse Sunlight
  - Monte Carlo Atmospheric Radiative Transfer model (Kobayashi)
  - Light transfer through canopies consider leaf clumping
- Coupled Carbon-Water for Better Stomatal Conductance Simlulations
  - Photosynthesis and Transpiration on Sun/Shade Leaf Fractions (dePury and Farquhar, 1996)
  - Photosynthesis of C3 and C4 vegetation considered
- Ecosystem Scaling Relations to parameterize models, based on remote sensing spatio-temporal inputs
  - Vcmax=f(N)=f(albedo) (Ollinger et al; Hollinger et al;Schulze et al.; Wright et al.
  - Seasonality in Vcmax is considered
- Model Predictions should Match Fluxes Measured at Ecosystem Scale hourly and seasonally.

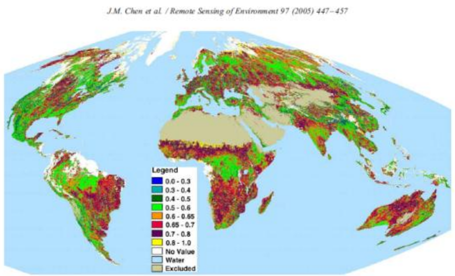
Physics wins; Biology is how it works.


BESS, Breathing-Earth Science Simulator
Atmospheric radiative transfer
Canopy photosynthesis, Evaporation, Radiative transfer
Soil evaporation
shade
sunlit
Beam PAR NIR
Diffuse PAR NIR
Rnet
LAI, Clumping-> canopy radiative transfer
Albdeo->Nitrogen -> Vcmax, Jmax
dePury & Farquhar two leaf Photosynthesis model
Surface conductance
Penman-Monteith evaporation model
Radiation at understory
Soil evaporation

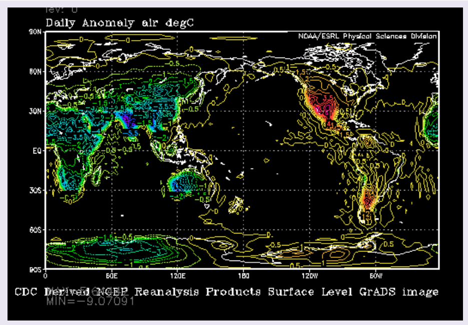
Imagery, Sensors, Models and Field Data



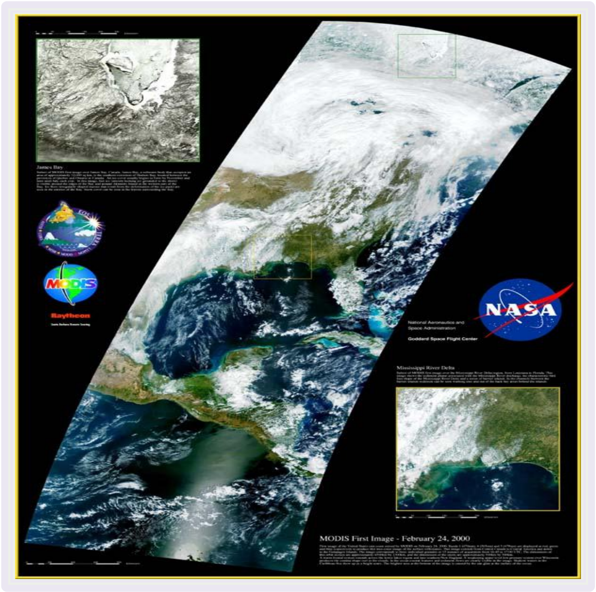
Climate classification
~1MB (1file)



Vegetative clumping
~5MB (1file)

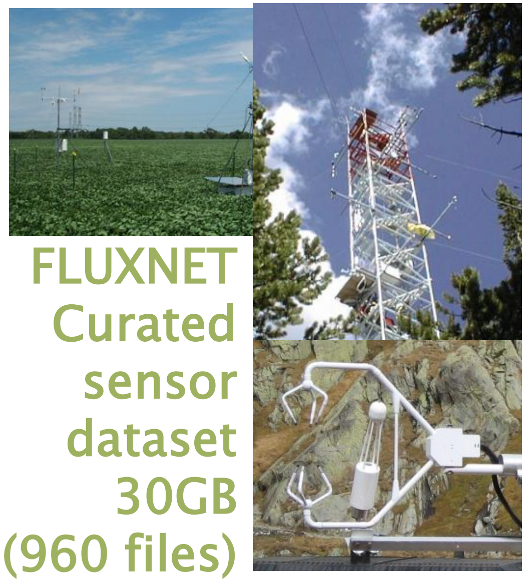


NCEP/NCAR
~100MB (4K files)

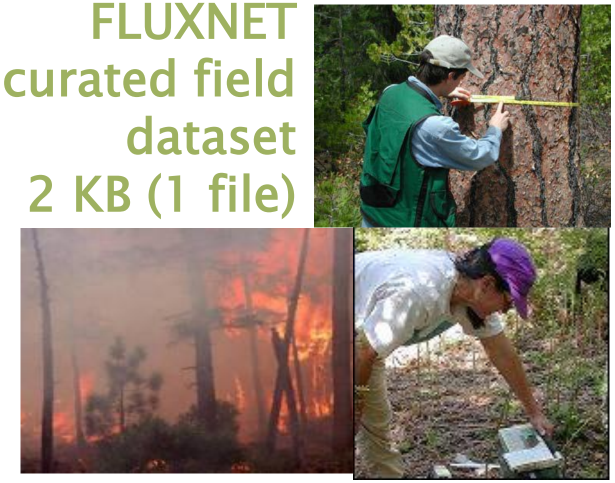


NASA MODIS imagery
archives
5 TB (600K files)

