Improving Global Optimization of Hydrologic Models

Terri S. Hogue
Hydrology and Water Resources Group
Civil and Environmental Engineering
UNIVERSITY OF CALIFORNIA
LOS ANGELES

Research Team and Collaborators!





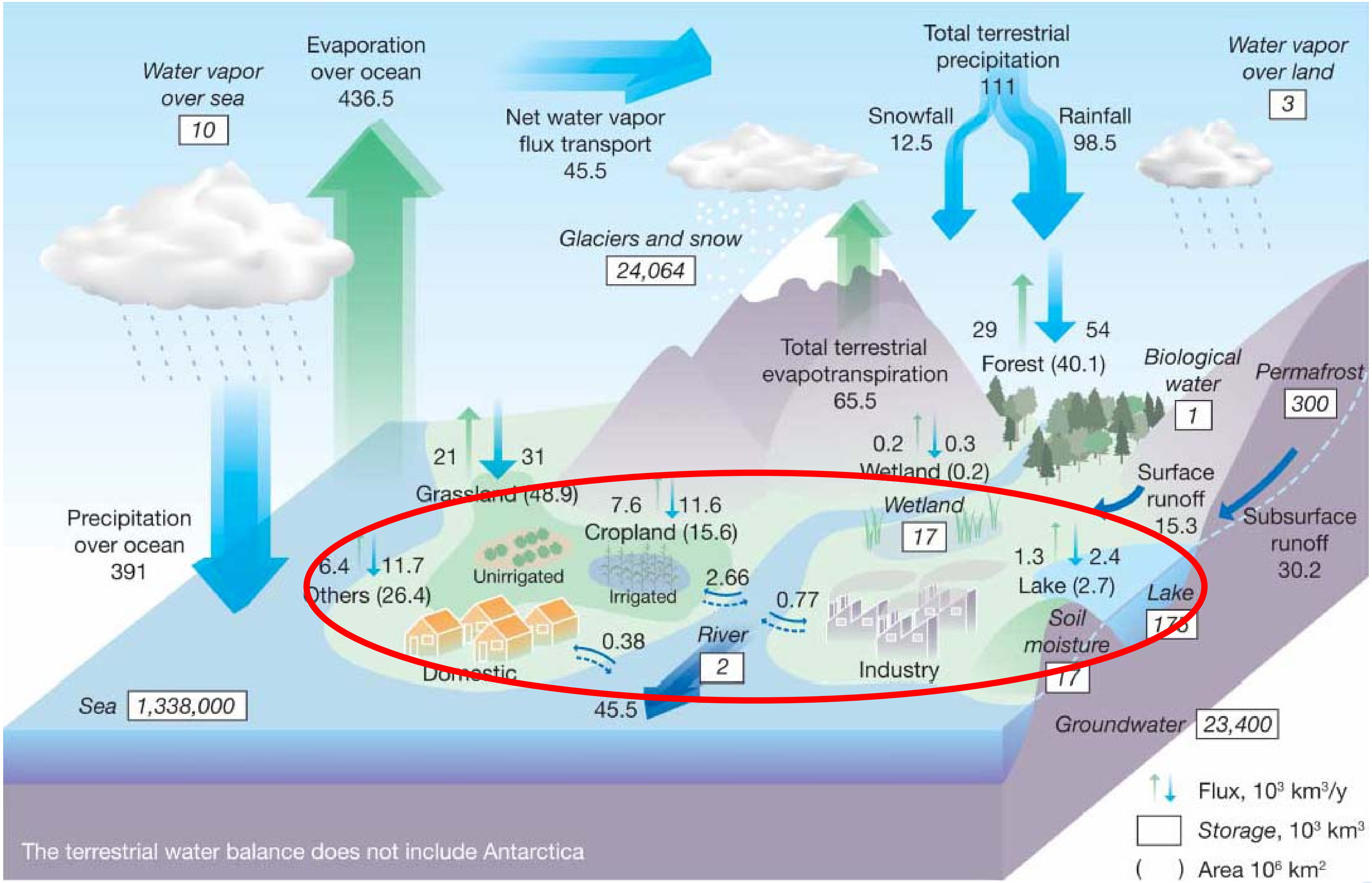




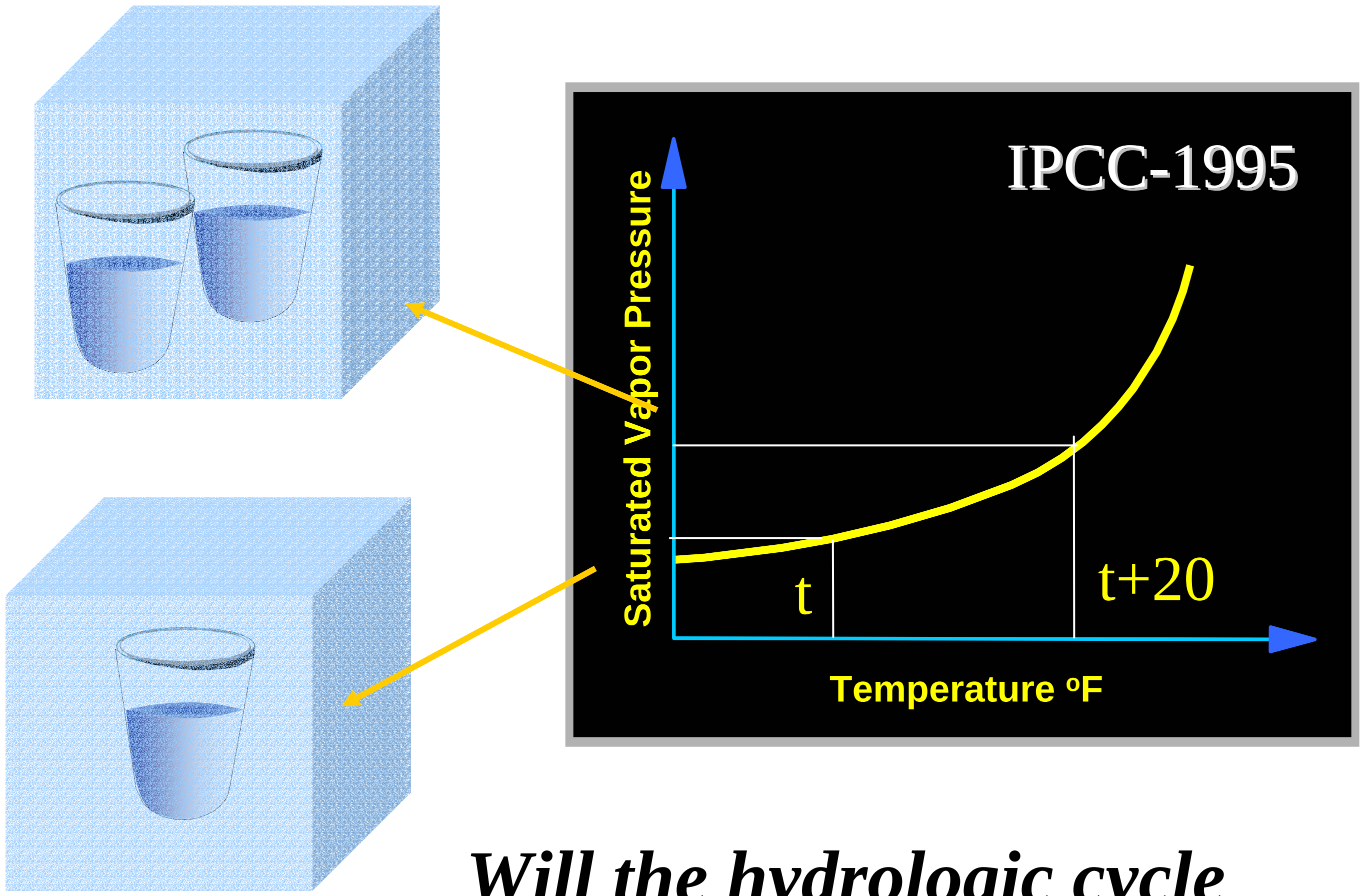

Dr. Jenny Jay, UCLA
Dr. Tom Meixner, U of Arizona
Dr. Laura Rademacher, U of Pacific

Global Hydrologic Fluxes

(Oki and Kanae, 2006)



Future Uncertainty…



Will the hydrologic cycle intensify as a result of climate change ?

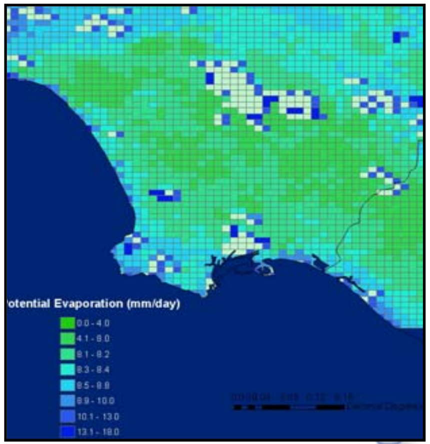
Research Themes

Changing Hydrologic Cycle

- o Wildfires
  - Geochemical - physical response
  - Modeling post-fire runoff
  - Contaminant transport

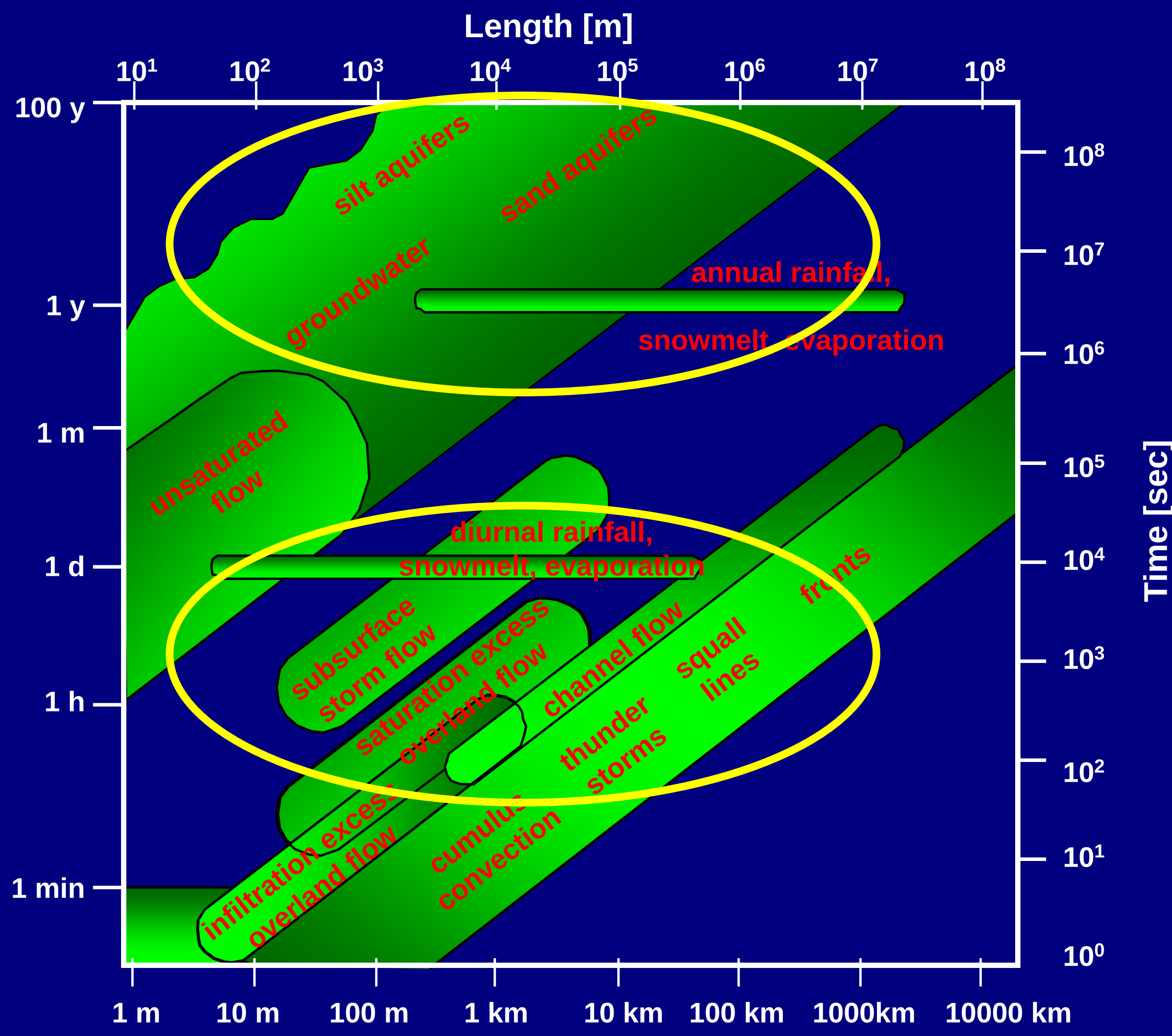
- o Urbanization
  - Energy and water fluxes in urbanized regions
  - Partitioning of flow regimes in developing watersheds
  - Atmospheric deposition and impact on "fringe" watersheds

- o Climate Impacts on Watersheds
  - Evaluating quantity and quality changes
  - Ecosystem response



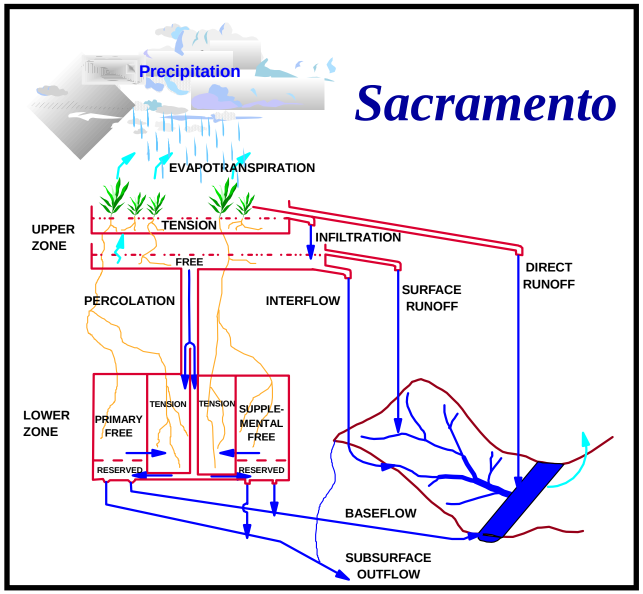
- o Improving tools for analysis/prediction
  - Remote sensing of land surface properties
    - MODIS ET model (stand-alone)
  - Model optimization and predictions
    land surface and hydrologic (operational)

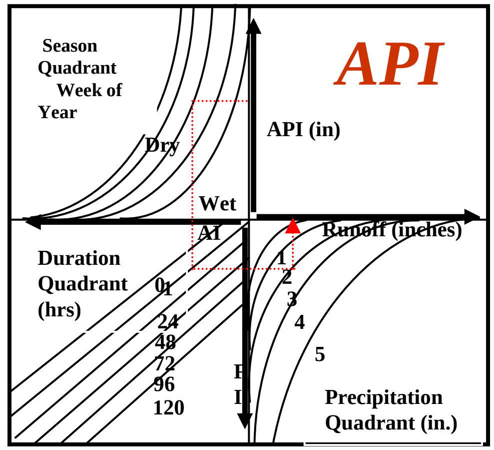
Spatial and Temporal Process Scales

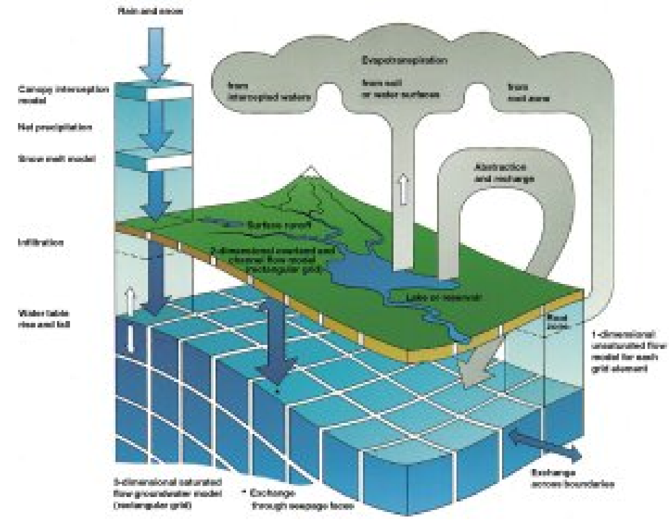


Source Bloschl, 1996

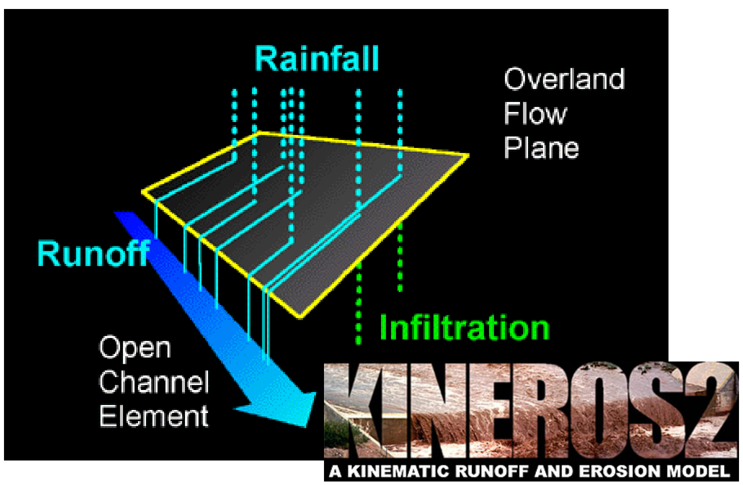
What Model? What Level of Complexity?