Sizes given are 1 US year
20 US year ~ 1 global land surface year



FLUXNET
Curated
sensor
dataset
30GB
(960 files)



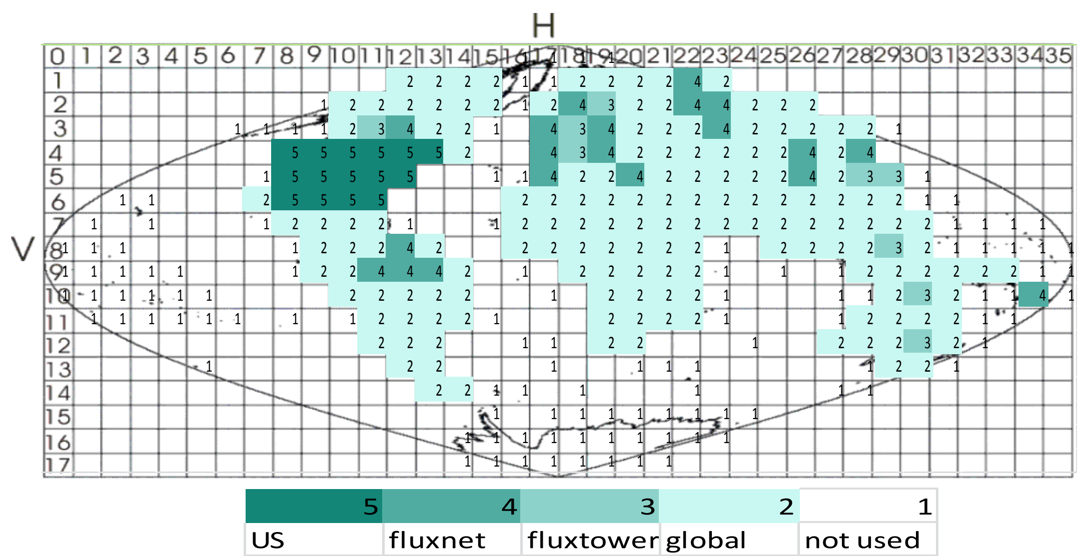
FLUXNET
curated field
dataset
2 KB (1 file)

The Resource Barrier:
Not Everyone Has Access to a Supercomputer

	#Source Files	Source Size	# Result Files	Result Size
USA (18)	21850	238 GB	27375	261 GB
FluxTower (3)	80670	993 GB	58400	210 GB
Global (3)	152670	2414 GB	352225	630 GB

1 Global year

- ~ 9000 hours or 375 days
  - 1 year takes 1 year !
  - With 200 cpus, that's > 2 days
- 8 TB download reduced to 2.4 TB by reprojection
  - How to know which are missed (and why) among 350K files?



US: 15 tiles FluxTower: 32 tiles Global: 194 tiles

The Tedium Barrier:

Extracting Data Drivers from Global Remote Sensing to Run the Model

MOD04	aerosol
MOD05	Precipitable water
MOD06	cloud
MOD07	Temperature, ozone
MCD43	albedo
MOD11	Skin temperature
MOD15	LAI
POLDER	Foliage clumping
MCD12Q	Land cover