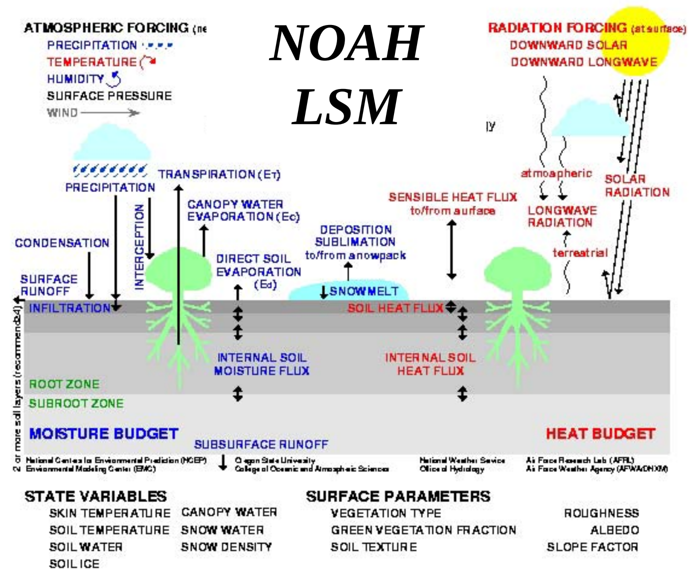




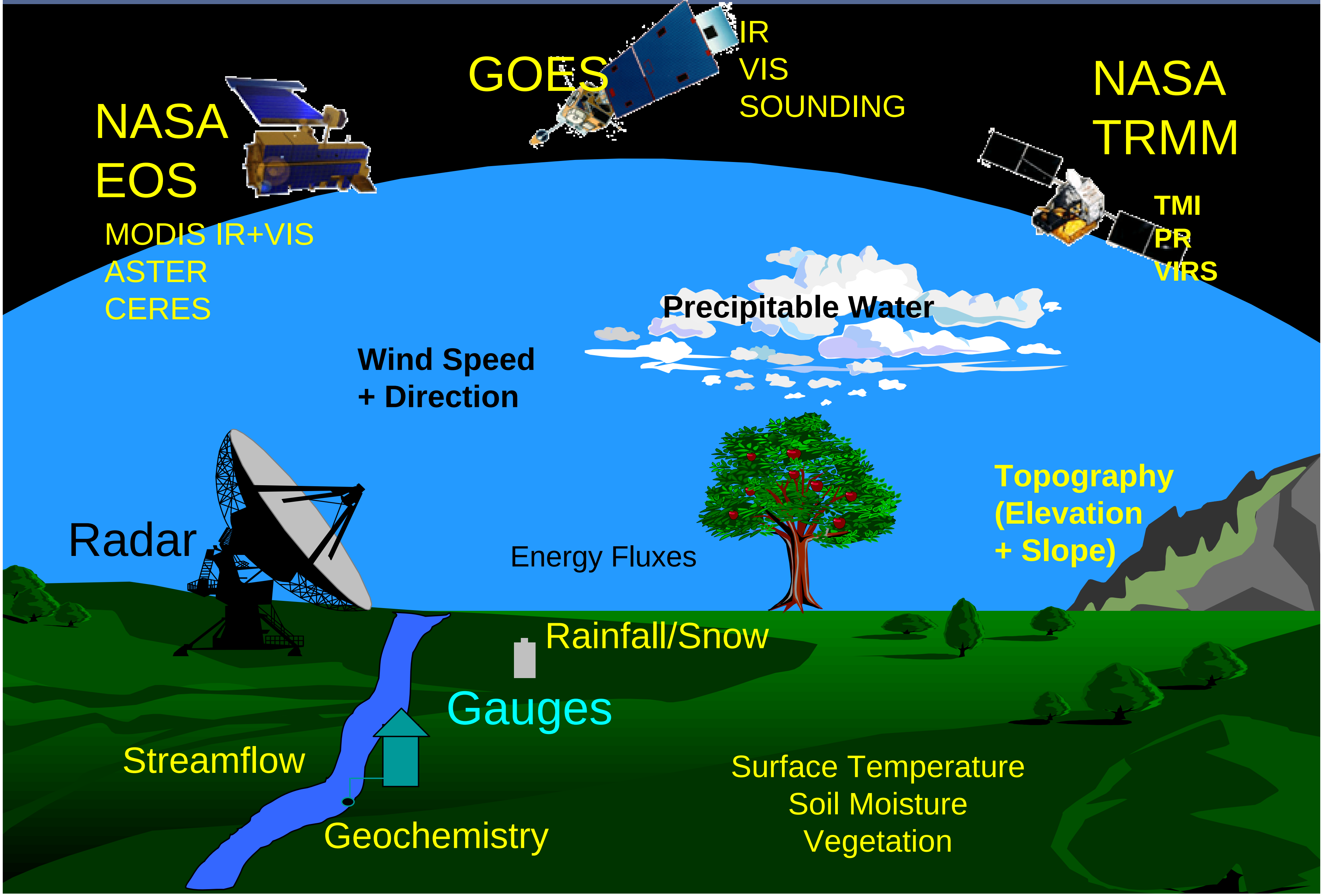


MIKE-SHE



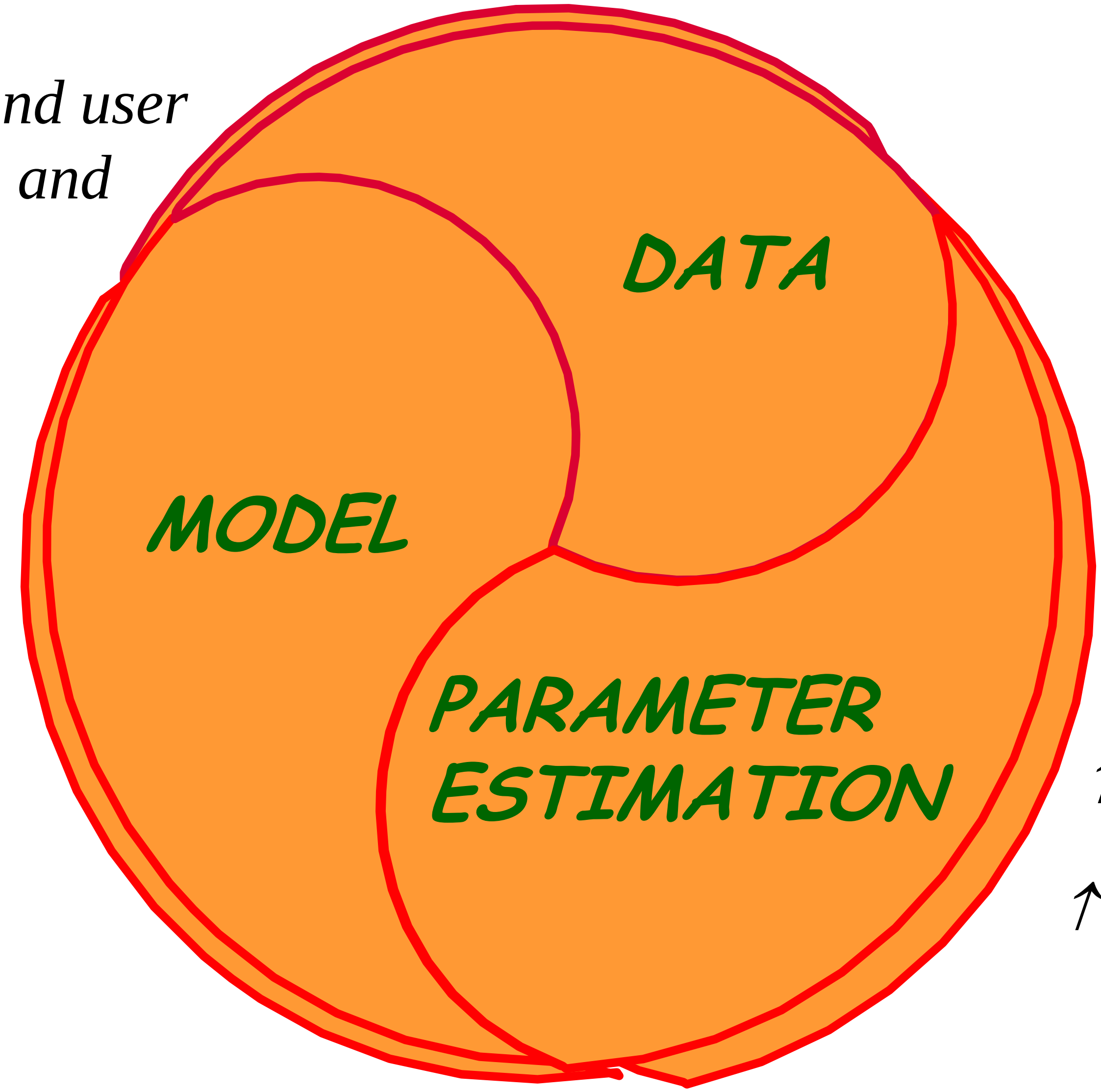


Data Sources for Hydrologic Modeling



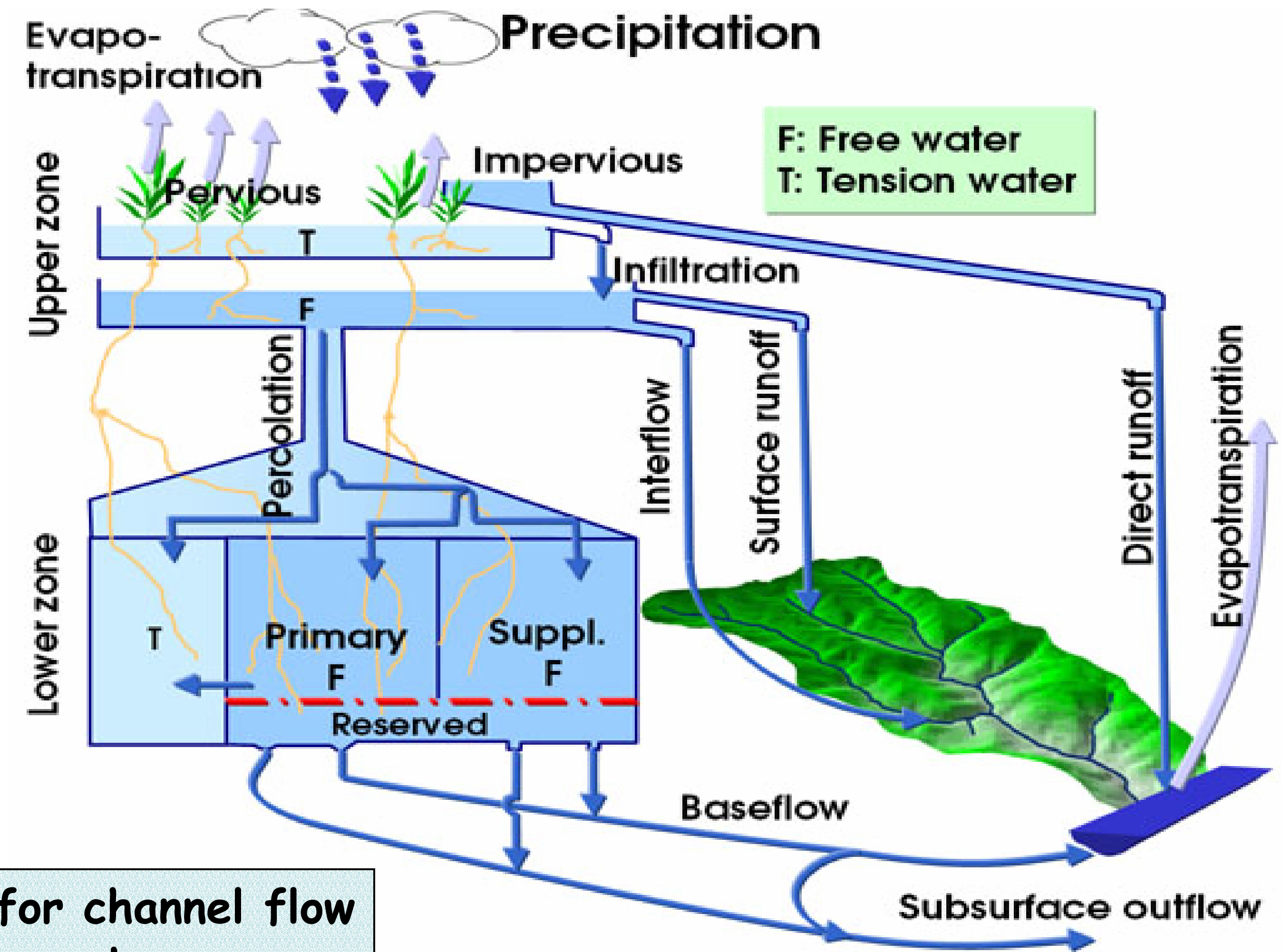
Modeling Components

Keeping in mind user sophistication and requirements



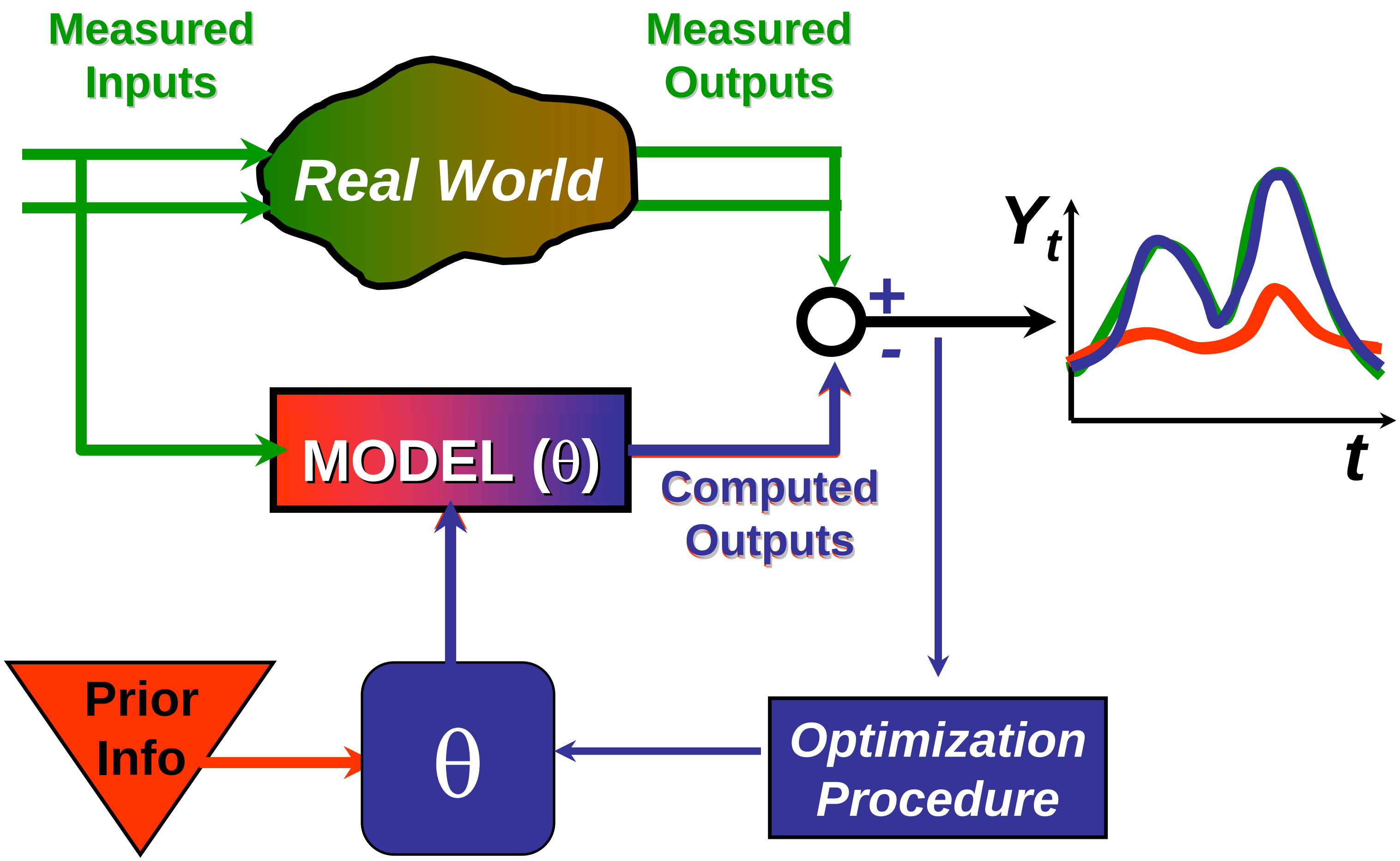
↑ Complexity
=
↑ Parameters

NWS Operational Model: SAC-SMA



5 pathways for channel flow
16 model parameters
5 model states

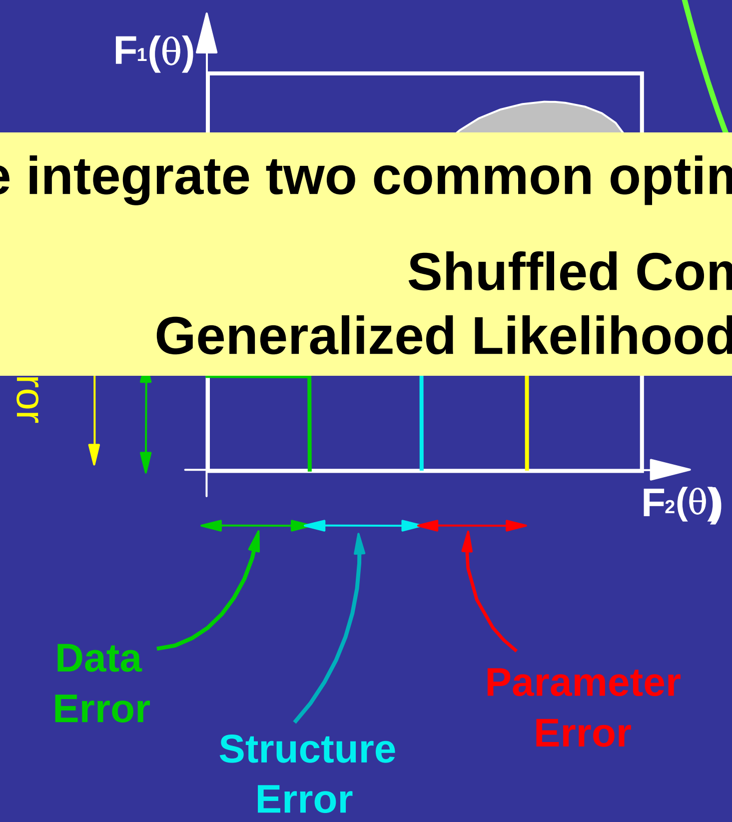
The Concept of Model Calibration



How much improvement is possible ?