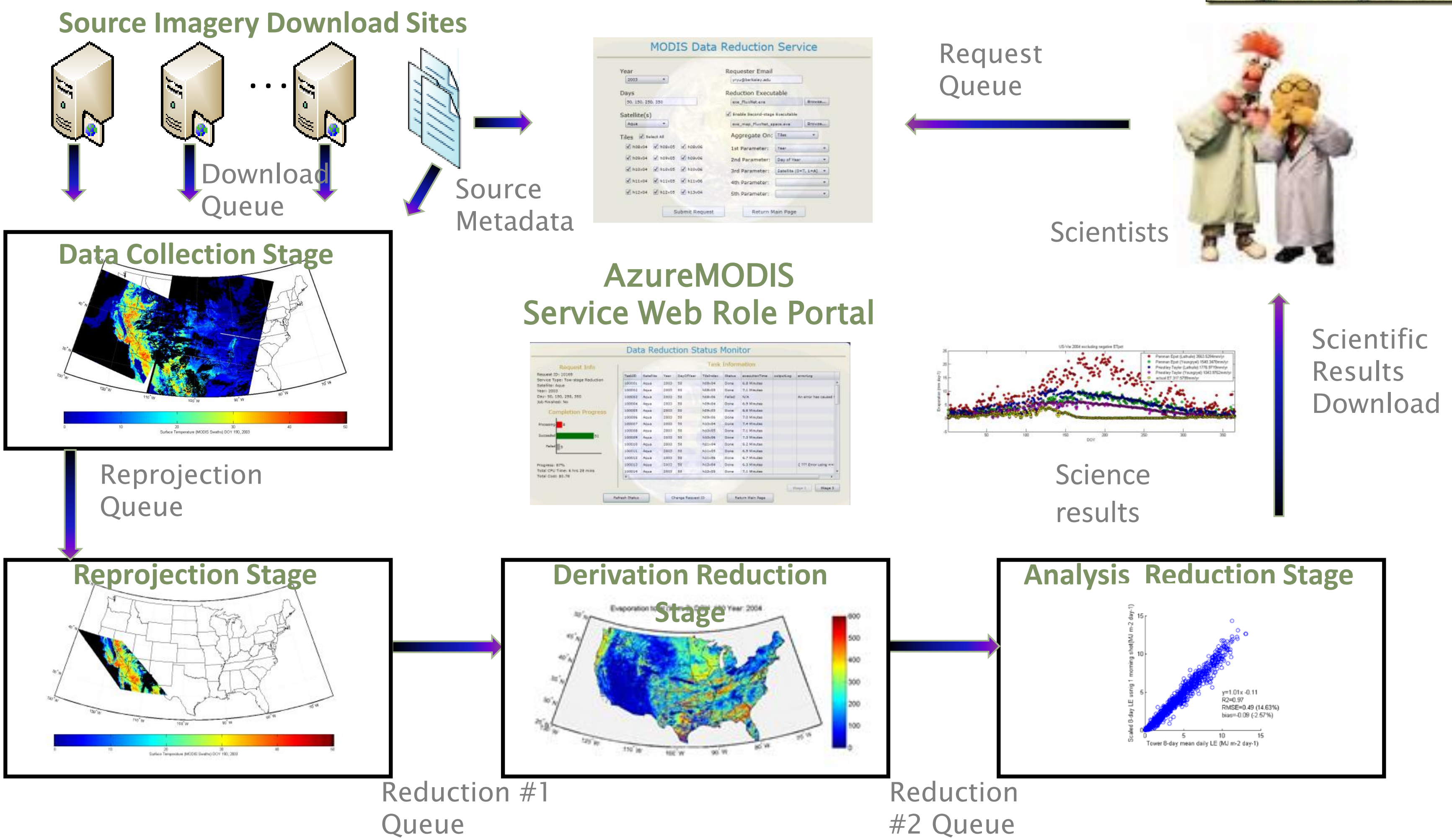
Atmospheric radiative transfer

Net radiation

Canopy radiative transfer

- Each science variable is associated with a MODIS product
  - Terra satellite products (eg MOD04) used preferentially as they tend to be day time observations
  - Aqua satellite products (eg MYD04) used when Terra products unavailable
  - MCD products are higher level products generated by a combination of Terra and Aqua
  - MOD44B from 2000 used throughout
- Each product is either swath or sinusoidal projection
  - Sinusoidal are ready to use
  - Groups of swath products must be reprojected to create a sinusoidal tile
- Each product has a recurrence interval of daily, 8 day, 16 day, annual

MODISAzure

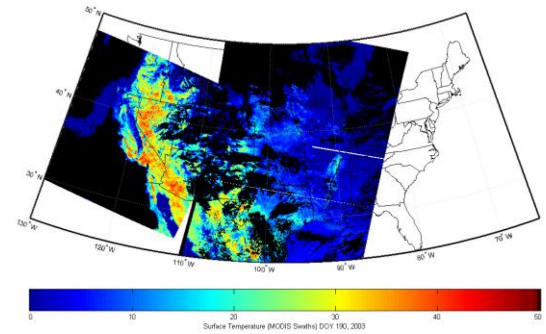


http://research.microsoft.com/en-us/projects/azure/azuremodis.aspx

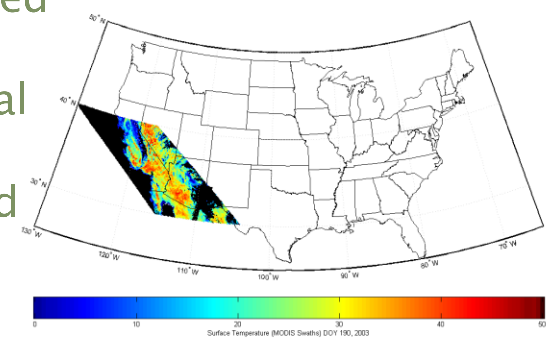
Tasks Performed with MODISAzure

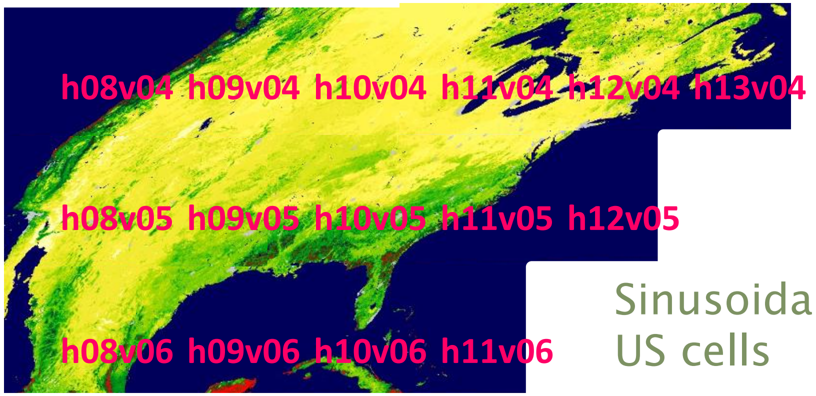
- Transparent Download
  - Example: downloading 4 swath product data files necessary for reprojection to one swath file
- Reprojection
  - Example: latitude-longitude swaths converted to sinusoidal cells.
- Spatial resampling
  - Example: converting from 1 KM to 5 KB pixels.
- Temporal resampling
  - Example: converting from daily observation to 8 day averages.
- Gap filling
  - Example: assigning values to pixels without data either due clouds or satellite outages
- Masking
  - Example: computing a land product or outside a spatial feature such as a watershed.

Source Data (Swath format)



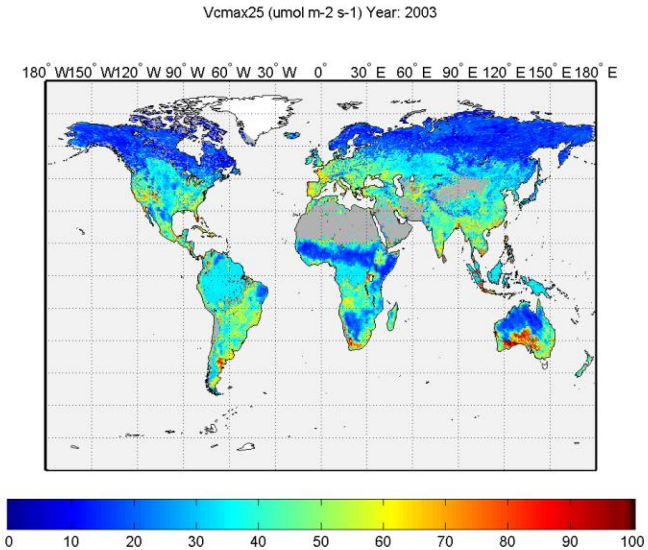
Reprojected Data (Sinusoidal format – equal land area pixel)