Total Error = Data Error + Model Structural Error + Parameter Specification Error

20 – 50 %



We integrate two common optimization routines in hydrologic sciences:

Shuffled Complex Evolution (SCE)
Generalized Likelihood Uncertainty Estimation (GLUE)

~ 30 %

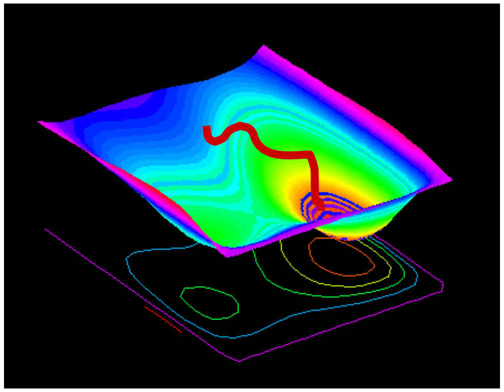
Estimated using Neural Networks

Shuffled Complex Evolution (SCE)

Duan et al., 1992; 1993

Global Optimization Scheme

Combines SIMPLEX search, genetic evolution, shuffling (sharing) of complex information



Requires:

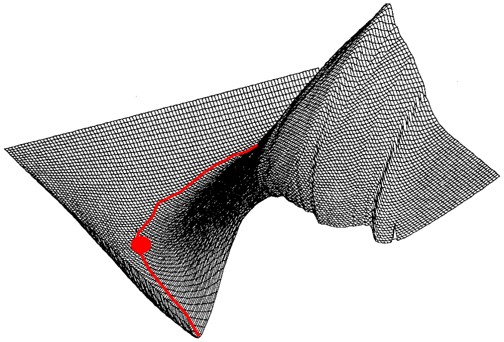
Historical observations

Objective function choice

Parameter constraints

PROCESS:

1. Define feasible parameter space
   a. Create m complexes with n points in each complex
   b. Randomly sample parameter space
2. Evolve each complex (using simplex method)
3. Shuffle complexes and re-search sample space
4. Determine if convergence criteria is satisfied
   "objective function threshold" – NSE, DRMS, etc
5. Loop through procedure until criteria met
6. Results in selection of single parameter set



<u>G</u>eneralized <u>L</u>ikelihood <u>U</u>ncertainty <u>E</u>stimation (GLUE)

Beven, 1993

Global Method to Estimate Behavioral Parameter Sets

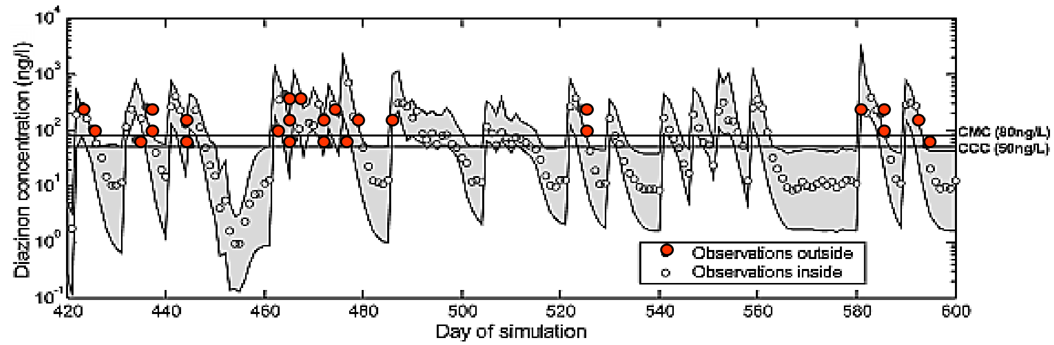
Random parameter sampling, selection of thresholds, produces "prediction bounds"

Requires:

Historical observations

Selection of likelihood fx.

Parameter constraints



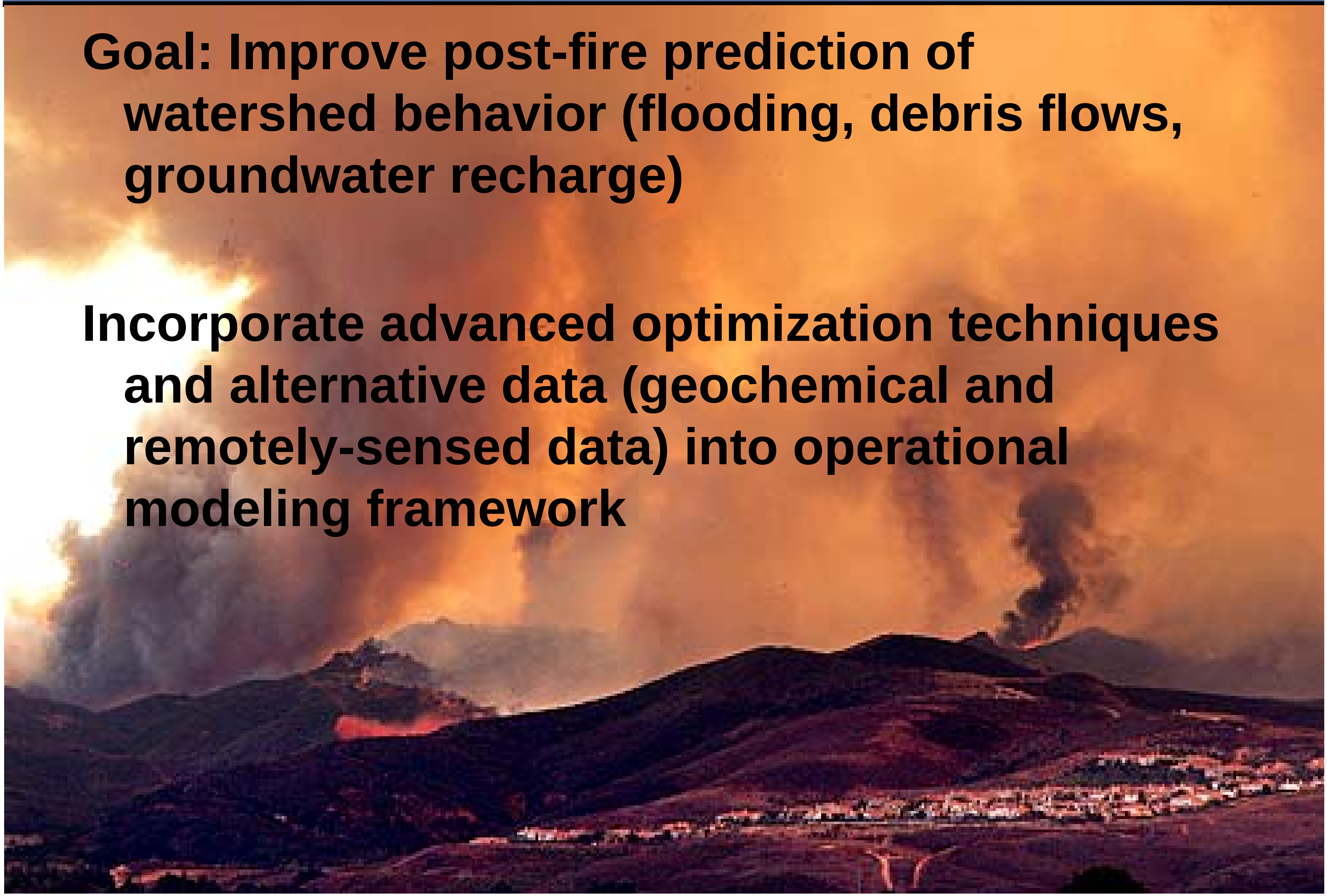
PROCESS:

1. **Define Feasible Parameter Space**
2. **Monte Carlo Simulations (Random, Latin Hypercube, etc.)**
   a. **Randomly sample parameter space (10,000x)**
   b. **Run model simulations with all par. sets**
   c. **Calculate Likelihood (theoretical or "obj. fx")**
3. **Determine Behavioral & Non-behavioral Sets**
   a. **Reject Non-behavioral via set "threshold"**
   i.e. NSE> 0.3 cfs, DRMS < 0.5 mg/L
4. **Cumulative Distribution Function**
   a. **Select "prediction uncertainty"**
   5 – 95 of cdf % (upper and lower bounds)

$$NSE = 1 - \frac{\sum_{t=1}^{N} (pred - obs)^2}{\sum_{t=1}^{N} \left(obs - \overline{obs}\right)^2}$$

Goal: Improve post-fire prediction of watershed behavior (flooding, debris flows, groundwater recharge)

Incorporate advanced optimization techniques and alternative data (geochemical and remotely-sensed data) into operational modeling framework



Improving Post-fire Flow Predictions

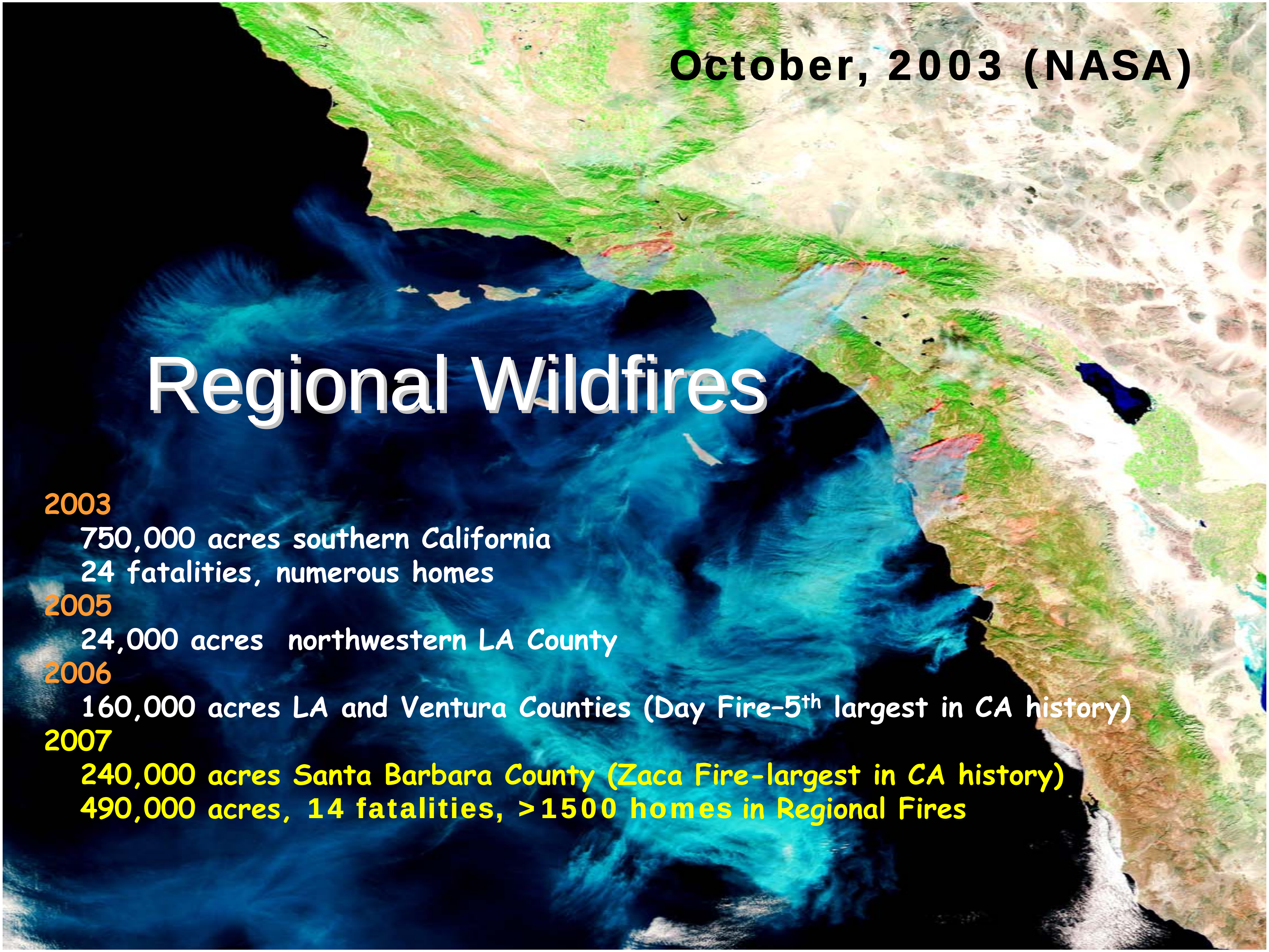
Limited work on "predicting" post-fire runoff at watershed scale (numerous plot-scale studies)

National Weather Service responsible for forecasting in burned watersheds

Can we improve performance / predictability of models?
- NWS – conceptual rainfall-runoff model
- USACE – HEC-HMS watershed model (variety modules)
- EPA – HSPF conceptual watershed model
- DHI – MIKE-SHE distributed physical model

TOOLS??

Optimization Algorithms, Sensitivity Analysis, Data Assimilation, Remote Sensing Data, Field Observations



October, 2003 (NASA)

Regional Wildfires

2003
- 750,000 acres southern California
- 24 fatalities, numerous homes

2005
- 24,000 acres northwestern LA County

2006
- 160,000 acres LA and Ventura Counties (Day Fire–5th largest in CA history)

2007
- 240,000 acres Santa Barbara County (Zaca Fire-largest in CA history)
- 490,000 acres, 14 fatalities, >1500 homes in Regional Fires

Impacts on Watersheds

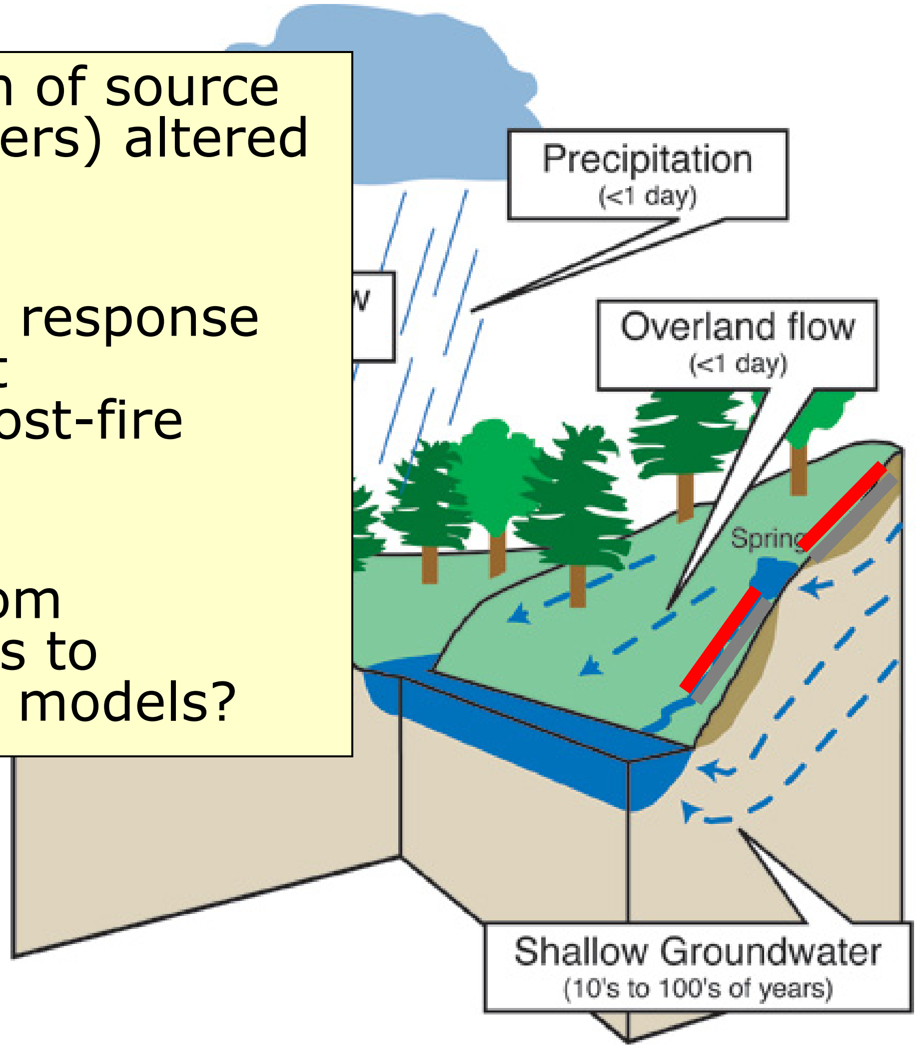
- Loss of Vegetation
- Decreased evapotranspiration
- Increased net solar radiation
- Hydrophobic layer formation
- Decreased permeability (ash)
- Altered flowpaths
- Decreased water quality
- Increased erosion
- Transport of metals and nutrients

Post-fire Hydrologic Behavior

How is the distribution of source waters (end-members) altered after fire?

Does the geochemical response support our current understanding of post-fire behavior?

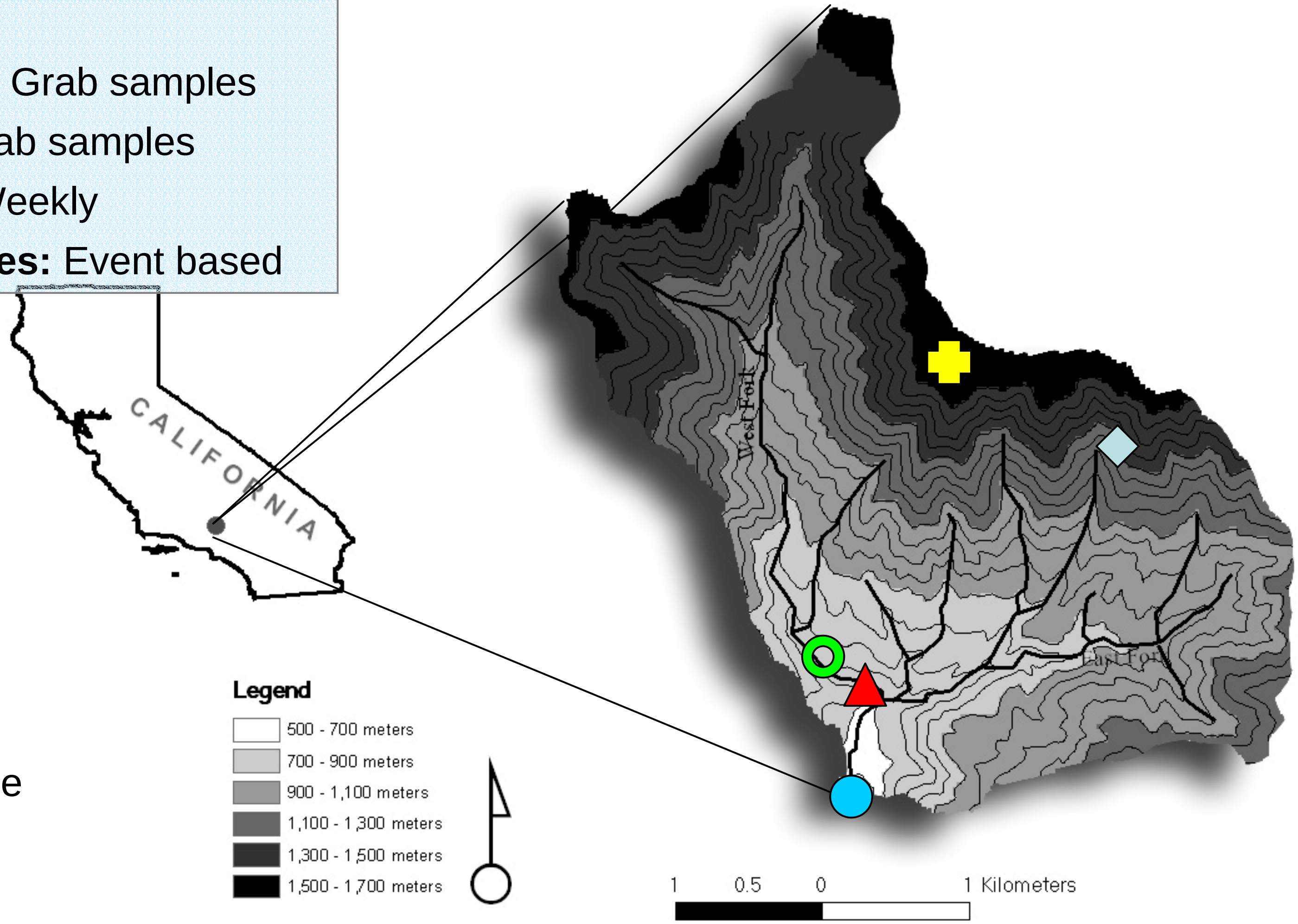
Can we use insight from geochemical models to improve hydrologic models?



2003 Fires – San Bernardino Mountains

Size: 14 sq. km
Land cover: 66% chaparral
Burned area: 97%

Lysimeter samples: Grab samples
Spring samples: Grab samples
Stream samples: Weekly
Precipitation samples: Event based



- Precipitation gage
- Lysimeters
- Spring samples
- Stream samples
- USGS stream gage

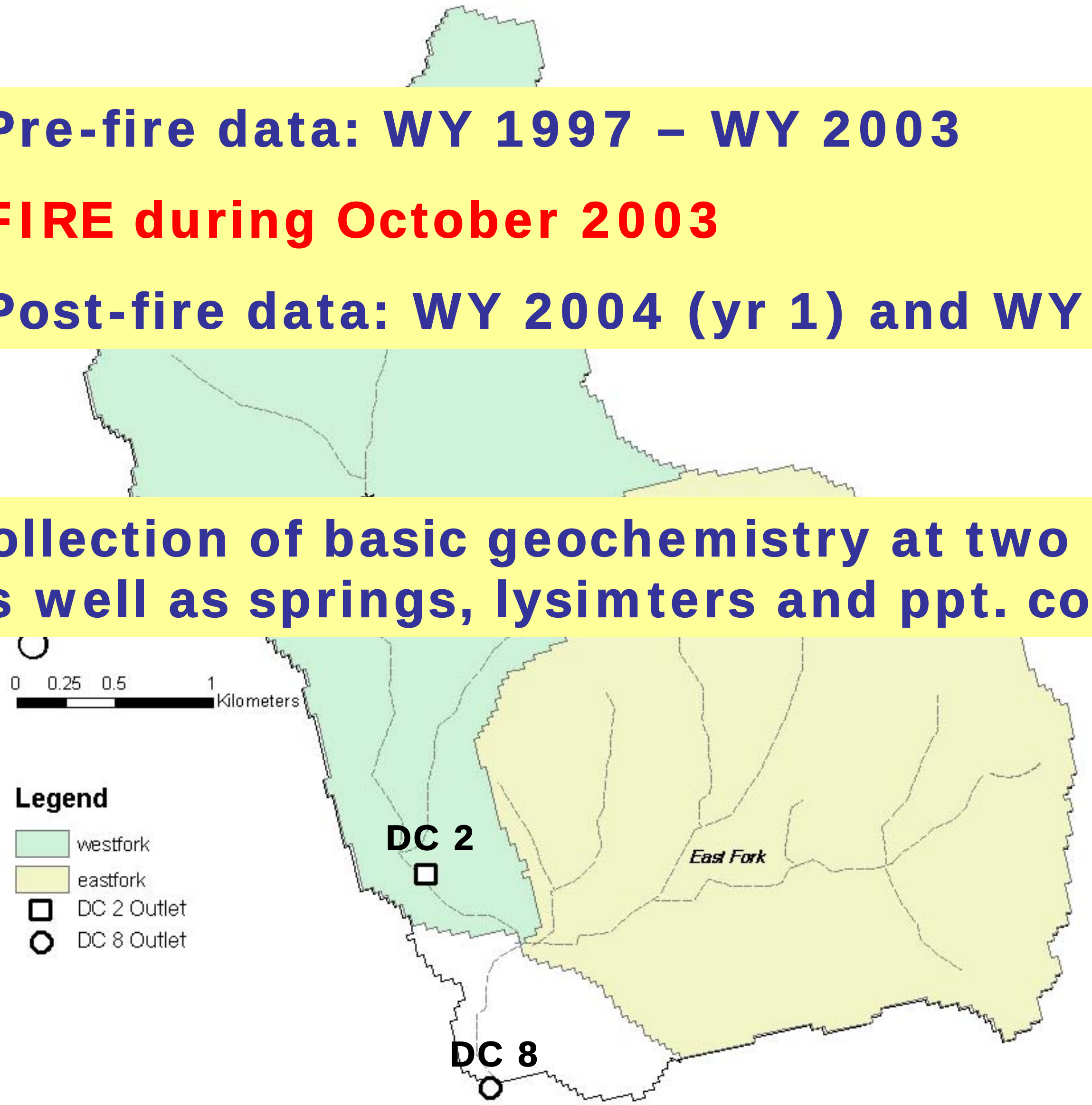
Devil’s Canyon – DC2 (West Fork), DC8 (outlet)

Pre-fire data: WY 1997 – WY 2003

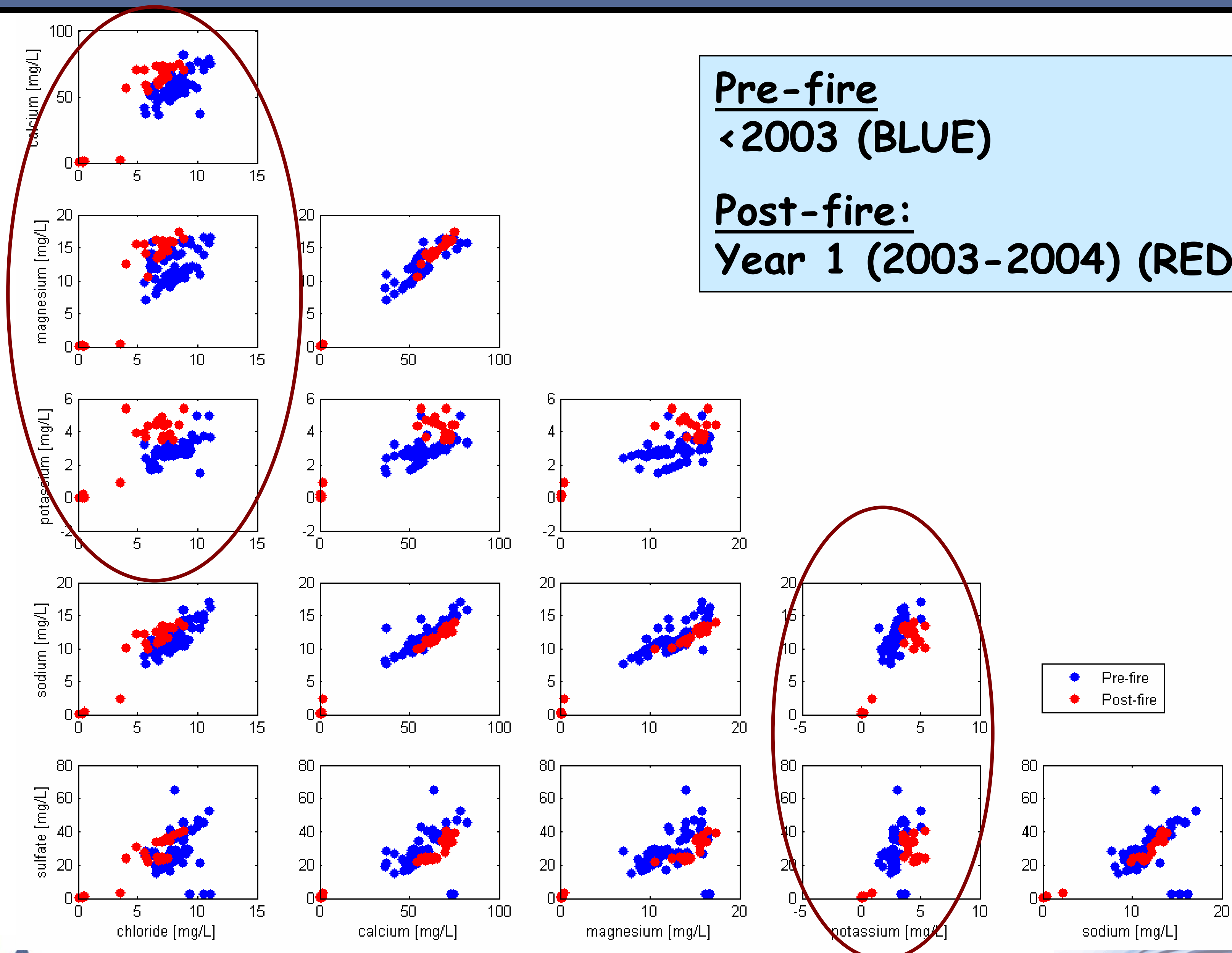
FIRE during October 2003

Post-fire data: WY 2004 (yr 1) and WY 2007 (yr 4)

Collection of basic geochemistry at two primary sites, as well as springs, lysimters and ppt. collectors



DC2: Pre and Post-fire Geochemistry



Pre-fire
<2003 (BLUE)

Post-fire:
Year 1 (2003-2004) (RED)

Geochemical Analysis

End Member Mixing Analysis (EMMA):

- **Stream water is a mixture of water from various sources (surface, lateral, baseflow, riparian, etc.)**
- **Each source has a unique geochemical signature (i.e. *end members*)**
- **Solutes used in EMMA are conservative**
- **End-members are time-invariant (when no LC change)**

Principal Component Analysis (PCA) used to reduce dimensionality

Simplifies data sets by linear transformation

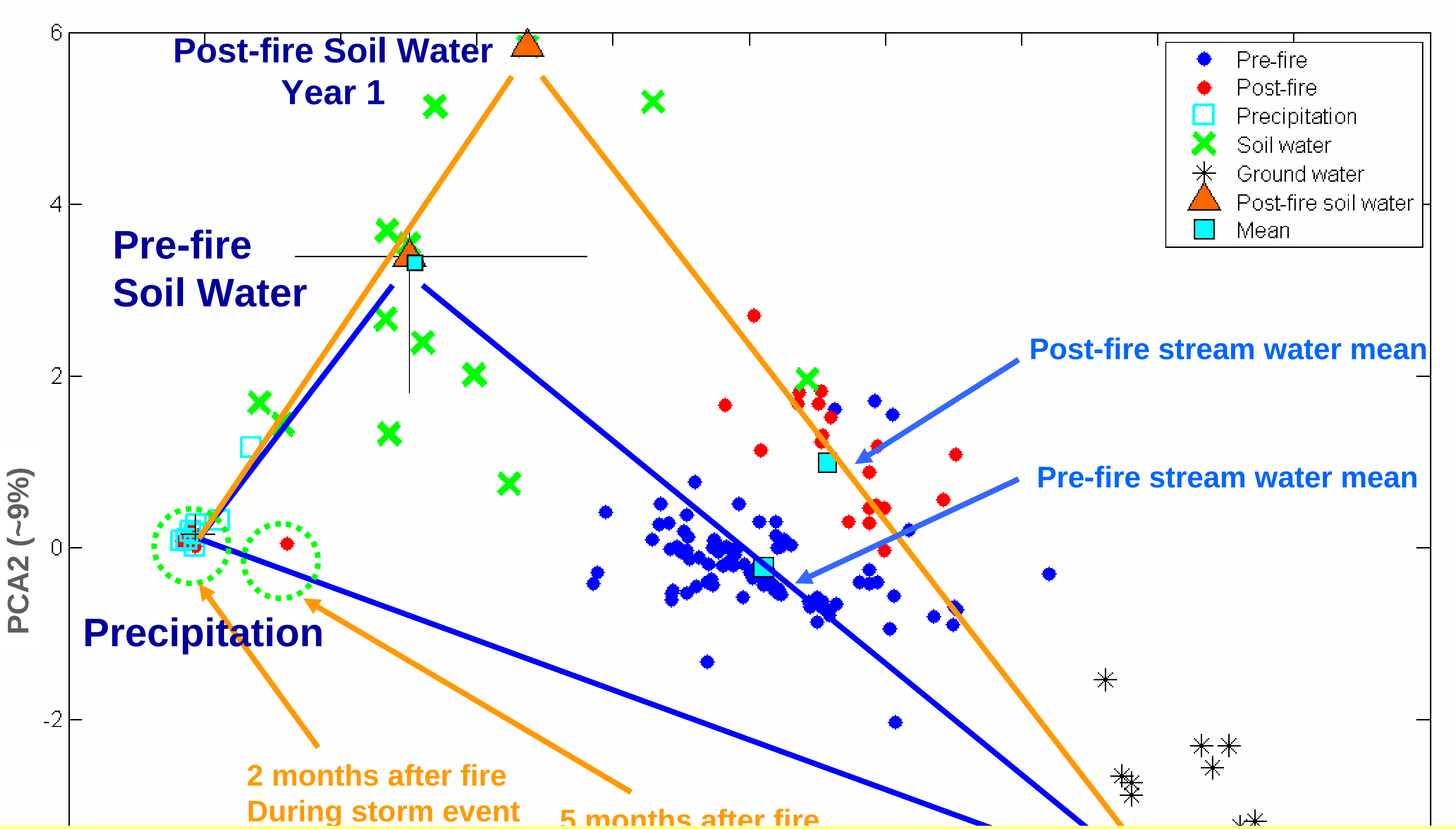
Chooses a new coordinate system based on correlation of the data

Plot streamwater and end-member values in PCA space

→ Determine relative (%) contribution of end members (baseflow, overland flow, etc.) to stream water

Jung et al., (2008) Hydrological Processes (in review)

DC2: PCA-EMMA Development



PRE-FIRE

Three observable end-members at both sites

POST-FIRE

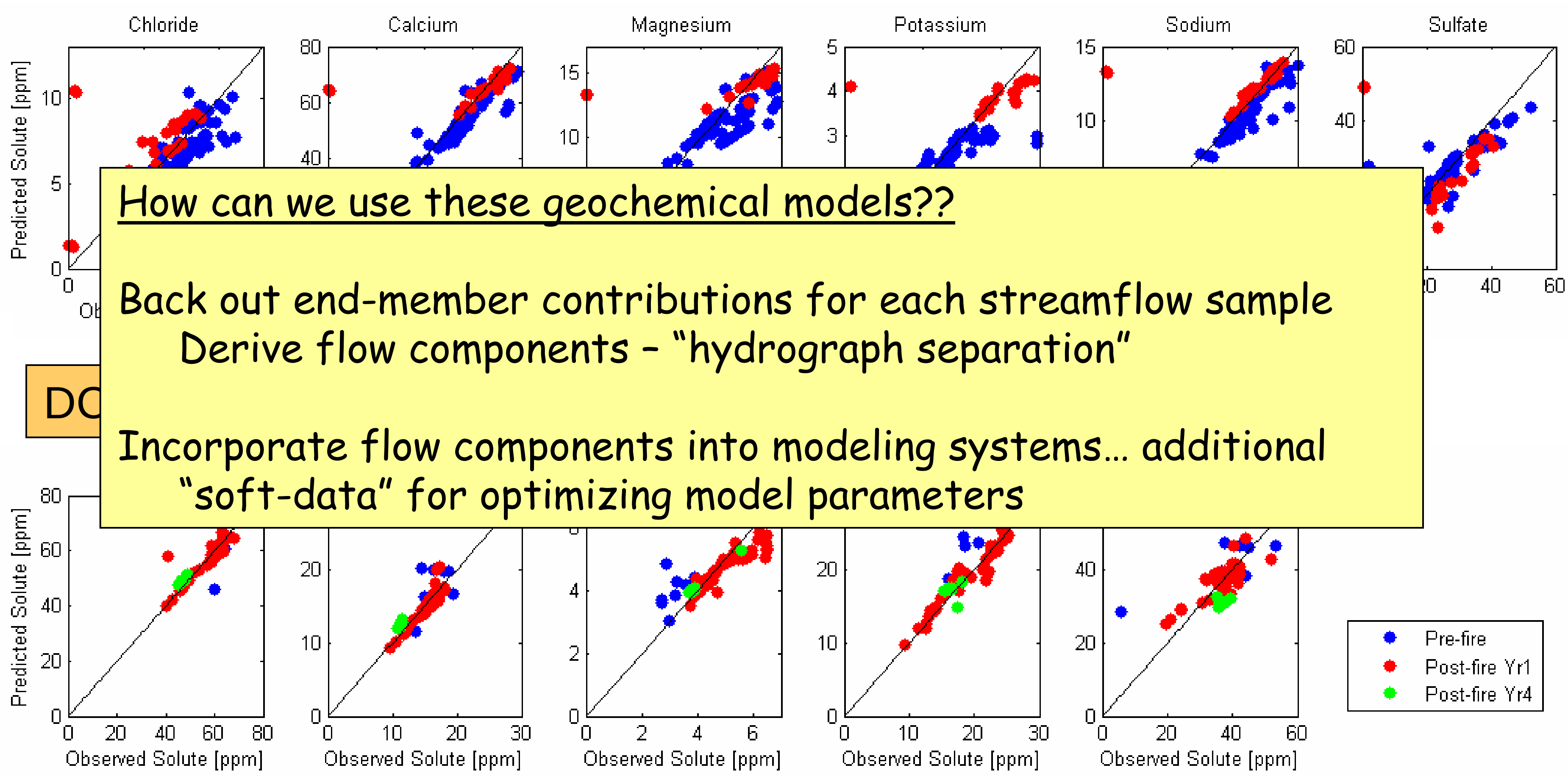
Early storm events dominated by overland flow (supports knowledge!)

Less soil water in interior point (reduced infiltration – supports!)