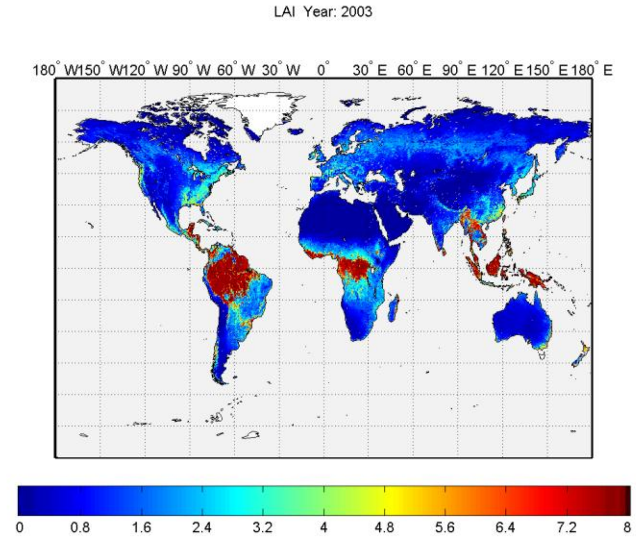


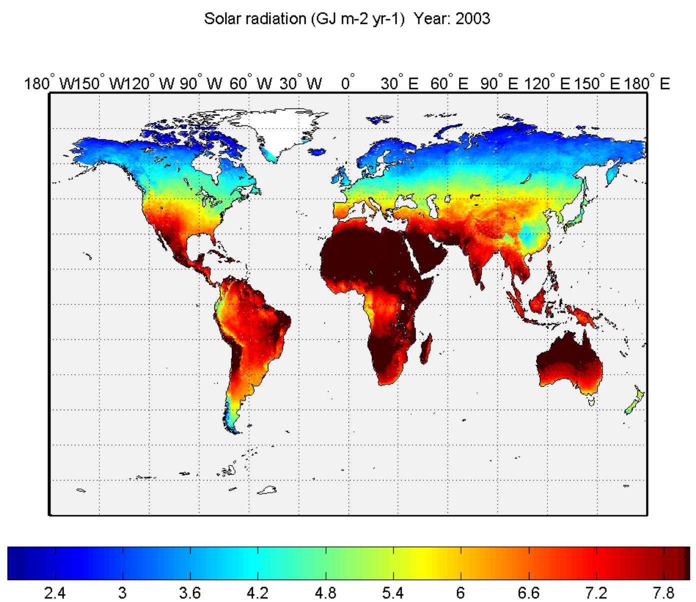


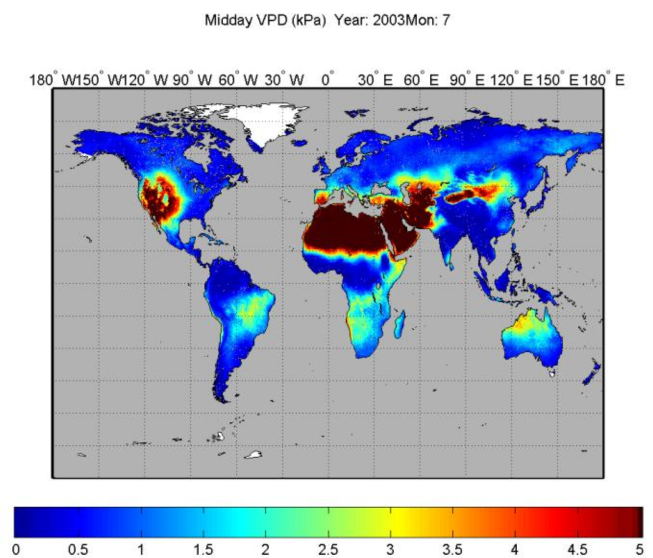
Sinusoidal US cells

Generated Data Layers for Model Forcing

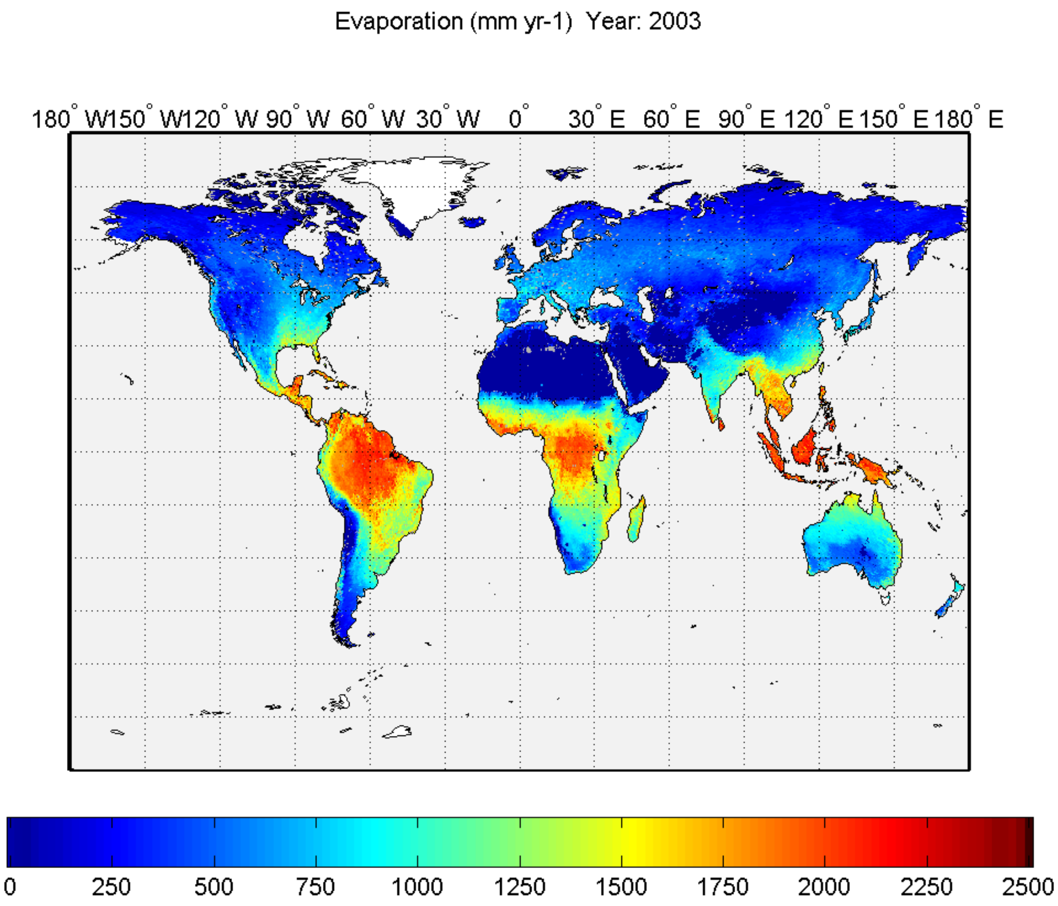






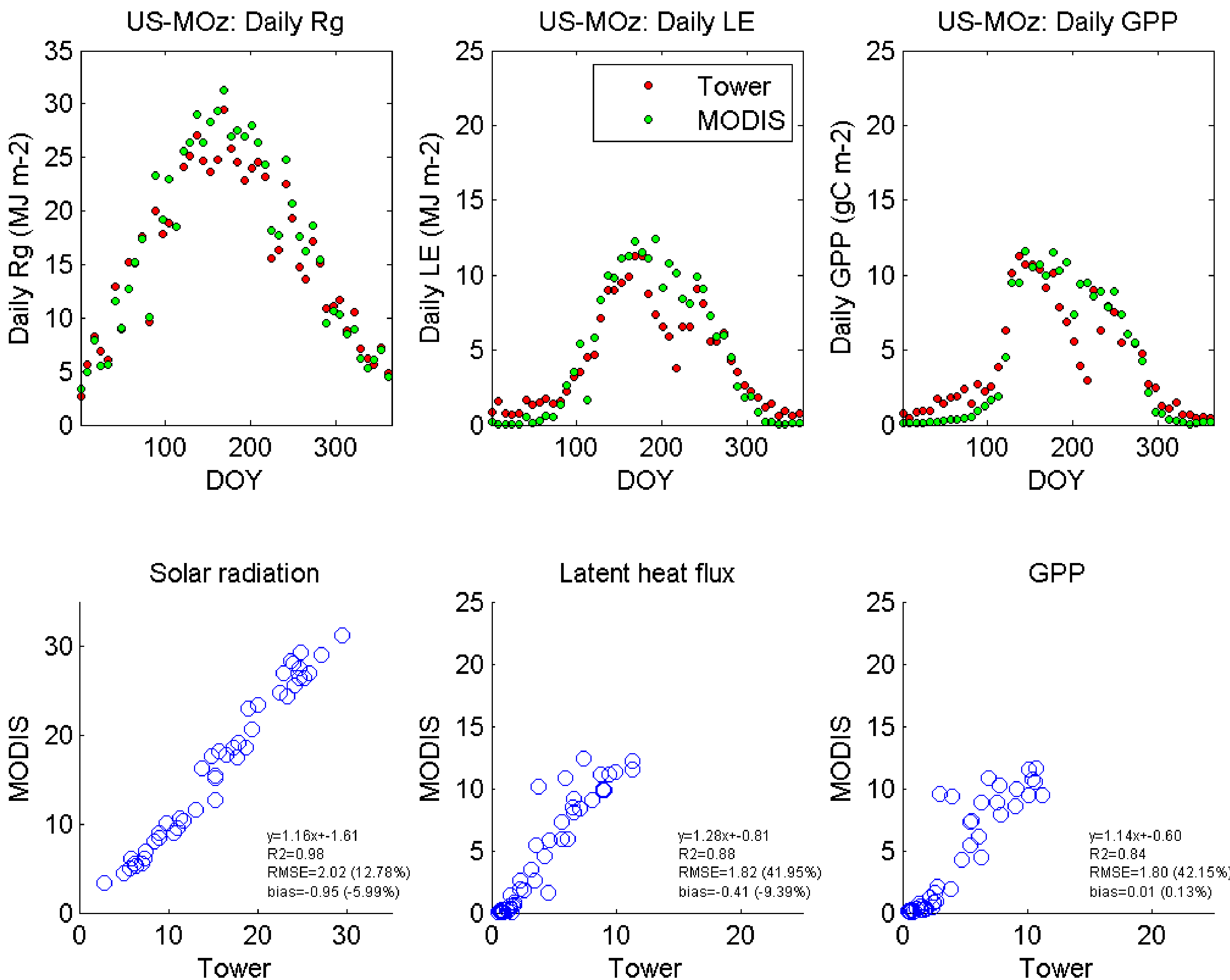


Global Evaporation at 1 to 5 km scale