Less GW at outlet (GW-dominated regime) – reduced infiltration is affecting this site differently (had less soil water influx originally)

Evaluation of Developed EMMA Model

DC2 Predictions



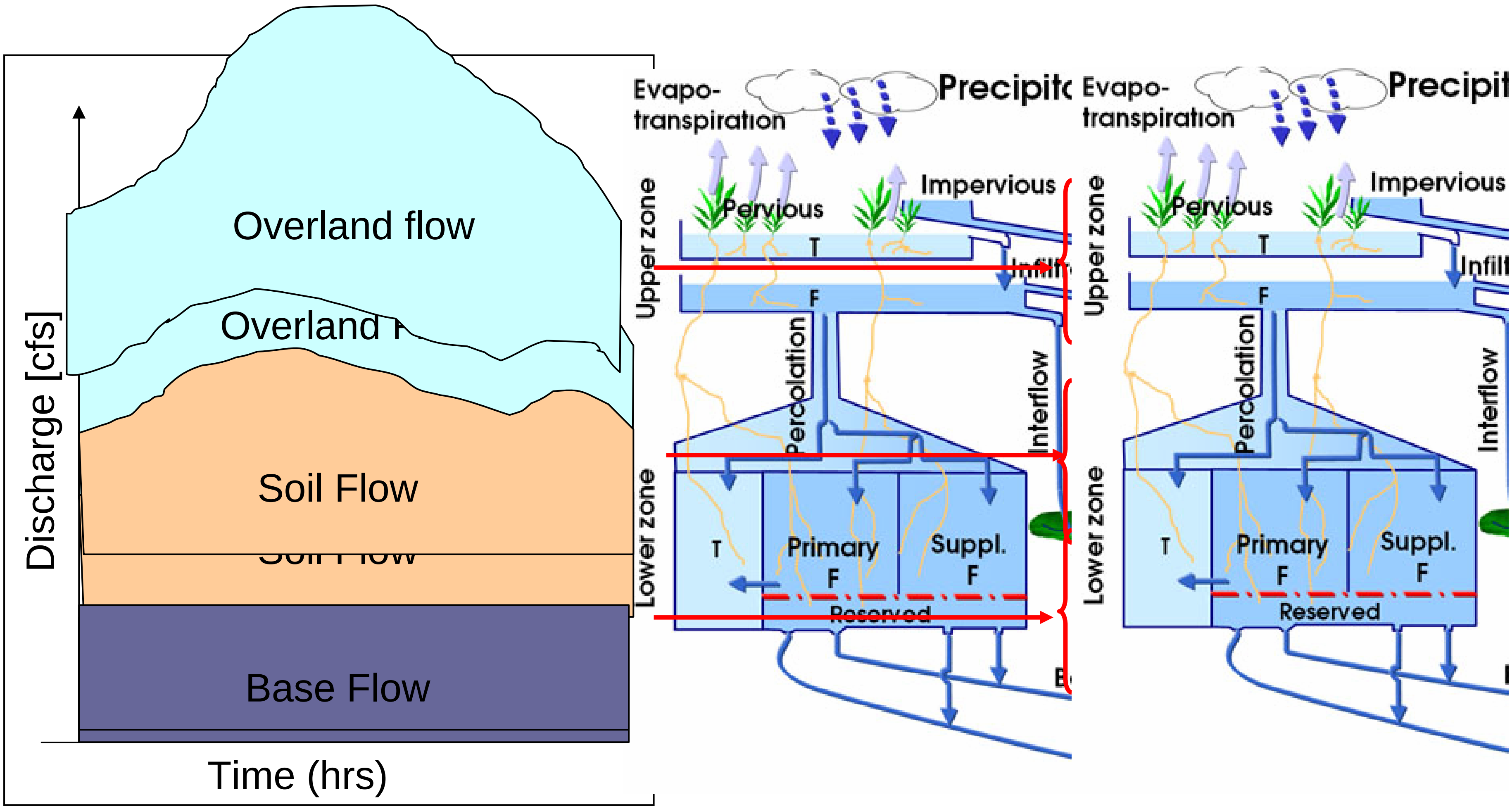
<u>How can we use these geochemical models??</u>

Back out end-member contributions for each streamflow sample
Derive flow components – "hydrograph separation"

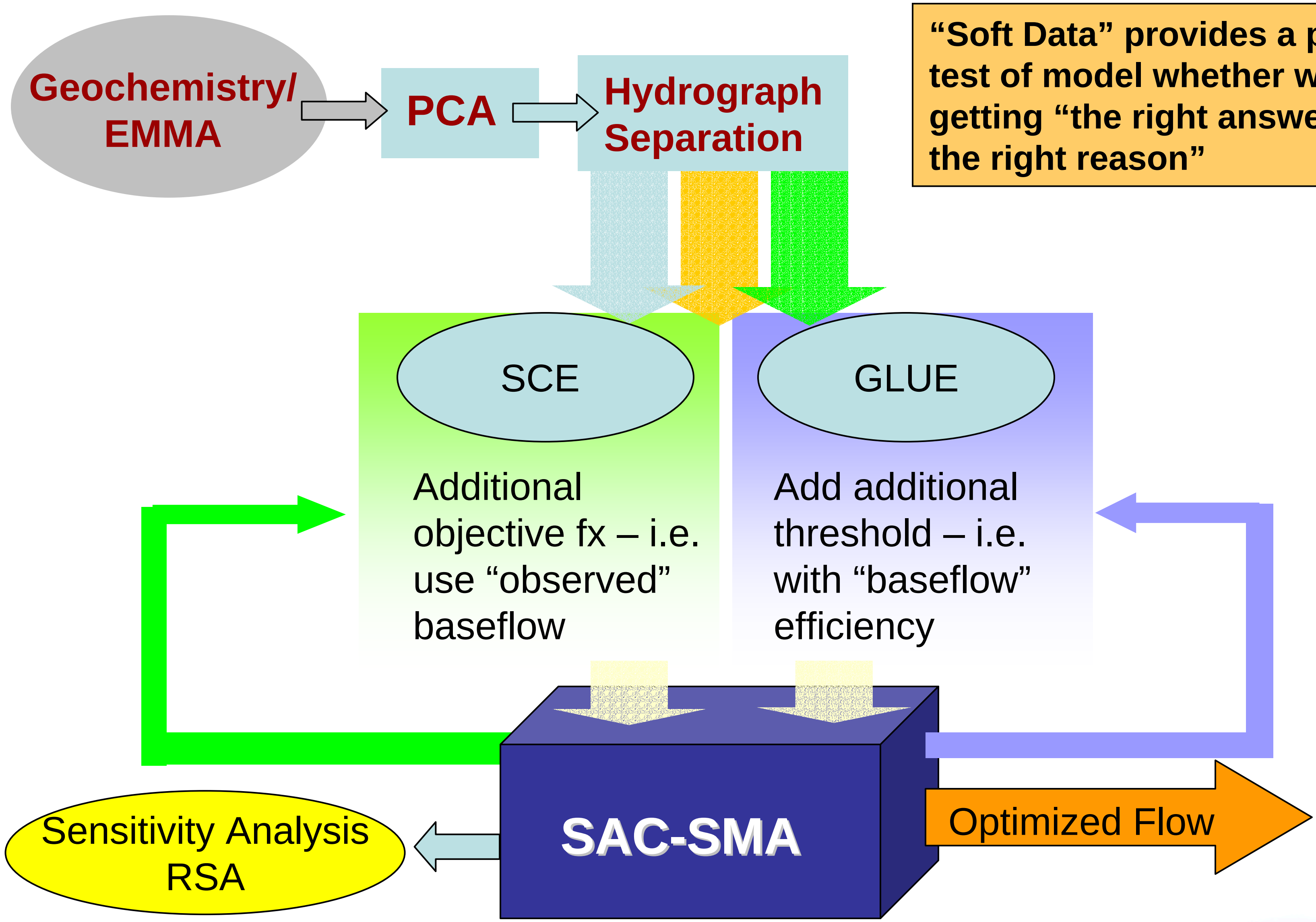
Incorporate flow components into modeling systems… additional "soft-data" for optimizing model parameters

Jung et al., (2008) Hydrological Processes (in review)

NWS Operational Model: SAC-SMA



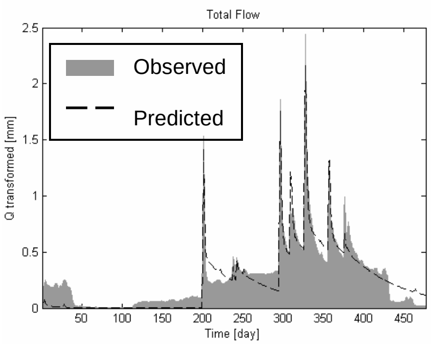
Incorporating Geochemical Data in Modeling

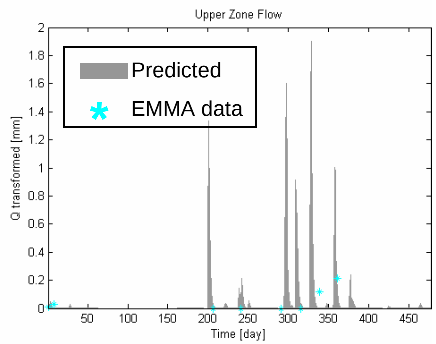


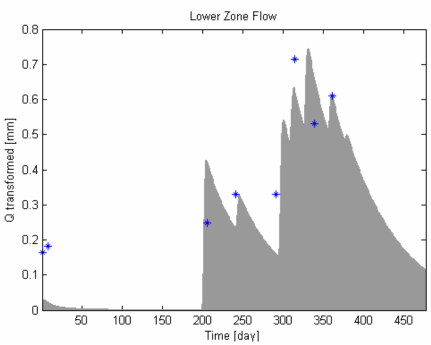
"Soft Data" provides a powerful test of model whether we are getting "the right answers for the right reason"

DC8 Pre-fire SCE

One-Step: RMSE



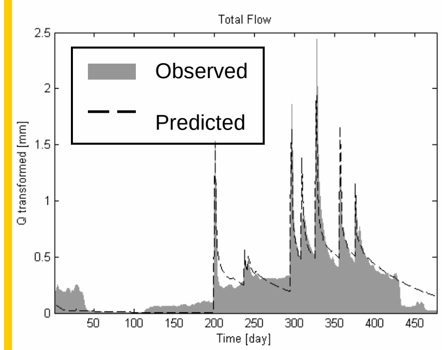


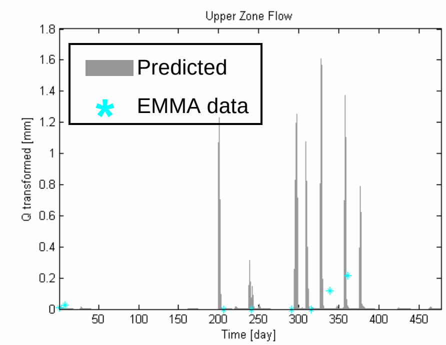


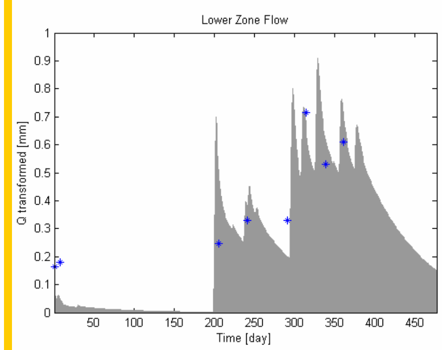
ONE STEP

	DRMS	BIAS	R2
Total	0.14	-0.10	0.90
Upper	0.06	-0.53	0.64
Lower	0.35	-0.95	0.04

Hydrograph Separation Method







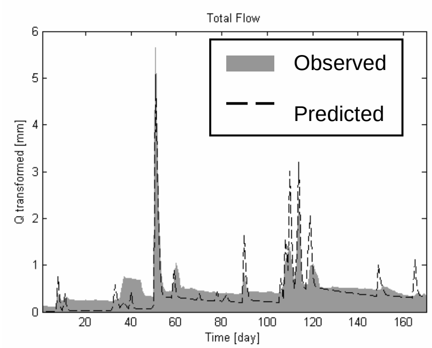
HYDROGRAPH SEPARATION

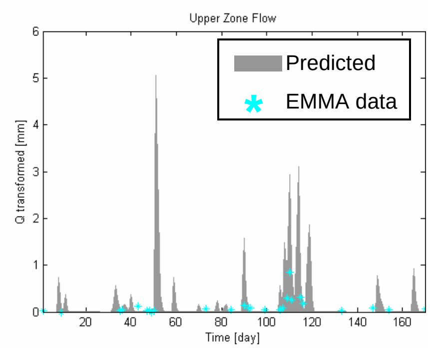
	DRMS	BIAS	R2
Total	0.18	0.01	0.84
Upper	0.08	-0.94	0.74
Lower	0.40	-0.95	0.03

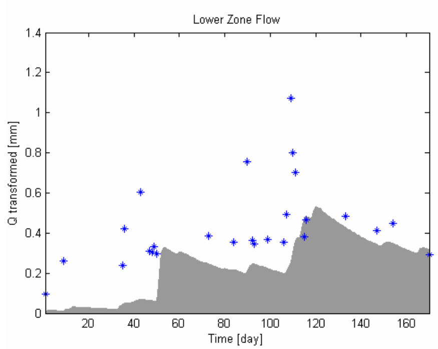
- Improved Qtotal (lower Bias) with additional data
- Slight change in baseflow performance
- Only 8 data points – but better match of flow behavior

DC8 Post-fire SCE

One-Step: RMSE



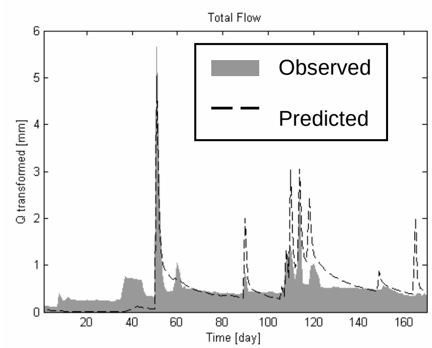


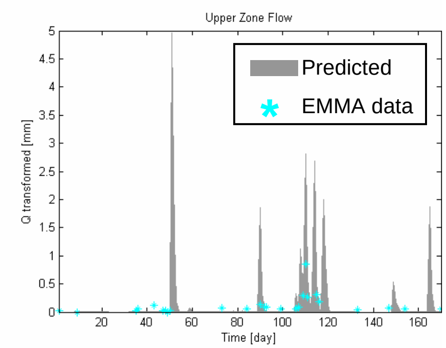


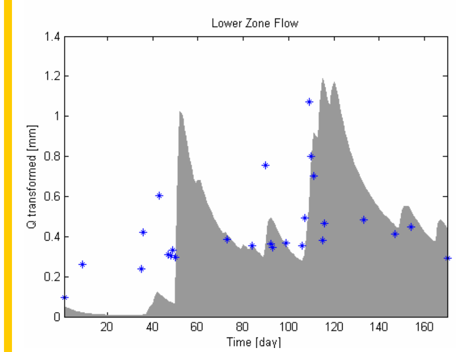
ONE STEP

	DRMS	BIAS	R2
Total	0.76	-0.14	0.86
Upper	1.33	3.07	0.85
Lower	0.37	-0.70	0.00

Hydrograph Separation Method







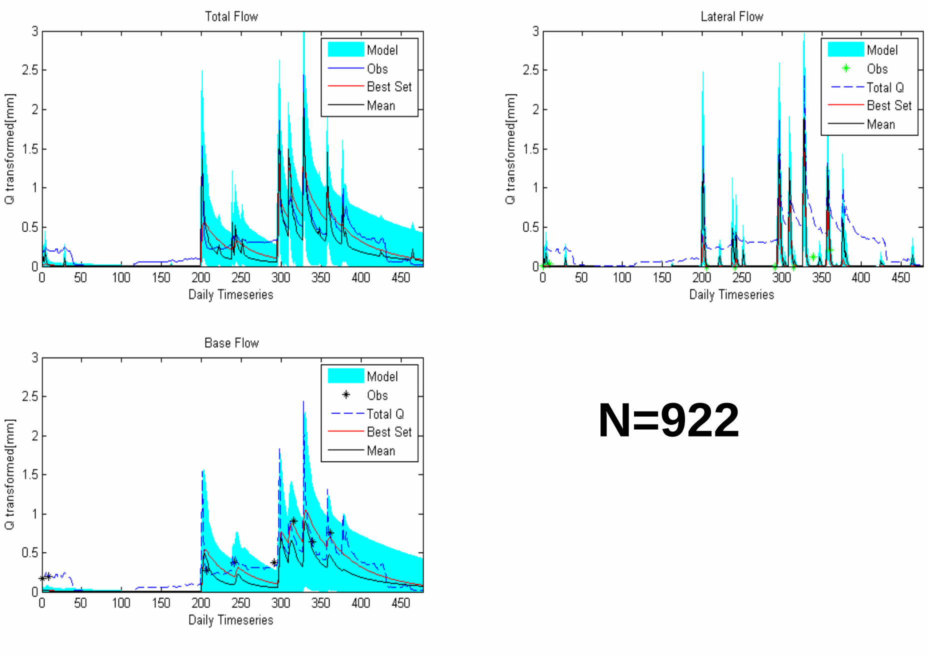
HYDROGRAPH SEPARATION

	DRMS	BIAS	R2
Total	0.89	0.09	0.81
Upper	1.29	2.56	0.75
Lower	0.67	-0.70	0.00

- Incorporation of additional data reduces bias in all components
- Q total bias also reduced

DC8 Pre-fire GLUE

Threshold: NSE > 0.3

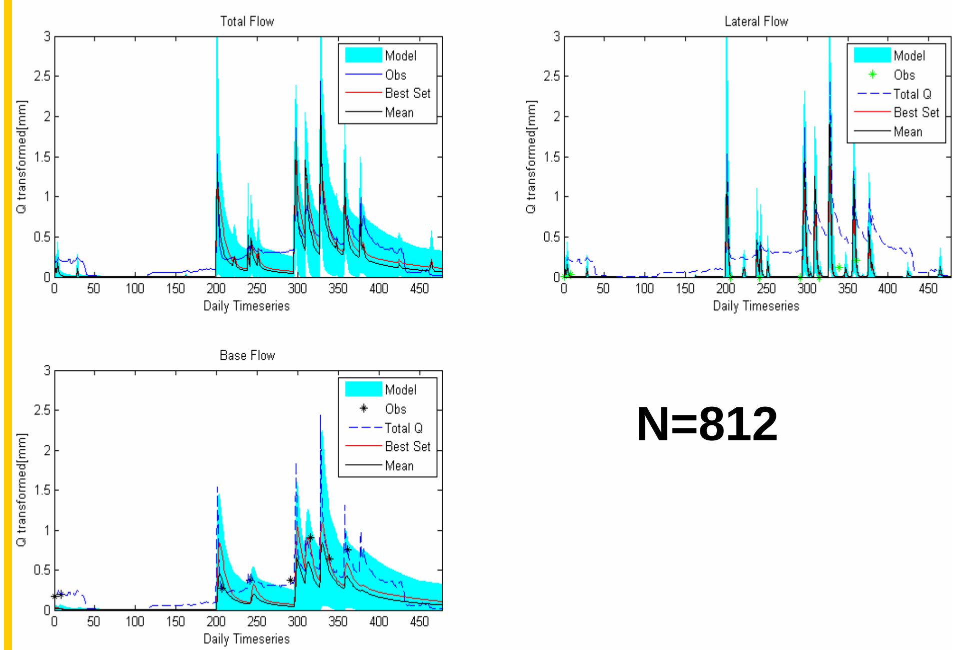


N=922

Ensemble Mean	RMSE	BIAS	R2
Total	0.2555	-0.4198	0.7692
Upper	0.0502	-0.1209	0.6142
Lower	0.2364	-0.7975	0.0376

Best set	RMSE	BIAS	R^2
Total	0.2565	-0.2571	0.7692
Upper	0.0741	-0.7552	0.5722
Lower	0.1871	-0.2271	0.0376

Threshold: NSE > 0.3 and DRMS_baseflow < 0.5



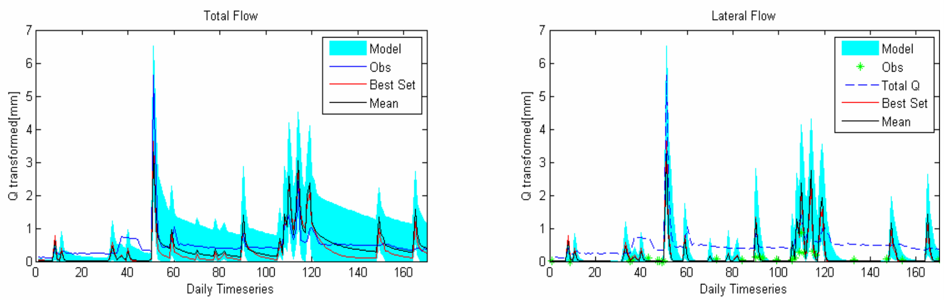
N=812

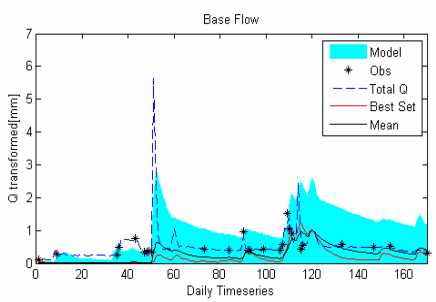
Ensemble Mean	RMSE	BIAS	R2
Total	0.2689	-0.4669	0.6915
Upper	0.0499	-0.1529	0.6157
Lower	0.2584	-1.1791	0.0650

Best set	RMSE	BIAS	R^2
Total	0.2663	-0.3496	0.6915
Upper	0.0525	-0.3742	0.6218
Lower	0.2403	-0.5478	0.0650

DC8 Post-fire GLUE

Threshold: NSE > 0.3



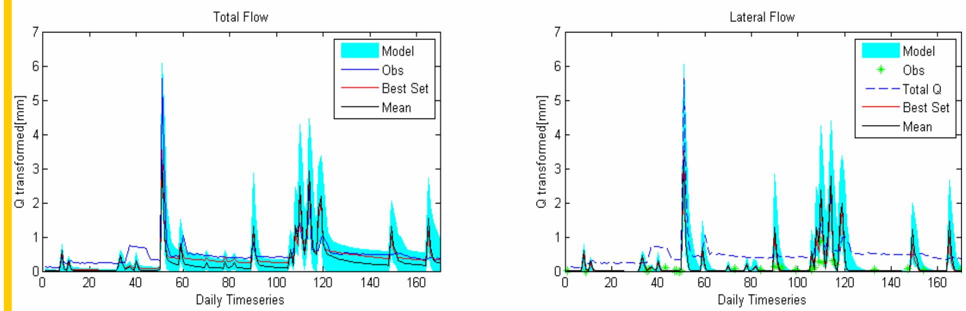


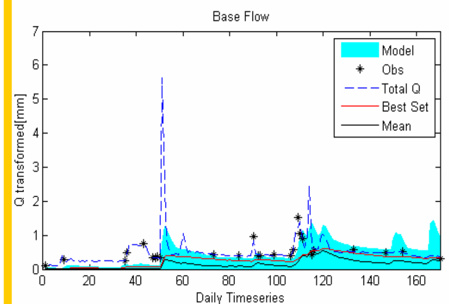
N=2560

Mean sets	RMSE	BIAS	R^2
Total	1.8162	-0.4476	0.2329
Upper	0.3839	0.9892	0.7773
Lower	0.4154	-0.5413	0.0153

Best sets	RMSE	BIAS	R^2
Total	1.8094	-0.5793	0.2329
Upper	0.3586	0.8718	0.7241
Lower	0.4598	-0.5632	0.0153

Threshold: NSE > 0.3 and DRMS_baseflow < 0.5

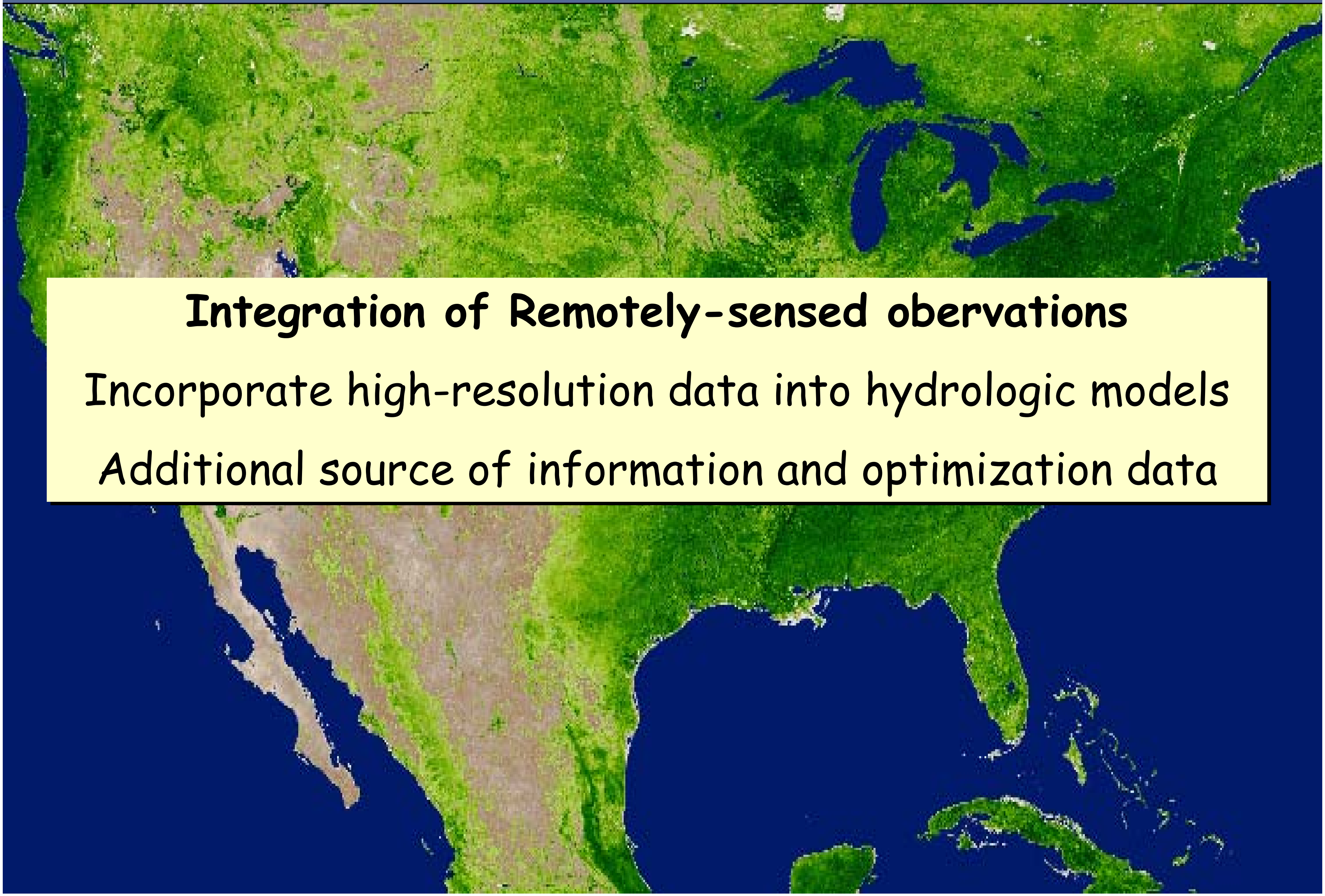




N=1144

Mean sets	RMSE	BIAS	R^2
Total	1.8370	-0.5962	0.3955
Upper	0.3843	1.0062	0.7865
Lower	0.4770	-0.7022	0.0001

Best sets	RMSE	BIAS	R^2
Total	1.7966	-0.4837	0.3955
Upper	0.3092	0.6518	0.8442
Lower	0.4058	-0.5310	0.0001



Integration of Remotely-sensed obervations

Incorporate high-resolution data into hydrologic models

Additional source of information and optimization data

Potential Evapotranspiration (PET) Model

- Priestley-Taylor's equation

$$LE = \alpha(Rn - G)\frac{\Delta}{\Delta + \gamma}$$

Δ the derivative of saturated vapor pressure (pa/K)

γ psychrometric constant (Pa/K)

α Priestly-Taylor parameter (1.26 in wet surface areas)

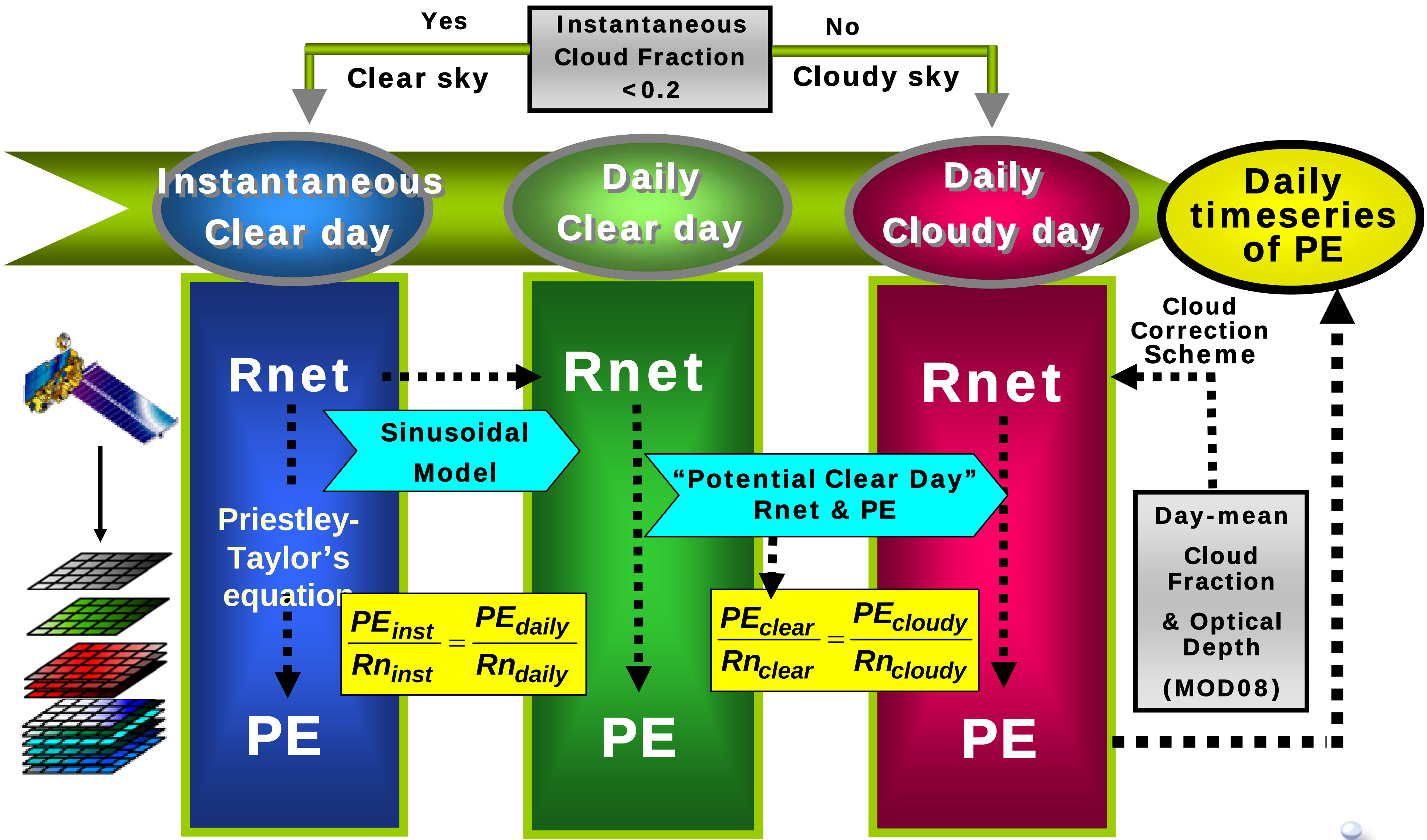
Required Inputs:

✓ Net Radiation (Rn) (~ Bisht et al.,2005)

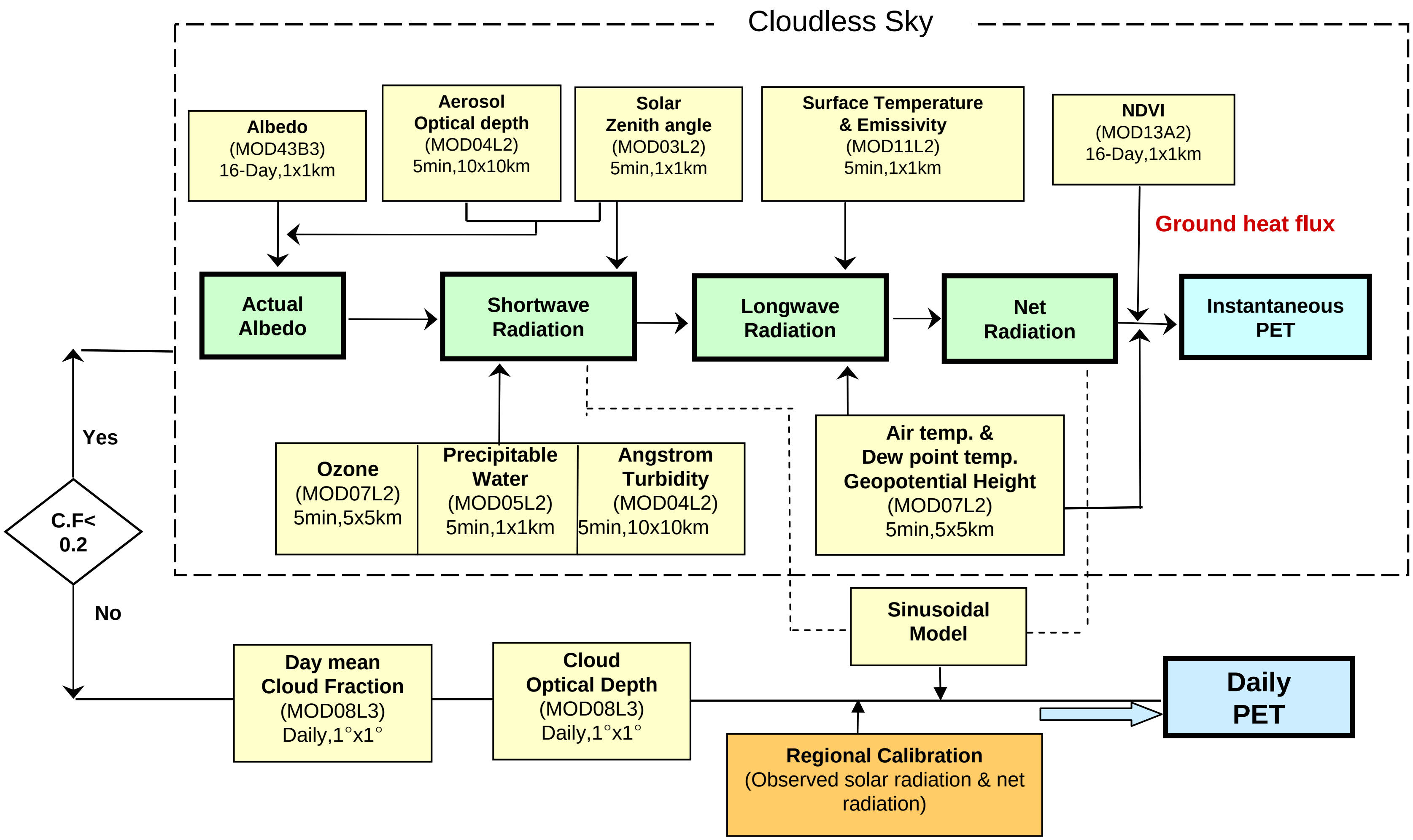
✓ Ground Heat Flux (G) $G = f(Rn, NDVI)$ (~ Moran et al.,1989)

✓ Air Temperature at screen level (Δ) $\Delta = f(T_a)$

Kim and Hogue (2007) Journal of Hydrometeorology in press


Yes
Instantaneous Cloud Fraction <0.2
No
Clear sky
Cloudy sky
Instantaneous Clear day
Daily Clear day
Daily Cloudy day
Daily timeseries of PE
Cloud Correction Scheme
Rnet
Rnet
Rnet
Sinusoidal Model
Priestley-Taylor's equation
"Potential Clear Day" Rnet & PE
$\frac{PE_{inst}}{Rn_{inst}} = \frac{PE_{daily}}{Rn_{daily}}$
$\frac{PE_{clear}}{Rn_{clear}} = \frac{PE_{cloudy}}{Rn_{cloudy}}$
Day-mean Cloud Fraction & Optical Depth (MOD08)
PE
PE
PE

Algorithm Flow Chart



13 MODIS variables → 1km Potential Evapotranspiration Product

Validation Sites

Selected four sites across the United States that have required variables for our study

Audubon Arizona (06/2002-12/2004)

Westville Oklahoma (01/2001-12/2004)

Bondville Illinois (01/2001-12/2004)