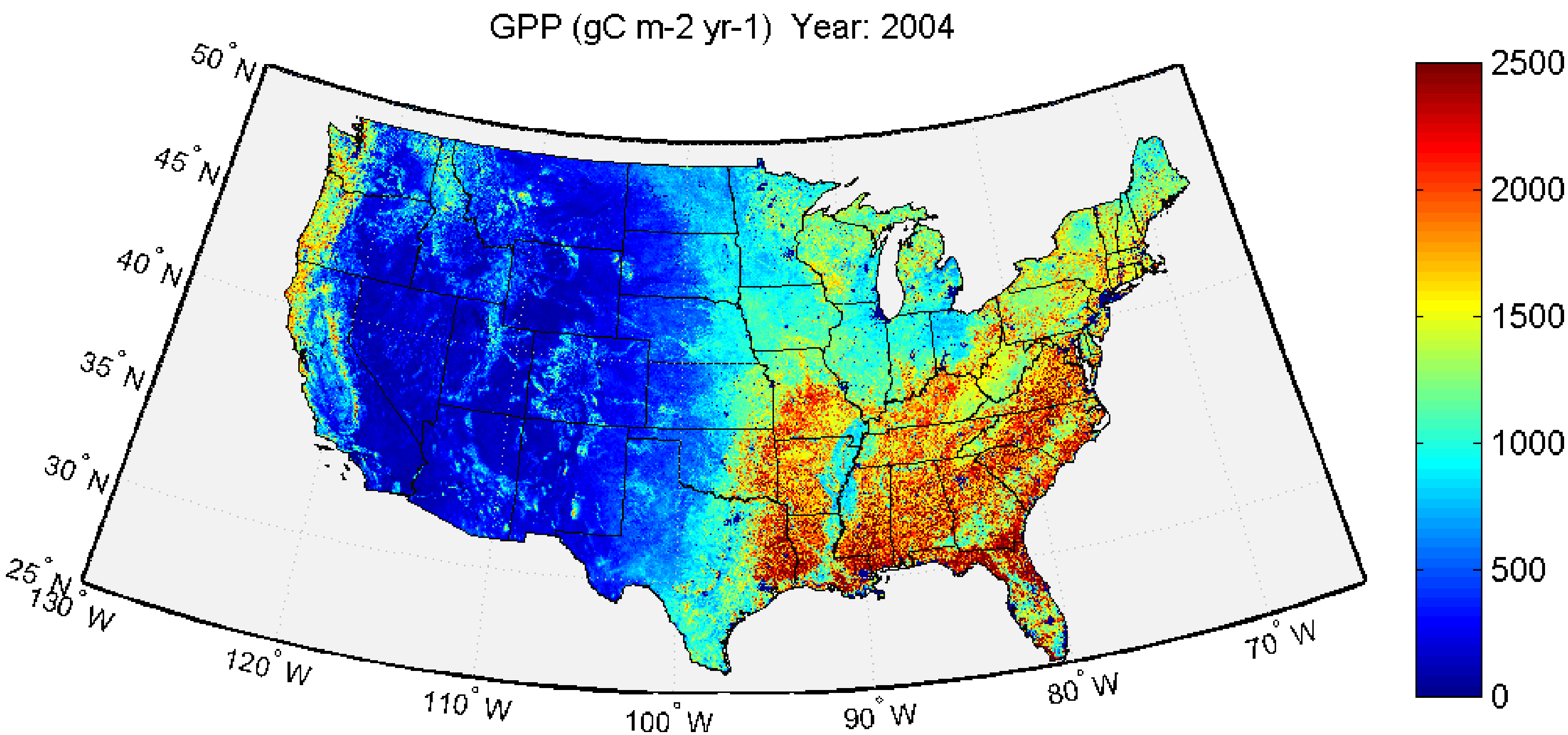
$\langle ET \rangle = 503$ mm/y $== 7.2\ 10^{13}\ m^3/y$

An Independent, Bottom-Up Alternative to Residuals
based on the Global Water Balance, ET = Precipitation – Runoff



US-MOz: Daily Rg
Daily Rg (MJ m-2)
DOY
US-MOz: Daily LE
Daily LE (MJ m-2)
Tower
MODIS
DOY
US-MOz: Daily GPP
Daily GPP (gC m-2)
DOY
Solar radiation
MODIS
Tower
y=1.16x+-1.61
R2=0.98
RMSE=2.02 (12.78%)
bias=-0.95 (-5.99%)
Latent heat flux
MODIS
Tower
y=1.28x+-0.81
R2=0.88
RMSE=1.82 (41.95%)
bias=-0.41 (-9.39%)
GPP
MODIS
Tower
y=1.14x+-0.60
R2=0.84
RMSE=1.80 (42.15%)
bias=0.01 (0.13%)