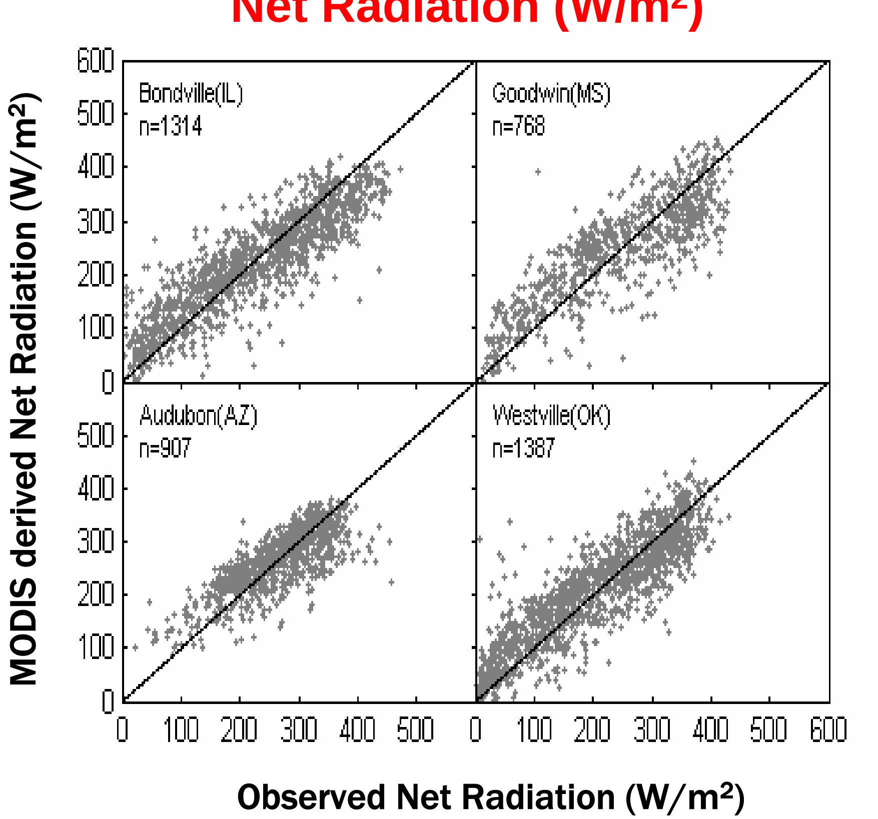
Goodwin Creek Mississippi (05/2002-12/2004)

- Remote Sensing Data
  MODIS subsetted over the core sites
- *in situ* Site Data
  AMERIFLUX, MESONET Networks

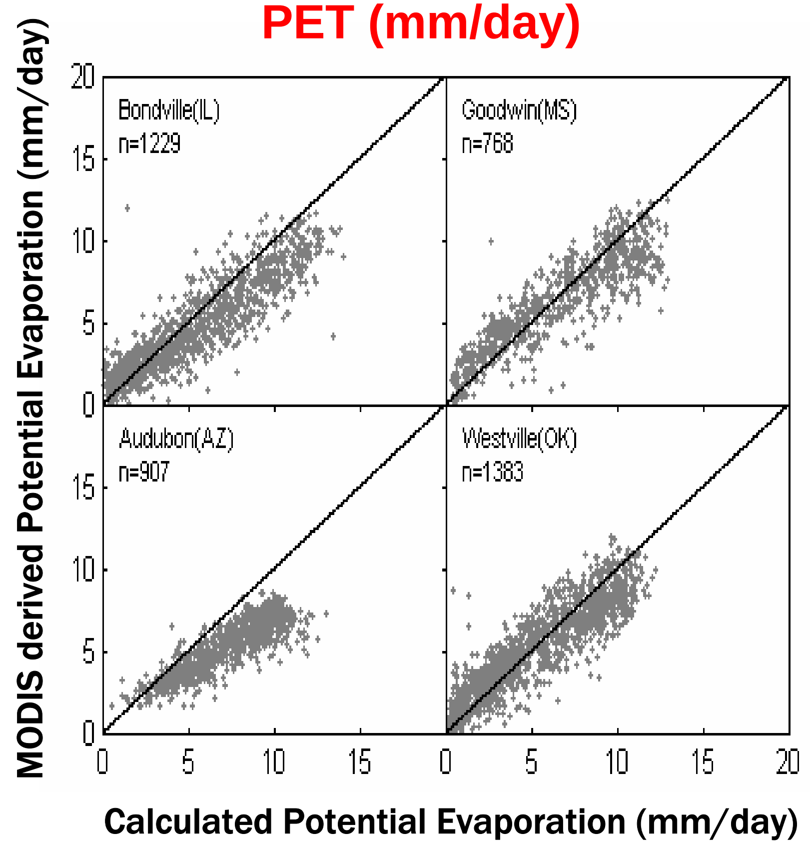
☆ **Ameriflux**
△ **Mesonet**

Daily Net Radiation & PET (all-sky conditions)

Net Radiation (W/m^2)



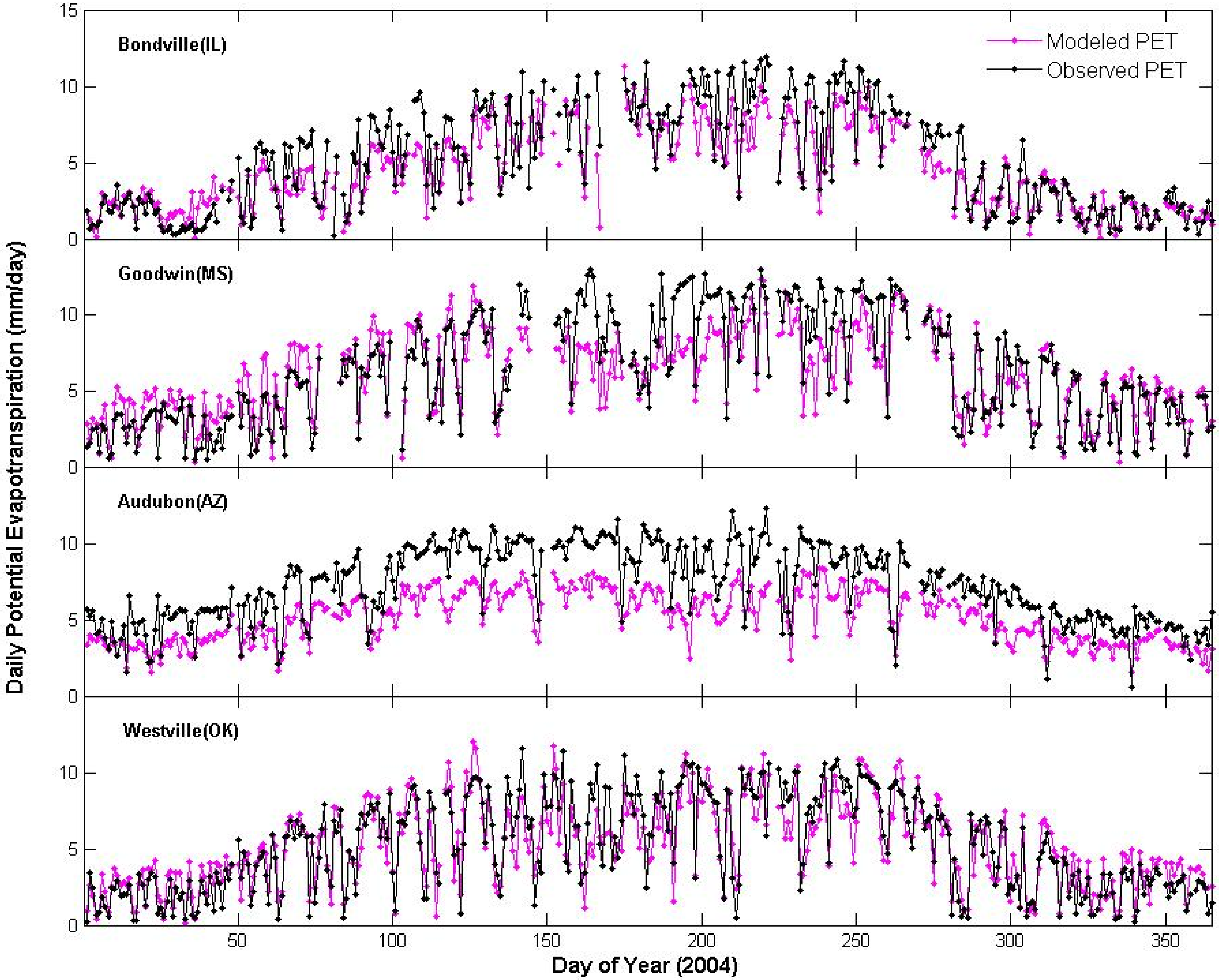
PET (mm/day)



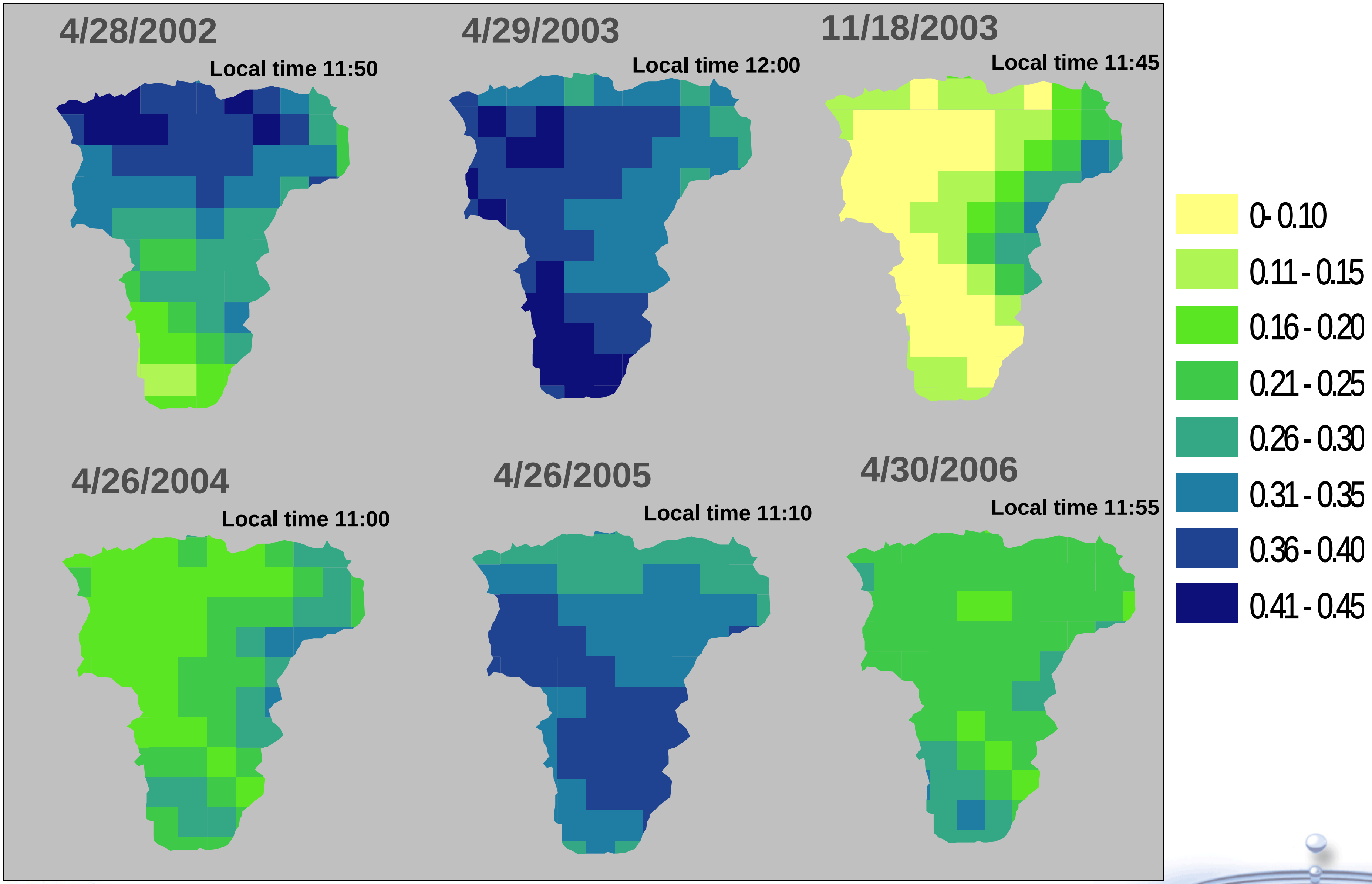
	Bondville (IL)	Goodwin (MS)	Audubon (AZ)	Westville (OK)
R^2 Bias RMSE	0.88 -1.63 55.74	0.84 3.14 61.3	0.78 1.05 45.01	0.87 6.67 52.2

	Bondville (IL)	Goodwin (MS)	Audubon (AZ)	Westville (OK)
R^2 Bias RMSE	0.9 -0.67 1.56	0.88 -0.28 1.68	0.86 -2.05 1.37	0.9 -0.07 1.43

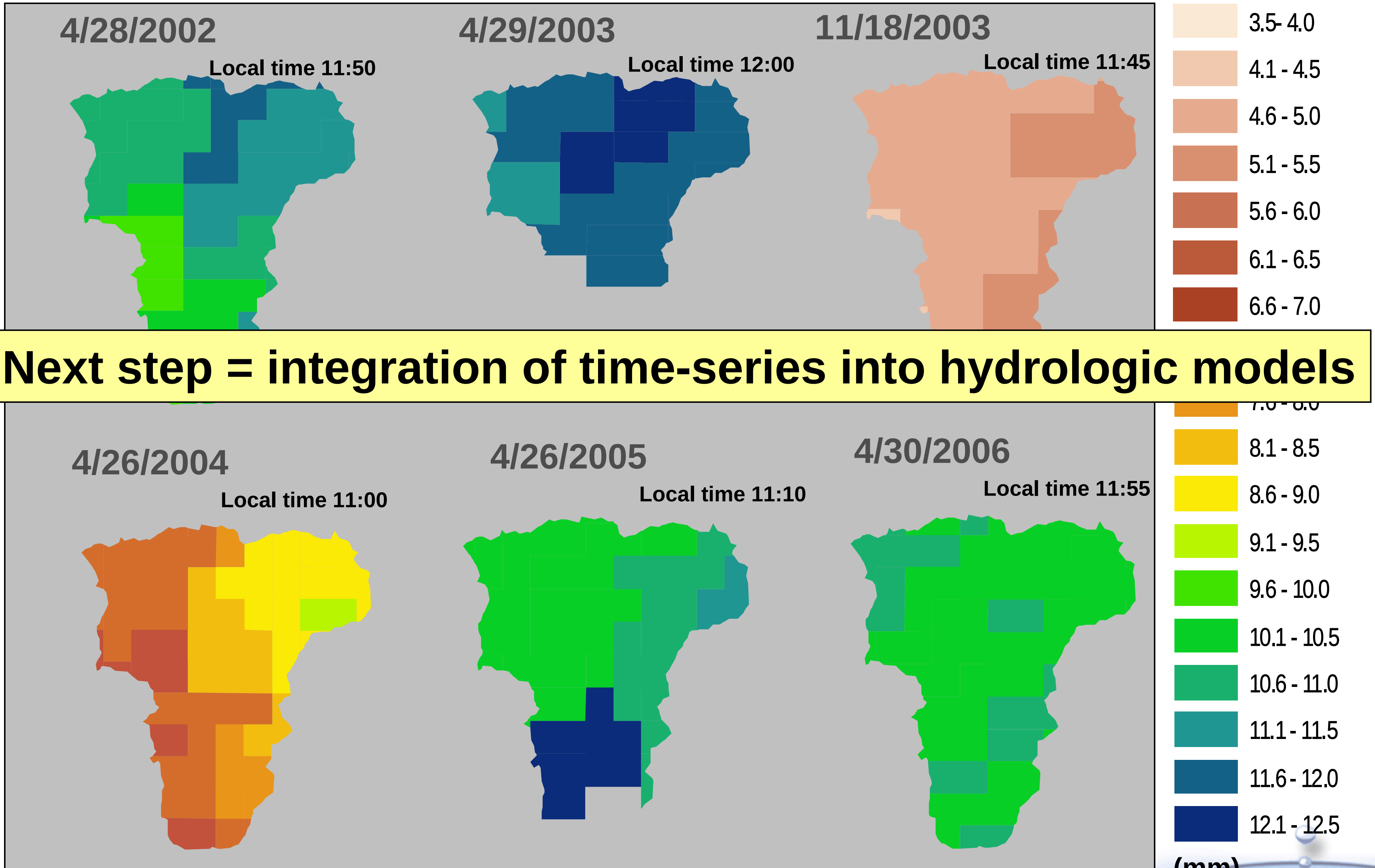
Daily Time Series of PET – Initial Study Sites



MODIS Enhanced Vegetation Index (EVI) in City Creek (CA)



Potential Evapotranspiration for City Creek (CA)

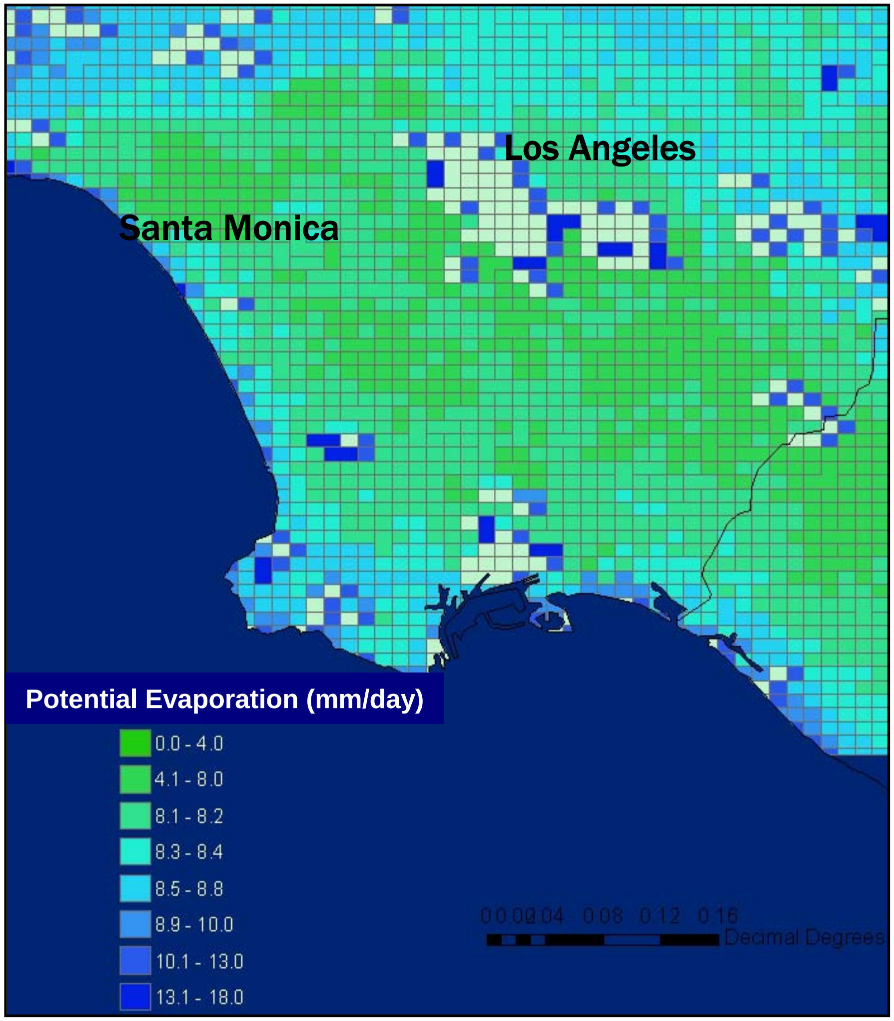


Next step = integration of time-series into hydrologic models

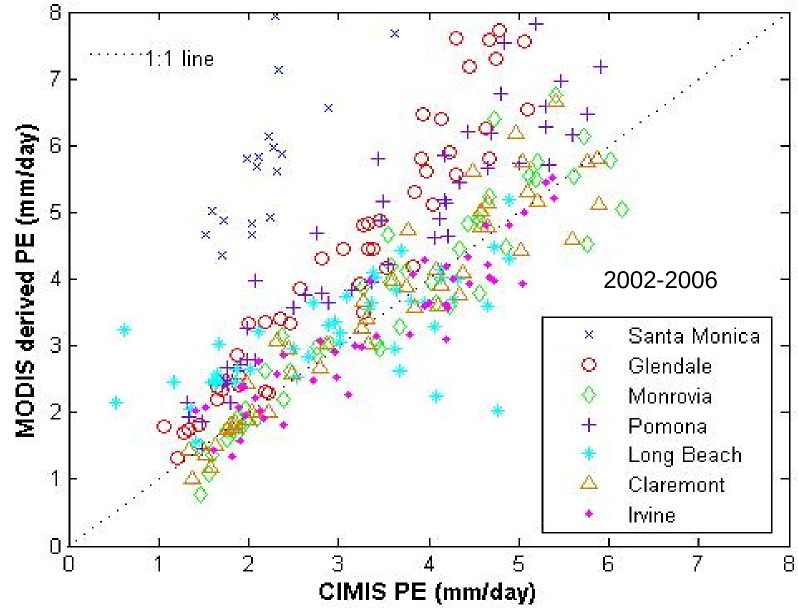
Regional-scale PET Estimation

Los Angeles – CA

01/06/2006 Local Time : 11:05



Comparison to ground-based CIMIS PET



Site	RMSE (mm/day)	Bias (mm/day)	Correlation
Santa Monica	5.448	5.713	0.932
Glendale	1.497	1.261	0.955
Monrovia	0.561	0.061	0.934
Pomona	1.225	1.072	0.945
Long Beach	0.916	0.236	0.674
Claremont	0.482	0.076	0.950
Irvine	0.458	-0.131	0.931

* Bias = MODIS derived PE-CIMIS PE

Concluding Remarks

- **THEME of our work: Integrate evolving data streams (hydrologic, geochemical, remotely-sensed) into hydrologic models to better understand and predict hydrologic response to change (wildfire, urbanization, climate)…**

- Need for **long-term, quality observational** networks (ground and remotely-sensed)

- Need for **novel approaches** to incorporate new data into models and **"validate" and "improve"** existing parameterizations

Funding Acknowledgements

NASA EOS, NASA Earth System Science (ESS) Fellowship, NASA-JPL SURP

NSF Hydrologic Sciences Program

NOAA – National Weather Service Hydrologic Laboratory

Calleguas Municipal Water District

NSF Science and Technology Center SAHRA at the University of Arizona

University of California Center for Water Resources

UCLA Faculty Grants Program, Graduate Research Mentorship Program and UC-LEADS

Hydrology and Water Resources at UCLA

Faculty - **Prof. William Yeh, Prof. Steve Margulis**
Adjuncts – Prof. Kendall, Prof. Sun

Research Areas

Surface Water Hydrology

- Rainfall-runoff and land surface modeling
- Watershed land-cover change studies
- Remote sensing of land surface parameters and processes
- Hydrometeorology and land-atmosphere interactions

Groundwater Hydrology

- Numerical simulation of groundwater flow and contaminant transport
- Inverse problems and experimental design
- Modeling and optimization of seawater intrusion barriers

Water Resources Engineering

- Optimization of large scale water resource systems
- Conjunctive use of surface water and groundwater

The M.S. program offers two options: i) 9 month comprehensive exam plan, and ii) a thesis option for those interested in research.

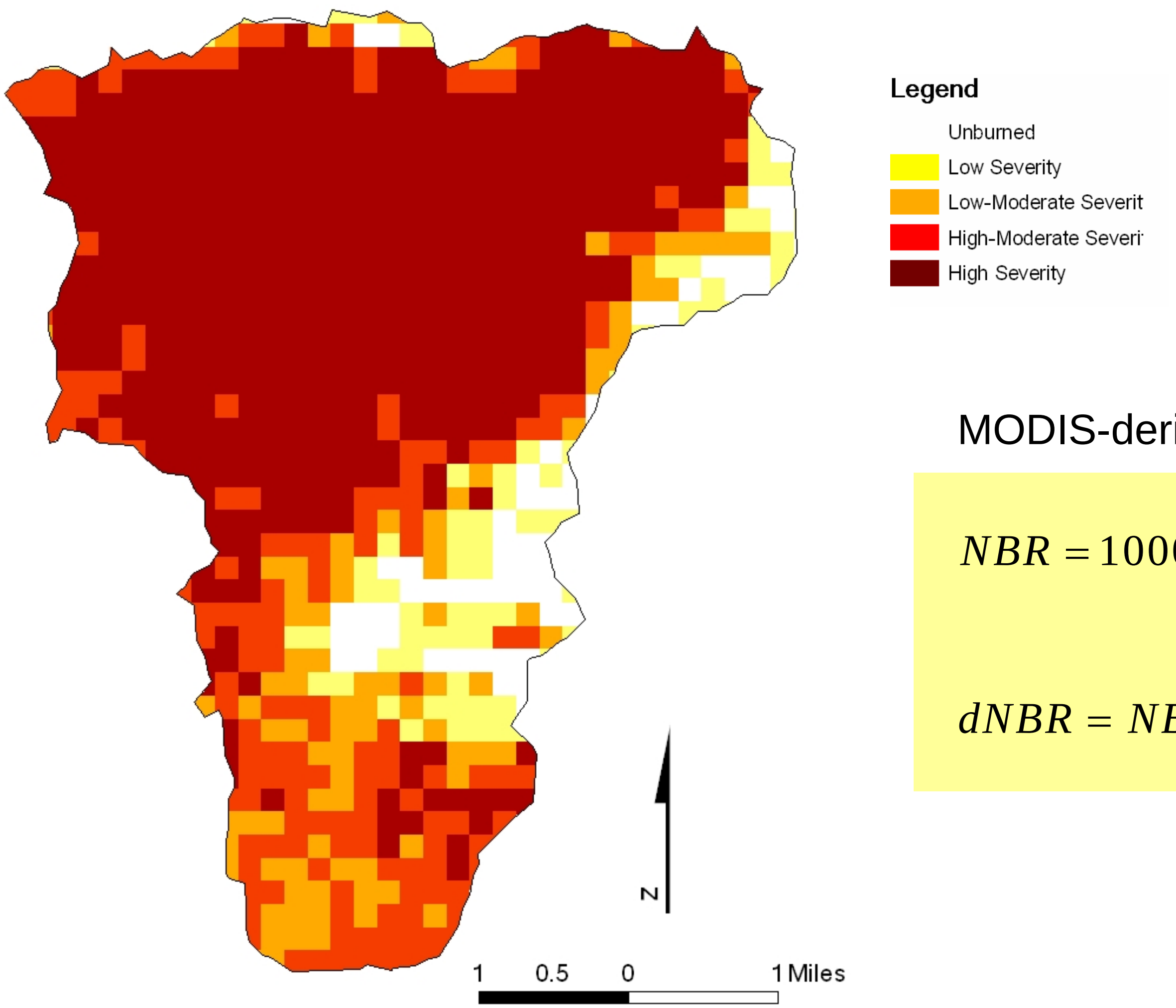
The Ph.D. program offers students the opportunity to perform in depth original research in the area of hydrology and water resources engineering, while obtaining breadth in other areas of study (e.g. environmental engineering, atmospheric science, geography, applied math, etc.).

Questions ?

South Sawyer Glacier, AK
September 9, 2007

Application to Post-fire Systems

City Creek Burn Severity



MODIS-derived Burn Severity

$$NBR = 1000\frac{(R_{788nm} - R_{2370nm})}{(R_{788nm} + R_{2370nm})}$$

$$dNBR = NBR_{pre} - NBR_{post}$$

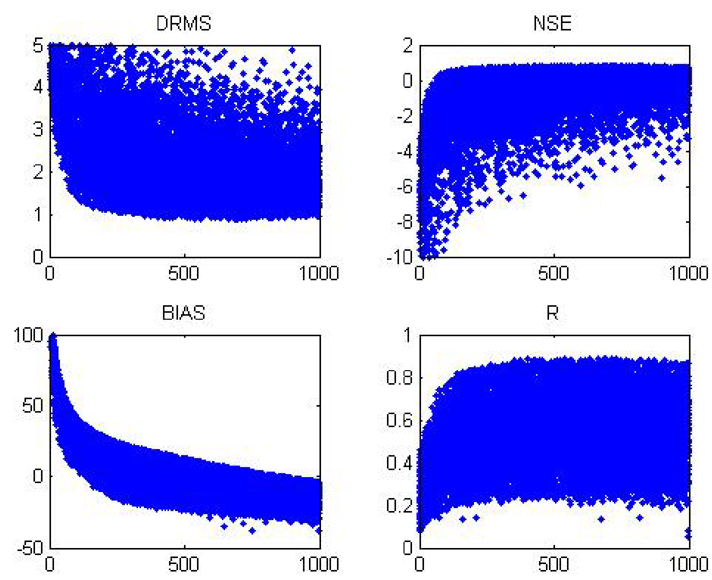
Regionalized Sensitivity Analysis (RSA)

- Global Sensitivity
- Monte-Carlo sampling of parameter space (16 pars)
  - All parameters randomly sampled at the same time (allows for parameter interaction)
  - 30,000 parameter sets
- Partitioning of response (threshold NSE > 0.3)

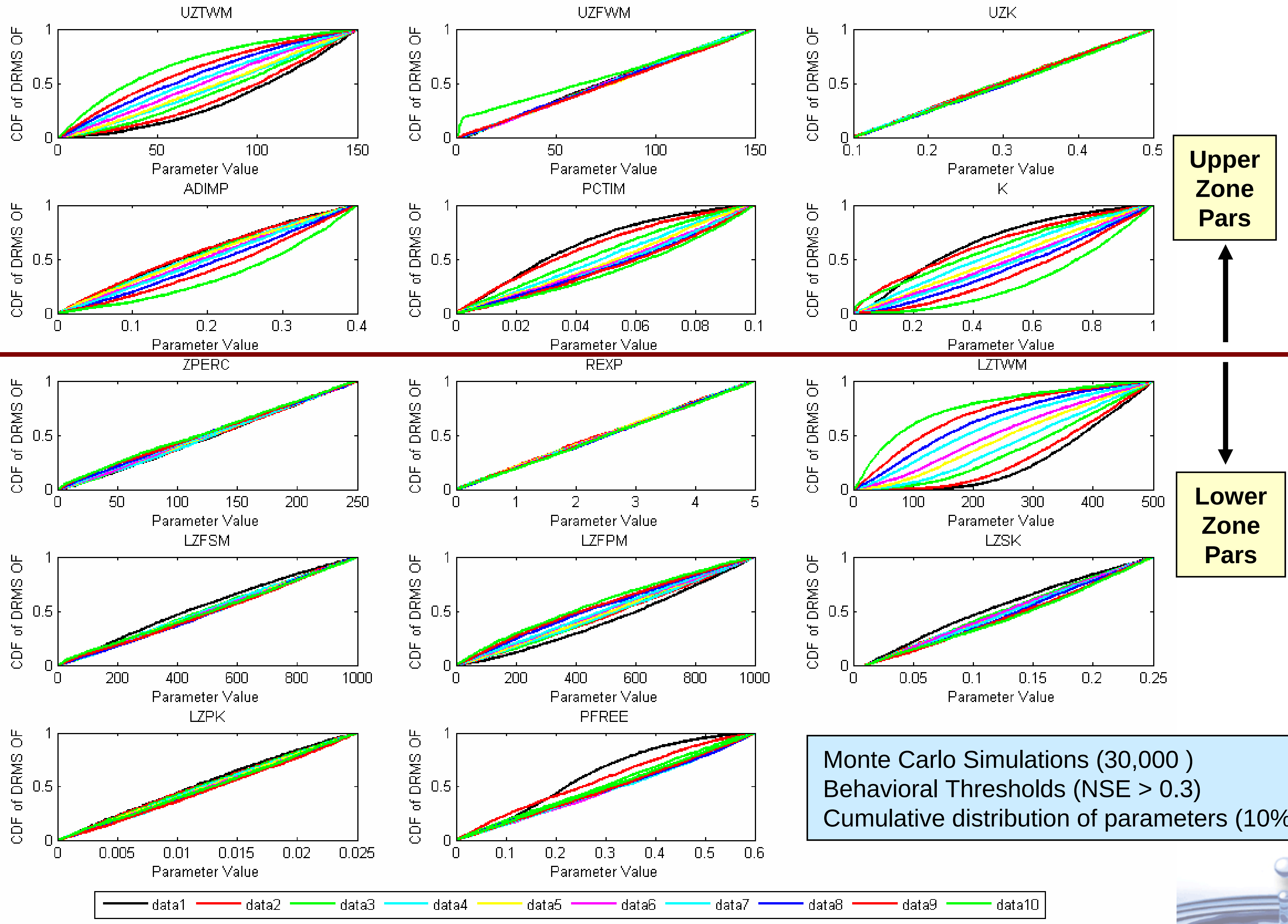
$$NSE = 1 - \frac{\sum_{t=1}^{N} (pred - obs)^2}{\sum_{t=1}^{N} \left(obs - \overline{obs}\right)^2}$$

  - Behavioral (acceptable)
  - Non-behavioral (unacceptable)

- Rescaling with behavioral sets and division into ten equal sets

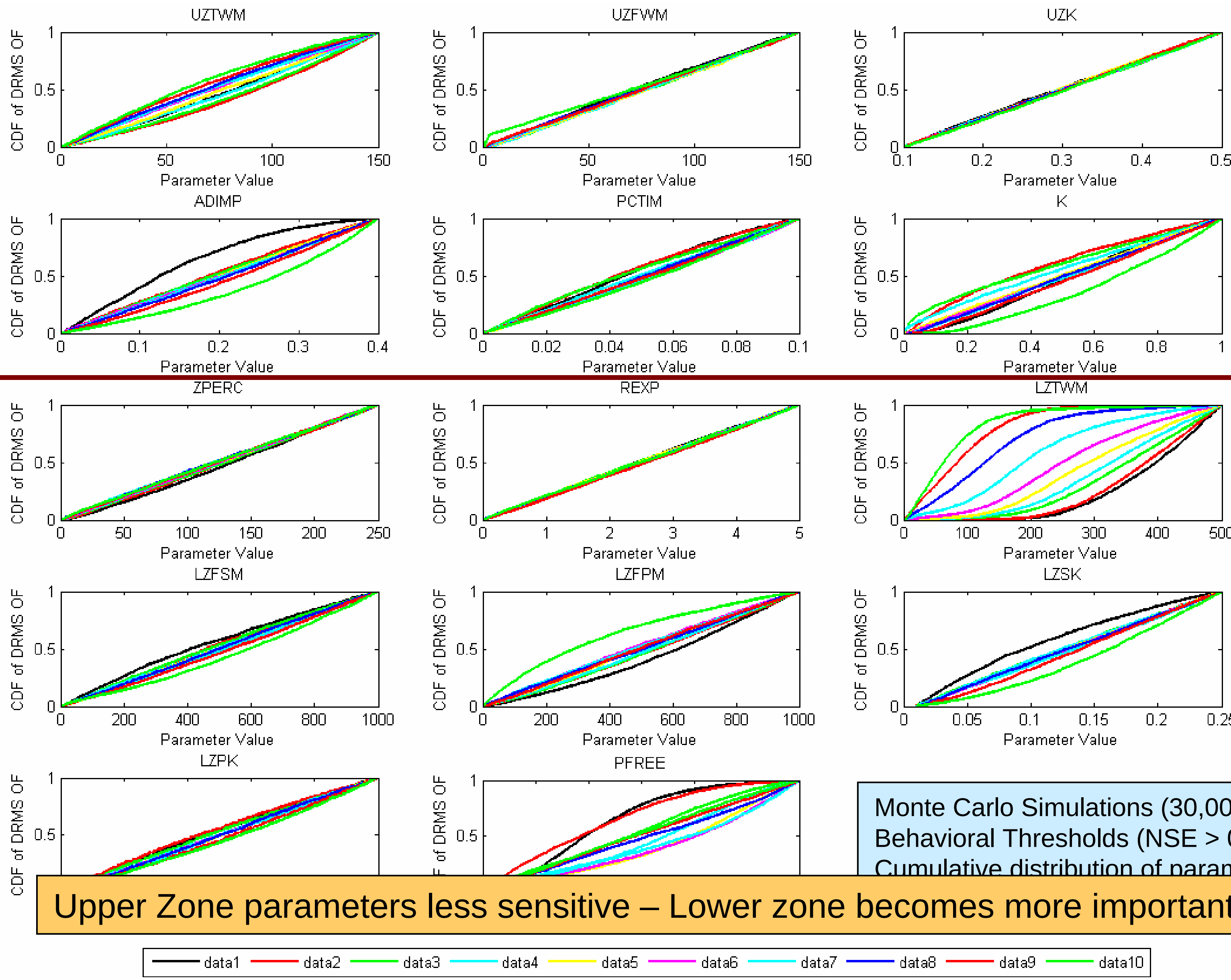


Pre-fire Parameter Sensitivity



Monte Carlo Simulations (30,000)
Behavioral Thresholds (NSE > 0.3)
Cumulative distribution of parameters (10%)

Post-fire Parameter Sensitivity



Upper Zone Pars

Lower Zone Pars

Monte Carlo Simulations (30,000)
Behavioral Thresholds (NSE > 0.3)
Cumulative distribution of parameters (10%)

Upper Zone parameters less sensitive – Lower zone becomes more important!!