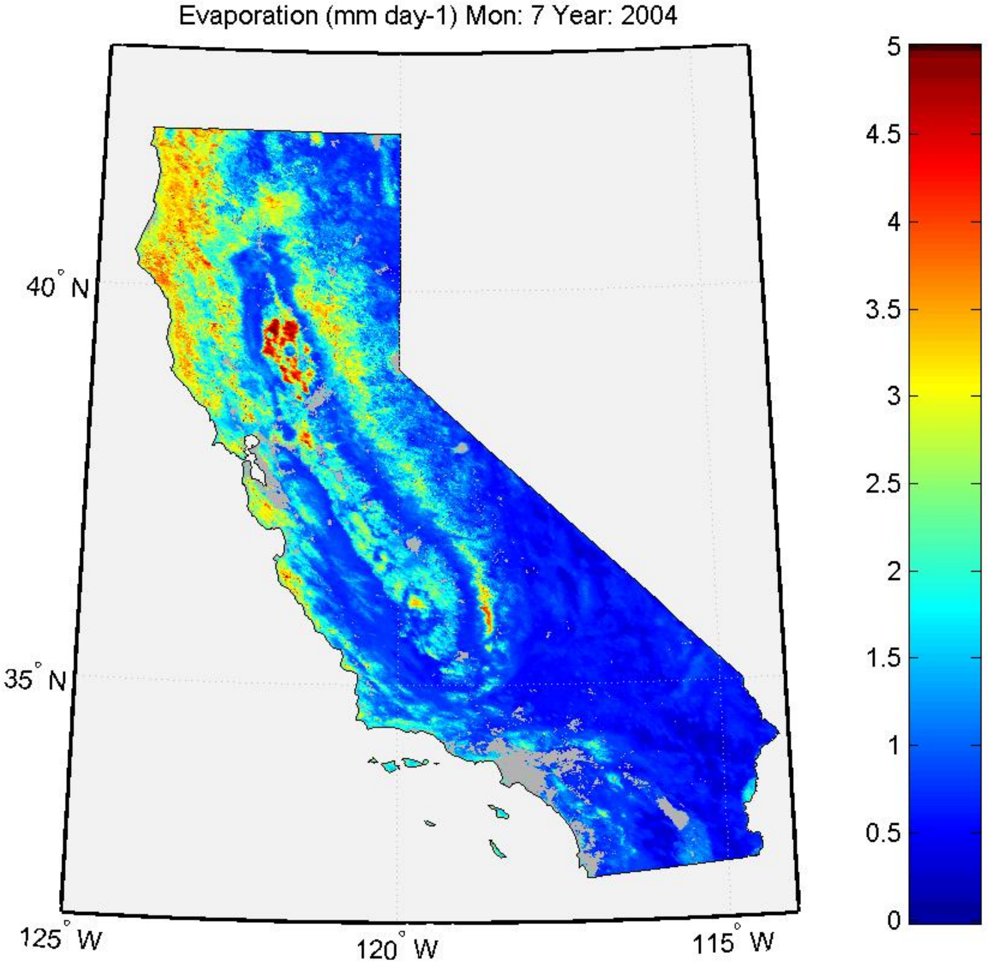
Gross Photosynthesis, GPP, Across the US



Lessons for Biofuel Production

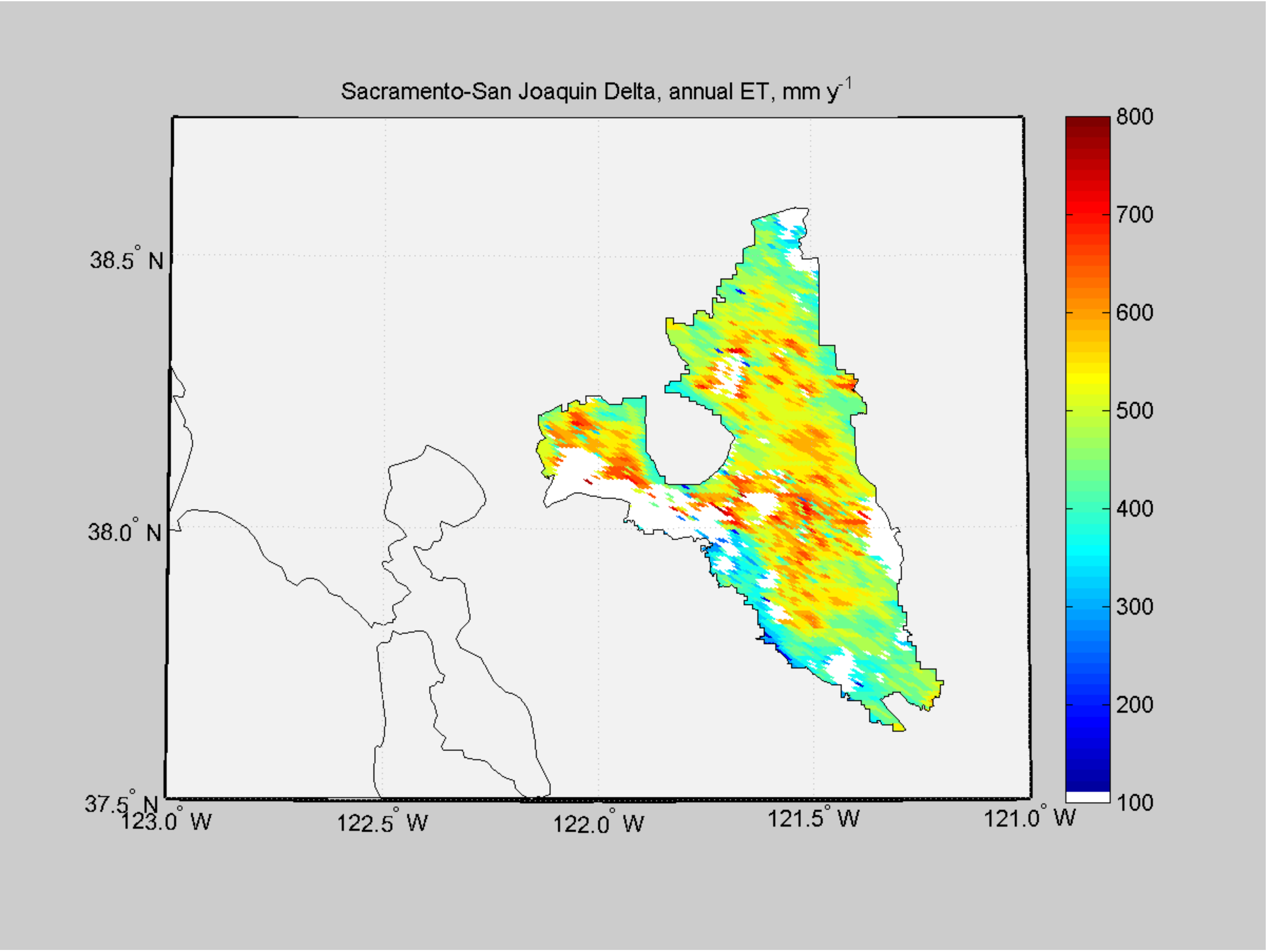
Less GPP in the Corn Belt, than the Adjacent Temperate Forests

California Evaporation at Peak of Summer Evaporative Demand

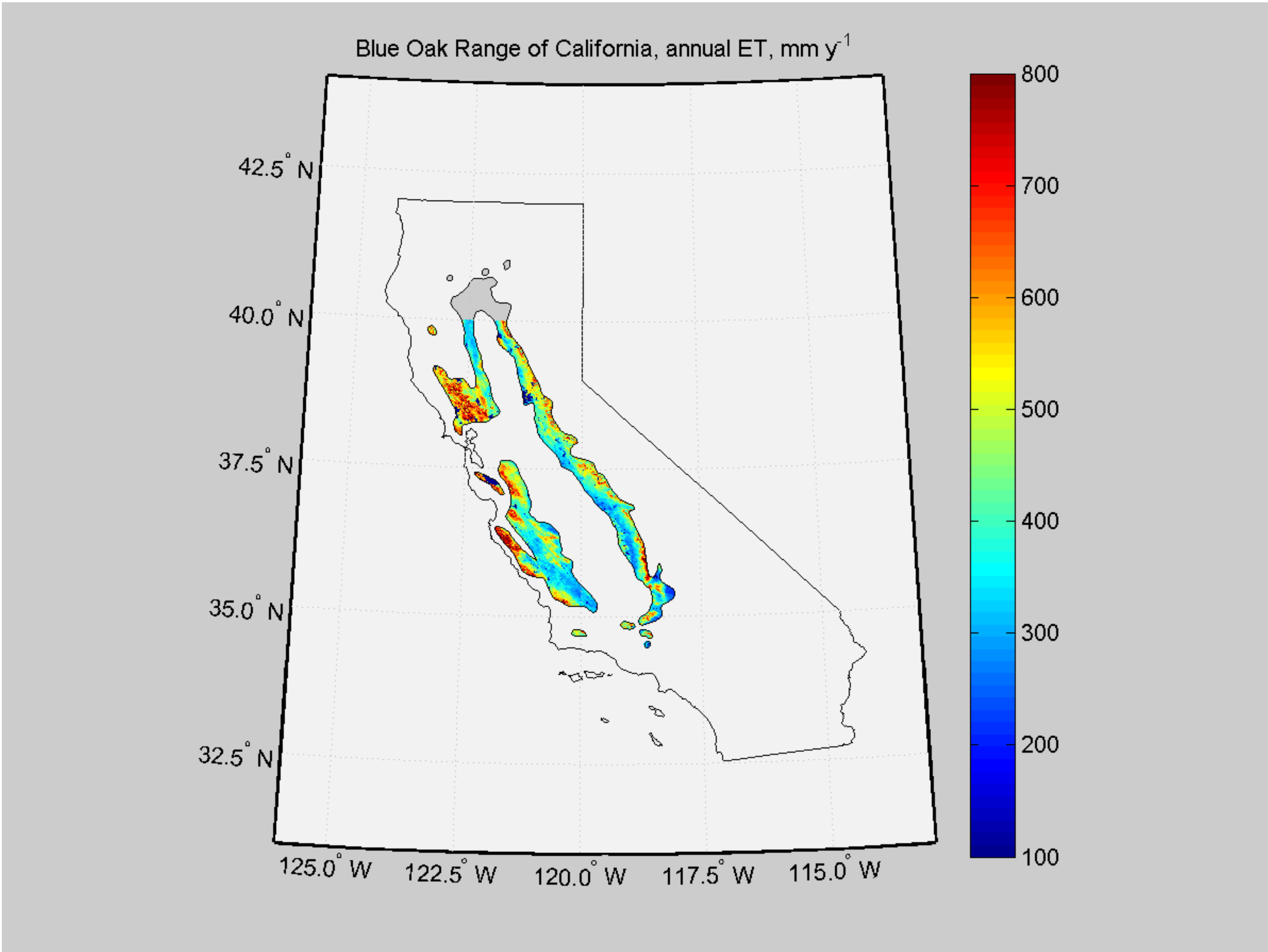


Shows ET 'Hot-Spot' in the Rice Growing Region of the Sacramento Valley

Water Management Issues: How Much Water is Lost from the Delta?



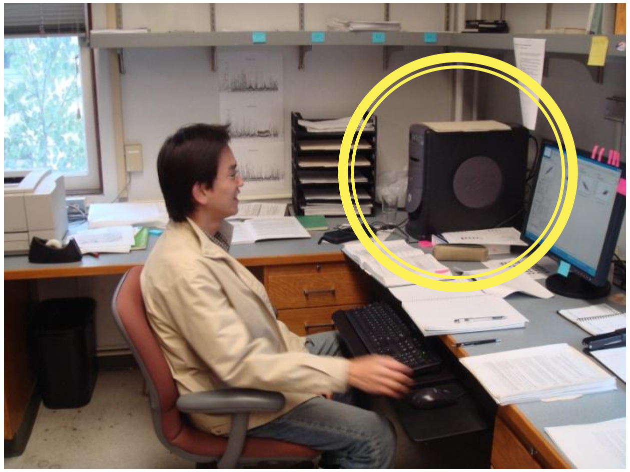
Regional Questions: How Water is Used by Deciduous vs Evergreen Oak Woodlands?



Conclusion

- Advances in Theory, Data Availability, Data Sharing and Computational Systems Enable us to Produce the Next-Generation of Globally-Integrated Products on the 'Breathing of the Earth'
- Data-Mining these Products has Much Potential for Regional and Locale Decision making on Environmental Management
- We're looking for help from additional collaborators with regional or other science knowledge

Youngryel was lonely with 1 PC



The Cavalry

It's the beginning of a great adventure